Identification of an intronic splicing regulatory element involved in auto-regulation of alternative splicing of *SCL33* pre-mRNA

Julie Thomas¹, Saiprasad G. Palusa¹, Kasavajhala V.S.K. Prasad¹,², Gul Shad Ali¹,³, Giridara-Kumar Surabhi¹, Asa Ben-Hur², Salah E. Abdel-Ghany¹ and Anireddy S.N. Reddy¹,*

¹Department of Biology, Program in Molecular Plant Biology, Program in Cell and Molecular Biology, Colorado State University, Fort Collins, CO 80523, USA, and
²Department of Computer Science and Program in Molecular Plant Biology, Colorado State University, Fort Collins, CO 80523, USA

Received 21 March 2012; Revised 11 August 2012; Accepted 20 August 2012; published online 29 October 2012.

*For correspondence (e-mail reddy@colostate.edu).

Present address: Institute for Genome Sciences and Policy, Department of Biology, Duke University, Durham, NC 27708, USA.

Present address: Mid-Florida Research & Education Center, Apopka, FL 32703, USA, and Department of Plant Pathology, Institute of Food & Agricultural Sciences, University of Florida, Gainesville, FL 32611, USA.

SUMMARY

In Arabidopsis, pre-mRNAs of serine/arginine-rich (SR) proteins undergo extensive alternative splicing (AS). However, little is known about the *cis*-elements and *trans*-acting proteins involved in regulating AS. Using a splicing reporter (GFP–intron–GFP), consisting of the GFP coding sequence interrupted by an alternatively spliced intron of *SCL33*, we investigated whether *cis*-elements within this intron are sufficient for AS, and which SR proteins are necessary for regulated AS. Expression of the splicing reporter in protoplasts faithfully produced all splice variants from the intron, suggesting that *cis*-elements required for AS reside within the intron. To determine which SR proteins are responsible for AS, the splicing pattern of the GFP–intron–GFP reporter was investigated in protoplasts of three single and three double mutants of SR genes. These analyses revealed that *SCL33* and a closely related paralog, *SCL30a*, are functionally redundant in generating specific splice variants from this intron. Furthermore, SCL33 protein bound to a conserved sequence in this intron, indicating auto-regulation of AS. Mutations in four GAAG repeats within the conserved region impaired generation of the same splice variants that are affected in the *scl33 scl30a* double mutant. In conclusion, we have identified the first intronic *cis*-element involved in AS of a plant SR gene, and elucidated a mechanism for auto-regulation of AS of this intron.

Keywords: Arabidopsis thaliana, alternative splicing, plants, pre-mRNA splicing, splicing regulatory elements, SR proteins, SCL33.

INTRODUCTION

Alternative splicing (AS), a mechanism for generating multiple transcripts from a single gene, contributes to transcriptome and proteome diversity (Kalsotra and Cooper, 2011). Splice variants of a gene may encode structurally and functionally different proteins that play important roles in an organism’s growth, development and diseases (Reddy, 2007; Kalsotra and Cooper, 2011). Recent genome-wide transcriptome sequencing (RNA-Seq) studies using next-generation sequencing indicate that AS is widespread in both animals and plants. Pre-mRNAs from approximately 60% of multi-exon genes in Arabidopsis (Filichkin *et al.*, 2010; Marquez *et al.*, 2012) and approximately 48% of multi-exon genes in rice (Lu *et al.*, 2010) undergo AS. In humans, pre-mRNAs from approximately 95% of intron-containing genes undergo AS, with specific isoforms in different tissue types (Pan *et al.*, 2008; Wang *et al.*, 2008). Mutations in *cis*-acting elements in RNA or *trans*-acting splicing factors cause mis-regulation of splicing, leading to numerous diseases in humans (Garcia-Blanco *et al.*, 2004; Kalsotra and Cooper, 2011). In plants, AS plays important roles in regulating several developmental processes and biotic and abiotic stress responses (reviewed by Ali and Reddy, 2008; Duque, 2011; Gassmann, 2008; Reddy, 2007; Reddy and Ali, 2011).
Although AS is ubiquitous in all multicellular organisms, the frequency of the various types of AS events differs between plants and animals. In plants, intron retention is the most prevalent AS event, whereas exon skipping is the most common in vertebrates (Reddy et al., 2012). The core splicing signals present at the exon–intron boundary (5′ splice site (5′SS)), the intron–exon boundary (3′ splice site (3′SS)), the polypyrimidine tract and the branch point sequence (BPS) are important for spliceosome assembly. Although there is significant conservation of these core elements across organisms, they alone are insufficient for constitutive splicing and AS. Other sequence elements in the pre-mRNA called splicing regulatory elements (SREs) bind to trans-acting splicing regulatory proteins. SREs, which are found in exons (exonic splicing enhancers/silencers) or in introns (ISE/ISS; intronic splicing enhancers/silencers), play critical roles in both constitutive splicing and AS (Chasin, 2007; Reddy, 2007; Day et al., 2012). The SREs, which are often short stretches of nucleotides (6–10 nt) (Xiao et al., 2007) function by recruiting trans-acting splicing factors that activate or suppress splice site recognition or spliceosome assembly (Chasin, 2007; Barash et al., 2010). These sequence elements, which are generally found in clusters or spaced at regular intervals, influence splice site choice through specific binding of splicing regulatory proteins such as serine/arginine-rich (SR) proteins or heterogeneous nuclear ribonucleoproteins (Long and Caceres, 2009).

In vertebrates, whose genes possess long introns, exon and intron specification is thought to occur through ‘exon definition’, which involves interaction of spliceosomal components with the downstream 5′SS and upstream 3′SS across the exon (Berget, 1995). In the alternative ‘intron definition’ model, which is thought to occur in organisms with short introns, interactions of spliceosomal proteins occur across the intron between factors recognizing the upstream 5′SS and the downstream 3′SS (Berget, 1995). Based on the presence of short introns and the high frequency of intron retention in plants (56% in Arabidopsis and 53.5% in rice, compared to 5% in humans), it is proposed that splice site recognition occurs predominantly by intron definition (Reddy et al., 2012). Early research on pre-mRNA splicing in plants showed that AU-rich or U-rich sequences, which are enriched in plant introns, are required for splicing (Filipowicz et al., 1995; Reddy, 2001; Schuler, 2008). A highly conserved putative AU-rich splicing regulatory cis-acting element identified in the gene encoding chloroplast-specific ascorbate peroxidase isozyme represents a plant cis-acting element that modulates tissue-specific AS (Yoshimura et al., 2002). The importance of the GC content in exons for efficient splicing (Carle-Urioste et al., 1997) and of an AG-rich exonic element capable of promoting downstream 5′SS selection have also been reported (McCullough and Schuler, 1997). Using computational tools, a number of putative hexameric exonic splicing enhancers were identified in Arabidopsis (Pertea et al., 2007). In animal systems, many SREs have been experimentally identified that bind splicing factors (Le Guiner et al., 2001; Oberstrass et al., 2005; Chasin, 2007; Fukumura et al., 2007; Jelen et al., 2007). A splicing code was assembled based on hundreds of known features involved in AS to predict exon-skipping events in animals (Barash et al., 2010; Rose et al., 2011). However, in plants, aside from a global analysis of gene structure and composition and mutational analysis of splice sites, there has been little experimental or computational analysis to uncover SREs (Isshiki et al., 2006; Reddy et al., 2012).

Serine/arginine-rich (SR) proteins are master regulators of constitutive splicing and AS, each probably regulating the splicing of hundreds to thousands of pre-mRNAs (Long and Caceres, 2009; Reddy and Ali, 2011). SR proteins regulate splicing by binding SREs through their N-terminal RNA recognition motif (RRM) domains that mediate RNA–protein interactions, and facilitating spliceosome assembly through their C-terminal arginine/serine-rich (RS) domains, which participate in protein–protein interactions and in some cases interact with RNA (Zahler et al., 1992; Caceres and Krainer, 1993; Le Guiner et al., 2001; Long and Caceres, 2009; Reddy and Ali, 2011). Analysis of pre-mRNA splicing of 18 Arabidopsis SR genes revealed extensive AS (Palusa et al., 2007). Remarkably, over 90 transcripts are produced from pre-mRNAs of 14 SR genes, representing a more than fivefold increase in the SR transcriptome, and many splice variants are the targets of nonsense-mediated decay (Palusa et al., 2007; Palusa and Reddy, 2010). Most of the AS variants produced from SR pre-mRNAs are generated by AS of introns, with the frequency of occurrence of AS events highest in the longest introns (Palusa et al., 2007). Most significantly, AS of some SR genes is controlled in a developmental and tissue-specific manner and altered in response to various stresses, suggesting that regulation of AS is probably important for development and stress responses (Reddy and Ali, 2011). Ectopic expression of RS2223 was shown to alter AS of its own pre-mRNA and other SR genes (atSRp30 and atSRp34), with pleiotropic changes in plant development (Kalyna et al., 2003). Over-expression of SR30 has also been shown to alter alternative splicing of its own pre-mRNA and that of other SR pre-mRNAs (Lopato et al., 1999). A loss-of-function mutant (sr45-1) of SR45, an SR-like gene, exhibited multiple developmental abnormalities (e.g. delayed flowering, reduced root growth, narrow leaves and altered number of petals and stamens), increased sensitivity to glucose and abscisic acid, and altered splicing patterns of several SR genes (Ali et al., 2007; Zhang and Mount, 2009; Carvalho et al., 2010). Interestingly, the long splice variant of SR45 complemented the flower petal phenotype, whereas the short isoform complemented the root growth phenotype (Zhang and Mount, 2009).

The lack of an in vitro splicing system derived from plant cells has hampered progress in understanding regulated
splicing in plants compared to animal systems. As pre-mRNAs of plant SR genes show extensive AS, they are excellent candidates to elucidate the mechanism(s) involved in regulated AS in plants. The pre-mRNAs of an Arabidopsis SR gene, SCL33, undergo AS and produce at least nine splice variants, eight of which are generated due to AS of the third intron (Palusa and Reddy, 2010, and this work). To determine whether the sequence elements within the intron are sufficient for regulation of AS, we developed an in vivo splicing assay using a splicing reporter in protoplasts. To identify which SR proteins are involved in regulated splicing of this intron, we analysed AS of the splicing reporter in protoplasts from three single and three double mutants of Arabidopsis SR genes. The results presented here show that all the signals necessary for AS of the third intron of SCL33 are present within the intron. We also show that two related SR proteins (SCL33 and SCL30a) are functionally redundant for the production of specific isoforms. Furthermore, using RNA binding studies, we identified a 92 nt region with multiple GAAG repeats in the SCL33 intron that binds to SCL33 protein, suggesting auto-regulation of AS of SCL33. Mutational analysis of the GAAG repeats confirmed the importance of these elements in AS.

RESULTS

Signals for alternative splicing of the SCL33 intron reside within the intron

We have previously shown that pre-mRNA from SCL33 undergoes AS in various tissues, including leaves, and all the splice variants except one are generated by intron retention, alternative 3’SS selection or by using both alternative 3’ and 5’SS (Palusa et al., 2007). The SCL33 third intron was chosen as a model to study AS regulation and identify splicing factors that regulate AS events using an in vivo protoplast system. We first tested whether the AS pattern of SCL33 in leaves is identical in mesophyll protoplasts by RT-PCR using RNA from leaves and protoplasts. As shown in Figure S1, the splicing pattern of SCL33 in protoplasts is similar to leaves, indicating that the protoplast system was suitable for studying the regulation of AS of this gene. To identify the splice sites used in producing these splice variants, cDNA prepared from protoplasts was used to amplify all splice variants from the SCL33 third intron, which were then sequenced. The sequence of all eight splice variants is presented in Figure S2.

To determine whether the signals necessary for AS of the third intron reside within the intron, we developed a splicing reporter construct that can be used to monitor AS of this intron in vivo using a protoplast system. The splicing reporter construct was produced by cloning the entire third intron of SCL33 (765 bp) into the coding region of GFP to generate a GFP–SCL33 intron–GFP (GFP–INT–GFP) construct (Figure 1bii). Splicing of the intron results in GFP fluorescence. Furthermore, the production of splice variants can be monitored by RT-PCR using GFP-specific primers. We transfected protoplasts with this construct or uninterrupted...
GFP (Figure 1ai) as a control. Expression of the splicing reporter in wild-type protoplasts showed GFP fluorescence (Figure 1bii), but less than the uninterrupted GFP (Figure 1aii), suggesting that the intron is excised properly from the splicing reporter.

We then determined whether the intron is alternatively spliced as in the endogenous gene, and whether the AS events and the sites used to generate the splice variants are identical to the native gene. To address these questions, we isolated RNA from protoplasts expressing the GFP and GFP–INT–GFP constructs, and performed RT-PCR using GFP-specific forward and reverse primers. As expected, a single product was seen in GFP-transfected protoplasts (Figure 1aiii). In GFP–INT–GFP-transfected protoplasts, all expected eight isoforms of the native SCL33 gene were obtained (Figure 1biii). To determine whether the splice variants are identical to those produced from the native gene, we cloned and sequenced all splice variants from the reporter gene (Figure S2). Alignment of isoforms generated from intron 3 in the native SCL33 gene with those generated from GFP–INT–GFP has revealed that all eight isoforms are identical to the splice forms from the endogenous gene (Figure 1biii, schematic diagram, and Figure S2). The same isoforms were obtained from leaves of stable transgenic lines expressing the GFP–INT–GFP construct, validating use of the transient expression system for AS. Isoform 8, the largest one, is produced by intron retention, and isoform 1 is the GFP product produced by complete removal of the intron. Three isoforms (5–7) are generated by alternative 5’SS selection, and isoforms 2–4 are produced by using both alternative 5’ and 3’SS. Among the eight forms, three isoforms (3, 4 and 6) share the same 5’SS for splicing of the second part of the intron, but have different 3’SS for splicing of the first part of the intron (Figure 1biii). These results demonstrate experimentally that the third intron of SCL33 has all the necessary signals to faithfully undergo AS, and suggest that sequences in other parts of SCL33 (exons or other introns) are not required for AS of this intron.

SCL33 protein binds to a 92 nt segment of the SCL33 intronic RNA

Some SR proteins in animals and plants are known to auto-regulate pre-mRNA splicing (Long and Caceres, 2009; Reddy and Ali, 2011). However, in plants, binding regions in native pre-mRNAs have not been identified for any SR proteins. To investigate whether the SCL33 protein interacts directly with intron 3 of SCL33, we performed electrophoretic mobility shift assays (EMSA) using various regions of labelled SCL33 intron (Figure 2a) and purified SCL33 protein. As shown in Figure 2(b), the 5’ side of the intron (P1) did not bind to the SCL33 protein, whereas the position of the P2 RNA shifted when recombinant SCL33 was added. The formation of RNA–protein complexes increased with increasing concentrations of purified protein. This result suggests the presence of at least one binding region for SCL33 in the 3’ segment (P2) of the intron. To map the binding region in P2, labelled RNA from two shorter fragments (P3 and P4) was

Figure 2. SCL33 protein binds to a specific region of the SCL33 intron. (a) Schematic diagram of various regions of the third intron of SCL33 that were used to generate RNA probes for EMSA. P1–P4 represent different parts of intron 3 as indicated. The number indicates the start and end nucleotide position for each probe relative to the intron 5’SS. Arrows and arrowheads show the 5’ and 3’SS of the various isoforms generated by AS, respectively. (b) EMSA with P1, P2, P3 and P4 RNA probes using purified SCL33 protein. Lane 1, free probe; lanes 2–5 have an increasing concentration of SCL33 (60, 120, 180 and 300 ng). Arrows indicate free probe, and the RNA–protein complexes are indicated by an arrowhead. (c) binding of SCL33 to P4 is competed for by cold P4. Lane 1, free probe; lane 2, probe + SCL33 protein (300 ng); lanes 3–7, as lane 2, with n increasing concentration of cold P4 (10 ×, 20 ×, 30 ×, 40 × and 50 ×).
used in an EMSA. Both RNAs bound to SCL33, and the extent of binding increased with an increase in protein concentration. As P4 is the smallest fragment (92 nt), we conclude that it contains a binding site for SCL33. To determine the specificity of P4 binding to SCL33, we performed a competition assay with an increasing amount of cold P4 RNA. As shown in Figure 2(c), protein–RNA complex formation was observed between the SCL33 protein and P4 RNA (lane 2), whereas addition of increasing concentrations of cold competitor RNA reduced binding. At a concentration of 50 × (lane 7), the competitor RNA completely abolished P4 binding, indicating that the interaction between SCL33 and this intron segment is specific. To further demonstrate the specificity of the SCL33 interaction with P4, we performed EMSA using P4 RNA and purified SR45, an SR-like protein (Golovkin and Reddy, 1999). The P4 RNA showed no binding to SR45 (Figure S3A). To further confirm the binding specificity between P4 and SCL33, we performed a competition assay in which increasing concentrations of cold P1 RNA were added to the labelled SCL33–P4 complex. Addition of excess P1 (50 ×) did not reduce the amount of SCL33–P4 complex (Figure S3B). Together, these results establish that SCL33 binds specifically to a 92 nt region in the SCL33 third intron, and suggest that SCL33 may auto-regulate AS by binding its own intronic RNA.

SCL30a is the closest paralog of SCL33 (Richardson et al., 2011), the third intron of SCL30a also undergoes AS, and the location of AS sites in some isoforms is similar to that of SCL33 (Palusa et al., 2007). To determine whether there is any sequence conservation between the third intron of SCL33 and that of SCL30a, we aligned the nucleotide sequences of these introns using TCOFFEE (http://tcoffee.vital-it.ch/). Interestingly, the 92 nt region bound to SCL33 is highly conserved (90% identity) between SCL33 and SCL30a, and all the alternative splice sites in both genes are near or within the conserved 92 nt region (see Figure 3). In addition, we found that four closely spaced purine-rich GAAG repeats, which are known exonic splicing regulatory elements in plants and animals (Chasin, 2007; Pertea et al., 2007), are found in the third intron of both
genes, suggesting that they may be important for SCL33 interaction and AS.

**SCL33 and SCL30a are functionally redundant in regulation of AS of SCL33 intron 3**

SR proteins in animals are known to regulate both constitutive splicing and AS by binding to SREs and recruiting spliceosomal components (Long and Caceres, 2009). Although many pre-mRNAs of plant SR genes are alternatively spliced (Palusa et al., 2007; Reddy and Ali, 2011), little is known about the trans-acting splicing factors that regulate AS in these genes. To identify the SR proteins that may regulate AS of SCL33, we used a genetic approach to address the role of three of the 18 Arabidopsis SR genes (Barta et al., 2010) in AS of SCL33. We identified loss-of-function T-DNA insertion mutants of scl33, sc35 and scl30a, with insertions in the exon of each gene (Figure 4a), and generated three double mutants (scl33 scl30a, scl33 sc35 and scl33 sc30a), as SR proteins, especially the closely related ones, have redundant functions. We confirmed homozygosity of the mutants by genomic PCR (Figure 4b) using primers for the genes and the T-DNA insert. The homozygous lines showed no expression of transcripts corresponding to the mutated gene(s) in single and double mutants (Figure 4c), suggesting that these mutants are complete knockouts. As the T-DNA insertion in the scl33 mutant is in the last exon, we performed RT-PCR using primers corresponding to exons flanking intron 3, which lie...
upstream of the T-DNA insertion site, to determine whether any splice variants are produced in the mutant. No splice variants were detected using these primers, confirming that scl33 is a complete loss-of-function mutant (Figure S4). We then used protoplasts from all the mutant lines to monitor AS of the splicing reporter to determine the functions of these SR proteins in AS.

Protoplasts from wild-type and three single and double mutants were transfected with the GFP–INT–GFP construct and splicing was analysed by RT-PCR. Splicing of this reporter in five of the six mutants was similar to that of wild-type. Only the scl33 scl30a double mutant showed altered splicing, with two isoforms (3 and 6) missing (Figure 5a). These results indicate that the SCL33 and SCL30a genes are functionally redundant, and generation of all splice variants from the SCL33 intron occurs in the presence of either SCL33 or SCL30a but not in the absence of both. As these two SR proteins are the closest paralogs and share 66% identity and 74% similarity in amino acid sequence, it is not surprising that the lack of one SR protein is compensated for by the other SR protein.

**The conserved GAAG repeats are required for producing specific isoforms**

As described above, alignment of the third intron of SCL33 and SCL30a, which undergoes AS, revealed considerable sequence conservation, especially at the 3' end where AS takes place, and contains four closely spaced GAAG sequence elements in the region that was shown to bind SCL33 (Figures 2 and 3). To test whether these GAAG elements are important for generation of one or more splice variants, we generated three mutants (M1&2, M3&4 and M1–4) in which the GAAG sequence is changed to CTTC (Figure S2B). The first two GAAG elements are mutated in M1&2, the last two elements are mutated in M3&4, and all four elements are mutated in M1–4. The wild-type SCL33

**Figure 5.** Analysis of AS of the SCL33 intron in wild-type (WT) and six mutants of Arabidopsis SR genes.
(a) Protoplasts from WT, three single (scl33, sc35, scl30a) and three double (sc35 scl30a, scl33 sc35, scl33 scl30a) mutants were transformed with the GFP–INT–GFP construct, and splice variants were analysed by RT-PCR using GFP-specific primers. The gel shows RT-PCR products, with changes in the splicing pattern in the double scl33 scl30a mutant compared to WT and other single and double mutants. The missing isoforms with sizes are shown along the side.

(b) Wild-type protoplasts were transformed with either GFP–INT–GFP or the mutant GFP–INT–GFP forms M1&2, M3&4 and M1–4, in which GAAG elements 1 and 2, 3 and 4, or all four, respectively, were mutated to CTTC (see Figure 4). (i) AS of the mutated GFP–INT–GFP (M1&2 and M3&4) and GFP–INT–GFP constructs in WT protoplasts. (ii) AS of mutated GFP–INT–GFP (M1–4) and GFP–INT–GFP constructs in WT protoplasts. (iii) Splicing of the endogenous SCL33 is not changed in protoplasts transformed with either the GFP–INT–GFP construct or the M1–4 mutated GFP–INT–GFP construct. The cyclophilin control for each experiment is shown below, and schematic diagrams of altered isoforms are shown on the right. Numbers on the right indicate the size of amplified products.
intron in the GFP–INT–GFP splicing reporter was replaced with the three mutated forms to monitor their splicing in wild-type protoplasts. In experiments comparing wild-type and mutated intron splicing, the M1&2 and M3&4 mutants, with the first or last two GAAG elements mutated, respectively, showed almost complete loss of isoform 3 and an increase in isoform 4 (Figure 5bi), whereas the M1–4 intron, with mutations in all four GAAG elements, resulted in loss of isoforms 3 and 6 and an increase in isoform 4 (Figure 5biii).

Analysis of AS of endogenous SCL33 by RT-PCR using the SCL33-specific primer showed no change in its splicing pattern (Figure 5biii), confirming that the splicing pattern is changed only in the mutated GFP–INT–GFP constructs. Interestingly, all three isoforms that are affected in M1–4 transfected cells have common 5' and 3' SS at the 3’ end of the intron (Figure 5biii, right). These results suggest that generation of these three isoforms requires a trans-acting splicing factor that recognizes the GAAG elements for accurate splicing of the 3' region of the intron. From this data, we also conclude that the number of GAAG elements plays a role in the reduction or increase of splice variants, as only two isoforms are affected when two GAAG elements are mutated at a time, but mutating all four GAAG elements affects three isoforms. Interestingly, the double mutant (scl33 scl30a) displays a similar splicing pattern to that for M1–4 GFP–INT–GFP, suggesting that the SCL33 or SCL30a proteins bind to these cis elements and are necessary for accurate splicing of the SCL33 intron. Binding of the SCL33 protein to the 92 nt fragment (Figure 2) supports this hypothesis.

DISCUSSION

Alternative splicing (AS) is highly prevalent in plants, and is thought to play an important role in increasing proteome diversity as well as in regulating gene expression at the post-transcriptional level (Reddy et al., 2012). Some of the regulators of AS are RNA-binding proteins, such as members of the SR family that bind to specific RNA sequences in pre-mRNA and assist in spliceosome assembly at weak splice sites and contribute to regulated AS. The pre-mRNAs from plant SR genes themselves undergo extensive AS, but the mechanisms that regulate AS are poorly understood. Splicing of pre-mRNA substrates in a cell extract, which has been used extensively in animals (Chasin, 2007; Long and Caceres, 2009), is a powerful method to study SREs and splicing factors. However, the lack of such a system derived from plant cells has hampered progress in this area in plants. The use of protoplasts to transiently express splicing reporters offers an alternative and powerful approach to identify trans-acting factors that regulate AS. Here we have developed an in vivo splicing reporter assay to study regulation of AS of the SCL33 intron. Use of this splicing reporter in protoplasts from mutant plants lacking one or more SR proteins or other putative splicing regulators offers a novel and tractable way to study the function of a given SR protein in AS.

Using in vivo splicing assays with a splicing reporter containing the SCL33 intron, we show that all splice variants from this intron are accurately generated, suggesting that all signals required for AS reside within the intron of the SCL33 gene. This supports the intron definition model for AS regulation of this intron. Using this reporter system, we performed further experimental characterization of the sequence elements located within this intron and identified SR proteins that are involved in regulating AS. The RNA–protein interaction studies and AS analysis in mutant protoplasts presented here indicate that SCL33 auto-regulates its AS. EMSA analysis revealed that the purified SCL33 protein binds to a 92 nt region in the SCL33 third intron. Four lines of evidence indicate that the observed interaction between the 92 nt fragment (P4) and SCL33 is specific. First, the 5' end of the intron (P1) did not bind SCL33. Second, binding of SCL33 to P4 is eliminated using excess cold RNA. Third, SR45, an SR-like protein, does not bind to this fragment (Figure S3A). Finally, binding of SCL33 to P4 was not affected by adding an excess of cold P1 RNA (Figure S3B). Although AS of the SCL33 intron is not altered in the scl33 mutant, specific isoforms are missing in the double mutant (scl33 scl30a) in which a closely-related paralog is also lost, suggesting that SCL33 and SCL30a have a redundant function. Mutations in the GAAG elements of the 92 nt segment in the SCL33 intron, which binds to SCL33 protein, resulted in an altered splicing pattern similar to that observed with the wild-type intron in the scl33 scl30a double mutant. This suggests that the GAAG element-containing intron sequence is critical for SCL33 binding and regulation of normal AS. Furthermore, the affected 5’SS at nucleotide position 604 is adjacent to the SCL33 binding region. A model illustrating the mechanism(s) by which SCL33 auto-regulates AS is presented in Figure 6. As the affected isoforms have the same 5’SS, it is possible that SCL33 binds to the 92 nt region in the middle of the intron and recruits U1 snRNP to the 5’SS either by directly interacting with one of the U1 snRNP (small nuclear ribonucleoprotein particle) proteins or other SR proteins that interact with U1 snRNP. There is prior evidence for such interactions with SCL33. We have previously shown that SCL33 directly interacts with U170K, one of the U1 snRNP proteins (Golovkin and Reddy, 1999) and also an SR-like protein (SR45) that is known to interact with U170K (Golovkin and Reddy, 1999; Reddy, 2007). Our results also suggest that SC35 alone or in combination with either SCL33 or SCL30a does not regulate AS of the SCL33 intron, as the pattern of AS in single (sc35) and double (sc35 scl33 or sc35 scl30a) mutants is not altered. The observation that only certain isoforms are affected in the scl33 scl30a double mutant indicates that other SR proteins may also be involved in regulating the splicing of this intron.
Many of the splice variants from SR genes have a premature termination codon and are targets of nonsense-mediated decay (Lareau et al., 2007; Palusa et al., 2007; Palusa and Reddy, 2010). In fact, all splice variants that contain all or any part of the third intron of SCL33 are potential targets of nonsense-mediated decay (Palusa et al., 2007), and some have been experimentally shown to be degraded by nonsense-mediated decay (Palusa and Reddy, 2010). The generation of splice variants by nonsense-mediated decay in SR proteins and several other RNA-binding proteins is auto-regulated, such that high levels of protein result in generation of premature termination codon-containing transcripts to tightly regulate the levels of splicing factors (Jumaa and Nielsen, 1997; Lopato et al., 1999; Sureau et al., 2001; Lareau et al., 2007; Schoning et al., 2008). It is likely that the auto-regulation of SCL33 AS that generates premature termination codon-containing isoforms plays a role in controlling the levels of functional SCL33 transcript and protein. The fine balance of the premature termination codon isoforms with respect to functional transcripts and proteins is an important feature in gene regulation (Mitrovich and Anderson, 2000; Sureau et al., 2001; Lareau et al., 2007; Schoning et al., 2008).

Mutational studies with GAAG elements in the SCL33 binding region indicate that the number and sequence of these repeats is important for some isoforms. In humans, studies have shown that the GAAGAA hexamer, the highest scoring exonic splicing enhancer motif, functions as an exonic splicing enhancer (Fairbrother et al., 2002), and such purine-rich elements are reported to function as exonic splicing enhancers in other vertebrates (Tack and Manley, 1995; Chasin, 2007). A computational analysis of the Arabidopsis exons for candidate exonic splicing enhancers identified GAAGAAAGAA as one the exonic splicing enhancers (Pertea et al., 2007), and our results show that GAAG repeats also function as intronic splicing regulators. In addition to the GAAG elements in the 92 nt fragment, there are more GAAG repeats on the 3’ side of the Arabidopsis SCL33 intron. To determine whether the third intron of SCL33 from other species also contains the GAAG repeats, we aligned the nucleotide sequences of the third intron from Arabidopsis thaliana, Capsella rubella, Brassica rapa and Populus trichocarpa. Remarkably, the 3’ end of the intron where all of the AS events take place in Arabidopsis is conserved. Furthermore, most of the GAAG repeats are highly conserved across various dicots (Figure S5). The 3’ region of the third intron of Brachypodium, a monocot, also contains multiple GAAG elements (Figure S6), suggesting conservation of this element in angiosperms. These findings, together with our experimental results, indicate that GAAG repeats function in regulating alternative splicing.

As intron retention is common in plants, we determined the presence of two or more GAAG repeats (with spacing of 3–15 nucleotides) in retained and constitutively spliced introns in the Arabidopsis genome. Among a total of 2780 introns that are retained (TAIR10 annotations, http://www.arabidopsis.org), we found that 59 have 2–12 GAAG repeats. A list of these genes along with the number of GAAG occurrences is presented in Table S2. This proportion (0.021) is statistically significantly higher than in constitutively spliced introns ($P < 2 \times 10^{-5}$ in a binomial exact test), in which 0.011 of introns have such a repeat. As retained introns tend to be shorter than constitutive introns, the $P$ value is conservative. This suggests that GAAG repeats may be one of the signals that contribute to intron retention.

The protoplast system with splicing reporters together with putative splicing factor mutants, as described here, may be employed to identify splicing regulators involved in AS. This, combined with in vitro RNA binding studies, may provide further insights into direct or indirect regulation of AS of a given pre-mRNA. Although transgenic lines over-expressing a splicing factor have been used to study the role of splicing regulators (Wang et al., 1996;
Lopato et al., 1999; Kalyna et al., 2003), the results from such experiments may not provide accurate insights as the effects of many splicing regulators are dose-dependent. The use of protoplasts from knockout mutants offers an alternative approach and has an advantage in that the cells lack one or more splicing regulators. As several SR proteins have paralogs, functional redundancy may also be addressed by using protoplasts from double or triple mutants as demonstrated here.

**EXPERIMENTAL PROCEDURES**

**Construction of a splicing reporter and generation of a transgenic line expressing the splicing reporter**

Genomic DNA was isolated from Arabidopsis thaliana Columbia (Col-0) using a plant DNAeasy kit (Qiagen, www.qiagen.com) and used as a template in PCR. The third intron of the SCL33 gene was amplified using Hot Start Pfu polymerase (Agilent, www.genomics.agilent.com) and intron-specific primers (SR33-IN-IF and SR33-IN-RP, see Table S1), and cloned into pGFP(GA5)II (obtained from Dr. Nam-Hai Chua, The Rockefeller University) at the MseI site within the coding region of GFP to generate the GFP–SCL33 INTRON–GFP (GFP–INT–GFP) construct driven by the CaMV 35S promoter. Correct orientation of the intron was verified by sequencing. The GFP–INT–GFP plasmids were used to transform leaf protoplasts obtained from the leaves of Arabidopsis wild-type, SC35 and SCL33 Arabidopsis T-DNA insertion lines for the genes SR33-IN-RP, see Table S1), and cloned into pGFP(GA5)II (obtained from genomics.agilent.com) and intron-specific primers (SR33-IN-FP and SR33-IN-RP, see Table S1), and cloned into pGFP(GA5)II (obtained from Dr. Nam-Hai Chua, The Rockefeller University) and used to transform Arabidopsis. Transgenic lines were selected on BASTA (Crescent Chemicals, www.creschem.com) (5 μg ml⁻¹), and F2 plants expressing GFP or GFP–INT–GFP were used to monitor splicing.

**Generation of single and double mutants of Arabidopsis**

The Arabidopsis T-DNA insertion lines for the genes SCL33 (SALK_058566), SC35 (SALK_033824), SCL33a (SALK_041849) in the Columbia background were obtained from the Arabidopsis Bio- logical Resource Center. The T-DNA insertion in each gene was verified by genomic PCR using a gene-specific primer and a T-DNA-specific primer (LBB1). Expression of the SR genes in the mutants was analysed by RT-PCR using gene-specific primers (Table S1). The following PCR conditions were used for genotyping: initial denaturation at 94°C for 2 min, followed by 29 cycles at 94°C for 30 sec, 56°C for 30 sec and 72°C for 1 min, with a final extension at 72°C for 10 min. DNase-treated RNA from 2-week-old seedlings from wild-type and mutant lines was used for RT-PCR analysis as described previously (Palusa et al., 2007). Wild-type and all mutant lines were grown under long-day conditions (16 h light/8 h dark; 100 μmol m⁻² s⁻¹ light intensity, 22°C). Three double mutants (sc35 sc310a, sc33 sc35 and sc33 sc310a) were generated by crossing the single mutants. All double mutants were genotyped for T-DNA insertion and homoygosity using genomic PCR and RT-PCR as described above.

**Analysis of GFP–INT–GFP splicing in Arabidopsis mesophyll protoplasts**

Splicing of GFP–INT–GFP pre-mRNA was analysed in mesophyll protoplasts obtained from the leaves of Arabidopsis wild-type, single (sc33, sc35 and sc310a) and double (sc35 sc310a, sc33 sc35 and sc33 sc310a) mutants. Protoplasts from rosette leaves of 3–4-week-old plants grown in a greenhouse at 22°C under 16 h light/8 h dark were prepared and transfected as described earlier (Yoo et al., 2007). Equal amounts (20 μg ml⁻¹) of the GFP or GFP–INT–GFP plasmids were used to transfect 2 ml protoplasts (2 × 10⁶) from wild-type, three single and three double mutants. Transfected protoplasts were incubated in Petri dishes in the growth chamber in the dark at 22°C for 15–16 h. The protoplasts were then visualized under a fluorescence microscope for GFP expression and used for RT-PCR analysis.

**RNA isolation and RT-PCR analysis**

RNA from transfected protoplasts, wild-type and mutant plants was isolated using an RNasey plant mini kit (Qiagen). On-column DNase (Qiagen) digestion was performed to remove any genomic or plasmid DNA contamination before cDNA synthesis. DNase-treated RNA (200 ng) was used to synthesize first-strand cDNA using Superscript II reverse transcriptase (Invitrogen, www.invitrogen.com), and 2 μl of the first-strand cDNA was used for PCR in a reaction volume of 20 μl. GFP-specific forward and reverse primers (Table S1) were used for amplification. To monitor the levels of SCL33 splice variants generated from the endogenous gene, SCL33 forward and reverse primers were used (Table S1). All splice variants generated from the SCL33 third intron of the endogenous gene in protoplasts were amplified using primers corresponding to exons 3 and 4 (Table S1). The PCR products were gel-purified, cloned into pCR2.1 TOPO (Invitrogen, www.invitrogen.com) vector and sequenced.

**Preparation of 32P-labelled RNA probes and cold competitor RNAs**

Intron 3 of SCL33 was divided into four parts (P1, P2, P3 and P4) and cloned into the pGEM4 vector (Promega, www.promega.com) using the BamHI and HindIII restriction sites. P1 consists of the first 421 nucleotides (nt), P2 consists of the remaining 344 nt fragment of the third intron, P3 consists of the first 208 nt of P2, and P4 consists of the first 92 nt of P2. All fragments were generated by PCR amplification using the primers listed in Table S1, and the clones were verified by sequencing. Each of these constructs were linearized by digesting with HindIII, and used as a template to prepare labelled P1, P2, P3 and P4 RNA probes as follows. Capped RNAs were transcribed in vitro and labelled with 45 μCi [³²P]UTP (800 Ci mmol⁻¹, Perkin-Elmer, www.perkinelmer.com) using SP6 RNA polymerase (Fermentas, www.fermentas.com) in the presence of 500 μM ATP, 500 μM CTP, 50 μM GTP, 50 μM UTP and 7mGpppG (7-methyl-diguanosine triphosphate) from linearized pGEM4I(Promega, www.promega.com) plasmid DNA templates, and gel-purified as previously described (Witusz and Shenk, 1988). Unlabelled competitor RNAs were generated in the same manner, but without 7mGpppG or radiolabelled nucleotide, and the concentrations of UTP and GTP were increased to 500 μM.

**Expression and purification of recombinant SCL33 and SR45 proteins**

The SCL33 clone in the pET32 expression vector was (Novagen, www.emdmillipore.com) used to prepare purified SCL33 protein as described by Golovkin and Reddy (1999) with minor modifications. The bacteria were grown at 37°C until they reached an OD₆₀₀ of 0.6, after which 0.5 mM isopropyl-β-D-thiogalactopyranoside was added and the culture was incubated for 4 h at 30°C to induce protein expression. Subsequently, the bacteria were centrifuged (5000 × g) at 4°C for five minutes, the pellet was resuspended in one-tenth of the culture volume of binding buffer (50 mM Tris HCl, 1 mM EDTA, 100 μg ml⁻¹ lysozyme and 0.1% Triton X-100) containing protease inhibitors, and
incubated at 4°C for 15 min. The sample was then sonicated, centrifuged, and the supernatant was collected. S-protein agarose beads were added to the supernatant, and the mixture was incubated for 1 h at 4°C. After washing the beads with binding buffer (3 washes for 5 min each), the bound protein was eluted using 0.2 M mercaptoethanol (pH 2.2) and neutralized by adding a 1/20th volume of 2 M Tris-base (pH 10.4). The eluted proteins were dialysed using phosphate buffer (10 mM Na2HP04, 2 mM KH2PO4, 2.7 mM KCl, 137 mM NaCl pH 7.4). SR45 was purified as described previously (Golovkin and Reddy, 1999).

Electrophoretic mobility shift assays

Internally radiolabelled RNAs (P1, P2, P3 and P4; 4-20 fmol) were incubated with increasing amounts of purified recombinant SCL33 in 20 units of RNase inhibitor (Inivitrogen), 0.15 mM spermidine and gel shift buffer (15 mM HEPES pH 7.9, 8% glycerol, 100 mM KCl and 2 mM MgCl2) for 5 min at 30°C in a 14 μl reaction. Following incubation, 4 μg μl-1 of heparin sulfate (Sigma, www.sigmaaldrich.com) was added to each reaction, the samples were then chilled on ice for 5 min, and 1.5 μl of 10 × loading dye (30% glycerol, 0.5% bromophenol blue, 0.5% xylene cyanol) was added. RNA–protein complexes were run on a 5% native polyacrylamide gel at room temperature in 1 × TBE (Tris-Borate-EDTA) buffer (200 V for 2–6 h). Gels were then dried, exposed to a phosphor screen, and visualized by phosphor imaging using a Storm 840 phosphor imager (Molecular Dynamics, www.amershambio.com).

Generation of mutations in the SCL33 intron

The shortest fragment (92 nt) of the SCL33 intron that bound to SCL33 protein contains four conserved GAAG elements. To test whether these sequences are important for AS, we mutated them, two at a time or all four, to CTTC using a QuikChange Lightning site-directed mutagenesis kit (Stratagene, www.stratagene.com) was added to each reaction, the samples were incubated with increasing amounts of purified recombinant SCL33 in 20 units of RNase inhibitor (Inivitrogen), 0.15 mM spermidine and gel shift buffer (15 mM HEPES pH 7.9, 8% glycerol, 100 mM KCl and 2 mM MgCl2) for 5 min at 30°C in a 14 μl reaction. Following incubation, 4 μg μl-1 of heparin sulfate (Sigma, www.sigmaaldrich.com) was added to each reaction, the samples were then chilled on ice for 5 min, and 1.5 μl of 10 × loading dye (30% glycerol, 0.5% bromophenol blue, 0.5% xylene cyanol) was added. RNA–protein complexes were run on a 5% native polyacrylamide gel at room temperature in 1 × TBE (Tris-Borate-EDTA) buffer (200 V for 2–6 h). Gels were then dried, exposed to a phosphor screen, and visualized by phosphor imaging using a Storm 840 phosphor imager (Molecular Dynamics, www.amershambio.com).

ACKNOWLEDGEMENTS

This work was supported by a grant from the US National Science Foundation. We thank the Arabidopsis Biological Resource Center for providing the T-DNA insertion lines of single mutants, and Irene S. Day (Department of Biology, Colorado State University, USA) for critically reading the manuscript. The authors declare no conflict of interest.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Alternative splicing of SCL33 pre-mRNA in 3-week-old leaves and protoplasts isolated from 3-week-old leaves.

Figure S2. Alignment of sequences of splice variants produced from the third intron of SCL33 in protoplasts with splice variants produced from the same intron in the GFP–INT–GFP splicing reporter, and sequences of wild-type and mutated 92 nt segments.

Figure S3. SR45 does not bind P4 RNA, and an excess amount of cold P1 RNA does not abolish binding of P4 RNA to SCL33.

Figure S4. Transcripts corresponding to the region prior to the T-DNA insertion site are not present in the sis/c33 mutant.

Figure S5. Alignment of the nucleotide sequence of intron 3 of SCL33 homologs from Arabidopsis thaliana (At), Capsella rubella (Cr), Brassica rapa (Br) and Populus trichocarpa (Pt).

Figure S6. Nucleotide sequence of SCL33 intron 3 of Brachypodium distachyon.

Table S1. List of primers used in genomic PCR, RT-PCR and in preparing all constructs used in this study.

Table S2. Introns with GAAG repeats in the Arabidopsis genome. Please note: As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. It is the responsibility of the author to ensure the accuracy and integrity of the online material.

REFERENCES


946  Julie Thomas et al.


