

Evolutionary conservation and regulation of particular alternative splicing events in plant SR proteins

Maria Kalyna*, Sergiy Lopato, Viktor Voronin and Andrea Barta*

Department of Medical Biochemistry, Max F. Perutz Laboratories, Medical University of Vienna, Dr Bohr-Gasse 9/3, A-1030 Vienna, Austria

Received May 9, 2006; Revised July 13, 2006; Accepted July 20, 2006

ABSTRACT

Alternative splicing is an important mechanism for fine tuning of gene expression at the post-transcriptional level. SR proteins govern splice site selection and spliceosome assembly. The Arabidopsis genome encodes 19 SR proteins, several of which have no orthologues in metazoan. Three of the plant specific subfamilies are characterized by the presence of a relatively long alternatively spliced intron located in their first RNA recognition motif, which potentially results in an extremely truncated protein. In atRSZ33, a member of the RS2Z subfamily, this alternative splicing event was shown to be autoregulated. Here we show that atRSp31, a member of the RS subfamily, does not autoregulate alternative splicing of its similarly positioned intron. Interestingly, this alternative splicing event is regulated by atRSZ33. We demonstrate that the positions of these long introns and their capability for alternative splicing are conserved from green algae to flowering plants. Moreover, in particular alternative splicing events the splicing signals are embedded into highly conserved sequences. In different taxa, these conserved sequences occur in at least one gene within a subfamily. The evolutionary preservation of alternative splice forms together with highly conserved intron features argues for additional functions hidden in the genes of these plant-specific SR proteins.

INTRODUCTION

The accuracy of intron excision from pre-mRNAs relies on the precise recognition of rather degenerate splicing signals.

Splice sites are defined by consensus sequences where only the two terminal nucleotides [GT at the acceptor (5') and AG at the donor (3') splice sites for the majority of introns] are highly conserved. The fidelity of the splicing process is attained through various exonic and intronic sequences acting as splicing enhancers or silencers thereby influencing splice site selection. The precise recognition of splice sites is achieved by the spliceosome, a large RNP particle consisting of 5 snRNPs and ~200 different proteins (1). Out of these, Ser/Arg-rich (SR) proteins play an important role in the regulation of splice site selection. SR proteins are a family of evolutionary conserved, structurally related, splicing factors which possess one or two RNA-recognition motifs (RRM) at the N-terminus and a C-terminal arginine/serine-rich (RS) domain (2,3). They are important for recognizing splice sites or exonic splicing enhancers and facilitate spliceosome assembly. It has been demonstrated that some SR proteins influence alternative splice site selection in quantitative manner (4). Consequently, SR proteins are one of the determinants of splicing pattern present in different cells and at specific developmental stages or conditions.

Alternative splicing is an important mechanism for the generation of proteome complexity and fine tuning of gene expression at the post-transcriptional level both in the metazoan and plant species. Differential intron removal allows production of splice variants which may code for distinct protein isoforms affecting their post-translational modification, subcellular localization and the ability to interact with their binding partners. In addition, alternative splicing can also influence translational control and stability of mRNA transcripts.

Our analysis of the *Arabidopsis* genome has revealed 19 SR proteins which is almost twice as much as in humans (5,6). They fall into seven subfamilies (6) some of which have orthologues in metazoan (SF2/ASF, 9G8, SC35) but interestingly three of them seem to be plant-specific (RS, RS2Z and SCL). Many of the SR protein genes are alternatively spliced both in metazoa and in *Arabidopsis* (7–16). It has been noted that alternative splicing occurs mainly in

*To whom correspondence should be addressed. Tel: +43 1 4277 61640; Fax: +43 1 4277 9616; Email: andrea.barta@meduniwien.ac.at

*Correspondence may also be addressed to Maria Kalyna. Tel: +43 1 4277 61642; Fax: +43 1 4277 9616; Email: mariya.kalyna@univie.ac.at

Present addresses:

Sergiy Lopato, Australian Center for Plant Functional Genomics, The University of Adelaide, PMB 1 Glen Osmond, South Australia, 5064, Australia
Viktor Voronin, Gregor Mendel-Institut, Dr Bohr-Gasse 3, A-1030 Vienna, Austria

© 2006 The Author(s).

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.0/uk/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

and around the long introns of the *Arabidopsis* SR genes ranging in size from ~400 to 1100 nt (6), while the typical size of plant introns is <150 nt (17). In *atSRp30* and *atSRp34/SR1*, two *Arabidopsis* homologues of human SF2/ASF, such a long intron is situated near the 3' end of the gene, and alternative splicing in this intron results in protein isoforms with shortened RS domains (12,15). Interestingly, all three plant-specific gene subfamilies possess a long intron at the beginning of the gene dividing the first RRM. Alternative splicing of this intron occurs in all three subfamilies and in a way that the potential protein is extremely truncated containing only a part of the RRM (6,13,14). Alternative splicing in the long intron of *atRSZ33*, a member of the RS2Z subfamily is autoregulated, as demonstrated by overexpression experiments (16). This autoregulation is crucial for correct gene expression levels, because overexpression of an intronless version of *atRSZ33* is lethal.

In this study, we extend the analysis of alternative splicing and its regulation to the plant-specific RS subfamily, focusing on *atRSp31*. We also show that *atRSZ33*, a member of the RS2Z subfamily, is involved in this regulation. In addition, we show that the position of the long intron and its capacity for alternative splicing is a characteristic feature of the plant-specific families of SR genes and is conserved from green algae to flowering plants. Moreover, sequences around the alternative splice sites in the long introns are much more conserved over large evolutionary distances than those in the region of the constitutive splice sites in the respective genes, indicating that they are under strong selective pressure.

MATERIALS AND METHODS

Accession numbers

Arabidopsis and rice sequences can be found at <http://www.ncbi.nlm.nih.gov> and <http://www.tigr.org> under following numbers: *atRSp31* (At3g61860), *atRSp31a* (At2g46610), *atRSp40* (At4g25500), *atRSp41* (At5g52040), *osRsp29* (Os04g02870), *osRsp33* (Os02g03040), *atRSZ32* (At3g53500), *atRSZ33* (At2g37340), *osRSZ36* (Os05g02880), *osRSZ37a* (Os01g06290), *osRSZ37b* (Os03g17710) and *osRSZ39* (Os05g07000). Accession numbers of *Arabidopsis*, rice, maize, *Pinus taeda*, *Physcomitrella patens* and *Chlamydomonas reinhardtii* transcripts are in Supplementary Data.

Sequence retrieval and analysis

To identify splice variants of *Arabidopsis* SR proteins, corresponding genomic sequences were used in BLAST search against EST database limited to *Arabidopsis thaliana* at <http://www.ncbi.nlm.nih.gov/BLAST/>. ESTs with matches to corresponding gene were selected and then aligned using Geneseq (18) at <http://www.plantgdb.org>.

To identify splicing events in rice, orthologues of *atRSp31* and *atRSZ33* were searched using *Arabidopsis* protein sequences at The TIGR Rice Database (<http://tigrblast.tigr.org>). Genomic sequences for retrieved proteins were used to get corresponding transcripts. Selected transcripts were aligned to genomic sequences using Geneseq.

Splice variants of maize orthologues of *atRSp31* were retrieved at Geneseq using genomic sequences of *zmRSp31A* and *zmRSp31B* (19).

To identify orthologues of either *atRSp31* or *atRSZ33* in the distant species *Arabidopsis* protein sequences were used to pre-screen translated EST databases at <http://www.ncbi.nlm.nih.gov/BLAST/> limited to Coniferophyta, Bryophyta or algae. A second screen was done using either *atRSp31* or *atRSZ33* protein sequences in a BLAST search limited to species detected in the pre-screen. EST sequences were assembled into contigs using the CAP2 program (20) at <http://www.infobiogen.fr>. Obtained contigs were analysed by blastx against *Arabidopsis* proteins. In addition, translated contigs were compared to *Arabidopsis* and rice SR proteins using ClustalW program (21) at <http://clustalw.genome.jp>. Shading of multiple alignment files was done with BOXSHADE at <http://www.ch.embnet.org/>.

Contigs were used to retrieve genomic sequences at NCBI and species specific databases (PHYSCObase, <http://moss.nibb.ac.jp>; ChlamyDB <http://www.chlamy.org/chlamydb.html>). Genomic sequences were used in screen for additional ESTs representing splice variants. Gene structures and alternative splicing events were finally analysed using Geneseq.

Plant growth and protoplast transformation

Wild-type plants of *A.thaliana* (Col-0) and plants overexpressing *atRSZ33* (16) were germinated and grown either on GM medium (22) or in soil under 16 h light/8 h dark photoperiod at 23°C.

Arabidopsis cell suspension growth conditions and protoplasts preparation and transformation were as described previously (23). Protoplasts were collected by centrifugation and frozen in liquid nitrogen for RNA isolation or resuspended in SDS-PAGE loading buffer for analysis by western blotting, 24 h after transformation.

RNA isolation and analysis of alternative splicing forms

Total RNA from *Arabidopsis* plants and transformed cell culture protoplasts was isolated using RNeasy Plant Mini Kit (Qiagen) and treated with DNase I (Promega).

cDNAs of mRNA isoforms of *atRSp31* were obtained by RT-PCR (24).

Primer 1, 5'-AAATGAGCTCCATTATGAAGTTTCTACTG-3', containing a SacI restriction site (underlined), was used to prime reverse transcriptase.

Primer 2, 5'-AACTGGATCCAGTCGTCGTCGTCGTCTAGGG-3', and primer 3, 5'-ATATAGGGATCCCATAA-GGTCTTCTTGGACTGGAG-3', both of which contain an additional BamHI restriction site, were used for PCR. Splicing events in the second intron were investigated by PCR using primers from the adjacent exons: primer 4, 5'-AATGAGCTCGAATTGCGAATTAAGATAAAG-3', and primer 5, 5'-ATAGGATCCTTTGCCATTCAACTGATAAC-3', containing a SacI and BamHI restriction site, respectively.

To analyse regulation of the splicing profile in *Arabidopsis* cell suspension protoplasts transfected with either *pHA-catRSp31* or *pHA-gatRSp31* (see below), reverse transcription was carried out using primer 5'-CTTTGAGTAGCTTCAAGGG-3'. PCR was performed using the same reverse primer and a direct primer either to the HA tag 5'-GATCCT-ACCCATATGACGTTCCAGATTACGCTA-3' (to detect

transgenic *atRSp31*), or 5'-CGAATTAAGATAAAGATG-↓AGGCCA-3' (to detect endogenous *atRSp31*; the position of the HA tag insertion in the pHA-catRSp31 or pHA-gatRSp31 is indicated by an arrow).

To control loading, RT-PCR of ubiquitin with following primers 5'-CTCCTTCTTTCTGGTAAACGT-3' and 5'-CTCCTTCTTTCTGGTAAACGT-3' (25) was used. All PCR products, except ubiquitin, were sequenced.

***In vitro* transcription and splicing in HeLa splicing extracts**

A DNA template for *in vitro* transcription of the complete second intron was obtained by PCR amplification on the genomic clone using primers 4 and 5. The PCR product was subcloned into the SacI and BamHI sites of the vector pSP65 (Promega), linearized with BamHI and used for *in vitro* transcription as described previously (26). RNA was purified and stored in aliquots at -70°C . Splicing reactions were carried out using the HeLa cell RNA Splicing System (Promega). Products of splicing were analysed by RT-PCR using primers 4 and 5 and sequenced.

Protein purifications

The coding region of *atRSp31* cDNA was amplified by PCR using primers 6, 5'-ATATCCATGGGGCCAGTGTTCGT-CGG-3', and primer 3, containing NcoI and BamHI restriction sites, respectively. Following cleavage of the PCR product with these enzymes, it was subcloned into the pET-3d expression vector (Novagen). To obtain the NcoI restriction site, the fourth nucleotide of the coding region was changed to G (see above primer 6, boldface G). Thus, in the expressed protein the second amino acid is changed from an arginine to a glycine. Induction of recombinant protein expression in the *Escherichia coli* strain BL21(DE3) pLysS was done according to the Novagen protocol. *atRSp31* was purified from inclusion bodies as described previously (12). The SR proteins from *Arabidopsis*, carrot and rabbit were purified using a two-step salt precipitation method as described in (26).

Preparation of constructs of HA-tagged *atRSp31* for protoplast transformation

The coding and genomic sequences of *atRSp31* were amplified using primers 5'-CATGCCATGGCTTACCCATATGACGTTCCAGATTACGCTAGGCCAGTGTTCGTCG-3' (NcoI site, underlined; HA epitope tag, in boldface) and 5'-AAAAGTGCAGCATTGATCAAGGCTTTCCTC-3' (PstI site is underlined). The resulting PCR fragments were cloned into NcoI and PstI sites of pDEDH-Nco (27). Next, translation initiation sequence of pDEDH-Nco was replaced by the 5'-untranslated region (5'-UTR) of *atRSp31*, which was amplified using primers 5'-CCCCGGGGAGTCGTCGTCGT-CGTCTAG-3' (XmaI site is underlined) and 5'-CATGCCA-TGGCTTTATCTTAATTTCGCAATTCC-3' (NcoI site is underlined). This yielded the pHA-catRSp31 and pHA-gatRSp31 constructs. The constructs were verified by sequencing.

Western blotting

Proteins were separated by SDS-PAGE on 15% gels and electroblotted on a Immobilon-P Transfer Membrane

(Millipore). Antibodies used were rat monoclonal anti-HA 3F10 (1:1000) (Roche Diagnostics) and goat anti-rat IgG horseradish peroxidase-conjugated (1:10 000) (Sigma-Aldrich). Blots were developed using ECL western blotting detection reagents (Amersham Biosciences).

RESULTS

Gene structure and alternative splicing forms of *atRSp31*

The five introns of *atRSp31* are arranged such that they predominantly separate protein domains (Figure 1). The first intron is situated in the 5'-UTR, whereas the much longer second intron divides the first RRM, separating the highly conserved RNP2 and RNP1 domains. The third and the fourth introns border the second RRM. Furthermore, the RS region is delimited by the fourth and fifth introns so that the last exon contains the remaining nine amino acids and the 3'-UTR. Previously, northern blot analysis of poly(A)⁺ RNA from various tissues of *A.thaliana* plants revealed at least three transcripts of *atRSp31* which were differentially expressed in roots, leaves, stems and flowers of wild-type plants (11). RT-PCR amplification using total RNA was performed, and the three main products of 1320, 1199 and 791 bp were subcloned and sequenced (Figure 1). The 791 bp form corresponds to the 'correctly' spliced cDNA (mRNA1) which was published earlier (11) and which gives rise to a full-length 31 kDa protein. The longest form (1320 bp) arises from the usage of an alternative 3' splice site in the long intron (mRNA3) and encodes a short protein of 71 amino acids (8.5 kDa) due to a new in frame stop codon in the included intron (Figure 1). The hypothetical protein contains the first 35 amino acids from the RRM1 domain comprising the conserved RNP2 and 36 amino acids from the included sequence of the second intron. In the 1199 bp product, the same alternative 3' splice site is used as in mRNA3 but an additional alternative 5' splice site is recognized in this intron resulting in a new alternative exon (mRNA2, Figure 1). However, this splicing event gives rise to the same potential protein product as the one encoded by mRNA3. Additionally, both mRNA2 and mRNA3 possess partial sequences of the third intron due to use of an alternative 3' splice site.

According to the RT-PCR analysis presented in the Figure 1, all three splice forms of *atRSp31* are expressed in *Arabidopsis* flowers. In contrast, in wild-type suspension protoplasts (Figure 2A, middle panel) and wild-type seedlings (Figure 2C, lane 1), only mRNA1 encoding the full-length protein is produced. Interestingly, alternative mRNA2 and mRNA3 are the only transcripts of *atRSp31* detected by northern blotting in the leaves and stems (11). Because expression of alternative splice forms is regulated in a tissue/organ-specific manner, it was interesting to check if a protein is produced from these alternatively spliced mRNAs. *Arabidopsis* cell culture protoplasts were transiently transfected with a genomic construct of *atRSp31* under the control of a 35S *CaMV* promoter. A hemagglutinin (HA)-tag was introduced just after the start codon to allow monitoring the protein production. As shown in Figure 2A, lane 2, upper panel, both alternatively spliced mRNAs could be detected,

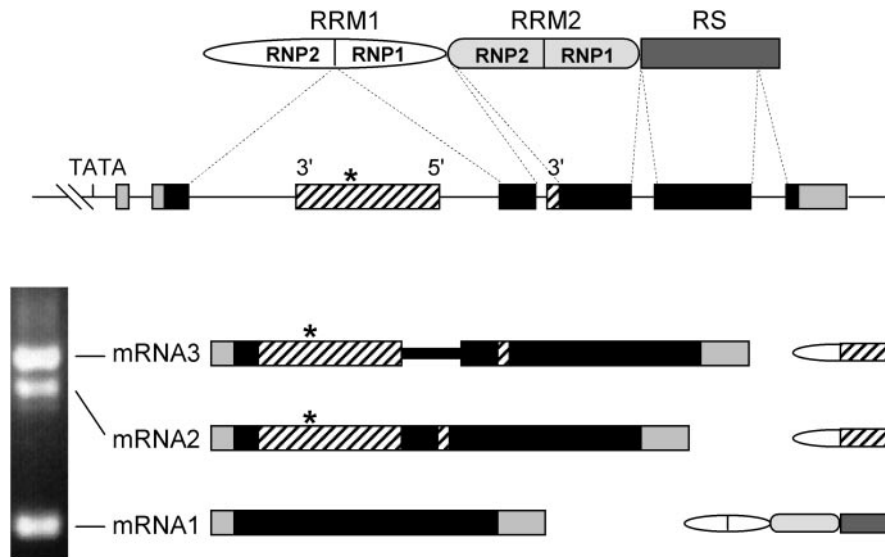


Figure 1. Schematic representation of atRSp31, the alternatively spliced products and the deduced proteins. The domain structure of the atRSp31 protein is shown; the RNA recognition motifs (RRM, containing the highly conserved RNP2 and RNP1 sequences) and the arginine/serine (RS) regions are indicated. Exons are illustrated as boxes and intron sequences are shown as lines. Exons containing the 5'- and 3'-UTRs are painted grey, the coding region black and alternatively spliced exons are striped. The thin lines connect the protein domains with the gene structure indicating where intron insertions occur. The star indicates the new stop codon in the alternatively spliced products. mRNA3 contains intronic sequences of the 3' alternatively spliced second and third intron; mRNA2 contains sequences of the 3' and 5' alternatively spliced second intron and the 3' alternatively spliced third intron; mRNA1 is the generic spliced mRNA. Protein structures for the different splice forms are shown to the right of the corresponding mRNA species. Parts of proteins generated by included intronic sequences are shown as striped boxes. RT-PCR analysis of RNA isolated from *Arabidopsis* flowers is shown next to schemes of corresponding splice variants.

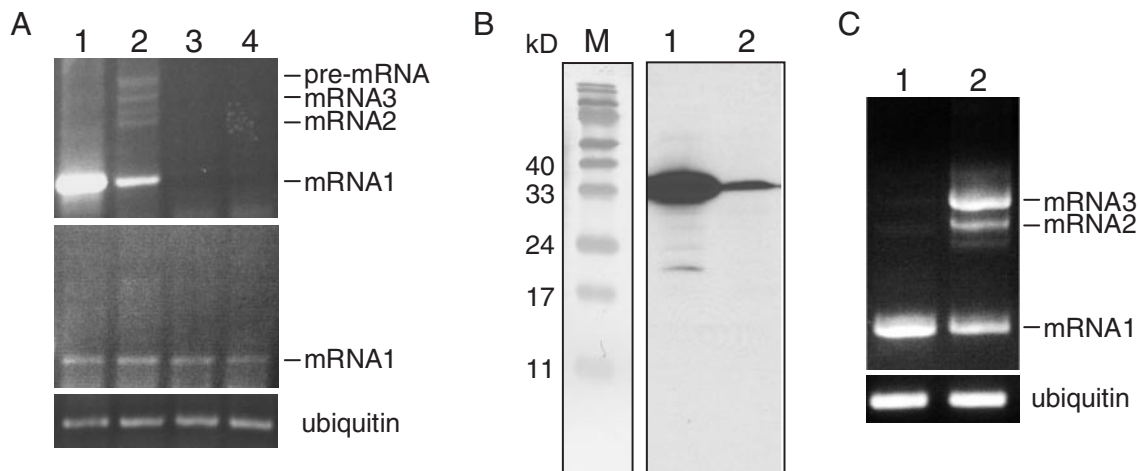


Figure 2. Regulation of alternative splicing in atRSp31. (A) Analysis of atRSp31 splicing in *Arabidopsis* cell culture protoplasts. Protoplasts were transiently transformed with either the cDNA or the genomic construct of atRSp31 under 35S CaMV promoter (lanes 1 and 2). Lanes 3 and 4: control transformations with empty vector and water, respectively. Total RNA was analysed by RT-PCR. Upper panel: detection of the transgenic atRSp31 using the primer to HA tag and the primer to the fourth exon of atRSp31. Middle panel: detection of the endogenous atRSp31 using the primer to the sequence which is disrupted in transgene by insertion of HA tag and the primer to the fourth exon of atRSp31. Lower panel: RT-PCR of ubiquitin used as a loading control. (B) Immunodetection of HA-tagged atRSp31. *Arabidopsis* cell culture protoplasts were transformed with HA-tagged the cDNA or the genomic construct of atRSp31 under 35S CaMV promoter (lanes 1 and 2, respectively). Total protein extracts were analysed by western blot using a monoclonal antibody against HA-tag. Lane M, molecular weight markers. (C) Regulation of alternative splicing in atRSp31 by atRSZ33. Upper panel: total RNA from wild type (lane 1) and seedlings overexpressing atRSZ33 (lane 2) was analysed by RT-PCR using the primers to the beginning of the second and the end of the fourth exons of atRSp31. Lower panel: RT-PCR of ubiquitin used as a loading control.

but no 8.5 kDa protein was visible in a western blot analysis with HA-antibodies (Figure 2B, lane 2). This experiment shows that under the conditions employed no protein from mRNA2 and mRNA3 is detectable.

Regulation of alternative splicing in atRSp31

Experiments with another *Arabidopsis* SR protein, atRSZ33, had uncovered that this protein regulates splicing of its own pre-mRNA by changing alternative splicing in the similarly

positioned long intron in the middle of the RRM (16). We therefore asked whether *atRSp31* might also regulate its own splicing. Previously, we had shown that recombinant *atRSp31* can stimulate splicing activity of a HeLa cell S100 extract (11). As no plant splicing extract is available, we used the HeLa cell system to test the ability of recombinant *atRSp31* to influence splice site choice in its own pre-mRNA. A construct containing the long intron of *atRSp31* with adjacent exon sequences was used for *in vitro* transcription (Supplementary Figure S1). Obtained pre-mRNA was spliced in a HeLa cell nuclear extract, and the RNA products were characterized by RT-PCR using primers from the adjacent exon regions. A time course of the splicing reaction (Supplementary Figure S1) revealed the production of correctly spliced mRNA but not of alternatively spliced mRNAs. Neither the addition of recombinant *atRSp31* nor of SR protein preparations from *Arabidopsis* (*arabSR*), carrot (*carSR*) or rabbit (*rabSR*) influenced this splicing pattern. These data show that the two splice sites of the long intron leading to the correct mRNA are strong splice sites in an animal splicing system. Therefore, if *atRSp31* would indeed be able to influence splicing of its own pre-mRNA *in vivo*, then more plant-specific splicing factors might be needed for autoregulation. Moreover, these alternative splicing events might be controlled by *cis*-acting sequences not present on this short pre-mRNA construct.

To test these possibilities we transformed *Arabidopsis* cell culture protoplasts transiently with either a genomic or cDNA construct of *atRSp31* containing an N-terminal HA-tag, and analysed splicing by RT-PCR. The RT reaction primer was from the end of the fourth exon, and the same reverse primer was used for the PCRs. To distinguish expression of transgenic or endogenous *atRSp31* transcripts, the direct primer was either to the HA-tag or to the sequence which is disrupted in the transgene by the insertion of the HA-tag, respectively. The upper panel on Figure 2A shows expression of the transgenic cDNA construct (lane 1) and the genomic construct (lane 2) of *atRSp31*. Lanes 3 and 4 are control transformations with empty vector or water, respectively. The middle panel shows the expression of endogenous *atRSp31* in the same samples, and in this case only correctly spliced mRNA was detected. Western blot analysis with an HA-tagged antibody confirmed that a protein is produced from both constructs (Figure 2B, lane 1, cDNA construct; lane 2, genomic construct). The amount of protein is much lower in the case of the genomic construct which conforms to the lower level of mRNA1. As the mRNA patterns of the endogenous transcripts were unchanged by overexpression of *atRSp31*, these experiments show that *atRSp31* does not influence splicing of its own pre-mRNA.

Because *atRSZ33* regulates splicing of a similarly positioned intron of its own pre-mRNA (16), we decided to check whether *atRSZ33* is capable to affect also splicing in the long intron of *atRSp31*. Total RNA from 10 days old wild-type seedlings and seedlings overexpressing *atRSZ33* was isolated and analysed by RT-PCR using primers to the second and fourth exons of *atRSp31*. In wild-type seedlings, mRNA1 coding for the full-length protein is the only splice form detected (Figure 2C, lane 1). Overexpression of *atRSZ33* drastically affected splicing in the long intron of

atRSp31 as mRNA1 is decreased and the production of the alternative splice variants mRNA2 and mRNA3 is greatly enhanced (Figure 2C, lane 2), strongly resembling the splicing pattern of *atRSp31* in wild-type flowers (compare to Figure 1). The effect of *atRSZ33* on alternative splicing of the long intron in *atRSp31* is similar to the one in its own pre-mRNA, suggesting that *atRSZ33* also regulates alternative splicing of *atRSp31*.

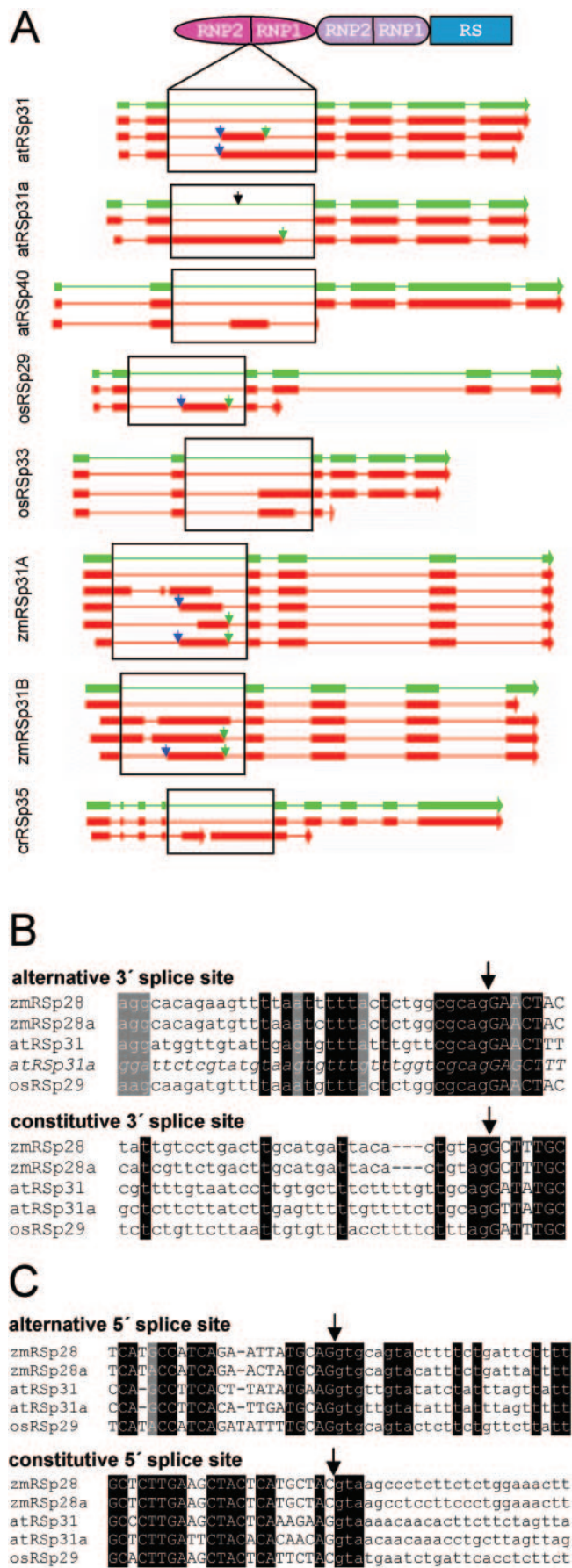
Alternative splicing in the *Arabidopsis* plant-specific RS subfamily

atRSp31 is a member of the plant-specific RS subfamily, which includes also *atRSp31a* (6), *atRSp40* and *atRSp41* (11). *atRSp31/atRSp31a* and *atRSp40/atRSp41* are two pairs of paralogues located at duplicated regions of the *Arabidopsis* genome (6). All four genes contain long introns at the same position between RNP2 and RNP1 of the first RRM. We were interested to analyse whether similar alternative splicing occurs in the other genes of this family as well. To identify potential alternative splicing events in *atRSp31a*, RT-PCR was performed with RNA from 6 days old seedlings, leaves, flowers and roots. Two main products were obtained (data not shown), one of them was the normally spliced mRNA and the other was a splice variant utilizing an alternative 5' splice site in the long intron of *atRSp31a* with the sequence matching the alternative 5' splice site in *atRSp31* (Figure 3A and C). We did not detect a splice form with the alternative 3' splice site in the long intron of *atRSp31a* although this sequence is almost identical to *atRSp31* (Figure 3B). *In silico* search in the available cDNA and EST databases revealed also only the transcript using the alternative 5' splice site (Supplementary Table S1). Because of the high sequence conservation of the 3' alternative splice sites in both genes, we suppose that this splice site might be utilized in tissues and/or conditions different to those used in our analysis. However, we cannot exclude that certain *cis*-elements necessary for the utilization of the alternative 3' splice site are absent in *atRSp31a*.

According to our previous data (11), only *atRSp40* but not *atRSp41* undergoes alternative splicing. In agreement with these data, database searches revealed alternative splice forms involving the long intron of *atRSp40* only (Supplementary Table S1). Similar to *atRSp31*, alternative 3' and 5' splicing occurs in the long intron of *atRSp40* (Figure 3A). However, comparison of the *atRSp40* and *atRSp31/atRSp31a* sequences did not reveal any sequence similarity at the regions of their alternative splice sites. There is no experimental indication for alternative splicing in the long intron of *atRSp41*, and analysis of its intronic sequence did not reveal any potential splice sites similar to those in *atRSp40* or in *atRSp31/atRSp31a*.

Evolutionary conservation of alternative splicing in the RS subfamily

Alternative splicing in most of the genes of the *Arabidopsis* RS subfamily prompted us to examine the splicing profile of their orthologues in other species. As rice genome sequencing is finished, we decided to analyse alternative splicing events in this species. A BLAST search showed that the rice genome encodes only two proteins which belong



to the RS subfamily, Os02g03040 and Os04g02870. In a recent paper also these two proteins from this subfamily were identified, osRSp29 (Os04g02870) and osRSp33 (Os02g03040) (28). Both rice genes contain long introns at the identical positions in the first RRM (Figure 4). Moreover, analysis of available transcript data revealed alternative splicing events similar to those found in *Arabidopsis* (Figure 3A and Supplementary Table S1). For *osRSp29*, we were able to retrieve a single type of differential transcript utilizing alternative 3' and 5' splice sites at the same time. In *osRSp33*, there are two types of splice variants, one which also utilizes alternative 3' and 5' splice sites simultaneously and one which uses only the alternative 3' splice site (Figure 3A and Supplementary Table S1). However, only *osRSp29* revealed a surprisingly high conservation of the sequences adjacent to the alternative 3' and 5' splice sites in the long introns. In comparison, sequence conservation around any constitutive splice site is poor (Figure 3B and C).

Recently, orthologues of atRSp31 in maize were identified and termed zmRSp31A and zmRSp31B. It has been shown that they produce multiple alternatively spliced transcripts (19). We compared the genomic sequences of the maize genes with their orthologues in *Arabidopsis* and rice, and found again the conserved sequences around the alternative 3' and 5' splice sites in the long introns (Figure 3B and C). Analysis performed using Geneseq revealed the splice variants described by (19) as well as some new splice forms (Supplementary Table S1). Interestingly, different combinations of both conserved and non-conserved splice sites are utilized in the long introns of maize genes from RS subfamily (Figure 3A).

As both monocot species possessed the conserved sequences utilized for alternative splicing in the long intron we wanted to investigate how far this feature was conserved in more distant taxa. To find proteins from the RS subfamily we used the atRSp31 protein sequence in BLAST searches versus translated EST databases of gymnosperms (*P.taeda*), bryophytes (*P.patens*) and algae (*C.reinhardtii*) (Materials and Methods).

We have identified five proteins that belong to the RS subfamily in *P.taeda* (Figure 4). In two of these genes, *ptRSp34* and *ptRSpNN*, the intron position is conserved and the intron is alternatively spliced (Supplementary Figure S2 and Table S2). Because we lack genomic sequences for *P.taeda* we do not know if the sequences surrounding these splice sites are still conserved.

Figure 3. Conservation of alternative splicing events in the RS plant-specific subfamily. (A) Gene structures and alternative splicing. Domain structure typical for this subfamily of proteins is shown on the top. Gene structures for each type of protein is shown in red. Long introns are boxed. Blue and green arrows indicate conserved alternative 3' and 5' splice sites, respectively. Black arrow in *atRSp31a* indicates the presence of the conserved sequence of the alternative 3' splice site; however, in this case alternative splicing is not supported by EST/cDNA data. (B) Alignments of the nucleotide sequences around 3' alternative splice site in the long introns and around the 3' splice site of a constitutively spliced exon. Sequence of *atRSp31a* at the alternative 3' splice site is shown in italics to indicate the lack of EST data supporting this splicing event. (C) Alignments of nucleotide sequences around the 5' alternative splice site in the long introns and around the 5' splice site of a constitutively spliced exon. Intronic and exonic sequences are shown in small and capital letters, respectively. Nucleotides identical at all positions are shaded.

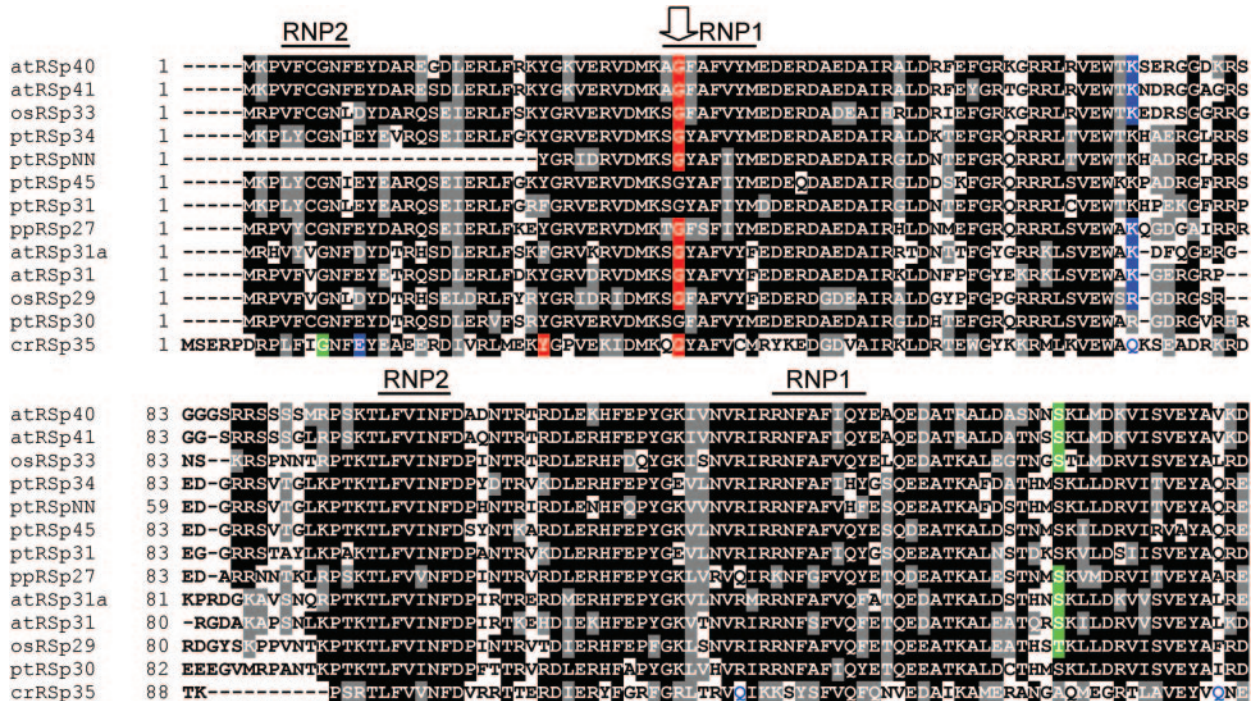


Figure 4. An alignment of protein sequences and intron positions in the RS plant-specific subfamily. Identical or similar amino acids are shaded in black or grey, respectively. RNP2 and RNP1 submotifs of the first and second RRM are overlaid. Positions and phases of introns are indicated by a color code: phase 0, blue; phase 1, red; phase 2, green. *P.taeda* intron positions are indicated only when supporting evidence is present. Position of the conserved alternatively spliced intron between RNP2 and RNP1 is indicated by an arrow. Alignment of the C-termini of proteins covering the RS region is not shown. at, *A.thaliana*; os, *O.sativa*; pt, *P.taeda*; pp, *P.patens*; cr, *C.reinhardtii*.

In *Physcomitrella*, a single protein from the RS subfamily was detected, which we named ppRSp27 (Figure 4). Comparison of the genomic sequence to ESTs revealed the presence of the long intron at the conserved position. However, no ESTs were found supporting alternative splicing in this gene. Alignment of the *Physcomitrella* and *Arabidopsis*/rice/maize genomic sequences also did not show any conserved alternative splice sites in the long intron.

We have also detected a single protein for the RS subfamily in *C.reinhardtii*, crRSp35 (Figure 4 and Supplementary Table S2). Interestingly, in crRSp35, the number and positions of the majority of introns are not conserved. The only introns which have conserved positions and phases in *Chlamydomonas* are the one corresponding to the long intron and the next one separating the RRM1 and RRM2. In addition, most introns in the *Chlamydomonas* gene have symmetrical phases (Figure 4). Nevertheless, existing EST data support alternative 3' splicing in the long intron of crRSp35 (Figure 3A and Supplementary Table S1). However, comparison of the genomic sequence of crRSp35 to the ones from other species shows that sequence at this alternative splice site is not conserved, and no other putative conserved alternative splice sites were detected in the long intron.

These analyses demonstrate that, in the RS subfamily, the presence of an alternatively spliced long intron at the position between RNP2 and RNP1 of the first RRM is highly conserved from green algae to angiosperms. Furthermore, the sequences of particular alternative splice sites in this intron are highly conserved between monocots and dicots (and possibly in gymnosperms) arguing for the evolution of

a conserved alternative splicing mechanism for these genes in higher plant species.

Evolutionary analysis of alternative splicing in the plant specific RS2Z subfamily

The fact that atRSZ33 could equally modulate the conserved alternative splicing in the long intron of atRSp31 as well as in its own pre-mRNA prompted us to investigate the evolutionary conservation of the long introns in the RRM of the plant specific RS2Z subfamily. Proteins of this subfamily are characterized by a single N-terminal RRM, two zinc knuckles, followed by an RS- and SP-rich domain (13). *Arabidopsis* has a paralogous pair of genes, atRSZ33/atRSZ32, which are located on duplicated regions of the *Arabidopsis* genome (6). We have published previously that alternative 3' splicing occurs in the long intron of atRSZ33 situated in the RRM similar to the one in atRSp31 (13,16). Analysing the second RS2Z gene in *Arabidopsis*, atRSZ32, we found that it is also alternatively spliced and produces similar splice variants as atRSZ33. One of them originates from the usage of the 3' alternative splice site in the long intron [Figure 5A and Supplementary Table S1 (accession nos AY095996, BX842118 and BX824622)]. Another splice variant found in both *Arabidopsis* genes is created by the same event in the long intron and the retention of the following intron [Figure 5A and Supplementary Table S1 (accession nos BX823668 and BX825053)].

The BLAST search for proteins of this subfamily in rice revealed four genes: osRSZ36 (Os05g02880), osRSZ37a

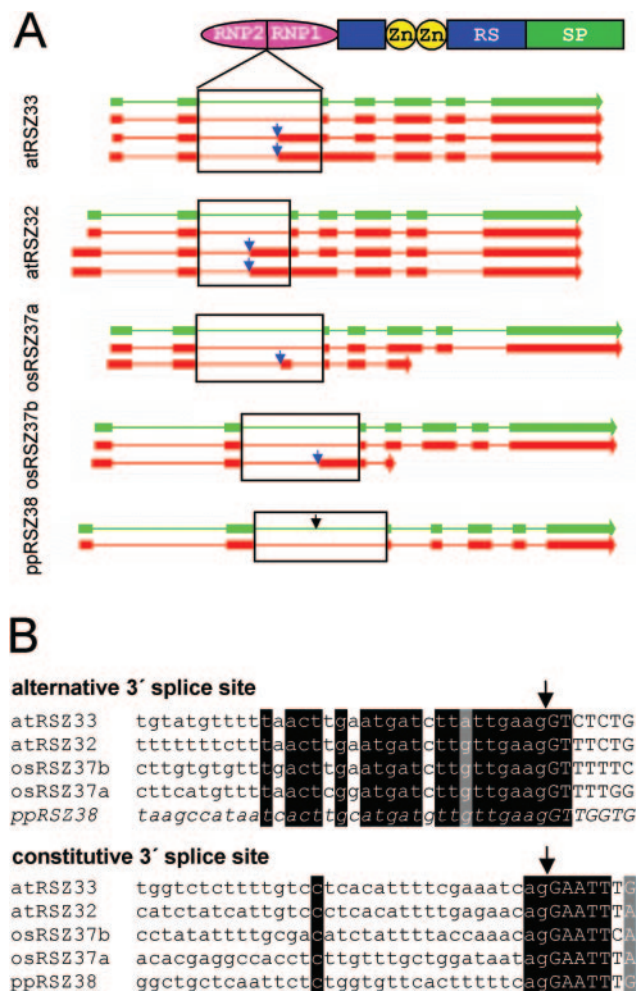


Figure 5. Conservation of alternative splicing events in the RS2Z subfamily (A) Gene structures and alternative splicing. Domain structure typical for this subfamily of proteins is shown on the top. Gene structures are shown in green. ESTs/cDNAs representing each type of splicing are shown in red. Long introns are boxed. Blue arrows indicate the conserved alternative 3' splice sites. The black arrow in *ppRSZ38* indicates the presence of the conserved sequences around the alternative 3' splice site; however, alternative splicing is not supported by EST data. (B) Alignments of nucleotide sequences around the 3' alternative splice site in the long introns and around a 3' constitutive splice site. Sequence of *ppRSZ38* at the alternative 3' splice site is shown in italics to indicate the lack of EST data supporting this splicing event. Intronic and exonic sequences are shown in small and capital letters, respectively. Nucleotides identical at all positions are shaded.

(Os01g06290), *osRSZ37b* (Os03g17710) and *osRSZ39* (Os05g07000) (28). All four rice genes contain long introns at the conserved position (Figure 6) and two of them, *osRSZ37a* and *osRSZ37b*, show alternative splicing in the long intron (Figure 5A and Supplementary Table S1).

Again the alignment of the genomic sequences of the *Arabidopsis* and rice genes within the RS2Z subfamily revealed that sequences around the alternative splice sites are highly conserved (Figure 5B), even more than those within the RS subfamily (compare to Figure 3B). The degree of conservation is much higher than around constitutive splice sites in any of the respective genes (Figure 5B).

Further, we retrieved sequences of RS2Z proteins in gymnosperms, bryophytes and algae in a similar way as for

the RS subfamily. The alignment of proteins which belong to RS2Z subfamily is shown in Figure 6. Intron positions and phases are conserved in all detected genes.

In *P.taeda*, we have found a protein, *ptRSZ35*, which belongs to this subfamily (Figure 6). Comparison of the ESTs revealed alternative splicing in the conserved position between RNP2 and RNP1 (Supplementary Figure S3 and Table S2). We conclude that the position and alternative splicing of this intron in RS2Z subfamily is conserved in angiosperms and gymnosperms; however, as genomic sequences are not available in *P.taeda* we cannot draw any conclusions about the conservation of the sequences preceding the alternative 3' splice site.

In *Physcomitrella*, at least one protein with all the domain features as well as exon-intron structure of the RS2Z subfamily is present, which we named *ppRSZ38* (Figure 6). This protein contains an unusual glycine-rich stretch in front of the RNP2. *ppRSZ38* contains a long intron at the conserved position. Alignment of genomic sequences revealed highly conserved alternative 3' splice site in the long intron of *ppRSZ38* almost identical to the corresponding *Arabidopsis*/rice sequences (Figure 5A and B). However, no EST was found in the available databases which would support alternative splicing in this intron. Whether this splice site is utilized in *Physcomitrella* needs to be proven.

Interestingly, in *C.reinhardtii* we found a single protein with an N-terminal RRM, an SP and RS regions, but no zinc knuckles (data not shown). This protein might therefore not be considered a true member of the RS2Z subfamily.

In summary, the RS2Z subfamily possesses a long intron in a conserved position in the RRM which contains intron sequences highly conserved from mosses to angiosperms. These conserved sequences are used for alternative splicing in monocots and dicots and most likely also in gymnosperms and mosses.

DISCUSSION

The SR protein family from *Arabidopsis* includes three plant-specific subfamilies (RS, RS2Z and SCL) characterized by the presence of an alternatively spliced relatively long intron in their N-terminal RRM (6). Analysis of the alternative splice forms in all three subfamilies shows that they encode truncated putative proteins which would contain only a part of the RRM due to premature termination codons (PTC) generated by the inclusion of intronic sequences (13,14). If produced, these proteins should have no influence on splicing activity as they lack both RNA and protein interaction domains. To investigate the significance of these alternative splicing events we have analysed the regulation of alternative splicing of *atRSp31* from the RS subfamily. We found that its alternative splicing is regulated by *atRSZ33*, a member of the RS2Z subfamily. As this regulation is reminiscent of the autoregulation of the similar long intron in *atRSZ33* (16) we were interested to follow the evolutionary conservation of these splicing events in both plant-specific subfamilies in different lineages.

Evolutionary conservation of splicing profiles

Our analysis shows that positions and phases of introns are conserved in the majority of the orthologous genes in the

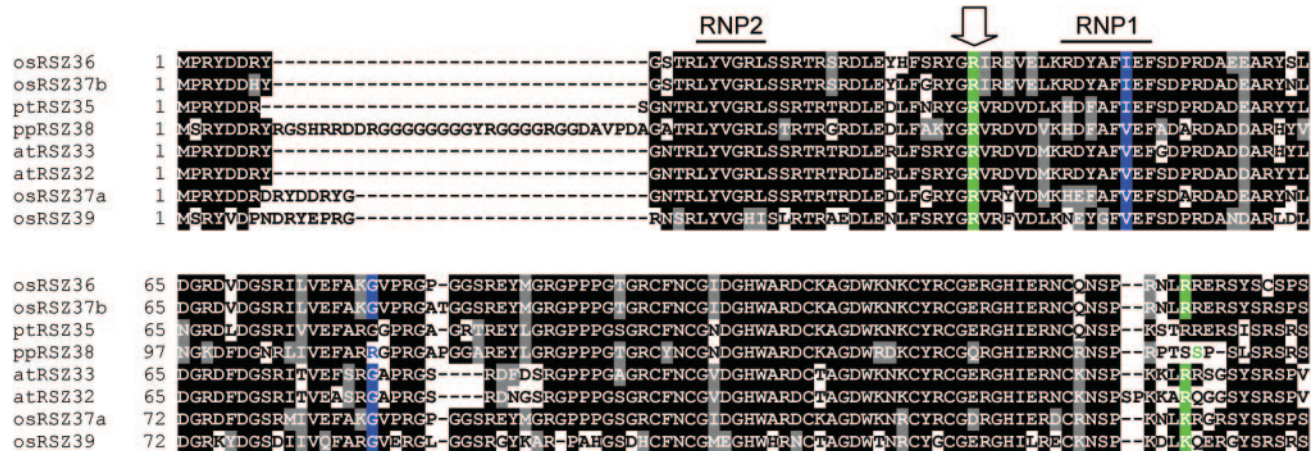


Figure 6. An alignment of protein sequences and intron positions in the RS2Z plant-specific subfamily. Identical or similar amino acids are shaded in black or grey, respectively. RNP2 and RNP1 submotifs of the RRM are overlined. Cysteine and histidine residues of zinc knuckles are underlined. Alignment of the C-termini of the proteins is not shown. Positions of introns are indicated by colour code: phase 0, blue; phase 2, green. *P.taeda* intron positions are indicated only when supporting evidence is present. Position of the conserved alternatively spliced intron between RNP2 and RNP1 is indicated by an arrow. at, *A.thaliana*; os, *O.sativa*; pt, *P.taeda*; pp, *P.patens*.

RS and RS2Z subfamilies. In addition, introns are arranged such that they predominantly separate protein domains. This is in agreement with the hypothesis of a modular assembly of functional or structural domains in the evolution of complex genes (29). The only exception is a *Chlamydomonas* member of RS subfamily, *crRSp35*. Beside two introns at conserved positions in the first RRM, this gene has additional introns, most of them are in symmetrical phases. At present it is not clear whether *Chlamydomonas* has gained the additional introns or if the exon–intron structure of the *Chlamydomonas* gene represents an ancestral arrangement and the other orthologues of the RS subfamily went through loss of introns.

One of the most conserved intron positions in both subfamilies is occupied by the relatively long intron separating RNP2 and RNP1. In addition to its particular length and conserved position, we have observed alternative splicing of these introns in some members of the plant-specific SR subfamilies. The functional significance of alternative transcripts can be best assessed by comparison of splicing profiles of orthologous genes in different species. Interestingly, in both subfamilies alternative splicing was found to be highly conserved in evolution in all species tested from angiosperms to green algae for the RS subfamily, and most likely down to the mosses for the RS2Z subfamily. In addition, a subset of alternative splicing events involves conserved 3' and/or 5' alternative splice sites, which are specific for each subfamily. These alternative splice sites conform to the splice site consensus; however, the surrounding sequences are much more conserved than in any of the constitutive splice sites in the corresponding genes. In the RS subfamily, the conserved intronic sequences are found both in monocots and dicots. In gymnosperms, the lack of genomic sequences does not allow to prove whether the observed alternative splicing events utilize the same conserved sequences. In contrast, the conserved intron sequences are present in the RS2Z family from angiosperms to mosses, although supporting experimental data are still missing for the latter class.

Taken together the evolutionary preservation of these particular alternative splicing events suggests a functional, yet unknown significance. During preparation of this manuscript, an *in silico* analysis of alternative splicing events in plant SR proteins revealed that also the SCL subfamily possess conserved alternative splicing events although no hint for conserved intron sequences was presented (30).

It is worth noting that not all members of the subfamily in a species possess this particular alternative splicing event. For example, in the *Arabidopsis* RS subfamily, two genes, *atRSp31* and *atRSp1a*, utilize these conserved alternative splicing signals, but not *atRSp40* and *atRSp41*. Similarly, both rice orthologues, *osRSp29* and *osRSp33*, are alternatively spliced, but only in *osRSp29* alternative splicing involves the conserved sequences. In maize, both orthologues, *zmRSp31A* and *zmRSp31B*, produce multiple splice variants [(19) and this study], and some of them are generated by the simultaneous use of both conserved and non-conserved alternative splice signals. Similarly, among the four rice genes in the RS2Z subfamily, two have the conserved alternative splicing event. It therefore seems that in a given species selective pressure has preserved conserved alternative splicing events at least in one member of the subfamily. These findings corroborate the significance of these highly conserved alternative splicing events. They might be also a good example for the notion that gene function is more often correlated with splicing profile similarity between orthologues in different species than with sequence similarity of paralogues in the same species (31).

Functional significance and regulation of the alternative splicing events

Alternative splicing is an important mechanism for generation of proteome diversity and regulation of gene expression at the post-transcriptional level. In both RS and RS2Z subfamilies, all conserved alternative splicing transcripts potentially encode extremely truncated proteins containing only a

part of the RRM and some sequences from the included introns due to the generation of PTC. Our current data from the analysis of *atRSp31* suggest that no truncated protein is made from these variant transcripts. Similarly, no truncated protein could be detected for *atGRP7* which also contains an alternatively spliced long intron in the RRM domain (32). For maize orthologues of the RS subfamily, it has been shown that PTC-containing alternatively spliced transcripts associate with polysomes, but there is no evidence of their translation (19). Even if the alternative splice forms are translated, these short proteins will certainly have no splicing activity as they lack a complete RRM and RS domains. These alternative splicing events are different from those leading to the production of diverse protein isoforms. For example, alternative splicing events in the *Arabidopsis* homologues of human ASF/SF2 occur in long introns located in the 3' ends of genes. Here, splice variants encode protein isoforms with a shortened RS domains (12,15), which might affect the phosphorylation status of the protein and/or its ability to interact with other proteins.

Generally, alternative splicing events preserving the reading frame are more conserved than those leading to frame-shifts. It has been shown that only 5% of cassette alternative exons are both conserved and have the potential to introduce PTC (33). It is therefore very unlikely that the evolutionary conserved alternative splicing events seen in the plant SR protein subfamilies have occurred just by chance as aberrant splicing events, as they are preserved in different subfamilies of genes and across distant taxa.

Given that no or only inactive proteins are made from the alternative transcripts, what functional significance could these conserved alternative splicing events in the long introns of the plant specific SR proteins have? Either the alternative splicing event is crucial for modulating the level of the splicing factor or the alternative transcript itself encodes an yet unknown activity. Although there is no evidence for the latter case, it has been shown previously that the presence of PTC in splice isoforms can promote nonsense-mediated decay (NMD) of such transcripts, and some splicing factors are subjected to such regulation (34,35). However, it has been shown recently by quantitative microarray profiling that NMD does not significantly influence the overall steady-state levels of transcripts with PTC generated by alternative splicing events (36). The alternative transcripts of *atRSp31* are readily detectable by northern blotting and in some tissues they represent the only type of transcripts (26). It is therefore unlikely that the variant transcripts of *atRSp31* are controlled by NMD, however, experimental proof is required.

Tight control of the *atRSZ33* protein levels by autoregulation of alternative splicing has been demonstrated previously (16). Autoregulation seems to be conserved in the RS2Z family as the rice homologue, *osRSZ36*, has also been shown to cause changes in its splicing pattern (28). Autoregulatory circuits have been shown for several non-plant splicing factors, such as *SXL* (37), *TRA-2* (38), *SWAP* (39) and *SRp20* (40). On the other hand, we have demonstrated here that overexpression of *atRSp31*, either *in vitro* or *in vivo*, does not stimulate usage of the alternative splice sites. Interestingly, also in rice none of the two homologues of *atRSp31*, *osRSp29* and *osRSp33*, can change the splicing pattern of their pre-mRNA (28). Together with this, we show that *atRSZ33* is involved in

the regulation of alternative splicing in the long intron of *atRSp31*. However, *atRSZ33* does not interact with *atRSp31* in pull-down assays (13) which would imply that regulation of splicing in the long intron of *atRSp31* by *atRSZ33* occurs without direct interaction of these two proteins.

How and why are these alternative splicing events regulated? It has been noted that similar to animals, many of the alternatively spliced genes in *Arabidopsis* encode proteins with regulatory functions and many stress related genes are alternatively spliced (41). A genome wide analysis in *A.thaliana* uncovered many alternative splicing events in splicing factors and several of them were strongly influenced by cold stress (42). This is especially interesting as these alternatively spliced splicing regulators might coordinate alternative splicing of a particular set of stress response genes. Our unpublished data indicate that expression of *atRSp31* is tightly regulated both transcriptionally and post-transcriptionally during plant development and in response to hormones, sugars and different light conditions. It would be extremely interesting to find out whether the highly conserved alternative transcripts of plant-specific SR genes have a specific role in development and/or environmental responses. Preservation of alternative splicing events that have often been attributed to accidents or erroneous action of splicing machinery argues for basic plant cell specific regulatory circuits established early in the plant evolution.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Zdravko Lorkovic and Arndt von Haeseler for helpful discussions and tree construction. This work was supported by Austrian SFB 1711 and 1712 and Vienna Science and Technology Fund LS 123 to AB. Funding to pay the Open Access publication charges for this article was provided by the Austrian Science Foundation.

Conflict of interest statement. None declared.

REFERENCES

1. Jurica, M.S. and Moore, M.J. (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell*, **12**, 5–14.
2. Fu, X.-D. (1995) The superfamily of arginine/serine-rich splicing factors. *RNA*, **1**, 663–680.
3. Graveley, B.R. (2000) Sorting out the complexity of SR protein functions. *RNA*, **6**, 1197–1211.
4. Smith, C.W. and Valcarcel, J. (2000) Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci.*, **25**, 381–388.
5. Lorkovic, Z.J. and Barta, A. (2002) Genome analysis: RNA recognition motif (RRM) and K homology (KH) domain RNA-binding proteins from the flowering plant *Arabidopsis thaliana*. *Nucleic Acids Res.*, **30**, 623–635.
6. Kalyna, M. and Barta, A. (2004) A plethora of plant serine/arginine-rich proteins: redundancy or evolution of novel gene functions? *Biochem. Soc. Trans.*, **32**, 561–564.
7. Ge, H., Zuo, P. and Manley, J.L. (1991) Primary structure of the human splicing factor ASF reveals similarities with Drosophila regulators. *Cell*, **66**, 373–382.
8. Sreaton, G.R., Caceres, J.F., Mayeda, A., Bell, M.V., Plebanski, M., Jackson, D.G., Bell, J.I. and Krainer, A.R. (1995) Identification and

- characterization of three members of the human SR family of pre-mRNA splicing factors. *EMBO J.*, **14**, 4336–4349.
9. Jumaa,H., Guenet,J.L. and Nielsen,P.J. (1997) Regulated expression and RNA processing of transcripts from the Srp20 splicing factor gene during the cell cycle. *Mol. Cell. Biol.*, **17**, 3116–3124.
 10. Lejeune,F., Cavaloc,Y. and Stevenin,J. (2001) Alternative splicing of intron 3 of the serine/arginine-rich protein 9G8 gene. Identification of flanking exonic splicing enhancers and involvement of 9G8 as a trans-acting factor. *J. Biol. Chem.*, **276**, 7850–7858.
 11. Lopato,S., Waigmann,E. and Barta,A. (1996) Characterization of a novel arginine/serine-rich splicing factor in Arabidopsis. *Plant Cell*, **8**, 2255–2264.
 12. Lopato,S., Kalyna,M., Dorner,S., Kobayashi,R., Krainer,A.R. and Barta,A. (1999) atSRp30, one of two SF2/ASF-like proteins from *Arabidopsis thaliana*, regulates splicing of specific plant genes. *Genes Dev.*, **13**, 987–1001.
 13. 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.
 14. Golovkin,M. and Reddy,A.S. (1999) An SC35-like protein and a novel serine/arginine-rich protein interact with *Arabidopsis* U1-70K protein. *J. Biol. Chem.*, **274**, 36428–36438.
 15. Lazar,G. and Goodman,H.M. (2000) The *Arabidopsis* splicing factor SR1 is regulated by alternative splicing. *Plant Mol. Biol.*, **42**, 571–581.
 16. Kalyna,M., Lopato,S. and Barta,A. (2003) Ectopic expression of atRSZ33 reveals its function in splicing and causes pleiotropic changes in development. *Mol. Biol. Cell*, **14**, 3565–3577.
 17. Lorkovic,Z.J., Wicczorek Kirk,D.A., Lambermon,M.H. and Filipowicz,W. (2000) Pre-mRNA splicing in higher plants. *Trends Plant Sci.*, **5**, 160–167.
 18. Usuka,J., Zhu,W. and Brendel,V. (2000) Optimal spliced alignment of homologous cDNA to a genomic DNA template. *Bioinformatics*, **16**, 203–211.
 19. Gupta,S., Wang,B.B., Stryker,G.A., Zanetti,M.E. and Lal,S.K. (2005) Two novel arginine/serine (SR) proteins in maize are differentially spliced and utilize non-canonical splice sites. *Biochim. Biophys. Acta*, **1728**, 105–114.
 20. Huang,X. (1996) An improved sequence assembly program. *Genomics*, **33**, 21–31.
 21. Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
 22. Valvekens,D., Van Montague,M. and Van Lijsbetbens,M. (1988) Agrobacterium tumefaciens-mediated transformation of *Arabidopsis thaliana* root explants using kanamycin selection. *Proc. Natl Acad. Sci. USA*, **85**, 5536–5540.
 23. Anthony,R.G., Henriques,R., Helfer,A., Meszaros,T., Rios,G., Testerink,C., Munnik,T., Deak,M., Koncz,C. and Bogre,L. (2004) A protein kinase target of a PDK1 signalling pathway is involved in root hair growth in *Arabidopsis*. *EMBO J.*, **23**, 572–581.
 24. In Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (eds) 1997, *The Polymerase Chain Reaction*. John Wiley & Sons, Inc., NY.
 25. Sablowski,R.W. and Meyerowitz,E.M. (1998) A homolog of NO APICAL MERISTEM is an immediate target of the floral homeotic genes APETALA3/PISTILLATA. *Cell*, **92**, 93–103.
 26. Lopato,S., Mayeda,A., Krainer,A.R. and Barta,A. (1996) Pre-mRNA splicing in plants: characterization of Ser/Arg splicing factors. *Proc. Natl Acad. Sci. USA*, **93**, 3074–3079.
 27. Lambermon,M.H., Simpson,G.G., Wicczorek Kirk,D.A., Hemmings-Mieszczak,M., Klahre,U. and Filipowicz,W. (2000) UBP1, a novel hnRNP-like protein that functions at multiple steps of higher plant nuclear pre-mRNA maturation. *EMBO J.*, **19**, 1638–1649.
 28. Isshiki,M., Tsumoto,A. and Shimamoto,K. (2006) The Serine/Arginine-rich protein family in rice plays important roles in constitutive and alternative splicing of pre-mRNA. *Plant Cell*, **18**, 146–158.
 29. Roy,S.W. and Gilbert,W. (2006) The evolution of spliceosomal introns: patterns, puzzles and progress. *Nature Rev. Genet.*, **7**, 211–221.
 30. Iida,K. and Go,M. (2006) Survey of conserved alternative splicing events of mRNAs encoding SR proteins in land plants. *Mol. Biol. Evol.*, **23**, 1085–1094.
 31. Västermark,Å., Shigemoto,Y., Abe,T. and Sugawara,H. (2004) Splicing profile based protein categorization between human and mouse genomes by use of the DDBJ web services. *Genome Inform.*, **15**, 13–20.
 32. Staiger,D., Zecca,L., Wicczorek Kirk,D.A., Apel,K. and Eckstein,L. (2003) The circadian clock regulated RNA-binding protein AtGRP7 autoregulates its expression by influencing alternative splicing of its own pre-mRNA. *Plant J.*, **33**, 361–371.
 33. Baek,D. and Green,P. (2005) Sequence conservation, relative isoform frequencies, and nonsense-mediated decay in evolutionarily conserved alternative splicing. *Proc. Natl Acad. Sci. USA*, **102**, 12813–12818.
 34. Morrison,M., Harris,K.S. and Roth,M.B. (1997) smg mutants affect the expression of alternatively spliced SR protein mRNAs in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA*, **94**, 9782–9785.
 35. Wollerton,M.C., Gooding,C., Wagner,E.J., Garcia-Blanco,M.A. and Smith,C.W. (2004) Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Mol. Cell*, **13**, 91–100.
 36. Pan,Q., Saltzman,A.L., Kim,Y.K., Misquitta,C., Shai,O., Maquat,L.E., Frey,B.J. and Blencowe,B.J. (2006) Quantitative microarray profiling provides evidence against widespread coupling of alternative splicing with nonsense-mediated mRNA decay to control gene expression. *Genes Dev.*, **20**, 153–158.
 37. Bell,L.R., Horabin,J.I., Schedl,P. and Cline,T.W. (1991) Positive autoregulation of sex-lethal by alternative splicing maintains the female determined state in *Drosophila*. *Cell*, **65**, 229–239.
 38. Mattox,W. and Baker,B.S. (1991) Autoregulation of the splicing of transcripts from the transformer-2 gene of *Drosophila*. *Genes Dev.*, **5**, 786–796.
 39. Zachar,Z., Chou,T.B. and Bingham,P.M. (1987) Evidence that a regulatory gene autoregulates splicing of its transcript. *EMBO J.*, **6**, 4105–4111.
 40. Jumaa,H. and Nielsen,P.J. (1997) The splicing factor SRp20 modifies splicing of its own mRNA and ASF/SF2 antagonizes this regulation. *EMBO J.*, **16**, 5077–5085.
 41. Kazan,K. (2003) Alternative splicing and proteome diversity in plants: the tip of the iceberg has just emerged. *Trends Plant Sci.*, **8**, 468–471.
 42. Iida,K., Seki,M., Sakurai,T., Satou,M., Akiyama,K., Toyoda,T., Konagaya,A. and Shinozaki,K. (2004) Genome-wide analysis of alternative pre-mRNA splicing in *Arabidopsis thaliana* based on full-length cDNA sequences. *Nucleic Acids Res.*, **32**, 5096–5103.