ABSTRACT

Title of Document: CHARACTERIZATION OF *ARABIDOPSIS THALIANA* SR PROTEIN GENES: MUTATIONS, ALTERNATIVE SPLICING, AND ESE SELECTION.

Jason Edmonds, Doctor of Philosophy, 2007

Directed By: Associate Professor, Stephen M Mount, Department of Cell Biology and Molecular Genetics

RNA processing in eukaryotes is a highly complex process requiring numerous steps and factors that can play roles in the regulation of functional protein production. SR proteins are a well-defined family of splicing factors identified by a conserved RNA Recognition Motif (RRM) and carboxyl-terminal arginine/serine (RS) repeats. SR proteins are known to bind to mRNA precursors via Exonic Splicing Enhancers, and to recruit U2AF and the U1 snRNP to promote splicing.

I have identified mutations in five *Arabidopsis thaliana* SR protein genes that result in altered phenotypes. Two (*scl28-1* and *srp31-1*) result in embryonic lethal phenotypes, while three others (*sc35-1, sr45-1*, and *srp30-1*) result in viable and fertile plants with a range of phenotypes.

I have also found that mutations in individual SR protein genes can effect the ability of a specific sequence to act as an ESE and hence affect splicing efficiency. Because 16

of the 20 *Arabidopsis thaliana* SR proteins themselves are alternatively spliced, I have looked for cross regulation using RT-PCR analysis of isoform accumulation in alternatively spliced SR protein genes. I found that SR proteins do, in fact, regulate the alternative splicing of gene targets and do so in both a gene and a tissue specific manner.

In order to begin to fully understand the relationship between individual SR proteins it is essential to know when and where they are expressed throughout development. I have studied the expression pattern of 16 of the 20 SR proteins in the roots of wild-type plants as well as *sc35-1, srp30-1,* and *sr45-1* mutants. I have identified both spatial and temporal expression patterns for these 16 proteins relative to specific tissues that compose the root.

CHARACTERIZATION OF *ARABIDOPSIS THALIANA* SR PROTEIN GENES:
MUTATIONS, ALTERNATIVE SPLICING, AND ESE SELECTION.

By

Jason Matthew Edmonds

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2007

Advisory Committee:
Professor Stephen M Mount, Chair
Professor Caren Chang
Professor Zhongchi Liu
Professor Todd Cooke
Professor Dorothy Beckett

Acknowledgements

I would like to first thank my advisor, Dr. Stephen Mount for his continued support in and out of the lab. He allowed me to work independently while also encouraging me to present my work both locally and before international audiences. I would also like to express my gratitude to Dr. Mount for challenging me to become a better scientist and a better person.

I would also like to thank my committee for following my progress with interest and providing invaluable insight and suggestions in my research and experiments.

A large portion of the early work performed in my graduate career could not have been accomplished as quickly and efficiently without the help of undergraduates working in our lab. Specifically, Tyjen Tsai, Christina Kary, and Sita Shablock took on portions of my project with the enthusiasm and care as if it was there own. I would also like to thank Dr. Xiaoning Zhang for her insightful suggestions and conversation as well as helping with the editing of my thesis.

I would also like to thank Mrs. Ruth Mausolf who is one of the nicest and hardest working people at the university. Equipment and supplies were always exactly as ordered and always arrived on time which is an essential part of research.

A very special thank you goes to Payam Sajedi, who helped me keep my sanity in the lab and who is by far the best undergraduate to ever work in a laboratory. Finally, I would like to thank all of my family and friends who supported and encouraged me during my graduate career. I cannot thank them enough.

Table of Contents

List of Tables

List of Figures

Abbreviations

AS   Alternative Splicing

ESE   Exon Splicing Enhancer

ESS   Exon Splicing Silencer

I/S ratio   Included to Skipping ratio

mRNA   messenger RNA

MS   Murashige and Skoog

RNAi   RNA interference

RRM   RNA Recognition Motif

RS domain   Arginine/Serine domain

RT-PCR   Reverse Transcriptase PCR

SR protein   Serine/Arginine-rich protein

U2AF   U2 snRNP auxiliary factor

UTR   Un Translated Region

Chapter 1: Introduction: pre-mRNA splicing and SR proteins

Splicing

In 1977, the complexity of eukaryotic genomes and gene regulation suddenly increased with the discovery that genes are divided into exons and introns. To add further complexity, it was also found in the same studies that introns are effectively excised from mRNA transcripts in the nucleus (Berget *et al.* 1977; Chow *et al.* 1977). The exact recognition of splicing signals and consequential processing is carried out by five small nuclear ribonucleoproteins (snRNPs) and approximately 300 proteins (Rappsilber *et al.* 2002; Zhou *et al.* 2002; Jurica *et al.* 2003), of which many are conserved in plants (Lorkovic *et al.* 2000; Reddy *et al.* 2001), with a large range of RNA-RNA and RNA-protein interactions (Moore *et al.* 1993; Nilsen *et al.* 1994). This discovery would lead to further research and alternative splicing would soon add to the complex intricacies of the knowledge of mRNA processing at that time. Additionally, alternative splicing is also involved in regulating gene products at the transcript level by creating truncated transcripts that are predetermined to undergo nonsense-mediated mRNA decay (NMD) (Le Hir *et al.* 2000; Reichert *et al.* 2002). The regulation of RNA processing and production is part of a mRNA "factory" composed of many protein and RNA factors (Maniatis *et al.* 2002).

The amount of alternative splicing varies among individual genes and from species to species. The greatest example of the potential of alternative splicing is in the *Drosophila* neurological gene *DSCAM*. From this one single gene, 30,016 alternatively spliced transcripts are potentially produced (Crayton *et al.* 2006) which

is more than twice the 13,600 genes in the entire *Drosophila* genome. Additionally, the amount of alternative splicing within an organism is relative to the complexity of the organism. For example, the number of human and *Arabidopsis* genes is roughly the same with 25,000 and 28,000 genes respectively. Available EST and cDNA data suggests that approximately 75% of all human multi-exon genes undergo alternative splicing (Johnson *et al.* 2003), and only 23.5% of Arabidopsis multi-exon genes are alternatively spliced (Campbell *et al.* 2006). One of the difficulties in determining the amount of alternative splicing in any organism is distinguishing between *bona fide* alternative splicing and aberrant splicing. Splicing in *Arabidopsis* is a good example of how difficult this determination can be. The most frequent type of alternative splicing in plants is intron retention. In fact, 45% of all plant alternatively spliced transcripts are intron retention forms (Figure1.1)(Haas *et al.* 2003; Campbell *et al.* 2006). It may be that these transcripts were incompletely spliced and entered into cDNA and EST libraries rather than products of alternative splicing. However, the retention of introns has been shown to be conserved across different plant species, giving validation to this form of alternative splicing in plants (Haas *et al.* 2003; Kalyna *et al.* 2006) as opposed to its merely being the result of splicing inefficiency. In addition to species conservation of the retained introns, analysis of 218 transcripts with the intron retained form of alternative splicing in *Arabidopsis* was carried out to determine the outcome of the alternatively spliced products (Haas *et al.* 2003). Of these 218 retained intron transcripts, only 99 produced truncated products that would expect to be degraded via NMD and the remainder of the 119 transcripts actually produce variant proteins. This gives biological support to the theory that these

Exon skipping/inclusion	5%
Alternative 3' splice site	23.9%
Alternative 5' splice site	11.4%
Mutually exclusive exons	rare
Intron retention	45%
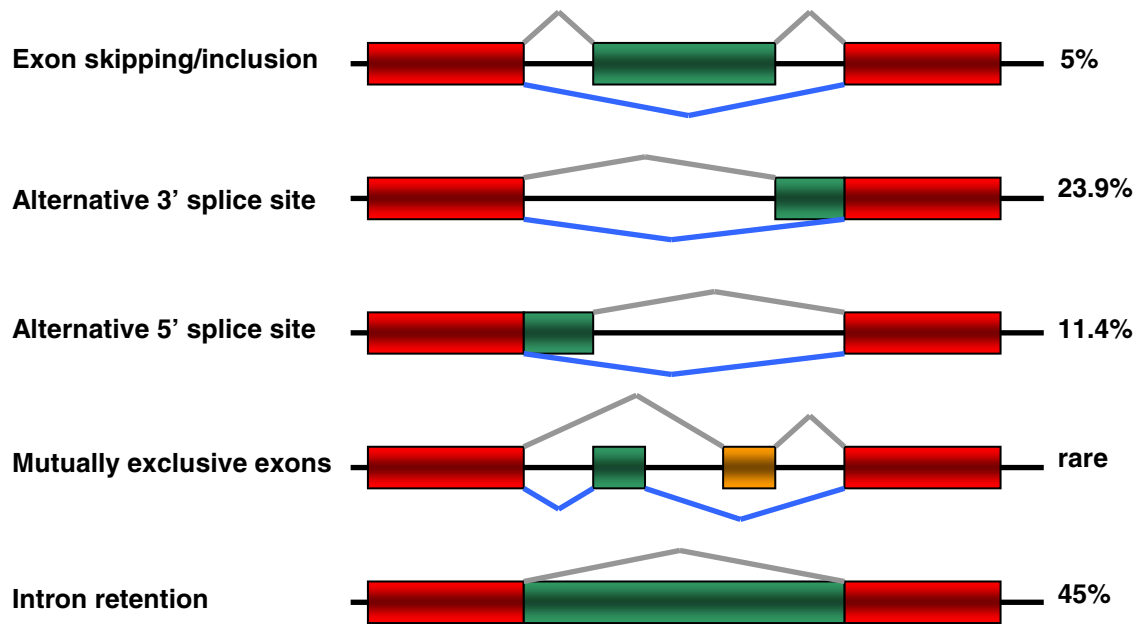
Figure 1.1 Five types of alternative splicing are observed in varying frequencies in *Arabidopsis thaliana*. Alternative splicing in *Arabidopsis thaliana* can occur in 5 distinct manners, with each occurring at different frequencies. The remaining ~15% of cases not shown here involve multiple termini. Red boxes indicate constitutive exons, green and yellow boxes indicate alternatively spliced regions, and black dashes indicate introns.

products are, in fact, the result of true alternative splicing events, and not incorrectly spliced mRNA (Haas *et al.* 2003; Kalyna *et al.* 2006).

While many diseases in humans such as juvenile spinal muscular atrophy, cystic fibrosis, and Marfan's syndrome can result from aberrant splicing, very little information is known about the functional significance of alternative splicing in plants (Reddy *et al.* 2001; Jordan *et al.* 2002). In fact, there are only two well studied plant genes that undergo alternative splicing and have biological functions associated with the resulting alternatively spliced products. One is the tobacco mosaic virus TMV resistance *N* gene (Dinesh-Kumar *et al.* 2000; Zhang *et al.* 2003) and the very well studied *FCA* gene in *Arabidopsis* (Macknight *et al.* 2002; Quesada *et al.* 2003). The tobacco *N* gene contains an alternatively spliced exon located within intron three. Unexposed plants, or plants that have been exposed to TMV for less than three hours, produce the N protein containing the alternatively spliced exon. However, when exposed to TMV for four hours or more, the alternatively spliced intron is not included. A truncated protein is produced, which leaves the plant resistant to the virus.

The *FCA* gene encodes an RNA binding protein that undergoes alternative splicing to negatively auto-regulate its own product, which, in turn, regulates flowering (Quesada *et al.* 2003; Simpson *et al.* 2003). The *FCA* gene naturally produces four unique mRNAs (Figure 1.2). Only one form, the gamma form, results in a functional protein that allows the plant to progress through floral transition after vernalization. The most abundant form, the alpha form with an alternatively included
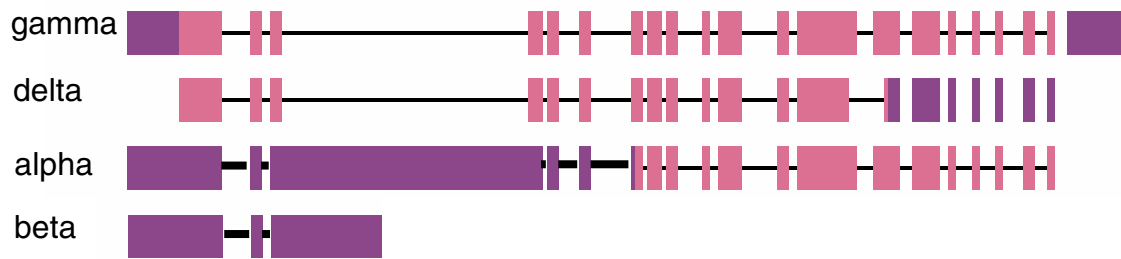
Figure 1.2 Flowering time in plants is regulated by alternative splicing of *FCA* gene. The *FCA* gene is capable of being spliced into four different mRNA isoforms, the gamma form being the only functional form. The delta isoform has a 13nt alteration in exon 13 that results in an out of frame 3' end. The alpha isoform, the most abundant of the four, encodes for a non functional protein resulting in the inclusion of intron three. The fourth isoform, beta, contains a partial intron three that encodes for a premature stop codon and is polyadenylated within this region. UTR regions are highlighted in purple, exons in pink, and introns are designated with dashes.

intron three, is non-functional. In addition to the *FCA* gene and the TMV resistance gene, there are only a few instances where effects of alternative splicing have been researched (Reddy *et al.* 2001).

The lesser amount of alternative splicing in plants does not necessarily demonstrate a lack of importance. While one could argue that the minimal amount of data available is a good indication of the relative significance of splicing in plants, it would be more appropriate to look at this viewpoint from the other angle. The lesser amount of splicing in plants makes it more difficult to study, as fewer targets are available with which to work. The fact that the major splicing signals such as intron/exon boundaries (Figure 1.3), exonic and intronic splicing signals (discussed in chapter three), and a high conservation of the major components of the spliceosome including SR proteins (discussed in great detail later in this chapter), are conserved between plants and animals, are indications of the significance alternative splicing plays in plants. If alternative splicing was not important in plants, the high conservation of both proposed and confirmed molecular function and protein composition would not exist to the observed degree. While not as prominent in plants as it is in animals, alternative splicing is equally important and essential for the viability of the organism.

SR proteins

SR protein genes are a highly conserved family of genes coding for RNA binding phosphoproteins, all of which are recognized by a single monoclonal antibody (mAb104) (Zahler *et al.* 1992). This family of non-snRNP proteins splicing
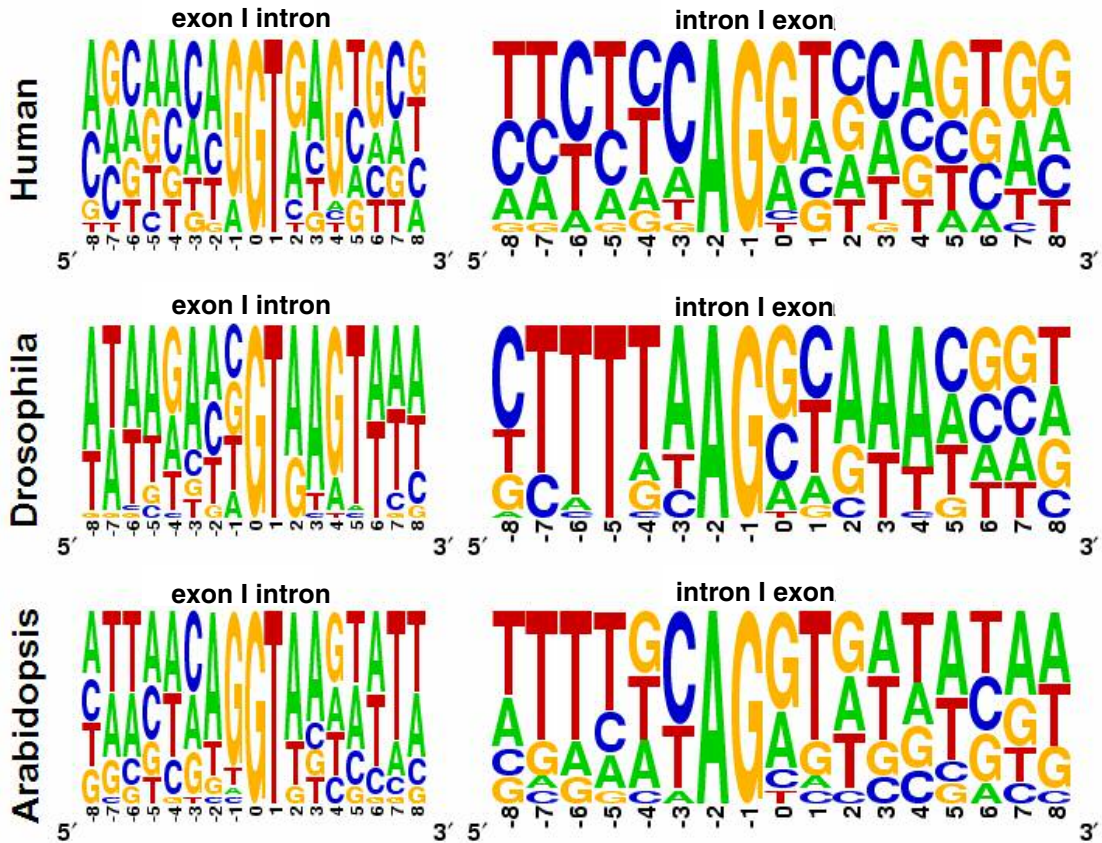
Figure 1.3 Splice site signals are conserved between animals and plants. Both 5'
and 3' splice site nucleotide composition of human, fly, and *Arabidopsis* is depicted
using Weblogo to generate the pictogram. The relative frequency of each nucleotide
is depicted by its height relative to the other nucleotides at each respective position.
This figure was generated using software available at http://weblogo.berkeley.edu/

factors carries out multiple roles in splicing (Graveley *et al.* 2000), the first being the recognition of exonic binding sites. Once bound to the pre-mRNA, SR proteins recruit components of the spliceosome to both the 5' splice site, AG/GURAGU (where the slash denotes the exon/intron boundary) and the 3' splice site, YAG/N. Through two transesterification reactions (Moore *et al.* 1993) in which splicing factors are bound to both splice sites and to the branch point, YNYURAC, the intron is removed from the transcript. During this multi-step process, it has been suggested that SR proteins play essential roles in many of these steps. SR proteins are major factors for facilitating both constitutive and alternative spicing (Ge *et al.* 1990; Krainer *et al.* 1990; Fu *et al.* 1992; Caceres *et al.* 1994). In constitutive splicing, SR proteins are believed to bind to an exon where they simultaneously interact with the U2AF, which is bound to the 3' splice site, and the 70K protein of the U1 snRNP, which is bound to the 5' splice site (Reed *et al.* 1996). Additional roles in alternative splicing, including splice site selection, are outlined in later chapters.

SR proteins are typically located in subnuclear domains called speckles (Misteli *et al.* 1997; Misteli *et al.* 1999; Lamond *et al.* 2003). However, the mammalian SR proteins ASF/SF2, SRp40, SC35, and SRp20 have been found to shuttle between the nucleus and cytoplasm depending on their phosphorylation state (Misteli *et al.* 1998a; Misteli *et al.* 1998b). Additionally, Misteli et al. found that the RS domain plays a major factor in the ability of SR proteins to associate and disassociate with each other within the nucleus. The fact that these proteins are capable of shuttling back and forth between these different environments and phosphorylation states through a precise mechanism, suggests that they may possibly

play a major role in mRNA export, mRNA stability, or mRNA translation (Lemaire *et al.* 2002; Sanford J. *et al.* 2004). SR protein genes themselves are alternatively spliced, and their pre-mRNAs undergo the same splicing process that they are responsible for facilitating as a means of crossregulation (Lopato *et al.* 1999; Kalyna *et al.* 2003). In fact, 15 of the 19 previously identified *Arabidopsis* SR protein genes are alternatively spliced. The number of alternatively spliced forms range from a single alternative isoform to at least eight alternative transcripts per gene (Palusa *et al.* 2007). The exact targets and roles for each individual SR protein, or even subfamily of SR proteins, are still unknown. *In vitro* splicing assays of animal SR proteins have shown redundant functions, as any human SR protein can complement a splicing-deficient cytoplasmic extract (Graveley *et al.* 2000). However, *in vivo* assays have shown that some SR proteins in animals are functionally redundant while others are not (Reddy *et al.* 2004).

SR proteins are composed of two fundamental structures, a highly conserved RNA Recognition Motif (RRM), located at the amino-terminal of the protein, and a carboxyl-terminal domain containing RS dipeptide repeats of various lengths. The RS domain of the SR protein gene family is commonly accepted as being responsible for protein-protein interactions between other RS domain containing splicing factors. The necessity of the RS domain to enable SR proteins to interact with other splicing factors was first shown by Wu & Maniatis (Wu JY 1993), Amrein *et al.* (Amrein 1994)*,* and Kohtz *et al.* (Wu JY *et al.* 1993; Amrein *et al.* 1994; Kohtz *et al.* 1994). In Yeast two-hybrid studies, it was shown that SR proteins were capable of interacting with each other, with the splicing regulators Tra and Tra2, and with other

components of the splicing machinery, including U1-70K and U2AF. Not only were

RS domains found to be necessary for protein-protein interactions, but were also

sufficient enough for these interactions to occur. Additionally, the RS domain of

SF2/ASF was shown to be sufficient for interacting with the *Drosophila* splicing

repressor RSF1 (Labourier *et al.* 1999). When artificially tethered to pre-mRNA, the

RS domains of ASF/SF2, SC35, RSp40, RSp55, RSp75, and 9G8 human SR proteins

were able to activate enhancer-dependent splicing, showing that these domains were

sufficient to recruit other members of the splicing machinery (Graveley BR *et al.*

1998). Conversely, in another study, the unphosphorylated RS domain of ASF/SF2

was shown to be unable to bind with the U1-70K splicing factor (Xiao *et al.* 1997).

The differences in these studies suggest that while the RS domain is clearly the

domain used for protein-protein interaction, the specificity within the RS domain

dictates the proteins with which it is capable of interacting. Graveley *et al.* tethered

the RS domains of six human SR proteins, ASF/SF2, SC35, RSp40, RSp55, RSp75,

and 9G8 to pre-mRNA, and assayed the ability for these domains to induce splicing.

They found that all six domains were capable of facilitating splicing but did so in

varying degrees of success (Graveley BR *et al.* 1998). Although the differences were

subtle, it is unknown what these differences, if any, would have in whole organisms.

The serine residues in RS domains of SR proteins have been shown to be

heavily phosphorylated, and this signal appears to determine the subcellular

localization of SR proteins. Heavily phosphorylated SR proteins are found in the

cytoplasm, while unphosphorylated SR proteins are nuclear (Gui *et al.* 1994; Colwill

et al. 1996). Additionally, the phosphorylation state of the RS domain also influences

the ability of SR proteins to interact with other proteins (Wang *et al.* 1998; Xiao *et al.*

1998). Both Wang *et al.* and Xiao *et al.* found that extensive phosphorylation of SR

protein ASF/SF2 dramatically increased its ability for interacting with splicing factor

U1-70K. However, Xiao further demonstrated that the phosphorylation state of this

specific SR protein did not affect splicing fidelity. The phosphorylation state of the

RS domain may be a strong determinant as to the function an individual SR protein

may perform, and potentially act as a mechanism for regulation. For example, the

Drosophila SR protein kinase DOA was found to phosphorylate SR protein *RBP1 in*

vivo (Du *et al.* 1998). Du *et al.* found that *Doa* mutations in flies interfere with the

splicing of *doublesex* pre-mRNA, suggesting that an unphosphorylated RBP1 is

unable to facilitate splicing of this particular gene transcript.

An important step in identifying the function of SR proteins is to understand

both the spatial and temporal activity of each individual SR protein. This additional

information may also lend insight into the amino acid conservation and possible

functional redundancy of the SR protein gene family among metazoans. Over-

expression studies identified the ability of individual *Arabidopsis* SR proteins such as

SRp30, RSZp33, and RSp31 to crossregulate the isoform production and

concentration of other SR protein genes. These studies also showed individual SR

proteins are expressed in different cell types within different plant tissues including

roots, leaves, cotyledons, and flowers (Lopato *et al.* 1996a; Lopato *et al.* 1999; Lazar

et al. 2000; Kalyna *et al.* 2003). Another lab, interested in nuclear organization of

splicing factors, found that the expression of specific YFP tagged SR proteins in the

nucleus was dependent upon cell type as well as developmental stage of the tissue

being observed (Golovkin *et al.* 1998; Ali *et al.* 2003; Fang *et al.* 2004). The different stages are discussed extensively in chapter 5.

Not only have comparisons been drawn between plants and animals with respect to subcellular localization (Bubulya *et al.* 2004; Fang *et al.* 2004), but also interactions within the SR protein gene family in addition to SR proteins and other splicing factors, have been demonstrated with protein-protein interactions. Animal SR proteins SC35, ASF/SF2, and SR38 have been shown to interact with the U1-70K splicing factor in yeast two-hybrid assays and *in vitro* experiments (Wu JY *et al.* 1993; Manley *et al.* 1996; Shin *et al.* 2004). In comparison, U1-70K has been shown to interact with a total of five *Arabidopsis* SR proteins, SCL33, SR45, SRp34/SR1, and plant specific SR proteins RSZ22 and RSZ21, in yeast two-hybrid assays (Golovkin *et al.* 1998; Golovkin *et al.* 1999; Lorkovic *et al.* 2004). While more SR proteins in *Arabidopsis* were found to bind with the U1-70K snRNP than animal SR proteins were capable of, it is not necessarily an indication that splicing in *Arabidopsis* is more complex than in animal systems. In fact, the larger total number of SR protein genes in *Arabidopsis* could potentially mean that each individual SR protein may have a more limited expression pattern with respect to temporal and spatial regulation. Thus, while the number of SR proteins involved in a single splicing event may not differ from that of animals, more individual SR proteins are capable of carrying out the same processes and functions that fewer SR proteins in animals are capable of facilitating.

Despite all of the similarities of SR proteins between plants and animals, there are some differences in cross-kingdom splicing ability. Proteins in *Arabidopsis*,

carrot, and tomato were identified by using the human ASF/SF2 antibody mAB104, and were found to be capable of complementing splicing-deficient HeLa S100 extracts (Lopato *et al.* 1996a; Lopato *et al.* 1999; Kalyna *et al.* 2003). However, the Arabidopsis ASF/SF2 homolog, SRp34/SR1, by itself, was incapable of complementing the S100 extracts but was able to promote splice site switching in mammalian nuclear extracts (Lazar *et al.* 1995). Although some plant SR proteins are capable of complementing splicing-deficient animal cellular extracts, plant introns are not recognized in these systems.

Many functions and interaction capabilities have been attributed to plant SR proteins, mostly based on their capability to interact in yeast two-hybrid screens or *in vitro* protein-protein interaction assays in plants (Graveley *et al.* 2000; Kalyna *et al.* 2004; Reddy *et al.* 2004; Kalyna *et al.* 2006) (Figure 1.4). This plethora of interactions and binding affinities requires SR proteins to be an active participant in many different steps of RNA processing from the initiation of splicing to aiding in translation. Unfortunately, little data is available to confirm these interactions *in vivo*. However, it is widely accepted that both plant and animal SR proteins identify and bind to exonic regulatory sequences, called Exon Splicing Enhancers (ESEs) to promote the use of 3' splice   sites (for further details see chapter four). The mechanism that performs this is quite uncertain and is a popular topic of research (Graveley *et al.* 2000). Some *in vitro* and *in vivo* studies are described in later chapters.

Regardless of how many unique functions SR proteins perform, whether they are actively involved in every step possible in pre-mRNA splicing or only serve to

13
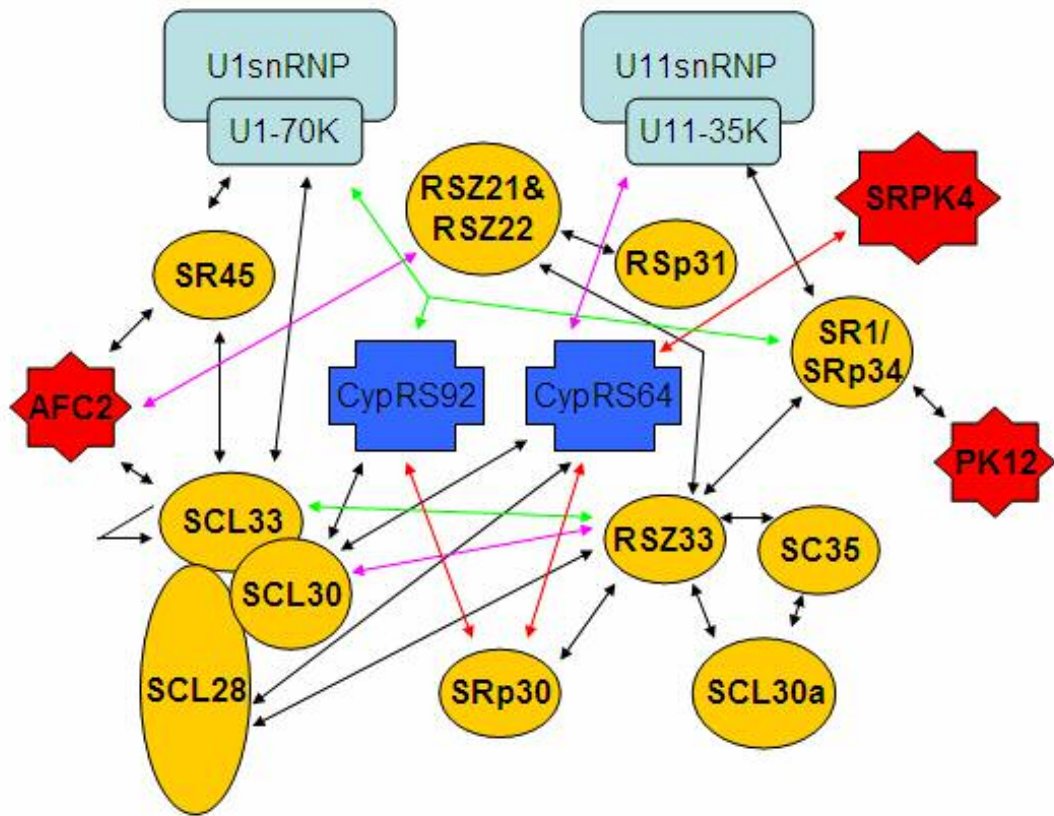
Figure 1.4. Network of interactions among serine/arginine-rich (SR) proteins and between SR proteins and other spliceosomal proteins. These interactions were identified using either yeast two-hybrid analysis and/or *in vitro* protein–protein interaction assays. SRZ21/RSZ21 and SRZ22/RSZ22 are two different proteins that showed similar interactions with other proteins. The interaction of SR33/SCL33 with itself is indicated by an arrow turning back on itself. SR33/SCL33 and RS31 interact only with SRZ21/RSZ21. All SR proteins are shown in yellow. U1 and U11 snRNP proteins are indicated in light blue. Protein kinases and cyclophilin-like proteins are shown in red and dark blue, respectively. SR34, an SR protein with a molecular mass of 34 kDa (also called SR1); SR33, an SR protein with a molecular mass of 33 kDa (also called SCL33 for SC35-Like protein 33); SRZ21 and SRZ22 are 21 and 22 kDa SR proteins with one zinc knuckle (also called RSZ21 and RSZ22, respectively); AFC2, a LAMMER-type protein kinase; PK12, a LAMMER-type kinase from tobacco; CypRS, cyclophilin-like protein with RS domain (Adapted from Reddy 2004). (Golovkin *et al.* 1998; Golovkin *et al.* 1999; Lopato *et al.* 2002; Savaldi-Goldstein S. *et al.* 2003; Lorkovic *et al.* 2004)

facilitate the initial step of pre-mRNA splicing, their contribution to the complexity of metazoan genomes is undebatable.

The SR protein RRM and subfamily organization

All known SR proteins contain two sequence features common to many proteins involved in RNA processing (Figure 1.5). One feature is long stretches of arginine/serine (RS) dipeptide repeats of various lengths, located at the carboxyl-terminus of the protein. This sequence allows for protein-protein interactions within the SR protein gene family, as well as between SR proteins and other splicing factors. (Wu JY *et al.* 1993; Kohtz *et al.* 1994; Graveley BR *et al.* 1998) While it is obvious that the low complexity RS domain of these proteins is of great significance, being conserved through evolution, there is contradictory evidence as to the exact significance of this feature. One group successfully performed domain swaps between mammalian SC35 and ASF/SF2 (Chandler *et al.* 1997). They concluded that RS domains in animals play little, if any, significant role in the specificity of individual SR proteins based on the inability of the engineered proteins to alter the splicing of mRNAs used to assay splicing ability (Chandler *et al.* 1997). However, another group showed that the length of the RS domain did affect functional splicing, but only moderately (Graveley *et al.* 1998). Using RS domains of various lengths tethered to an RNA binding protein, Graveley et al. (1998), determined that the length and number of RS dipeptide repeats had little effect on splicing efficiency. The two experiments were fundamentally different in design and objective, thus making the results difficult to compare adequately. However, recent rice data has also shown that

16

•MSSRSSRTVYVGNLPGDIREREVEDLFSKYGPVVQIDLKVPPRPPGY
AFVEFDDAR DAEDA HGRDGYDFDGHRLRVELAHGGRRSSDDTRGS
FNGGGRGGGGRGRGDGGSRGPSRRSEFRVLVTGLPSSASWQDLKDH
MRKGGDVCFSQVYRDARGTTGVVDYTCYEDMKYALKKLDDTEFRNA
FSNGYVRVREYDSRKDSRSPSRGRSYSKSRSRSRGRSVSRSRSRSR
SRSRSPKAKSSRRSPAKSTSRSPGPRSKSRSPSPRRSRSRSRSPLPS
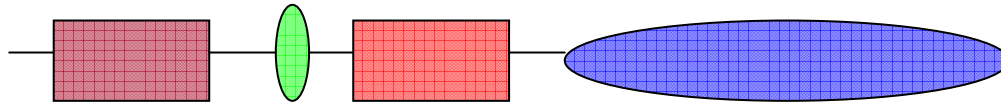
RRM1*   G-rich   RRM2   RS domain

Human ASF/SF2

Figure 1.5 Human ASF/SF2 SR protein. SR proteins are composed of two distinct features, the first class RRM (RRM1), which contains the SR protein defining DAEDA sequence (outlined in box), and an RS domain composed of RS dipeptide repeats of various lengths. SR proteins may also contain either a 2nd class RRM characteristic of that found in ASF/SF2, B52, or a 3rd class of RRM found only in the plant specific RSp31 subfamily, in addition to the first class RRM. SR proteins may also contain zinc fingers and zinc knuckles as well as low complexity domains such as G-rich regions.

17

substituting RS domains between OsRSp29 and OsSCL26 has little effect on the ability of the chimeric SR proteins to affect splice site selection of an alternatively spliced target gene in rice protoplasts (Isshiki *et al.* 2006). It is certain that the RS domains of homologous SR protein gene subfamilies have maintained compositional integrity and are likely to have some degree of effect on the fidelity of splicing, as well as its specificity.

In addition to the RS domain, all SR proteins contain a conserved RNA Recognition Motif (RRM) at the N-terminus of the polypeptide, typically beginning within the first few residues (Figure 1.5). While all RRM domains share some sequence features, there is a unique signature within the first class RRM of SR proteins that distinguishes it from other RNA binding proteins. Located at position 53 of the SR protein consensus sequence is a DAEDA motif that distinguishes SR proteins from other RNA binding proteins (Figure 1.6). In fact, of the hundreds of RRM-containing proteins identified in humans, *Drosophilia*, C. *elegans*, and *Arabidopsis*, a very small fraction belong to the SR protein gene family. Using the first class RRM as the defining sequence feature of SR proteins, it has been found that the human genome contains nine SR protein genes, both *Drosophila* and *C. elegans* contain six, and 19 SR protein genes have been previously identified in *Arabidopsis.*

While the amino acid sequence of each of the SR proteins is similar, certain features within the RRM allow the SR protein gene family to be assembled into subfamilies based on the differences that are conserved among proteins within a particular subfamily (Figure 1.7). These distinguishing features are apparent from the first residue of the first class RRM. For example, the SC35 subfamily has a 5 residue
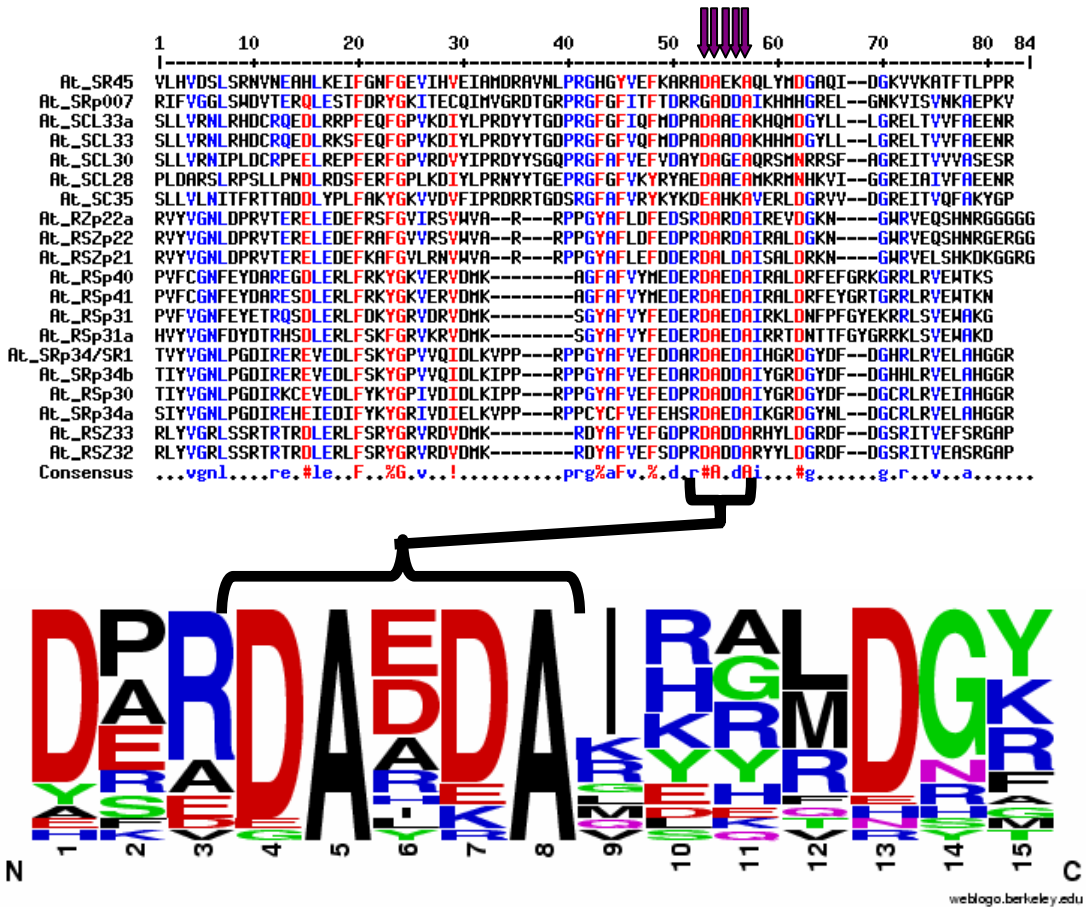
Figure 1.6 Multiple alignment of the first class RRM of all 20 SR proteins and amino acid frequency of the family-defining DAEDA motif. SR proteins RRMs contain residues that are found in all RRM domains, but also exhibit a unique DAEDA signature peptide sequence (arrows) which is found only in SR proteins. Two questionable SR proteins, SR45 and SRp007, do fit the consensus sequence as well as previously identified SR proteins. However, these proteins contain the invariant alanine at positions five and eight in the pictogram in addition to other conserved amino acids at the other locations of the DAEDA consensus sequence.

9G8 Subfamily

```
RVYVGNLDPRVTERELEDEFRSFGVIR--SV---WVARRPPGYAFLDFEDSRDARDAIRE
RVYVGNLDPRVTERELEDEFRAFGVVR--SV---WVARRPPGYAFLDFEDPRDARDAIRA
RVYVGNLDPRVTERELEDEFKAFGVLR--NV---WVARRPPGYAFLEFDDERDALDAISA
KVYVGNLGTGAGKGELERAFSYYGPLR--TV---WIARNPPRFAFVEFEDPRDAEDAVRG
KVYVGDLGNNARKNDLEYVFGAYGSLR--SV---WIARNPPGFAFVEFESARDAADAVRG
KVYVGNLGSSASKHEIEGAFAKYGPLR--NV---WVARNPPGFAFVEFEDRRDAEDATRA
KVYVGNLGSSASKYEIENAFSKYGPLR--NV---WVARNPPGFAFVEFEDRRDAEDATRG
KVYVGGLPSDATSQELEEIFDRFGRIR--KV---WVARRPPGFAFVEYDDVRDAEDAVRA
```

ASF/SF2 Subfamily

```
TVYVGNLPGDIREREVEDLFSKYGPVVQIDL---KVPPRPPGYAFVEFDDARDAEDAIHG
TIYVGNLPGDIREREVEDLFSKYGPVVQIDL---KIPPRPPGYAFVEFEDARDADDAIYG
TIYVGNLPGDIRKCEVEDLFYKYGPIVDIDL---KIPPRPPGYAFVEFEDPRDADDAIYG
SIYVGNLPGDIREHEIEDIFYKYGRIVDIEL---KVPPRPPCYCFVEFEHSRDAEDAIKG
RIYVGNLPPDIRTKDIEDVFYKYGAIRDIDL---KNRRGGPPFAFVEFEDPRDAEDAVYG
RIYVGNLPTDVREKDLEDLFYKYGRIREIEL---KNRHGLVPFAFVRFEDPRDAEDAIYG
RIYVGNLPPDIRTKDIQDLFHKFGKVTFVDL---KNRRG-PPFAFVEFEDARDADDAVKA
KVYVGNLPGDVREKEVEDIFHKYGRIKYVDI---K-SGRGPAFAFVEFEDHRDAEDAVRA
```

SC35 Subfamily

```
SLLVRNLRHDCRQEDLRRPFEQFGPVKDIYLPRDYYTGDPRGFGFIQFMDPADAAEAKHQ
SLLVRNLRHDCRQEDLRKSFEQFGPVKDIYLPRDYYTGDPRGFGFVQFMDPADAADAKHH
PLDARSLRPSLLPNDLRDSFERFGPLKDIYLPRNYYTGEPRGFGFVKYRYAEDAAEAMKR
SLLVRNIPLDCRPEELREPFERFGPVRDVYIPRDYYSGQPRGFAFVEFVDAYDAGEAQRS
SLLVLNITFRTTADDLYPLFAKYGKVVDVFIPRDRRTGDSRGFAFVRYKYKDEAHKAVER
SLKVDNLTYRTSPDTLRRVFEKYGRVGDVYIPRDRYTKESRGFAFVRFHDKRDAEDAMDA
SLKVDNLTYRTTPEDLRRVFERCGEVGDIYIPRDRYTRESRGFAFVRFYDKRDAEDALEA
SLKIDNLSYQTTPNDLRRTFERYGDIGDVHIPRDKYSRQSKGFGFVRFYERRDAEHALDR
```

Figure 1.7 First class RRM alignment of first 57 amino acids in Humans,

Drosophila, C. elegans, **and** *Arabidopsis* **9G8, ASF/SF2, and SC35 subfamilies.**

All SR proteins contain sequences that are highly conserved (in green) throughout the

entire SR protein gene family including the defining DAEDA sequence (last five

amino acids in each sequence). However, there are sequence features within the

RRM that define distinct subfamilies within the SR protein gene family and are

highlighted in blue and yellow in the 9G8 Subfamily, pink and purple of the ASF/SF2

subfamily, and blue and red in the SC35 subfamily.

SLKVD consensus sequence as well as a PRD signature sequence at positions 33-35.

However, the ASF/SF2 and 9G8 subfamilies carry a VYVG motif at positions two

through five and when aligned with the SC35 subfamily, contain a gap at the SC35

PRD location. The ASF/SF2 subfamily contains a PGD motif at positions eight

through ten while the SC35 and 9G8 have unrelated amino acids at that location.

The 9G8 subfamily contains a RSVWV motif at positions 27-31 and an ARR

sequence at positions 37-39, while the SC35 and ASF/SF2 subfamilies contain

unrelated residues at the same location. Not only are these sequences consistent

within an individual organism, but they are also conserved between plants and

animals (Figure 1.7).

Additionally, not only is the peptide sequence highly conserved between

species, but the manner in which the SR proteins themselves undergo alternative

splicing is highly conserved. A recent study by Iida *et al*. (2006) analyzed the peptide

structure of three different SR protein subfamilies: SC35-like (which the authors

distinguish as a separate subfamily from SC35), ASF/SF2, and 9G8. The latter two

subfamilies are present in animals, and all three subfamilies occur in *Arabidopsis*,

rice, and moss. They found that the peptide sequence and the different alternatively

spliced isoforms are conserved (Iida *et al.* 2006). When analyzing the exon/intron

boundaries of the SC35, SC35-like, and the 9G8 subfamilies within *Arabidopsis,*

moss, and rice, they found that the exon/intron boundaries of each respective

subfamily were conserved across species, and the exact nucleotide location of the

alternative splice sites were also conserved. The Barta lab found similar results

looking at conservation of exon/intron boundaries and alternative splicing of long

exons in green algae and flowering plants (Kalyna *et al.* 2006). Kalyna *et al.*

performed a similar analysis using *Arabidopsis*, rice, maize, and alga and found that

the exon/intron boundaries of RSp31 and RSp31a were highly conserved.

Additionally, alternative splice sites within these organisms were conserved to the

exact nucleotide location. The divergence between moss and flowering plants

occurred over 400 million years ago (Nishiyama *et al.* 2003) and the evolutionary

pressure for such a conservation must have been similar in order for all plant species

to maintain the exact same subfamilies, further suggesting the importance of not only

SR proteins as a whole, but also the division of these protein genes into a specific

organizational system.

The distribution of subfamilies among different organisms varies greatly

depending upon the specific subfamily (Figure 1.8). Some subfamilies such as the

SC35 and ASF/SF2 are present in both plants and animals while plant specific

subfamilies, RSZp22 and RSZp33 for example, and animal specific subfamilies, B52

and SRp20, exist. Some subfamilies are even more complicated when they consist of

proteins which are very similar, or related to, a core SR protein found in multiple

organisms. For example, a single core SC35 SR protein is found in humans,

nematodes, fruit flies, and *Arabidopsis*. However, while those organisms encode a

single SC35 SR protein, the *Arabidopsis* genome contains an additional four SC35-

like (SCL) protein genes. Unlike SC35, which is found in most eukaryotic systems,

there are subfamilies that appear to be kingdom specific. For example, while you will

find B52 orthologs in animals, you will not find one in plants. Conversely, there are

three subfamilies in plants (RSZp22/22a, RSZp33/33a, and the four member RSp31

SC35--core	1	1	1	1
related	0	0	0	4
B52--	3	3	1	0
ASF--	2	1	1	4
SRp20--	2	1	3	0
atRSZp33--	0	0	0	2
atRSZp22--	0	0	0	4
atRS31--	0	0	0	4

Figure 1.8 The distribution of SR protein subfamilies among different organisms. Some SR proteins are conserved in both plants and animals such as the SC35 core SR protein and ASF/SF2. However, the B52 and SRp20 subfamilies are found only in animals, while the RSZp33, RSZp22, and RS31 subfamilies are found only in plants.

subfamily) that are not present in animals. (Figure 1.8) In *Arabidopsis*, there are a total of seven subfamilies of SR protein genes, five of which are composed of multiple members, and the remaining two subfamilies, SR45 and SRp007, are composed of only a single SR protein gene (Figure 1.9).

In addition to the family-defining first class of RRM, the ASF/SF2 and RSp31 subfamilies contain additional conserved classes of RRM in addition to the first class RRM. These SR proteins have two RRMs. The second class of RRM is found in the ASF/SF2 and B52 subfamilies while the third class is found in the plant-specific RSp31 subfamily. While not all of the two-RRM SR protein genes have been analyzed at the molecular level in animals, it has been found that both domains of *Drosophilia* ASF/SF2 and B52, which is orthologous to mammalian SRp40, SRp55, and SRp75, are required for specific splicing (Tacke *et al.* 1995; Shi *et al.* 1997). However, the third class RRM in the rice SR protein gene, RSp29, is totally dispensable for regulation of the splicing of target genes in protoplast co-transfection assays (Isshiki *et al.* 2006). Other distinguishing features exist within SR protein subfamilies maintaining sequence integrity across kingdoms. For example, zinc fingers and zinc knuckles are found in the 9G8-like subfamily as well as a plant specific subfamily. Low complexity domains such as glycine-rich regions, proline and serine rich regions, and PRG repeats are also subfamily specific. (Figure 1.9)

Introduction to SRp007

Using the conserved sequence motifs of the first class RRM, 19 SR proteins have been previously identified in *Arabidopsis thaliana* (Kalyna *et al.* 2004; Wang *et al.* 2004). One *Arabidopsis* gene, At1g60650, was omitted from previously published
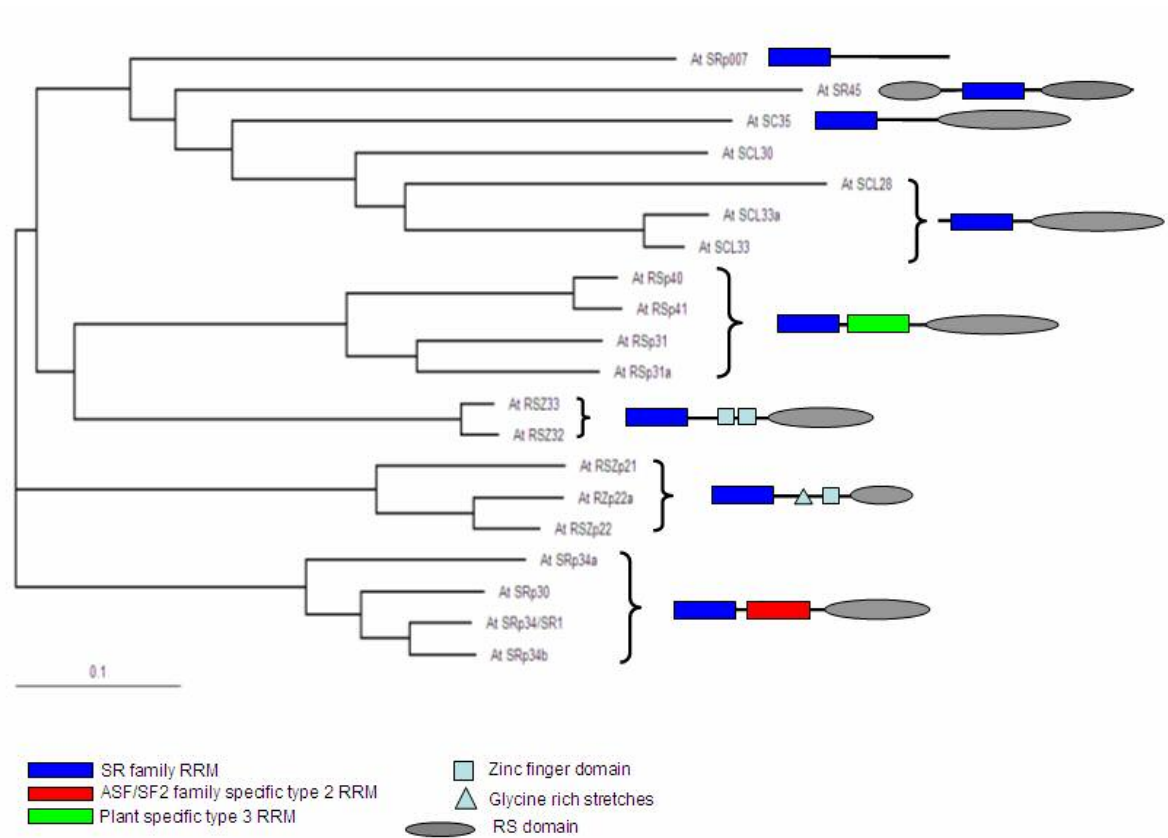
Figure 1.9 SR protein phylogenic tree based on first class RRM. All SR proteins consist of a first class RRM unique to only the SR protein gene family. In addition to the first class RRM (blue box), SR proteins also contain of a C-terminal RS dipeptide repeat domain of various lengths (grey oval). Subfamilies also have unique features such as a second class RRM (red box) found in the ASF/SF2 subfamily, a third plant specific class of RRM (green box), zinc fingers (light blue square), and low complexity domains such as glycine rich regions (light blue triangle).

lists of SR protein genes. This omission is understandable, for the protein resulting

from this gene lacks an RS domain. In addition, the RRM of this protein only loosely

resembles that of other benchmark SR proteins. However, the most highly conserved

residues, the DAEDA consensus sequence within the first class RRM that defines the

SR protein gene family, and is only found in SR proteins, is found within this gene.

Additionally, several other invariant residues found within the SR protein consensus

sequence are also present in the predicted sequence of At1G60650. In fact, the

alignment of the 19 previously identified SR protein genes and At1G60650 clearly

shows that At1G60650 more closely resembles the consensus sequence, including the

invariant and significant peptides, than other known SR protein genes. (Figure 1.5)

Based on the conserved peptides and overall similarity to known SR proteins, I

propose that At1g60650 be considered the 20th *Arabidopsis* SR protein gene. Due to

its elusive nature in earlier publications, as well as its ability to masquerade with the

SR protein gene family, I have named this gene *SRp007*.

While the inclusion *SRp007* in the SR protein gene family may be debatable,

it is not the first gene to incompletely assimilate with the consensus peptide structure

of the SR protein gene family. It is also necessary to point out that *SR45* is unique to

the SR protein gene family for several reasons. SR45 is the only known SR protein to

have an N-terminal RS domain as well as the usual C-terminal domain. The *SR45*

gene is absent in animals, but has two homologs in rice (Iida *et al.* 2006), and may

therefore have some plant specific functions that require the dual RS domain. In

addition to the unique RS domain arrangement, *SR45* is highly related to another

family of RNA processing genes. No other plant or animal SR protein shows this type

26

of relationship to another gene family. Additionally, there are groups of SR protein-like splicing factors, such as RNPS1 in *Arabidopsis* and *FUS* gene family found in humans, which inhibit splicing. SR45 and SRp007 could have diverged from and are also similar to this family of splicing regulators.

This phenomenon may be partially explained by a few theories. In addition to possible non-splicing duties, *SR45* may play a very minor role in plant specific RNA splicing. However, the function *SR45* plays is important enough that evolution has selected for the retention of this functional protein to act as an SR protein. This is seen by the effect of the mutant on plant development and its ability to alter splicing which will be documented in subsequent chapters. Conversely, this effect could be an instance similar to that seen with *SRp007*: *SR45* could have been a functional or non-functional member of another family of proteins but evolved to take the form and possible functions of a SR protein. Other proteins involved in RNA processing contain RS domains similar to those seen in SR proteins in addition to RNA binding domains. With all of the structural similarities to SR proteins and the lack of a close homolog, it is very likely that this is an instance of an unrelated protein having gone through selection and now, structurally and functionally, has become a bona fide SR protein gene.

SRp007 could be an example of similar situation which is observed with SR45, in that it could have been a member of another protein gene family, but has evolved to acquire the functions of an SR protein gene. Unlike any other SR protein, SRp007 is unique in that it lacks the RS domain from which the name of the gene family is derived. However, the highly conserved RRM is still present and is

potentially capable of performing RNA binding functions. The lack of an RS domain poses a fundamental problem in that SR proteins are thought to bind to ESEs and through their RS domains, interact with other factors to facilitate splicing. While probably incapable of recruiting the same factors with which *bona fide* SR proteins interact, SRp007 may still bind to the same regulatory sequences. This could be significant in that it could serve the function of displacing the antagonistic hnRNP proteins, a factor that competes with SR proteins and inhibits proper splicing. SRp007 may function as a temporary competitor of hnRNP proteins, or other splicing inhibitors, and then be displaced by another SR protein capable of recruiting other splicing machinery. Similarly, the SC35-like SR proteins are similar to human SRp38, which belongs to the SR protein family, but inhibits splicing. SRp007 may have an RRM similar to that of SR proteins, but because it may have a different function such as negatively regulating splicing, may have no need for the RS domain found in SR proteins. SRp007 does contain a long peptide stretch at the carboxyl terminus of the protein for which no function is known. It is quite possible that this stretch of the peptide does, in fact, perform a function in splicing unlike any of the other known SR proteins. Of course both of these hypotheses, blocking antagonistic proteins from binding to pre-mRNA and effectively not allowing binding sites for SR proteins to facilitate alternative splicing, or providing some unknown function in splicing, are purely speculation and no experimental evidence exists to date to support them. However, determining if SRp007 has any role in alternative splicing could lead to a whole new classification of SR proteins and potentially aid in understanding the exact mechanism of constitutive and alternative splicing.

Chapter 2: Mutations in *Arabidopsis thaliana* SR protein genes: phenotypes

Introduction

Analyzing the phenotypes of viable SR protein gene mutants *in vivo* has proven difficult. In mammalian systems, the Fu lab found that mutations in two genes, ASF/SF2 (Xu *et al.* 2005) and SC35 (Wang *et al.* 2001), as being early embryonic lethal (Reviewed Mount 1997). However, targeting heart tissue with a non-functional *SC35* gene proved more rewarding, as mice lived a normal lifespan while only being afflicted with dilated cardiomyopathy (Ding Jian-Hua *et al.* 2004). The *Drosophila* B52 gene was found to be essential for viability (Ring HZ *et al.* 1994; Peng *et al.* 1995), as was ASF/SF2 in chicken B-cell lines (Wang J *et al.* 1996). Deletion experiments of SRp20 in mice resulted in lethality at the blastocyst stage (Jumaa H *et al.* 1999). The Caceres group (Longman *et al.* 2000) used RNAi to knock down individual and multiple SR protein genes in *C. elegans*. Their study revealed similar results to those seen in mice and chicken cell lines with respect to ASF/SF2, producing an early embryonic phenotype. However, knocking down any of the other SR protein genes individually, or in any other combination, produced viable animals with a broad range of phenotypes (Longman *et al.* 2000).   To date, there have been no published data of knockouts or knockdowns of *Arabidopsis thaliana* SR protein genes. However, the Barta lab has performed over-expression experiments by fusing cDNA of several SR protein genes to the 35S promoter

resulting in severe developmental abnormalities in both embryos and young plants suggesting possible roles for individual SR protein genes in *Arabidopsis* (Kalyna *et al.* 2003).

It may be more fruitful to identify mutants in *Arabidopsis* than in animals for several reasons. One potential limiting factor in the inability to identify null viable mutants in mammalian systems is that there are a total of only nine SR protein genes in a genome consisting of 20,000-25000 genes. *Arabidopsis,* on the other hand, encodes over twice as many SR protein genes and has a genome encoding approximately 28,000 genes of which substantially fewer genes are predicted to undergo alternative splicing. While there are sure to be specific classes of ESEs in *Arabidopsis* that are recognized by specific SR proteins, there is a possibility of some overlapping functions within the *Arabidopsis* SR protein gene family that will permit viability in mutants. In addition to being a much less complex organism in comparison to humans, the *Arabidopsis* genome has undergone at least two duplications in evolution. Because of the duplication of the *Arabidopsis* genome, there may be fewer unique targets than the genome size indicates. Additionally, the large number of SR proteins in *Arabidopsis*, in comparison to animals, may result in some redundancy of functions or specific gene targets. This may allow plants to avoid lethality in the presence of a SR protein gene mutant, a fate that mammalian mutants are not able to evade.

In addition to the biological and molecular advantages *Arabidopsis* has over its animal counterparts, a vast collection of T-DNA insertion lines are already available, making *Arabidopsis* an attractive model system for genetic studies. The

Salk T-DNA insertion collection consists of over 200 unique lines in which mutations occur in 19 of the 20 SR protein genes (Alonso J.M *et al.* 2003; SALK *et al.* 2007). Analyzing these insertion lines may allow us to identify mutants with viable phenotypes and provide plants for investigation of molecular interactions between SR proteins and other splicing factors of whole organisms *in vivo*. In this chapter, I will report the identification of five mutant SR protein gene plant lines of which three are viable and two are lethal. The three viable mutants have a range of phenotypes ranging from a very mild delayed bolting phenotype to more severe developmental and morphological abnormalities.

<u>Results</u>

In order to identify possible SR protein gene mutants, I have obtained over 125 unique T-DNA insertion lines containing reported insertions located in the promoter, 5'UTR, or exonic regions in 19 of the 20 SR protein genes with RSp40; the only SR protein gene lacking an insertion within the described locations. Additional lines are available with reported insertions in introns and 3' UTR regions. However, insertions located within the 3'UTR or introns are unlikely to affect transcript production, and thus were not chosen for this study. Of the 125 Salk T-DNA insertion lines available at the initiation of this project, I have examined 45 Salk lines to various degrees (Table 2.1). Lines were initially screened for kanamycin resistance (the marker for T-DNA insertions). After the kanamycin selection, plants were genotyped via PCR an some products were clones and sequenced. From the 45 lines

Locus ID	Gene Name	Insertion Line	Status	Insertion location
At5g64200	SC35	SALK_080090.50.05.x 3'UTR	#	
		SALK_124563.49.55.x intron	# ¥	tcaat/tcaacatgaagcttctttct
		SALK_033824.31.85.x exon	# β ¥	tcaat/tattcgcaaagtatggaaag
At3g61860	RSp31	SALK_029586.34.15.X exon	#	
		SALK_014656.36.25.n 5'UTR	!	
		SALK_085956.38.20.x 5'UTR	# α ¥	tcaat/ttgacttaaaaaagcccgtc
At5g52040	RSp41	SALK_063076.37.55.x exon	#	
		SALK_035417.55.00.x promoter	# α ¥	tcaat/ttgctttacaccacaactag
		SALK_095222.12.80.x promoter	#	
At1g09140	SRp30	SALK_055239.38.25.X intron	#	
		SALK_116746.31.95.x 5'UTR	# α	
		SALK_116747.46.00.x 5'UTR	#	
		SALK_132986.49.85.x 5'UTR	€	
		SALK_029105.23.05.x promoter	# β ¥	tcaat/gtgtttccgatctagtgtct
At4g02430	SRp34b	SALK_055412.55.00.x exon	# ¥	tcaat/cagctacgaggacatgaaat
At1g02840	SRp34 (SR1)	SALK_102166.53.25.x exon	# ¥	
		SAIL_146_F04 5'UTR	€	
		SALK_010894.53.10.x 5'UTR	# α ¥	tcaat/tatatactatagtgtggcga
At3g49430	SRp34a	SALK_087841.52.95.x promoter	#	
At1g23860	RSZp21	SALK_114234 5'UTR	# ¥	ctaat/ttaccgaaaaataagggacc
		SAIL_172_B08 5'UTR	€	
		SAIL_614_B04 5'UTR	€	
At4g31580	RSZp22	SAIL_196_D11 promoter	#	
		SALK_012172.33.50.x 5'UTR	#	
At2g24590	RSZp22a	SALK_023090.18.65.x exon	#	
		SALK_094265.53.20.x 5'UTR	€	
		SALK_120507.26.50.x 5'UTR	@	
		SALK_120506.42.10.x 5'UTR	#	
		SALK_120497.38.80.x 5'UTR	€	
At3g55460	SCL30	SALK_029353.56.00.x promoter	# ¥	tcaat/ttgtctaagttgtgttagaa
		SAIL_113_A11 5'UTR	€	
At2g37340	RSZ33	SALK_083782.51.65.x 3'UTR	#	
		SALK_051523.50.50.x promoter	€	
		SALK_051610.50.50.x promoter	Ω	
		SALK_028009.56.00.x promoter	€	
At3g53500	RSZ32	SALK_031147.49.40.x promoter	@	
		SALK_076681.54.00.x promoter	@	
At1g60650	SRp007	SALK_042217.55.75.x promoter	#	
		SALK_001328.41.55.x exon	# ¥	tcaat/ccgcttcgacagtggagaca
At1g55310	SCL33	SALK_071319.26.30.x promoter	@	
At3g13570	SCL33a	SALK_041849.55.00.x exon	# ¥	tcaat/cagtttggtcccgtcaagga
At5g18810	SCL28	SALK_061417.21.45.x promoter	! ¥	
		SALK_144574.54.50.x promoter	€	
At1g16610	SRp45	SALK_004132.54.75.x exon	# β ¥	tcaat/cctgattctccccatcgccg
		SALK_152739.32.15.x promoter	Ω ¥	tcaat/atttatactttataaaatgc

tcaat is the last 5 nucleotides of left border
 # PCR verified homozygous viable

! Homozygous lethal insertion
@ Kanamycin resistant but unable to genotype via PCR
α RNA produced at WT levels
β RNA levels significantly knocked down
€ Kanamycin sensitive plants, no PCR performed
Ω Heterozygous only plants found. Homozygous lethality not certain
¥ insertion position independently sequenced

Table 2.1 SALK T-DNA insertion lines screened examined in this study.

Name	forward	reverse	promoter forward	promoter reverse
SC35	TCCAATTGCACAAATAACGGAA	TGTTAGAGATGTACCGGTGTTTGTA	TTTTCGTTTGTTACGTGATTTCCCC	CTCCCACTAGTTCACCTAATCATCATA
SRp102/RSp31a				
RSp31	CAAGATGATAAAGGCAGGTATGAA	TTGTGGCAAAAAGAGTGAAGA	TTAGCCGTTGGATCTGAAGCTCAGT	CTGCTGCATATCTCGTTTACCCTAGAA
RSp40				
RSp41	TTGGTTGAATTGAAATGGATAA	AAACTTCTGATCTTGTCGACCATC	TGGGGAGAGAAGTTTAATCTGGTGA	TATCAACCTCTCAACCTTGCCGTA
SRp30	TCTGGTCTGGTCTGGTACTCC	GTTTAGTTTAGGATAATCATAAAAGACAC		
SRp34b	GGTCGTGGCGGTCGT	AGCATACCCTCGTGGAGACA	CCTCACTCACATCAGAAGCCAATT	TAACCATCACGGCCATAAATTGCAT
SRp34 (SR1)	GATTAACGAAACACAAATAGGAGGA	AAAGACATATCTCAAGGCAAAAGA	TCTCAGCATTTAAAAGATGTGGAGG	CTCCTTTACGCATGTCATCCTAATGTA
SRp34a			CGGTACGGATTCATTCACCTCTCT	AGCTTTCCGCATATGATCCTGAAAA
RSZp21			GTCGATTTTGTAAAAGGAGGTAAGTG	GAGAAAAGTACTTTTACCCATTGGC
RSZp22	CAACCTGCTTTACGGTTAGG	TGTTCAACCCTCCAGCCATT	ATCATGCTATACTGGGCGATATTGC	ATTAGCATATGGCACTTCCTCACGC
RSZp22a	AGACAAGTAACACTCAGCTCCG	TTACTGAACGTGAACTCGAGG	CTGACCAAACTGAACCGTATCAAAA	GGGCATCTCTTGAATCTTCAAAATC
SCL30	CTGAATGGAAGCAGTAATAAGTGAAG	TGAGAACAAAAAATCATAAACATACTC	TTGTAATATTGTGCGAAGTGATTTTGA	TGATTTGAAATGACCCTTAATATCCTT
RSZ33	ACGTAATGTGAAGTTGAGAAACACAC	TCTGTATTTTGATATTTGTGTTCTTGA	GGATAATAGGCGGACTTTGATTTGTTG	AAAGTCTCGTCCGTCCAGGTAATGT
RSZ32			ATTGTTTTCGCTCTCTAGATTAGGGA	GGGAAAATAGTTGTCTACGGGACATG
SRp007	TTCATCTCGAGGGAAGGG	GATTTGAGATTGCATTACCTTTTTG		
SCL33	ACTCCTTGTCCTTTTCTTCAGG	GCAACTATCATAATAACCTCAGATTAT	TATTTCCCTCCCTAAAGGTTTCAAG	CAACTTCACTTCAATTCTTCACTACCA
SCL33a	TCCAAAGAACTCTAATTACTTCATTG	AAGGAGCTACACGCCATCA	TAGATATTTTCATTTTTTAATCAACCA	GGTCTCTTCAACTCTCTAGATTCACTC
SCL28			GAGTCAAAGAATGAACTAACAATGG	GCAAACCAAAAAGCTACTTACAACG
SRp45	TCTCTTTCCAATTATGAAATGCAC	AGAGAGACCCCAGAAAAAAGTTTA	AGCATCAAGTTGACTAAATCACTAGC	CGCAATTTCGACATGTATAACTTCACG
LBb3	GCTTCGTCGAACTCTCTCAG			

Table 2.2 Gene-specific primer pairs used for genotyping SALK T-DNA insertion lines. Gene-specific primer pairs were designed to amplify approximately 500bp upstream and downstream of the reported SALK T-DNA insertion site for each respective gene in addition to the T-DNA specific primer LBb3.

GENE	Upstream Forward Primer	Upstream Reverse Primer
SRp30	ATGAGTAGCCGATGGAATCGTACGA	TTCAGAGAAGCAGACATCTCCAGCT
SC35	TCGGAAGGTCAGGTCCA	GCTACGCCTTTCAGGAGACAT
SR45	TGTTAATCTTCCAAGAGGACATGG	GTCTCCACCTCCAAGGAGACT
SRp34b	TTGTTTCGATCTATTGACGTTTCG	TCCTTTACGCATGTGATCCTT
RSp41	ATGAAGCCTGTCTTTTGCGGAAACT	GTGCCTCATATTGGATGAATGCAAA
RSp31	GAGCGTAGTAGTCGTCGTCGTC	GGAATTGCCGACGAACACTGG
RSZp22	TGTGTACGTCGGAAAATTTGGA	TCTTGCACGAGGGCTGTAACT
SCL30	ATGAGGAGATACAGTCCGCCTTATTA	AGAGATTGAACGTGAACGAGACCGT
SCL33a	TTCAGTAATTGTTTTATGGCTTGG	ATGACGTAAGTTGCGAACCAA
SR1	AAAAAGCTCGACGACACAGAG	TGGAGACGGTGACCTTGAC

	Downstream Forward primer	Downstream Reverse Primer
SC35	TACCTTCCTAGGGATTACTATACTGGA	TGCGAACCCTTGAGCCATTAT
SR45	GGCGGTGAACTGCTGCT	GTCTAGGAGGTGTATCACCGC
SRp34b	TACGAGGACATGAAATATGCGC	TGAAAAAATTTCCAAAGGAAGA
RSp31	AGTCGGATCTGGAACGGTTGT	CTTTCAACGCATACTCCACGG
SR1	AAAAAGCTCGACGACACAGAG	CGATGGACTCCTAGTGTGGA

Table 2.3 RT-PCR primers for verification of loss of transcript production in SALK T-DNA insertion lines. Plants homozygous for a T-DNA insertion were analyzed for RNA production. Gene specific primers were designed to amplify the transcript which codes upstream or downstream from the insertion location.

that were screened, only five resulted in visible non-wild-type-like phenotypes. Gene

specific primer pairs were designed to amplify a region that included 500bp, both

upstream and downstream, of the reported insertion site (Table 2.2). Additionally, a

T-DNA-specific primer pair was designed approximately 150bp from the end of the

left border, which would be used in genotyping all of the individual Salk lines. To

genotype plants from a specific Salk line, a multiplex PCR was performed containing

the two respective primer pairs in addition to the T-DNA-specific primer LBb3

(Table 2.3).

Plants that carried the wild-type allele produced a band of approximately

1000bp. Additionally, plants carrying the T-DNA insertion allele produced a band of

only 650 bp. In a single PCR reaction, both wild-type and insertion alleles can be

identified. Insertion lines that were successfully genotyped as homozygous for their

respective T-DNA insertion were then sequenced to verify the insertion location. To

examine the effect of T-DNA insertions on gene expression, RT-PCR was performed

on viable plants identified as homozygous for the T-DNA insertion to determine if the

insertion alters mRNA production from the respective SR protein gene (Figure 2.1).

This screening process produced three unique viable lines, each containing an

insertion in either *SC35, SRp30*, or *SR45* protein genes for which we were able to

identify plants homozygous for the respective insertion. The resulting plant lines will

be referred to as *sc35-1*, *srp30-1*, and *sr45-1*. In addition to the three viable mutants,

this screen also identified an additional two lines with insertions located in SR protein

genes, SCL28 or RSp31, in which a homozygous insertion located in sequences

upstream of the coding regions in either gene results in embryonic lethality
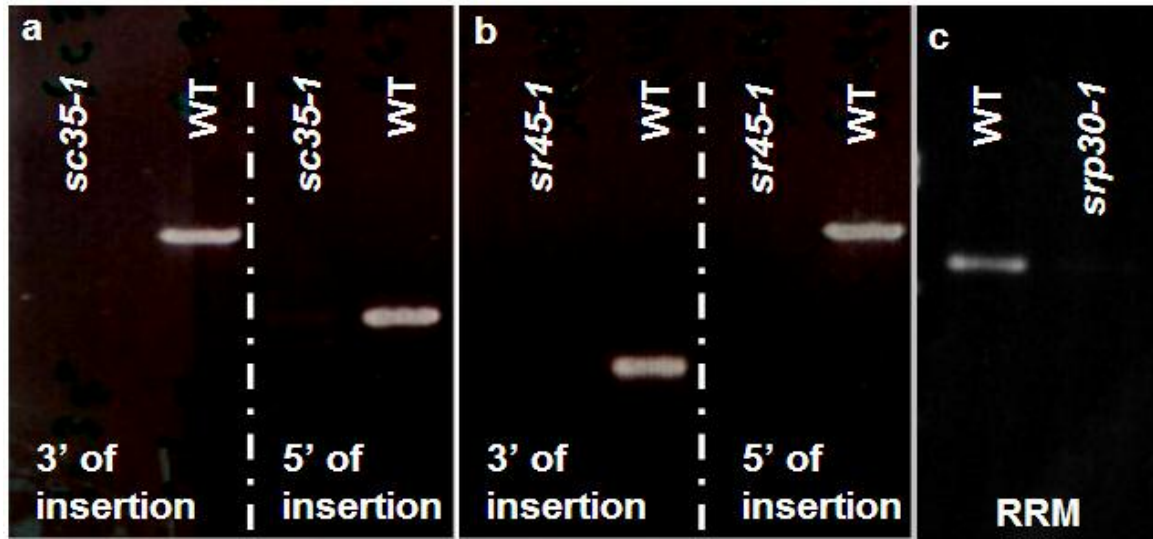
Figure 2.1 T-DNA insertions in *SC35, SR45*, and *SRp30* disrupt RNA

production. Plants homozygous for the Salk T-DNA insertions Salk_033824 and

Salk_004132 are located in the coding region of SC35 and SR45 respectively. Primer

pairs were designed to amplify regions both 3' and 5' of the insertion location to

determine if the insertion disrupted the entire gene function (Table 2.3). Both SC35

(a) and SR45 (b) do not produce transcript when homozygous for their respective T-

DNA insertion. Salk_ 029105 is located upstream of the SRp30 3'UTR. For that

reason, only a single primer pair was designed to amplify the sequence that codes for

the RRM. An agarose gel shows that a very small amount, roughly 5%-10%, of

transcript is still being produced in this insertion line (c). Tissue samples were taken

from whole plants at the 6-8 leaf stage.

(Figure 2.2). These alleles will be referred to as *scl28-1* and *rsp31-1* respectively.

SCL28: *scl28-1* is embryonic lethal

A single T-DNA insertion line, Salk_061417, located upstream of the coding sequence, has been identified in the SR protein gene SCL28. Plants identified as heterozygous for the insertion and wild-type allele were allowed to self-fertilize. Developing siliques, which resulted from this self-fertilization, were opened and seeds contained within were examined. Of the 1017 seeds that were scored, 256 (25.2%) displayed a phenotype common to early embryonic lethality (Figure 2.3). In young siliques, seeds doomed for early embryonic lethality were white in color but otherwise indistinguishable from wild-type seeds. As the siliques age and begin to senesce, the white seeds dimple, shrink, and eventually become semi-transparent. The viable seeds progress to maturity. Genotyping viable seedlings by PCR resulted in a 1:2 ratio of wild type (266 plants) to heterozygous (495 plants) for the insertion.

Plants heterozygous for the *scl28-1* allele were transformed with a *SCL28* transgene composed of *GFP* fused to the carboxyl-terminus of a PCR amplified *SCL28* native gene. T2 plants homozygous for the T-DNA insertion and containing the transgene were allowed to self-pollinate. Green siliques were opened and 100% of the seeds were observed to be green. Seeds removed from siliques that have senesced were identical in shape, size, and color to that of wild-type seeds. Plants identified as homozygous for the *scl8-1* allele and containing the SCL28:GFP fusion transgene displayed no visible phenotype and were wild-type in appearance.

a

SCl28
! SALK_061417

SC35
SALK_033824

SR45
SALK_004132

SRp30
SALK_029105

RSp31
!# SALK_014656

! Embryonic lethal
insertion location sequenced
  physical Insertion location

b

SC35
SALK_033824

SR45
SALK_004132
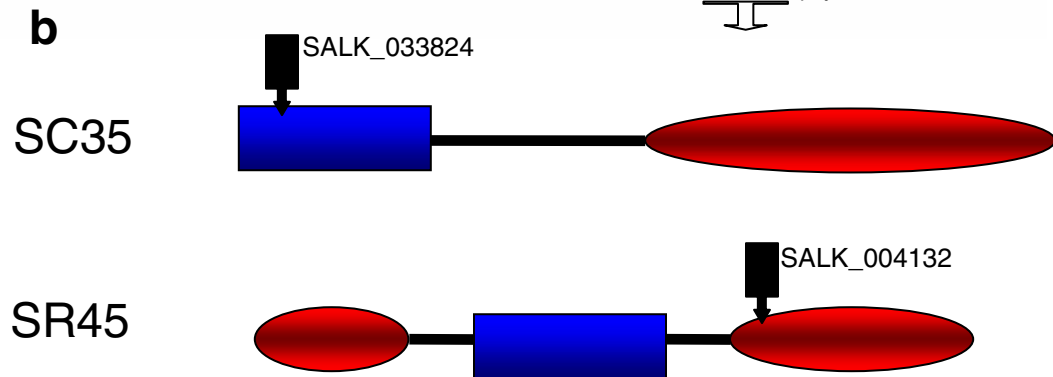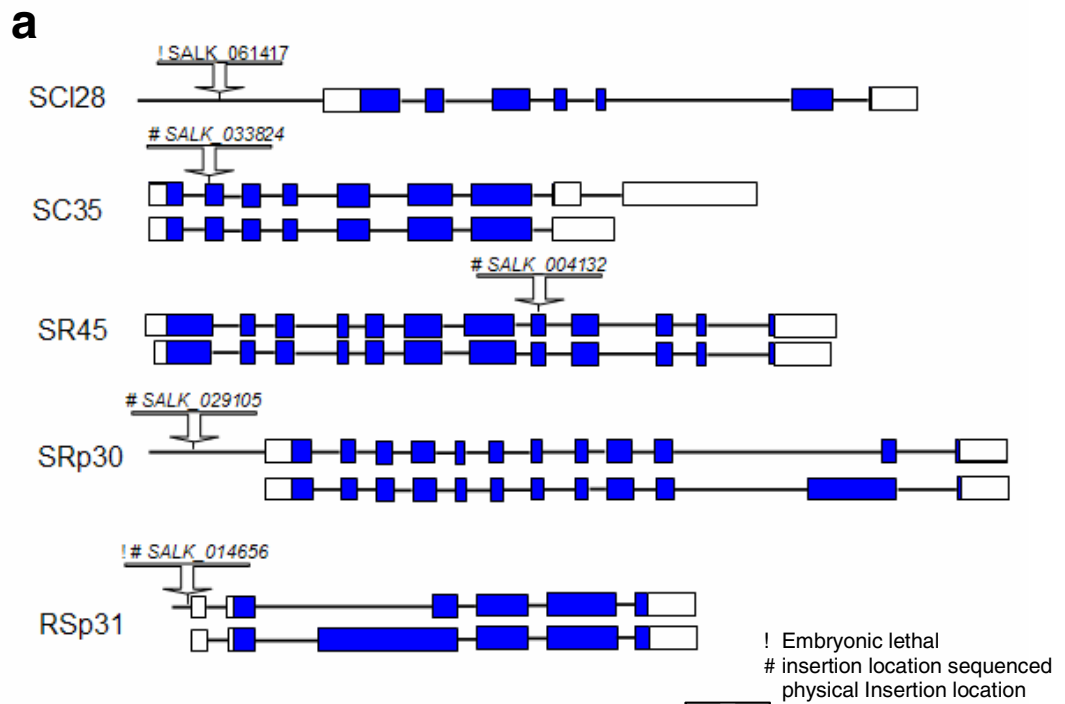
Figure 2.2 T-DNA insertion locations of 5 SR protein gene mutants.

SALK T-DNA insertion lines were screened for phenotypes and individual lines were

identified as null, embryonic lethal, or having RNA levels significantly knocked

down. Insertion locations were sequenced for 4 of the 5 lines to verify reported

insertion position. (a) Gene structures consist of UTRs (white boxes), exons (blue

boxes), and introns (dashes). Genes that produce alternatively spliced isoforms are

indicated by all confirmed existing structures. (b) Protein structures are represented

for SC35 and SR45, the two SR proteins where an insertion is located within the

coding region of their respective gene. The T-DNA of line SALK_033824 is located

within the RRM of SC35 while the T-DNA of line SALK_004132 is located within

the second RS domain of SR45.
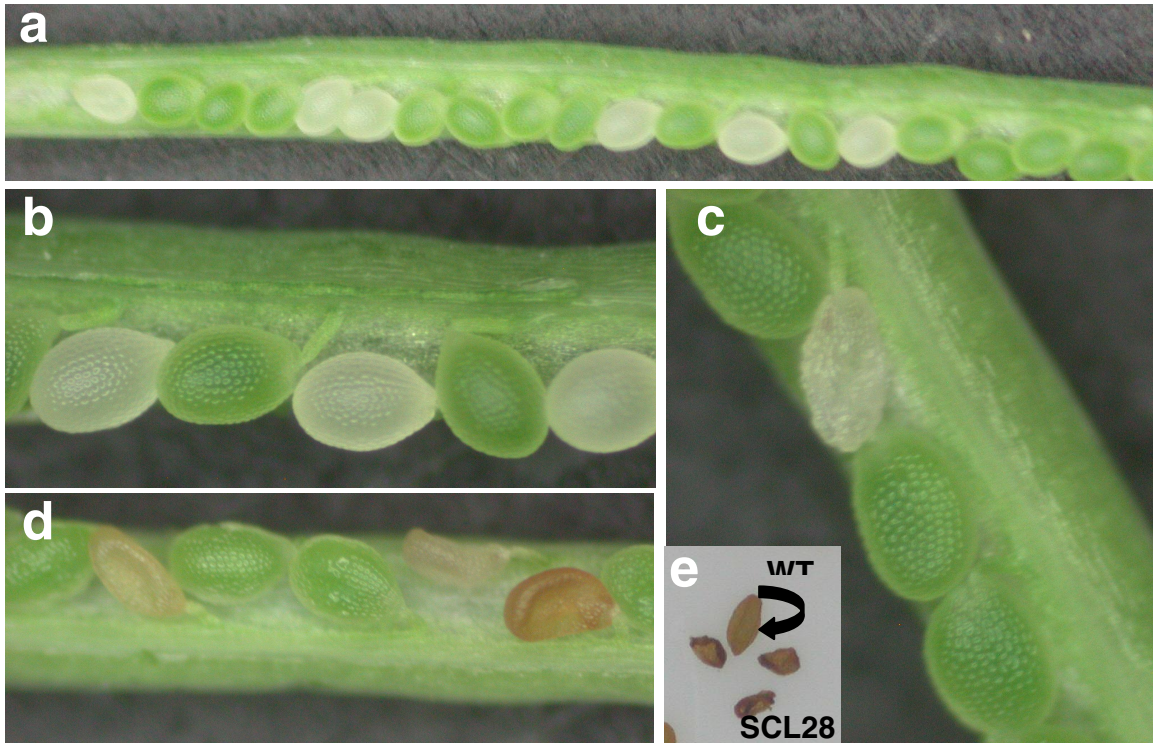
Figure 2.3 *scl28* mutants are embryonic lethal and seeds display a classic embryonic lethal phenotype. Plants heterozygous for the SALK T-DNA insertion line 061417 were allowed to self pollinate and resulting siliques contained white seeds and green seeds (a, b) with a ratio of 1:3 respectively. As the siliques aged, the white seeds began to dimple and shrink (c) and eventually turn brown and "deflate" (d). In addition to ¾ of seeds wild type in appearance, Mature siliques resulted in ¼ dead seeds which are a fraction of the size of wild-type seeds and are almost transparent (e).

RSp31: Homozygous *rsp31-1* fails to germinate

The *rsp31-1* homozygous mutant plants (containing the Salk_014656 T-DNA insertion in the promoter region of the *RSp31* gene) fail to germinate. Seeds from heterozygotes plated on MS media or planted in soil have a germination rate of approximately 75% of a total of 295 seeds germinated from a total of 389 planted. Genotyping by performing PCR on the viable seedlings identifies a ratio of 1:2 homozygous wild-type (101 plants) to heterozygous for the mutant allele (194 plants). No plants homozygous for the T-DNA insertion were found. Unlike the *scl28-1* mutant, seeds produced by a self-fertilizing heterozygous plant display no externally visible phenotype. Green siliques, when opened, are full of green seeds with no empty spaces or other evidence of an early embryonic lethal phenotype. I conclude that rsp31-1 seeds are visually indistinguishable from wild-type seeds at all stages of development up to the point of germination.

SRp30: Homozygous *srp30-1* plants show developmental defects

Salk line, Salk_029105, contains a T-DNA insertion located in the promoter region of SR protein gene *SRp30* approximately 300bp upstream of the 5' UTR. Homozygous viable mutant lines have been identified, and RT-PCR analysis has confirmed the transcript level of *SRp30* in rosette leaves is greatly reduced. Visualizing band intensity on an agarose gel reveals that *srp30-1* mutants produce roughly 5%-10% of the level of transcript seen in leaves of wild-type plants (Fig2.1c) a dramatic loss of transcript leading to functional protein. Although not a null mutation, srp30-1 results in viable plants with a multitude of phenotypes. At the 8-leaf stage, the rosette

a **b** **c** srp30-1 **d** WT
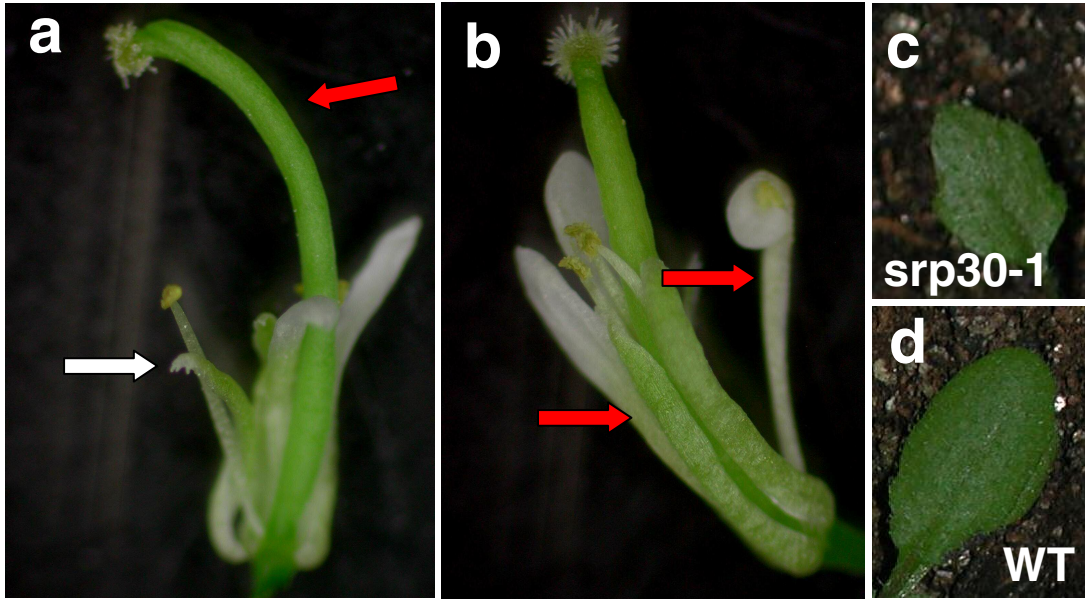
e SRp30 vs. WT rosette diameter from +/- selfing plant at 19 days, 8-10 rosette leaves
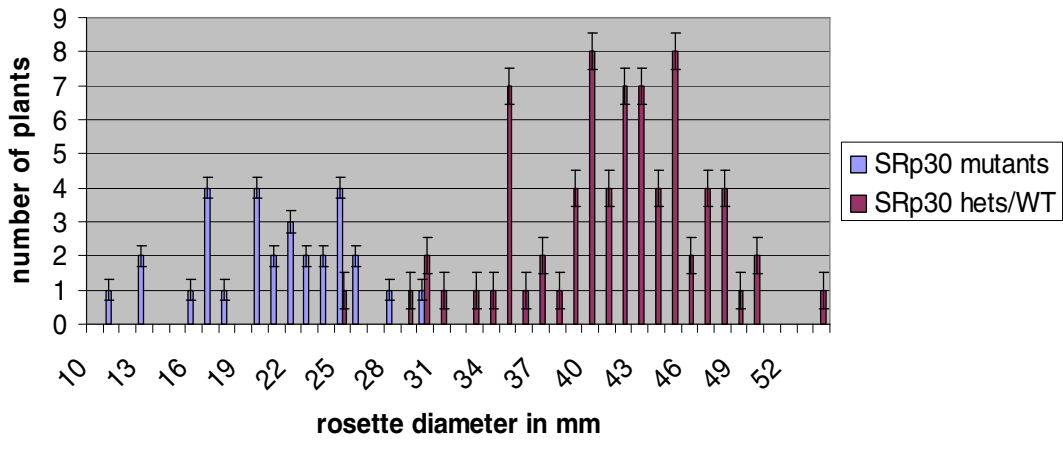
Figure 2.4 *srp30-1* **mutant plants display a range of phenotypes.** Flowers of

srp30-1 mutants display a very interesting LEUNIG-Like phenotype (Liu *et al.* 1995;

Franks *et al.* 2006) (a, b) where sepals, petals, and carpals are elongated and narrow

(red arrows). Additionally, carpeloid tissue can be found on the end of the sepals

(white arrow). Rosette leaves of *srp30-1* mutants are angular and serrated (c) in

comparison to wild-type rosette leaves (d). Finally, allowing a plant genotyped as

heterozygous to self-fertilize results in two segregating phenotype groups with respect

to rosette size (e). Plants that genotype as being homozygous for the mutation have

an average rosette diameter of ~23mm compared, where plants genotyped as wild-

type or heterozygous for the mutation having an average rosette diameter of ~45mm.

diameter of the *srp30-1* mutants measures only 21mm compared to the wild-type

plants, which have a diameter of 42.1mm at the same stage (Figure 2.4e). In addition

to the overall size of the rosette, the leaves of the *srp30-1* mutant plants display a

serrated rosette leaf phenotype (Figure 2.4c). The small stature of the mutants,

compared to wild-type plants, is consistent throughout the entire life of the mutants.

Plants reaching seed-setting age are also much shorter in stature, reaching heights of

~40% to that of wild-type plants. Flowers of *srp30-1* plants display multiple

interesting phenotypes. Plants homozygous for the mutant allele show a very high

degree of both male and female sterility. This is partially due to a dramatic lack of

pollen production and deformed female sex organs. The carpel of the *srp30-1* mutant

is elongated and much narrower than that found in a wild-type flower (Figure 2.4a).

In addition, flowers of these mutants show another very interesting *leunig*-like

phenotype: carpeloid sepals, elongated and narrow sepals, petals, and anthers (Liu *et

al.* 1995; Franks *et al.* 2006), all of which may be partially responsible for producing

very few seeds in mature plants (Figures 2.4a, 2.4b). Plants heterozygous for the

mutant allele were transformed with a *SRp30* native gene with a 3' *GFP* fusion. The

resulting T2 plants, which were genotyped as homozygous for the genomic mutation

and carrying the SRp30:GFP transgene, were phenotypically indistinguishable from

wild-type plants.

SR45: *sr45-1* plants show growth defects

Salk_004132 has a T-DNA insertion in exon eight of the *SR45* gene.

Homozygotes for this insertion do not accumulate full length mRNA transcripts, and

fully functional SR45 proteins are unlikely to be generated, so I have designated this mutant allele as *sr45-1*. Although RT-PCR analysis confirms that low levels of transcript 5' of the insertion are being produced, no transcript 3' of the insertion location is found in these same plants (Figure 2.1). In the absence of a full length transcript, it is highly unlikely that these plants synthesize functional proteins, and are functionally null for the SR45 protein.

The effect of the loss of protein production is apparent in the *sr45-1* mutant plants. Although germination and leaf number progress in parallel with wild-type plants, the *sr45-1* mutants are typically smaller in size with respect to both plant height and rosette diameter, prior to wild-type bolting time. In addition, longer rosette leaves are curled downward toward the underside of the leaf, and are narrower in width than those of wild-type plants. The bolting time of the mutant plant is also affected by the disruption of the SR45 gene. Under 12 hours of light, 37 of 37 wild-type plants bolted 19 days after germination at the 8-10 rosette leaf stage. However, it took 29 days (30 bolted after 29 days, 5 after 31 days, and three after 28 days) for the *sr45-1* null plants to bolt. The *sr45-1* mutants also accumulated approximately 22 rosette leaves (34 plants accumulated 22 rosette leaves at the time of bolting, 6 accumulated 24 rosette leaves) on average before bolting.

Despite having multiple developmental abnormalities, the plants are otherwise healthy. They mature through development and are able to set seed, although at a lower frequency due to sterility and physical incompatibility of reproductive organs. Similar to the rest of the plant, the flowers of *sr45-1* mutants are also proportionately smaller than those found on wild-type plants. There is a small

a: sr45, WT

b: WT, sr45

c: WT

d: sr45

e: Rosette diameter of 6 week old plants

number of plants

70
60
50
40
30
20
10
0
-10
-20

10 20 30 40 50 60 70 80 90 100 110 120 130 140 150

diameter in millimeters
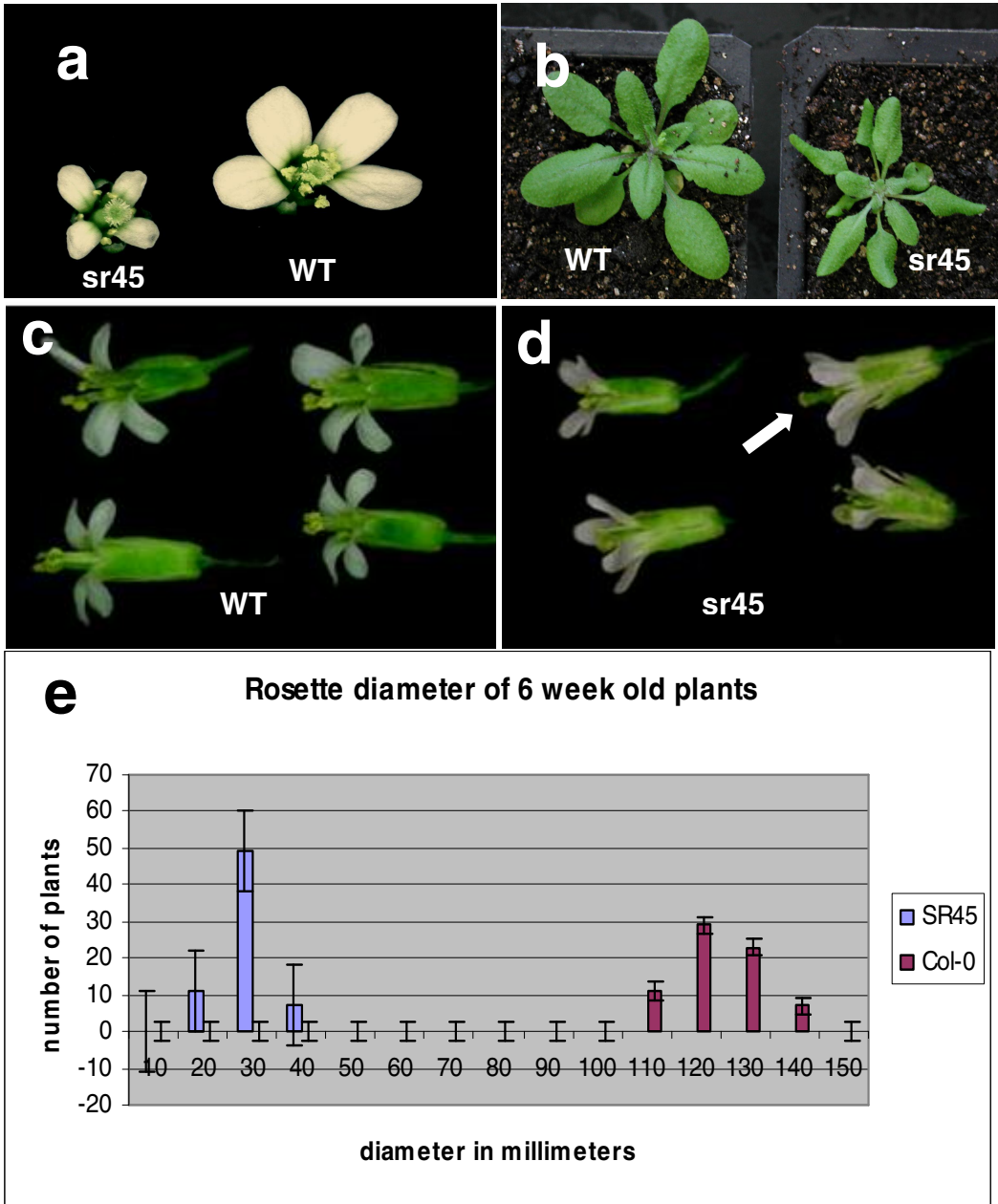
SR45
Col-0

Figure (2.5) *sr45-1* **null mutants are homozygous viable displaying visible phenotypes in total plant size, floral size and morphology, and both male and female sterility**. *sr45-1* mutants are visibly smaller than wild-type plants (b, e) of the same age. Additionally, the rosette leaves of *sr45-1* mutants curl down along the edges and tips of the rosette leaves (b). The flowers of the *sr45-1* mutants (d) are significantly smaller than the flowers from a wild-type plant of the same age (a). When viewed from the side, it is also obvious that the flowers from the *sr45-1* mutants (d) do not fully expand and open as do wild-type flowers (c). In some flowers of the *sr45-1* mutant, the carpel grows faster and emerges from the flower before the stamens are able to develop pollen and successfully self-pollinate the carpel (d arrow).

degree of male sterility, as less pollen is observed on the anthers of some flowers. Additionally, the pollen tends to be darker in color compared to that found on wild-type flowers. In some flowers, the carpel elongates and emerges from the floral bud before pollen matures on the anthers, which inhibits the flower from ever self-pollinating. The degree of pollen production and sexual organ compatibility is variable, not only from plant to plant, but also from flower to flower (Figure 2.5). When planted on MS plates and under 16 hours of light, the *sr45-1* mutants will begin to germinate, the radicle successfully pushes through the seed coat, but the tissue dies well before cotyledons are visible. Likewise, when planted on soil with 16 hours of light, only 20% (76/387) of planted seeds will result in mature plants. Seeds that do not develop into plants arrest and die at the same stage as seen on MS plates. When grown under 24 hours of light, the germination rate of the *sr45-1* mutants is similar to that of wild-type plants. Unfortunately, multiple attempts to clone the SR45 gene failed and thus, the mutant phenotype was never rescued in the *sr45-1* mutant line.

SC35 displays a mild phenotype

Due to its close homology to animal SR protein genes, a null allele of the SC35 gene is of special interest and is subject to several known T-DNA insertion lines located within the RRM. One particular line, SALK_033824, in which the T-DNA insertion is located in exon three of SC35, gives rise to a homozygous viable mutant that I will refer to as *sc35-1*. Loss of transcript was verified by performing RT-PCR with primers designed both upstream and downstream of the insertion location. The *sc35-1* mutant gives rise to seeds which germinate at a frequency

49

consistent with wild-type. However, under 24 hours of light, the *sc35-1* mutant is delayed in bolting. Although the mutant plants flower at the same time as wild-type plants, the floral meristem remains at the base of the rosette, and bolting does not occur until approximately three days after wild-type plants bolt. However, shoots of *sc35-1* null mutants do, in fact, reach a height equal to that of wild-type plants and are indistinguishable from wild-type plants at maturity. Unlike the *srp30-1* and *sr45-1* mutants, *sc35-1* mutants did not display any kind of floral phenotype. Plants homozygous for the mutant allele were transformed with a *SC35* transgene composed of *GFP* fused to the c-terminus of a PCR amplified *SC35* native gene. The resulting T2 plants that were genotyped as homozygous for the genomic mutation and carrying the *SC35-GFP* transgene were phenotypically indistinguishable from wild-type plants. The rescue of the wild-type phenotype with the *SC35* transgene confirms that the T-DNA insertion is responsible for the *sc35-1* mutant phenotype.

Discussion

In addition to being required for constitutive splicing, SR proteins are also essential for alternative splicing to occur. With the large amount of alternative splicing that occurs in mammalian systems, it is understandable that SR protein gene mutants in animals have been found to be early embryonic lethal. This observation, combined with the small number of mammalian SR protein genes, suggests that each individual SR protein has a unique set of targets with little complimentary function among the SR protein gene family in animals. However, the larger number of SR protein genes in *Arabidopsis thaliana*, combined with a significantly smaller

50

percentage of alternatively spliced genes, makes identifying viable mutants possible. While I did identify viable mutants for SR protein genes *SRp30, SC35, and SR45*, lethal mutants were found for two genes, *SCL28* and *RSp31*. This coincides with the data obtained from mammals that states that some SR proteins are absolutely required for the viability of the organism. In addition to the three mutants described above, I did identify four additional insertions located in RSp31, RSp41, SRp30, and SRp34/SR1 which were homozygous for their respective insertion but produced RNA at WT levels (Table 2.1).

The *sr45-1* mutant displays many different phenotypes in most organs throughout almost every developmental stage. Mutants have a low germination rate, deformed leaves, overall smaller size, and deformed flowers. The range of mutant phenotypes suggests that SR45 plays a global role in mRNA processing with respect to spatial and temporal expression. However, the viability of the mutant also suggests that while SR45 has a range of targets throughout the plant, the targets are probably limited to a very small number of non-essential target genes. Conversely, one or more of the other 19 SR proteins may be able to complement some of the SR45 protein functions, enabling the plant to maintain viability. It is also necessary to point out that SR45 is unique to the SR protein gene family for several reasons. As we have discussed previously, SR45 is the only known SR protein to have an N-terminal RS domain, as well as being highly related to another family of RNA processing genes. Other possible unknown functions not associated with pre-mRNA splicing could, in fact, make a contribution to the phenotypes observed.

Of the three viable SR protein gene mutants I identified, *srp30-1* may give the most insight into potential gene targets of an individual SR protein. Mutants display an overall smaller structure to wild-type plants, as well as serrated leaves, either of which could be the result of improper splicing of any number of unknown genes. However, *srp30-1* mutants have a striking floral phenotype similar to that seen in plants mutant for the gene *LEUNIG* (Liu *et al.* 1995; Franks *et al.* 2006). Both mutants share severe floral deformities including low pollen production, sterility, long and narrow sepals and petals, and sepals containing carpeloid tissue. It is quite clear that SRp30 plays an essential role, either directly or indirectly, in mRNA processing of genes essential for proper floral development. In future investigations designed to identify genes that are specifically targeted by individual SR protein genes, or groups of specific SR protein genes, *LEUNIG* and gene products known to associate with *LEUNIG* will provide a place to start.

The *sc35-1* mutant does not affect the reproductive ability of the plant, even though no transcript is being produced. However, the lack of an obvious mutant phenotype may be explained by studying the SC35 subfamily more closely. The SC35 subfamily contains four homologs of SC35, bringing the total number of members to five genes, or 25% of the total *Arabidopsis* SR protein genes. It is possible that, since the four members of the SC35-like subfamily resulted from the duplication of the *Arabidopsis* genome, they may have a high amount of overlapping functions of SC35. Therefore, they may target a smaller and more specific group of genes, while SC35 actually plays a lesser and potentially redundant role to the SC35-like genes. Plants mutant for the SCL28 SR protein gene, one of the four SC35-like

genes, are embryonic lethal, giving this theory some credibility beyond speculation based on sequence motif.

Materials and Methods

T-DNA Insertion Lines. T-DNA insertion lines were constructed by the Salk Institute Genomic Analysis Laboratory (http://signal.salk.edu/) and seed stocks were obtained from the ABRC located at The Ohio State University http://www.biosci.ohio-state.edu/pcmb/Facilities/abrc/abrchome.htm).

Primer Design and Genotyping. Gene specific primers were designed for each gene containing a SALK T-DNA insertion (table 2.1). Forward and reverse primers were created ~500bp upstream and ~500bp downstream of the reported insertion site. Multiplex PCR was performed using gene specific primes and an insertion primer, (table 3.2) creating bands of ~1000bp and 650bp for wild-type and insertion alleles respectively. PCR was performed using an Eppendorph gradient mastercycler with 40 cycles of 95 degree denaturing for 15 seconds, 55 degree annealing for 20 seconds, and 72 degree elongation period for 60 seconds. PCR product was analyzed on a 1% agarose gel, and homozygous mutant plants were selected for analysis.

Kanamycin Screening. Before performing PCR, insertion lines were first selected for Kanamycin resistance by plating seeds on plates containing 50ug/ml Kanamycin and 4.71g Murashige and Skoog (MS) salts. Plates were first held at four

degrees Celsius for 48 hours in the dark. Then they were transferred to a growth

chamber where plants were grown at 16 hours of light and 20 degrees Celsius. Plants

resistant to the antibiotic were transplanted onto soil after 14 days growth on MS

media.

Growth Conditions. Seeds potted in soil, to be grown for phenotype

observation and maturation, were grown for 16 hours light at 20 degrees Celsius after

48 hours kept at four degrees and 24 hours of dark.

Sequencing. PCR products from mutants, verified as homozygous for its

respective T-DNA insertion, were sent to Integrated DNA Technologies

(http://www.idtdna.com) for sequencing.

RNA Analysis of Homozygous Mutants. RNA was extracted from 10-day

old leaves using a Qiagen RNeasy kit following manufacturer's instructions. Primers

pairs were designed for each individual gene to detect product both upstream and

downstream of the insertion location (Table 2.2). Reverse transcription was

performed on 100 picograms of total RNA using Invitrogen's RT III two-step RT-

PCR kit following the manufacturer's instructions. PCR was performed using an

Eppendorf gradient mastercycler with 35 cycles of 95 degree denaturing for 15

seconds, 55 degree annealing for 20 seconds, and 72 degree elongation period for 30

seconds. PCR product was analyzed on a 3% agarose gel.

Chapter 3: Mutations in *Arabidopsis thaliana* SR protein genes: alterations in the activity of Exonic Splicing Enhancers

Introduction to Exon Splicing Enhancers

The initiation step of spliceosome recruitment starts with SR proteins identifying and binding to short 6-10 cis-acting nucleotide exonic sequences known as Exon Splicing Enhancers (ESEs). These regulatory sequences are typically located within exons adjacent to introns with weak splicing signals and are capable of compensating for weak exon/intron boundary signals such as weak pyrimidine tracts (Tian *et al.* 1994; Lorson *et al.* 2000). Additionally, strong ESEs have been shown to neutralize splicing inhibitory signals, and they are dominant over Exon Splicing Silencers (Chew *et al.* 1999; Kan *et al.* 1999) that would otherwise result in mis-spliced RNA products. The importance of ESEs is well established, yet very little specificity concerning these sequences is known.

Originally described as degenerate in nucleotide composition (Liu *et al.* 1998), this sequence of regulatory nucleotides determines not only the location of SR protein binding, but also distinguishes which SR protein(s) are needed to facilitate an individual splicing event. In animal systems, both *in vivo* and *in vitro* assays have demonstrated that individual SR proteins bind to unique classes of sequences. Multiple *in vitro* selection systems (Rian *et al.* 1995; Liu *et al.* 1998; Schaal *et al.* 1999), using mammalian SR proteins, have been performed with random sequences in an attempt to identify potential nucleotide regulatory elements that SR proteins recognize and bind. The pool of sequences that were identified by functional

selections with individual SR proteins (Tuerk *et al.* 1990; Liu *et al.* 1998; Liu *et al.* 2000), have been combined in a single online source open to the public, called ESE-Finder, and is used as an ESE annotation tool (Cartegni *et al.* 2003). In addition to this *in vitro* approach, an additional computational analysis performed by Fairbrother *et al.*, identified potential ESEs by comparing hexamer sequences that occur more frequently in internal exons than in unspliced RNA of human genes. Hexamers in human genes fulfilling this requirement were designated as potential ESEs. This data collection is also available as another online tool used to predict ESEs in sequences submitted to the ESE-Finder server (Fairbrother *et al.* 2004a). From the computational approach, not only were distinct groups of potential ESEs with specific nucleotide composition identified, but also these potential splicing signals have been found to have a high degree of functional ability in human genes (Fairbrother *et al.* 2004b).

The vast majority of work attempting to identify functional ESE sequences based on *in vitro* and *in silico* analysis has been limited mostly to animal model organisms. However, it is very reasonable to believe that the high degree of similarity between plant and animal SR proteins, with respect to both form and proposed function, implies that sequences in plants that are predicted to function as ESEs can be tested in *Arabidopsis,* and are expected to affect splicing with the same degree of success as seen in animals.

Analyzing such sequences *in planta* can be performed by using conceptual tools already developed in mammalian systems. Screening for Exonic Splicing Silencers (ESSs) in animal cell lines has been performed by Wang *et al.* in cell

culture lines using an ESS dependent splicing construct (Wang *et al.* 2004). In this study, Wang *et al.* were able to create a construct in which expression of a reporter gene was dependent upon the skipping of an exon due to the decanucleotides engineered into the construct. For each decanucleotide to be tested, a unique construct was created varying only by the specific sequence being examined.

The experimental system used in mammalian studies was easily applied to my research working with *Arabidopsis*. Our lab has developed an assay, conceptually similar to that already applied to animal systems, to test nucleotide 9mers for *in vivo* ESE function (Mount *et al.* in preparation). Similar to the animal experiments (Wang *et al.* 2004), inclusion or skipping of an exon in a synthetic gene is dependent upon the sequence engineered into the tester exon (Figure 3.1). This assay in *Arabidopsis* allows, mutants that are null for individual SR proteins (SC35, SRp30, and SR45) to be examined in whole organisms. By taking advantage of identified null SR protein gene mutants in *Arabidopsis*, I am able to study ESE selection, among other molecular interactions, in whole mutant plants. While the data obtained will give insight into the ability of predicted sequences to function as ESEs, it will also identify what effect individual mutants have on the ability of these 9mers to function as regulatory sequences.

For this study, I transformed each of four different plant lines (wild-type, *srp30-1, sc35-1*, and *sr45-1*) with 15 unique constructs (Figure 3.2), each containing a different 9mer sequence. These 9mers were previously selected by the statistical criteria outlined in Mount *et al.* (in preparation) and ESE activity was observed. Like the mammalian assay, the splicing of a synthetic gene is completely dependent upon

Mount et al., in preparation
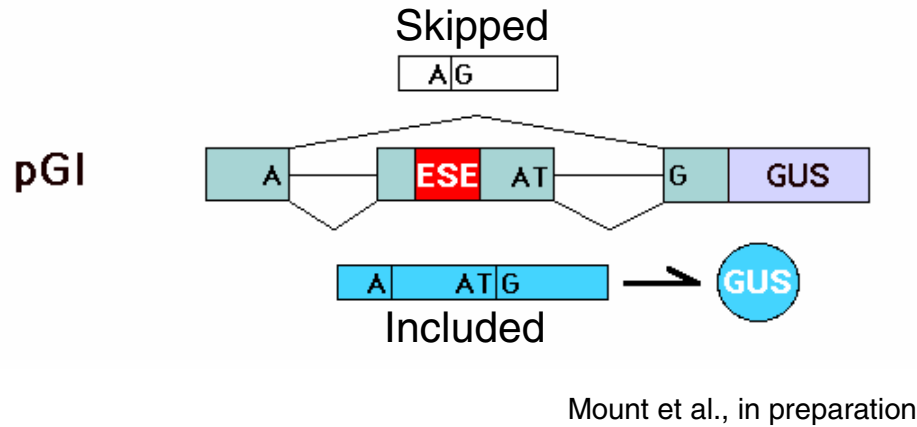
Figure 3.1 Diagram of synthetic gene engineered for testing ESE function of 9mer sequences. Sequences selected for ESE testing were cloned into the pGI vector created by Mount et al. (In preparation). In this assay, an alternatively spliced exon is either included or skipped depending on the ability of the sequence being tested to function as an ESE. If the sequence does function as an ESE, the alternatively spliced exon is included into the transcript and a functional GUS protein is produced. If the sequence does not function as an ESE then the exon is skipped and GUS is not produced. (Mount *et al.* in preparation)

the engineered sequence to act as a regulatory element. While this assay will confirm whether or not these sequences actually function as ESEs *in planta*, it will also potentially identify whether or not mutant plants *srp30-1, sc35-1,* or *sr45-1* have affected the ability of these sequences to function as regulatory elements in comparison to wild-type plants.

My experiments in screening a small number of sequences for ESE function in mutants have identified five sequences which, in mutants, are significantly altered in their ability to affect splicing in comparison to wild-type plants. Of these five sequences, two were affected by multiple mutant lines. More surprisingly, three of the five sequences were shown to increase in splicing efficiency in the presence of an individual SR protein gene mutant. These results show that SR protein gene mutants do affect alternative splicing through regulatory sequences such as Exon Splicing Enhancers.

Results

Sequences used as potential ESEs were chosen by identifying nucleotide sequences, nine bases in length, which occur in *Arabidopsis* internal exons more frequently than are found in unspliced RNAs (Pertea *et al.* 2007), in addition to sequences identified in animal functional selection experiments discussed earlier in this chapter. Using these sequences, Justin Benoit created individual constructs, each containing a single nine base pair sequence to be tested. These constructs varied from each other by only the nine nucleotide stretch engineered into the tester construct. A

total of 15 individual 9mers were tested for ESE activity in wild-type and *srp30-1, sc35-1*, or *sr45-1* SR protein gene mutants.

The resulting transcript (included and skipped products) level was analyzed by real time PCR. The ratio of intron inclusion isoform to intron skipping isoform was used to indicate ESE activity in each sample. As previously observed, nine of the 15 sequences predicted to be splicing factors did, in fact, induce exon inclusion in wild-type plants (Figure 3.3). One 9mer, ESE74, produced approximately equal amounts of skipped and included products. Of the remaining five sequences tested, only one sequence, ESE101, resulted in an I/S ratio of <0.1. While the eight sequences that induced more exon inclusion than skipping typically resulted in inclusion to skipping ratios of 10 or less, one sequence, ESE34, is of great interest due to its exceptionally high inclusion ratio. ESE34 (TGCCGCTGG) differs from ESE24 (AGCTGCTGG) at only two positions. However, those two base pair substitutions had a dramatic effect on exon inclusion. ESE34 resulted in an inclusion to skipped ratio of 5800, while ESE24 had an included to skipped ratio of only 6.4. (Figures 3.2, 3.3, and 3.4)

Analyzing the effect of SR protein gene mutants on the ability of plants to process the test sequences as ESEs also proved very interesting. In five of the 15 constructs assayed, lines with mutations in one of the three mutant genes altered the I/S ratios 10 fold or greater compared to that observed in wild-type plants (Figure 3.2). In the presence of the *sc35-1* mutation, the ratio of inclusion to skipped isoforms of the reporter exon containing ESE24 (AGCTGCTGG) drops from 6.4 in wild-type plants, to 0.22. Mutations in either *srp30-1* or *sr45-1* do not affect the ratio in favor of ESE24 exon skipping. Wild-type lines transformed with the synthetic

60

ESE #	Sequence	WT	srp30-1	sc35-1	sr45-1
24	AGCTGCTGG	6.44	28.2	0.220	19.6
34	TGCCGCTGG	5800	59.8	22.3	1650
51	GAGGATTGA	1.78	1.40	1.88	3.04
53	TGAATCGTC	0.370	1.27	2.40	.0990
54	ATCGTCACA	0.620	4.99	3.70	0.440
57	GGGTCCAAA	0.750	.00640	27.7	.000200
65	TGCAGATGA	0.210	1.63	4.29	0.950
69	CACCAAGAA	2.87	1.38	3.86	4.78
74	GCTTGGTTC	0.910	6.22	6.87	1.52
82	TTAAAGCTG	8.11	5.95	3.53	4.67
89	GAAGAGAAA	9.35	15.8	7.62	3.43
90	GAAAAGAAA	0.140	0.680	0.0680	1.93
100	TTGGTGCAA	9.77	19.7	7.87	16.0
101	GAGAATTGA	0.0360	0.0260	0.180	0.100
105	CGATCTACG	2.05	4.01	1.71	0.810

Figure 3.2 Mutant SR protein gene alleles altered splicing of predicted ESEs by 10 fold or greater in 5 of the 15 sequences tested. Real Time PCR was performed for each ESE lined into the four plant genotypes. The ratio of exon inclusion to exon skipping was then calculated and reported on this chart.
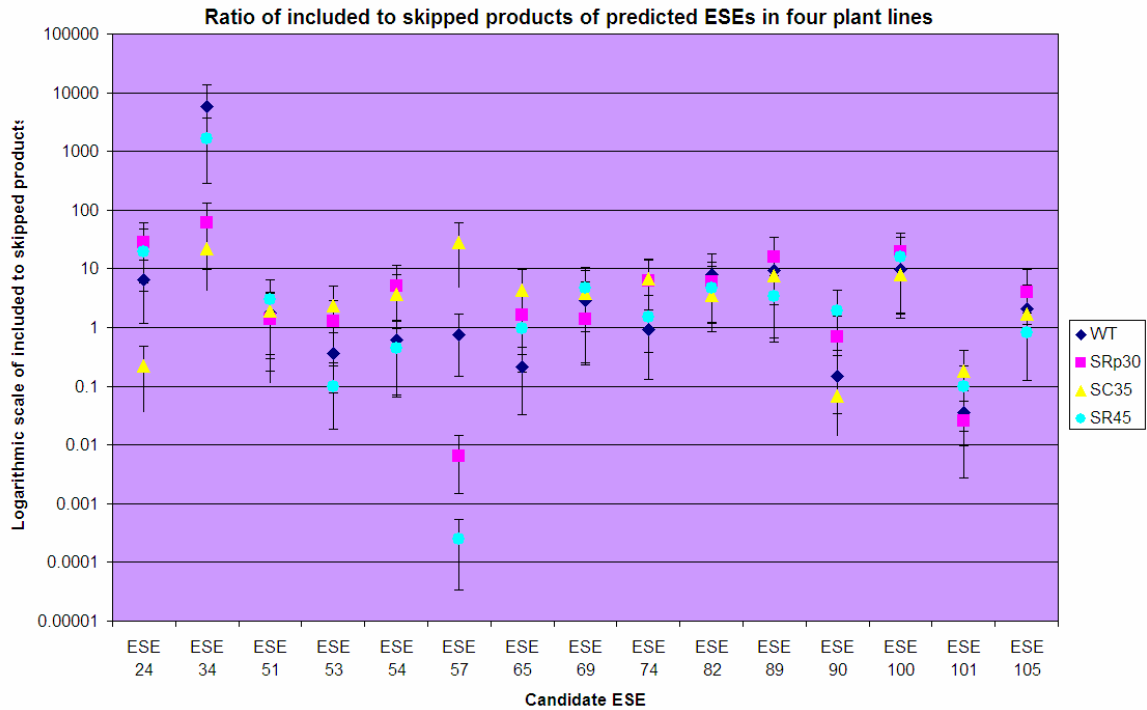
Figure 3.3 Logarithmic representation of the ratio of Included to Skipped products resulting from candidate ESEs in 4 plant lines. A graphical representation in including error bars of data shown in Figure 3.2. The individual data points are the ratio of exon inclusion to exon skipping as measured by Real Time RT-PCR. Each column represents a specific ESE transformed into each of the four genetic plant lines. Data obtained from wild-type plants is depicted by the blue diamond, *srp30-1* plants by the pink square, *sc35-1* plants by the yellow triangle, and the *sr45-1* plants by the light blue circle.
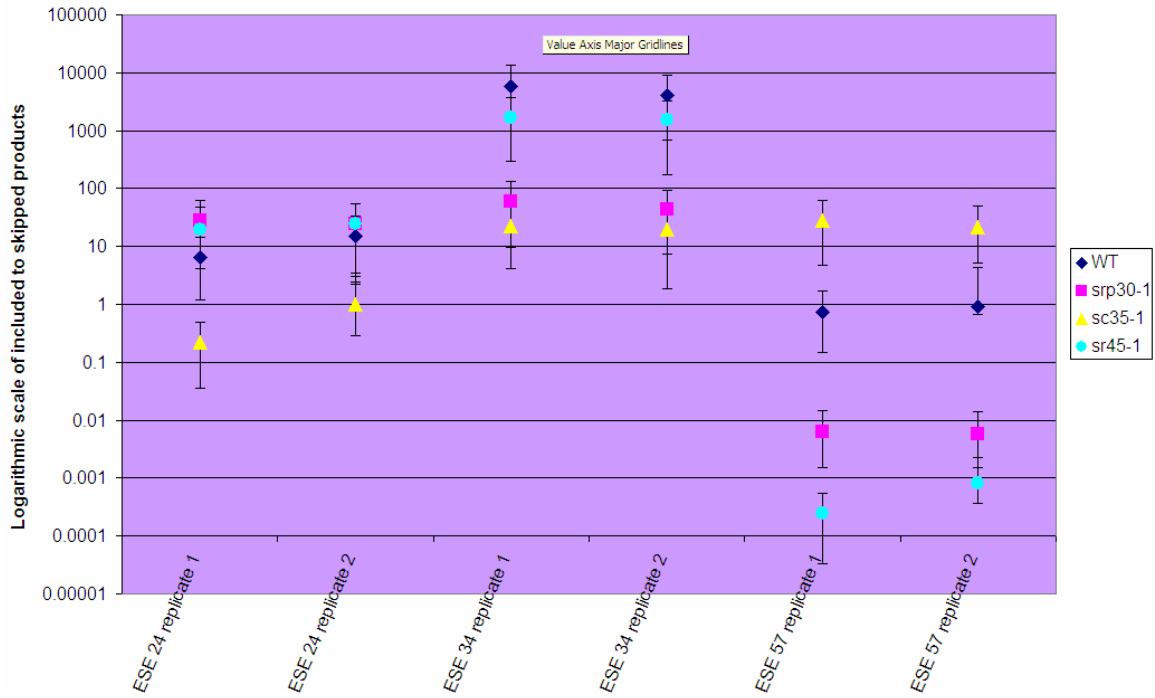
Figure 3.4 Biological replicates of ESE24, ESE34, and ESE57. Shown here are biological replicates for ESE24, ESE34, and ESE57, the three sequences affected the most by an SR protein gene mutant. A minimum of two biological replicates were performed on all samples. Each replicate recorded was performed using an RNA sample from an independently transformed T1 plant, and replicates were performed on different days. The consistency between individual replicates verifies that the data are precise, reproducible, and dependent solely upon the construct. RNA was extracted from the first and second rosette leaf from 8-10 rosette leaf stage.

gene containing ESE34 result in almost entirely included product with a ration of over 5800:1. While *sr45-1* mutants transformed with the same construct closely mimics the wild-type I/S ratio, *srp30-1* and *sc35-1* mutants shifted the ratio closer to equal amounts of included to skipped product, measuring 60:1 and 22:1 respectively. ESE57, GGGTCCAAA, is unique in this study in that all three mutant plant lines affected the ability of ESE57 to function as an ESE. Plants carrying the *srp30-1* or *sr45-1* alleles produced more skipped product resulting in included to skipped ratios of 0.0064 and 0.0002, respectively. These ratios are a significant contrast in comparison to wild-type plants with an included to skipped ratio of 0.75. Interestingly, *sc35-1* mutants actually increased the ability of ESE57 to function as an ESE, producing an included to skipped ratio of 28:1. When transformed with the tester gene containing ESE65, *sc35-1* mutants have an I/S ratio of 4.3 compared to the 0.2 of wild type. The other mutants process the tester gene in roughly the same manner as wild type. The final ESE tester sequence to show a statistically significant alteration in inclusion to skipping ratios, ESE90, was apparent only in *sc35-1* mutant plants. While wild-type plants produce a ratio of included to skipped products of 0.14, *sr45-1* mutant plants produce a ratio of nearly 2.0 (Figure 3.3).

Discussion

Of the 15 individual sequences that were tested, eight sequences induced exon inclusion, one produced nearly identical ratios, five had ratios of <0.1 inclusion to skipping. One interesting observation was between ESE34 and a very similar sequence, ESE24, which differs only by two nucleotides. Although very similar,

those two base pair substitutions had a dramatic effect on exon inclusion versus exon skipping ratio. ESE34 was the sequence with the greatest effect on inclusion, producing an included to skipped ratio of 5800, while ESE24 had an included to skipped ratio of only six.

Mutants also had a significant effect on the ability of the ESE tester sequences to function as ESEs. Of the 15 sequences, five constructs showed significant alteration in ESE activity in at least one of the mutants. One surprising outcome came from testing the ESE34 9mer sequence. This sequence was identified as a very strong ESE in wild-type and *sr45-1* plants, with I/S ratios of 5800 and 1650 respectively. However, I/S ratio of the same sequence was reduced to approximately 60 and 22 in *srp30-1* and *sc35-1* mutants respectively. What makes this data interesting is not only did a single SR protein mutant dramatically reduce the splicing efficiency but either of two different mutations could produce this result. The same holds true for ESE57 where the sequence is roughly a neutral sequence in wild type, producing almost the same amount of included as skipped product, while *srp30-1* or *sr45-1* mutants have I/S ratios of .0064 and .0002 respectively. SR proteins have been proven to interact with other splicing factors in addition to themselves through protein-protein interactions, and it has also been shown in animals that some SR proteins can, in fact, complement cellular extracts mutant for other SR proteins. However, while speculated upon, it has not been shown that two specific individual SR proteins are required for a single splicing event to occur.

There are several explanations for this result. First, both proteins could be binding directly to the RNA and recruiting unique splicing factors independent of

each other, and additional splicing factors are needed for the splicing machinery to process the transcript properly. While yeast-two-hybrid experiments have been performed with individual *Arabidopsis* SR proteins, the data provides only some of the potential binding partners. There is no empirical evidence suggesting the exact factors with which any individual SR protein does associate in the initial recruiting process. Additionally, each unique SR protein may have different factors to which it binds. Another mechanism is that SR proteins work through multiple indirect interactions to facilitate the alternative splicing of a pre-mRNA. This could include two or more different proteins binding to a single RNA and together recruiting a single splicing factor. Additionally, one SR protein may bind to RNA and through protein-protein interactions with other SR proteins that are not directly associated to the target RNA, recruit the required splicing machinery. Of course, there is always the possibility that any or all combinations of the above scenarios are executed in parallel or at different points in time.

In addition to ESE34 and ESE57, the ability of test sequence ESE24 to act as an ESE was also significantly altered in a mutant plant. While wild-type plants and plants mutant for *srp30-1* or *sr45-1* have I/S ratios of 28 and 20 respectively, only *sc35-1* mutants actually reduced the ability of ESE24 to function as an ESE, recording an included to skipped ratio of 0.22. While no evidence suggests that SC35 is directly responsible for this result, it does show that SC35 plays an important role in the use of ESE24 sequence as an alternative splicing target.

The current accepted theory for the function of SR proteins is that they initiate splicing by recognizing and binding to ESEs that recruit other splicing factors.

Assuming that ESEs are present and are not blocked by inhibitory factors recruited by an Exon Splicing Silencers (ESS) or other negative regulators, the presence of an SR protein should encourage splicing and exon inclusion to take place. With this thought as a guide, conventional wisdom suggests that in the absence of an SR protein, sequences that are normally recognized as ESEs but that SR protein will not be recognized, and the result will be products in which the exon is skipped. All of the data in this experiment seemingly follows this consensus except for when *sc35-1* mutants are transformed with the ESE57 tester sequence. In this case, the lack of a functioning SC35 protein actually increases the relative amount of included product, producing an included to skipped ratio of 28:1 compared to wild-type plants that produced a ratio of 0.75:1. This result is counterintuitive to the current understanding of how SR proteins may function at a molecular level. While initially surprising, there is evidence which could explain this phenomena. Other labs have shown that the over-expression of individual SR proteins can alter the splicing pattern of different SR proteins, suggesting that SR protein genes do undergo a level of regulation (Lopato *et al.* 1999). An individual sequence functioning as a stronger ESE in the presence of a null mutant could, in fact, be a result of this altered autoregulation. An SR protein gene or groups of SR protein genes, which normally are not producing functional proteins due to the transcriptional regulation of SC35, are now expressed at higher functional levels, and thus are able to process targets containing the ESE57 sequence at a higher frequency. While SC35 is not directly involved in the recognition of ESE57, it may still be able to effectively control what

sequences are identified as splicing factors, through crossregulation of other SR proteins that do directly interact with the ESE57 sequence.

The ESE assay provided some unexpected results and insight into possible functions of SR proteins and their ability to identify sequences as ESE and process the pre-mRNA accordingly. Through this assay, I observed multiple individual mutants resulting in a dramatic shift in splicing pattern of the test sequences. Splicing models suggest that individual SR proteins bind to RNA, recruit unknown splicing factors which, in turn, are able to recruit the essential snRNPs for splicing to take place. Unfortunately, the unknown splicing factors have yet to be identified with certainty. However, several groups have performed yeast-two-hybrid assays with individual SR proteins, and they have shown protein-protein interactions among SR proteins as well as between SR proteins and other RNA processing factors (Golovkin *et al.* 1998; Golovkin *et al.* 1999; Lopato *et al.* 2002; Lorkovic *et al.* 2004). This data demonstrates the potential for protein-protein interactions within the SR protein gene family as well as with outsiders. Additionally, I learned that plants null for individual SR proteins can also increase the ability of a sequence to act as an ESE, probably through crossregulation between SR proteins. These observations, in combination with previous research, give creditability to the proposition that SR proteins may play several other important roles in alternative and constitutive splicing beyond the initial step of binding to RNA. The mechanism of identifying sequences as ESEs is more complex than expected.

An obvious solution to elucidate some answers from this complex mechanism of multiple SR proteins being required for a single splicing event, would be to

identify more mutants and test many more ESEs. However, there was one incredibly interesting observation which would require no additional mutants or ESEs. The ability of ESE34 to function as an ESE was greatly reduced with a mutation in either the SRp30 or SC35 gene. In this instance, each mutant had a similar effect on the ESE activity of this sequence. It would be very interesting to see what effect a double mutant would have on this sequence. If both SR proteins are performing the same exact function, the reduction of ESE activity simply may be a dose dependent response. A double mutant would have more of an additive effect on the exon inclusion versus exon skipping ratio, and it would result in a much greater reduction than a mutation in either of those genes alone. Conversely, if these proteins are performing two separate functions for a single processing event, and splicing is inhibited by a loss of function in either one, then a double mutation would have little, if any, additional effect because loss of either would have already inhibited this event.

Although more information and better insight into potential functions of individual SR proteins can be further investigated with materials currently available in our lab, the future direction of this research is really dependent on additional materials and mutants. Additional individual SR protein gene mutants and plant lines mutant for multiple SR protein genes are essential for further investigation. Additional 9mer test sequences are needed in order to begin to determine specific classes of ESEs recognized by individual SR protein genes. Simply put, this assay needs to be expanded in numbers by almost every aspect. This study analyzed the ability of 15 unique sequences to function as ESEs in wild-type plants as well as in the SR protein gene mutants I identified. The 15 sequences tested represent a very

69

small fraction of the 262,144 possible unique nucleotide sequences which could theoretically be analyzed in our assay. Another limitation of this assay is the confines of the tester gene and location of the engineered 9mer within the synthetic gene. Moving the location of the test sequence within the alternative exon could dramatically alter the ability of any specific 9mer sequence to function as an ESE. However, testing all 262,144 possible 9mer sequences in all possible locations within our assay is impractical if not impossible. To aid in sequences chosen to be tested, future experiments in our lab will use results from an alternative splicing microarray currently in development. Not only will this microarray chip provide an immense amount of data regarding the effects of SR protein mutants on alternative splicing, but will also provide information about potential splicing signals that can then be independently tested in assays such as that used in this report.

Materials and Methods

ESE Vectors. ESE vectors were obtained from Justin Benoit and will be described in Mount et. al (in preparation). Sequences selected for *in vivo* analysis were cloned into an expression vector in which the inclusion or exclusion of a synthetic exon was dependent upon the sequences cloned into the vector.

Plant Transformation. Plants used for transformation were allowed to grow at 20 degrees and 16 hour light. Primary shoots were cut to induce auxiliary shoot formation and growth. Plants were then prepared for transformation via *Agrobacterium* using the floral dip method (Desfeux *et al.* 2000).

70

Transformant Selection. Seeds were collected from plants that underwent transformation and approximately 250 T1 seeds were planted on soil. After two days of four degrees Celsius under 24 hours dark, pots were then moved to a growth chamber and allowed to germinate at 20 degrees Celsius and 24 hour light. At the 4 leaf stage, seedlings were sprayed with Finale™ herbicide. Plants surviving the herbicide selection were thinned down to five plants per line and were allowed to self-fertilize for seed stock.

RNA Analysis of Homozygous Mutants. RNA was extracted from 10 day old leaves using a Qiagen RNeasy kit following manufacturer's instructions. cDNA was synthesized from 100 picograms of total RNA using the Taqman™ First Step RT-PCR and following the instructions provided by the manufacturer. Each cDNA was subject to Real Time PCR and was run in triplicates of three individual reactions to detect products created from exon skipping, exon included, and the standardizing *GUS* gene. Primers for detecting Skipped product were (Forward: CGCCGATATTACAGGATATAGAAAAG, Reverse: CGGCGAACTGATCGTTAAAACT), Included product (Forward: CGCCGATATTACAGGATATAGAAAAA, Reverse: AATTGCCCGGCTTTCTTGTAA), and Gus standardization primers (Forward: CAGTGTGCATGGCTGGATATG, Reverse: CCCTTTCTTGTTACCGCCAA). Real Time PCR was performed on a Roche Light Cycler 480 with a PCR program of five minutes denature at 95 degrees followed by 55 cycles of 95 degrees for 10

seconds, 48 degrees for 10 seconds, and 72 degrees for 20 seconds. Amplification

cycles were followed by two minutes of constant 72 degrees.

Chapter 4: Mutations in *Arabidopsis thaliana* SR protein genes: Cross regulation

Introduction

Identifying regulatory sequences such as ESEs is crucial in studying the physiological roles of SR proteins. However, finding the precise genetic targets is one of the ultimate goals in studying this gene family. Unfortunately, a lack of viable mutants has limited the attempts to identify potential gene targets to overexpression studies. This approach, while capable of altering the splicing of individual transcripts, may do so in a manner that is unlikely to occur *in vivo*, or produce levels of gene expression in a target gene which the organism may never actually encounter.

Attempts to overexpress individual SR proteins in *Drosophila* were futile, as high levels of B52 have been shown to be lethal at the early larval stage (Kraus *et al.* 1994). To add to the complexity, identifying potential targets of individual SR proteins is both conceptually and experimentally difficult. This is because genes are temporally and spatially expressed and are under the regulation of not only molecular processes, but environmental factors as well. Identifying a native non-splicing factor gene that is alternatively spliced, and studying the alterations of isoform production in the presence of an overexpressed SR protein, has yet to be published. However, there are some co-expression studies in which known genomic alternative splicing events have been engineered into constructs and transformed into cell lines with an additional construct overexpressing SR proteins. For example, the chicken fibronectin (FN) protein undergoes a regulated splicing event in Exon EIIIA during

the development of chicken limbs (Bennett *et al.* 1991; Burton-Wurster *et al.* 1997).

Liang *et al*. (2005) has created a construct containing the alternatively spliced *FN*

EIIIA exon and flanking exons. This construct was then transformed into chicken

cells of wild-type animals as well as cells overexpressing individual animal SR

proteins (Liang *et al.* 2005). The data obtained from this experiment proved that

different overexpressed SR proteins affected the processing of the two different

alternatively spliced isoforms to different degrees (Liang *et al.* 2005). In their

experiments, Liang *et al*. (2005) found that the overexpression of SRp75 induces

inclusion of the alternatively spliced EIIIA exon, but not to the same degree seen in

SRp40. However, there is no way of knowing how the native gene would be

processed in the whole chicken, or even a cell line mutant for an individual SR

protein. To add to the complexity of this situation, even if it were possible to identify

a cell line or animal mutant for an individual SR protein gene, a task which has yet to

be accomplished, there is no evidence that any one specific individual SR protein

gene mutant would even have an affect on splicing of this particular transgene.

The lethal effect of overexpressing individual SR proteins in animals, in

combination with identifying an individual gene that can easily be studied in most

environmental conditions, seems an ever daunting task. However, the Barta lab has

used *Arabidopsis* to overcome some of the obstacles that have been insurmountable in

animal systems. This lab has successfully overexpressed individual SR protein genes

in *Arabidopsis* by engineering constructs assembled from both the cDNA and

genomic sequence while under the control of the TMV promoter. Despite

developmental phenotypes, including degrees of sterility, the plants are completely

74

viable (Lopato *et al.* 2002; Kalyna *et al.* 2003). More importantly, they have chosen

to use SR proteins as their targets for alternative splicing. As previously mentioned,

there are 20 individual SR protein genes in *Arabidopsis,* of which 16 (15 identified by

Palusa *et al.* 2007 and SRp007) are regulated by alternative splicing. Taking

advantage of the crossregulation of SR proteins will aid greatly in identifying targets

of individual SR proteins due to the small, finite number of genes. Additionally,

these targets belong to a single gene family and undergo documented alternative

splicing while presumably essential for the correct development of the organism.

Additionally, unlike in animal cell lines, *Arabidopsis* has the added advantage of

having many organs and cell types in which the alternative splicing of the target gene

set can be analyzed.

Due to the lack of mutants available at the time, the Barta lab performed

overexpression assays, but were very successful in finding altered splicing patterns in

their transgenic lines in response to overexpressing individual SR protein genes

(Lopato *et al.* 1999; Kalyna *et al.* 2003). In one experiment, *SRp30* was

overexpressed and found to autoregulate both the total amount, and splice variants, of

the native *SRp30* mRNA. Additionally, transgenic lines overexpressing the *SRp30*

protein gene induced a dramatic shift in alternative splicing of both *RSp31* and

SRp34/SR1 SR protein genes pre-mRNA (Lopato *et al.* 1999). A similar experiment

was performed using transgenic lines overexpressing *RSZ33*, a plant specific SR

protein gene. Overexpressing *RSZ33* gave very similar results to that of the *SRp30*

overexpression experiment, in that *RSZ33* was found to regulate the alternative

splicing of its own native transcript, in addition to influencing the transcript level and

alternatively spliced isoforms of *SRp30* and *SRp34/SR1* (Kalyna *et al.* 2003). These experiments were successful in demonstrating that SR protein genes, when overexpressed *in planta*, were capable of significantly altering the isoform production of genomic targets. More importantly, *bona fide* genetic targets were identified in normal growth conditions and allow for a starting point in studying gene regulation of SR proteins. This represents a great advance over animal systems where splicing studies are still limited to synthetic genes engineered into constructs.

Similarly to the Barta lab, I have chosen to use SR protein genes as targets for alternative splicing. However, I studied this regulation in loss-of-function mutants for individual SR protein genes, which was essential for this study. The data from this experiment, using a reverse genetics approach, demonstrated that an individual SR protein does, in fact, regulate another SR protein gene at the transcript level through alternative splicing of the target pre-mRNA. Each of the three mutant lines did affect the alternative splicing of at least one other SR protein gene in rosette leaves, flowers, or roots.

Results

RNA was extracted from the rosette leaves, flowers, and roots of *sc35-1*, *srp30-1*, and *sr45-1* mutants, and wild-type plants. RT-PCR was performed on the RNA from each tissue type of all four plant lines. RT-PCR analysis of alternatively spliced SR protein gene transcripts proved quite interesting and clearly demonstrated that individual SR proteins do, in fact, regulate both the expression pattern and mRNA levels of other individual SR protein genes. Based upon the overexpression

76

studies performed by other labs, alteration of isoform production of individual SR

protein genes was expected to be observed in some of the tissues. However, the

specific SR protein gene(s) and tissue in which the gene is expressed had not been

identified. This analysis revealed not only that some SR protein gene transcripts are

altered in plants mutant for another individual SR protein.

RSp41, SCL30, and RSZ32. Sixteen of the 20 *Arabidopsis thaliana* SR

protein genes are known to undergo alternative splicing (15 documented in Paulsa et

al. 2007 and SRp007). These may be autoregulated or crossregulated by other SR

proteins in a manner that is currently unknown. In this study, I found that the

transcript level and isoform type of *RSp41, SCL30*, and *RSZ32* in leaves, flowers, and

roots, are unaffected by *srp30-1, sc35-1,* or *sr45-1* mutants (Figure 4.1). These three

genes produce only a single isoform, the transcript that encodes the full-length, and

presumably functional, protein.

RSp31, RSZ33, and SRp34a. Individual SR protein gene transcripts of

RSp31, RSZp33, and *SRp34a*, did show alterations in the full length transcript

expression levels in response to mutants (Figure 4.2). However, isoform type was

unaffected by the *srp30-1, sc35-1*, and *sr45-1* SR protein gene mutants. While

expression levels of *SRp34a* is unaffected in rosette leaves and roots, the amount of

SRp34a transcript is upregulated in flowers of *srp30-1* and *sr45-1* mutants.

Transcript levels of *RSp31* were similar to that of *SRp34a*. While expression

appeared to be unaffected by the mutants in rosette leaves and roots, the amount of
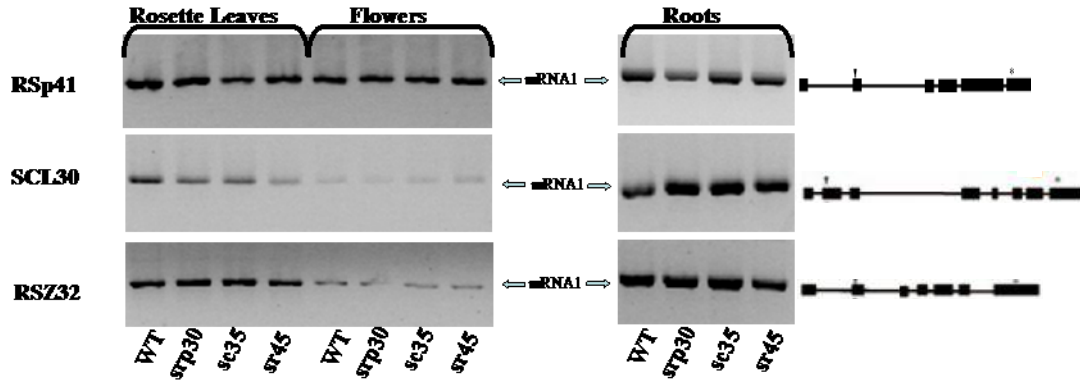
Figure 4.1 Alternative splicing and expression levels of SR protein genes *RSp41, SCL30*, and *RSZ32* transcripts are unaffected in *srp30-1, sc35-1*, and *sr45-1* mutants. RT-PCR was performed on whole RNA extractions of 10 day-old tissue using gene specific primers for RSp41, SCL30, and RSZ32 (Table 4.1). Equal amounts of RNA were used in the reverse transcription reaction (Figure 4.6) and 10% of the resulting cDNAs were used as template for the subsequent PCR reaction.

RSp31 transcript is increased in flowers of *srp30-1* and *sr45-1* mutants and the same is observed of the *SRp34a* transcript. Individual mutants did not affect the isoform accumulation of *RSZp33* transcripts. In rosette leaves, each of the three mutants increased the amount of the transcript that encodes for the full length protein, while in flowers and roots, the *srp30-1* mutant did, in fact, downregulate the amount of the same transcript that accumulated in the tissue.

SCL33a. The production of the *SCL33a* transcript was relatively unaffected by the mutant lines used in this study, and gene transcript was undetectable in roots in any of the four plant lines (Figure 4.3c). Many attempts to PCR amplify cDNA from multiple RNA samples simply did not produce PCR products. Performing RT-PCR using the same reagents and primer pairs with RNA extracted from rosette leaves and flowers reliably produced PCR products on every attempt. With this being said, I am confident that *SCL33a* simply is not expressed in roots and *srp30-1, sc35-1*, and *sr45-1* mutants have no effect, either directly or indirectly, on *SCL33a* expression or regulation in roots. The only other difference observed was a slight alteration of mRNA isoform accumulation in flowers. Wild type plants accumulate a second mRNA isoform in flowers, which is absent in flowers of all three mutants, as well as in wild-type and mutant rosette leaves.

RSp31a. The alternatively spliced transcripts of the *RSp31a* gene were significantly altered by the *srp30-1* and *sr45-1* mutants (Figure 4.3a). In roots and flowers of the *sr45-1* mutant, the *RSp31a* mRNA1 transcript was reduced to levels
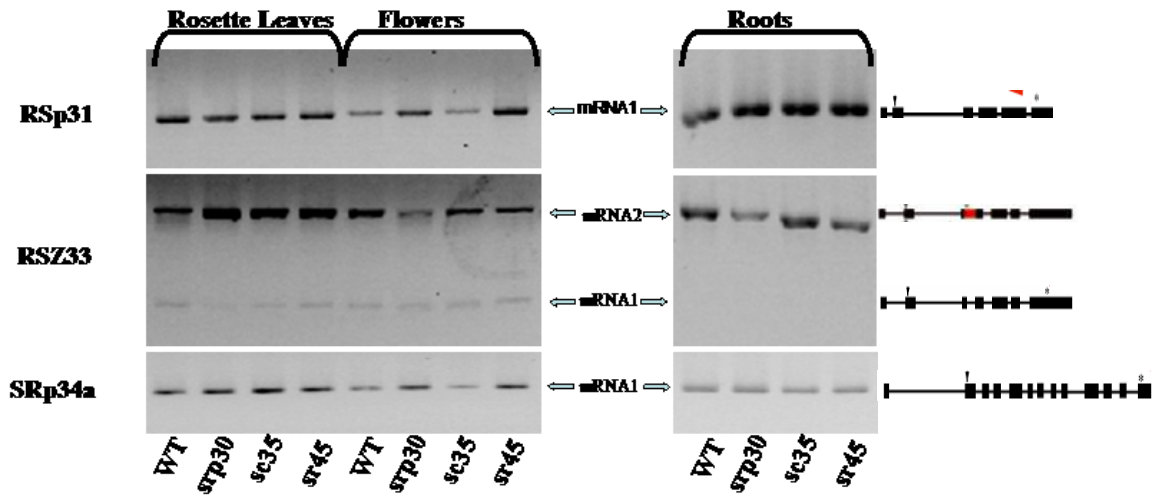
Figure 4.2 Expression levels of *RSp31, RSZ33*, and *SRp34a* are reduced in SR protein mutants *srp30-1, sc35-1*, and *sr45-1* in a tissue and mutant specific manner. RT-PCR was performed on whole RNA extractions of 10 day-old tissue using gene specific primers for RSp31, RSZ33, and SRp34a (Table 4.1). Equal amounts of RNA were used in the reverse transcription reaction (Figure 4.6) and 10% of the resulting cDNAs were used as template for the subsequent PCR reaction.
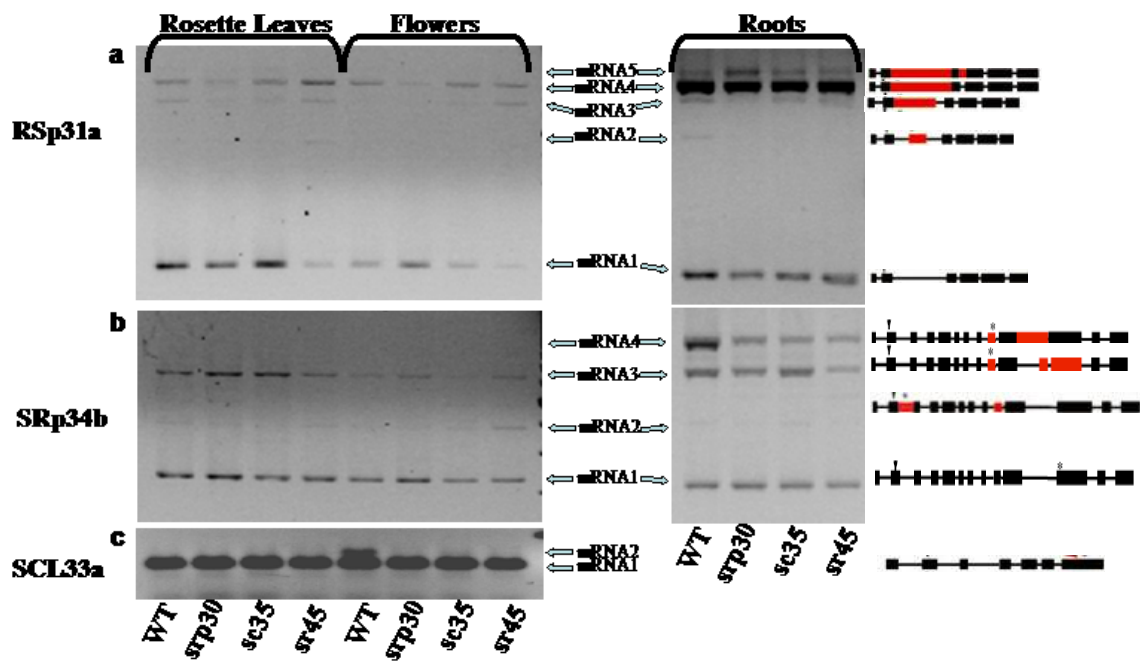
Figure 4.3 Expression levels and patterns of *RSp31a, SRp34b*, and *SCL33a* mRNAs are altered by mutations in *srp30-1, sc35-1*, and *sr45-1* SR protein genes. RT-PCR was performed on whole RNA extractions of 10 day-old tissue using gene specific primers for RSp31a, SRp34b, and SCL33a (Table 4.1). Equal amounts of RNA were used in the reverse transcription reaction (Figure 4.6) and 10% of the resulting cDNAs were used as template for the subsequent PCR reaction.

just barely detectable by RT-PCR. Additionally, in leaves of *sr45-1* mutants, mRNA4 was expressed at higher levels. In flowers, mRNA3 was present in the *sr45-1* mutant while completely absent in the other three plant lines. The general splicing pattern in roots differed from rosette leaves and flowers greatly. For instance, mRNA4, which is not seen in rosette leaves or flowers, is expressed at high levels compared to the functional-protein coding transcript mRNA1 in all four lines. Also, in roots, mRNA5 is present, while it is absent in leaves and flowers. However, the *srp30-1* mutant was the only mutant to affect the alternative splicing of the *RSp31a* transcript in roots. In this mutant, mRNA3 is absent, while mRNA5 is expressed at much higher levels than observed in the other three plant lines, clearly indicating that *SRp30* either directly or indirectly plays an important role in the regulation of the *SRp31a* gene in roots.

SRp34b. The SRp34b transcript was mostly unaffected by the three mutants used in this study, and very minor alterations in mRNA transcript were observed (Figure 4.3b). In flowers of the *sr45-1* mutant, a small amount of mRNA2 was observed, which is absent in the other three plant lines. Additionally, the *sr45-1* mutant produced slightly lower expression levels of mRNA3 in roots. In wild-type roots, mRNA4 is expressed at very high levels, while the same transcript is produced at much lower levels in all three mutants.

SRp30. The *SRp30* transcript level and isoform production was mostly unaffected by the three mutants used in this study (Figure 4.4a). The only alteration

observed was in the rosette leaves of the *sr45-1* mutant. In this line, mRNA2 is produced at almost the same level as the functional isoform mRNA1. The other three lines lack mRNA2 and produce twice the amount of mRNA1 that is observed in the rosette leaves of the *sr45-1* mutant.

SCL33. The *SCL33* mRNA was altered only in roots, and was done so in all three mutant lines (Figure 4.4b). The *sc35-1* mutant appeared to produce all five isoforms at slightly higher levels than wild type. Conversely, the *sr45-1* mutant produced all five isoforms at slightly lower levels. However, neither of these two mutants appeared to alter the ratio of the different isoforms that accumulated in the tissue of the roots. The *srp30-1* mutant was the only mutant that significantly altered the isoform ratio and overall accumulation of the *SCL33* transcript. The full-length mRNA2 transcript, while expressed at very low levels compared to that of wild type and the other two mutants, is the only transcript in this mutant expressed at a higher level. mRNA1 and mRNA3 are completely absent in the root, while mRNA4 and mRNA5 are barely detectable.

SR1. The *SR1* transcript remained mostly unaffected in all three tissues in all three mutant lines (Figure 4.4c). However, although no alteration in splicing is observed in rosette leaves and roots, a high accumulation of mRNA2 is detectable in flowers of the *sr45-1* mutant. Also of note is an increase of mRNA1 in flowers of all three mutants in comparison to that of wild-type flowers.
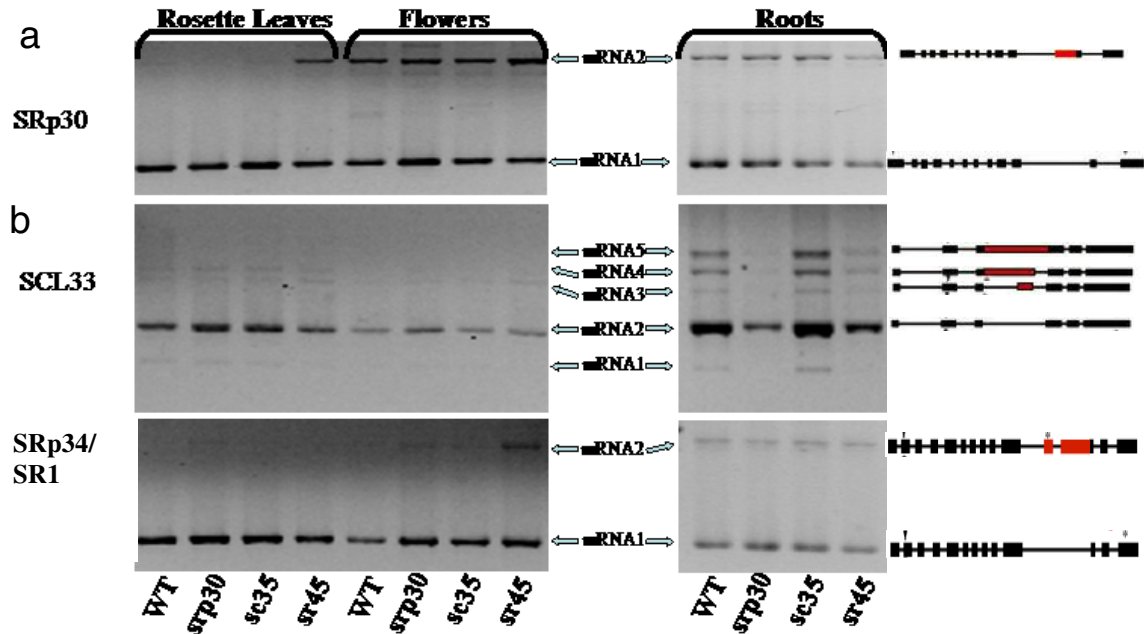
Figure 4.4 Expression levels and patterns of *SRp30, SCL33*, and *SR1* mRNAs are altered by mutations in *srp30-1* and *sr45-1* SR protein genes. RT-PCR was performed on whole RNA extractions of 10 day-old tissue using gene specific primers for RSp31, RSZ33, and SRp34a (Table 4.1). Equal amounts of RNA were used in the reverse transcription reaction (Figure 4.6) and 10% of the resulting cDNAs were used as template for the subsequent PCR reaction.

RSp40. The mutant lines appeared to have no affect on alternative splicing of the *SRp40* transcript in roots, although two additional mRNAs were found in this tissue but were completely absent in leaves and flowers of all four plant lines (Figure 4.5a). In rosette leaves of the *sc35-1* mutant, the *RSp40* transcript is not detectable, while the other two mutants have no effect on the transcript production compared to wild-type plants. Similarly, in flowers, the *RSp40* transcript is not detectable in the *srp30-1* mutant, and the other two mutants have no effect on the *RSp40* mRNA compared to the wild-type line.

SR45. The SR45 SR protein gene transcript was affected by only the *srp30-1* mutant(Figure 4.5b). While no change in alternative splicing pattern is observed in rosette leaves, the *SR45* mRNA1 level is increased while the mRNA2 level is decreased in the flowers and roots of the *srp30-1* mutant. The *sc35-1* mutant has no effect on *SR45* splicing pattern, and no product is detected in *sr45-1* mutant rosette leaves or flowers. However, a very low level of transcript is observed in roots of the *sr45-1* mutant.

SC35. The *SC35* splicing pattern was unaffected by any of the mutants in rosette leaves (Figure 4.5c). However, both flowers and roots contained transcripts in mutant lines that differed significantly from what was observed in wild-type plants. In the flowers of *srp30-1* mutants, mRNA1 was elevated by approximately three fold compared to that of wild-type. The *sr45-1* mutant had no effect on the alternative splicing of *SC35* in flowers. The alternative splicing of *SC35* was most affected in

85
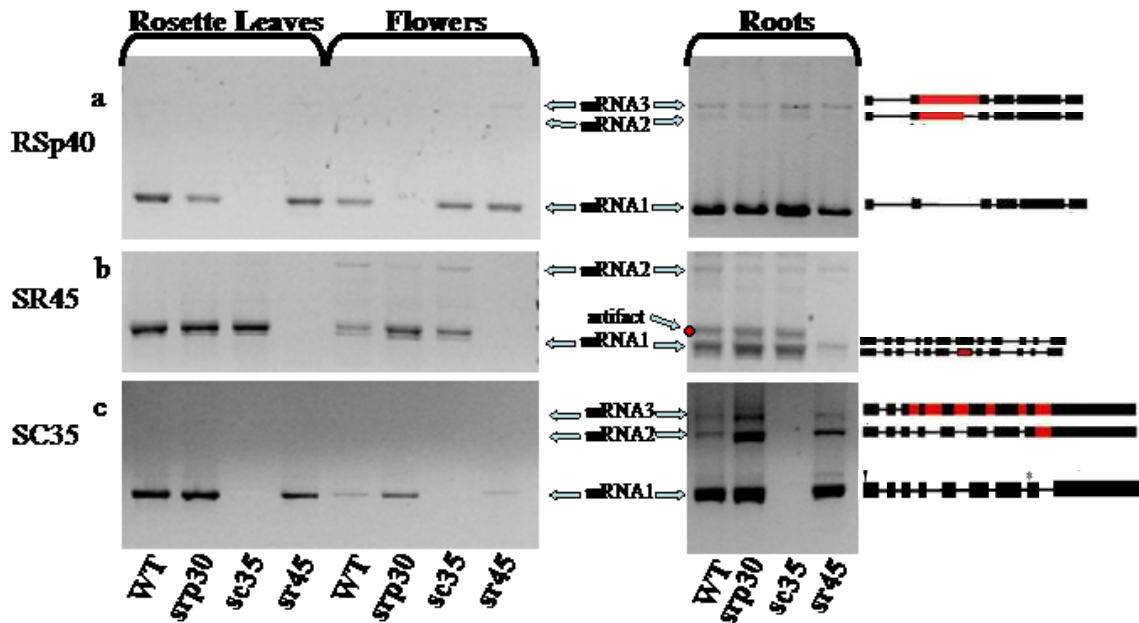
Figure 4.5 Expression levels and patterns of *SRp40, SR45*, and *SC35* mRNAs are altered by mutations in *srp30-1* and *sr45-1* SR protein genes. RT-PCR was performed on whole RNA extractions of 10 day-old tissue using gene specific primers for RSp31, RSZ33, and SRp34a (Table 4.1). Equal amounts of RNA were used in the reverse transcription reaction (Figure 4.6) and 10% of the resulting cDNAs were used as template for the subsequent PCR reaction.

the roots of both *srp30-1* and *sr45-1* mutants. In roots, two additional mRNAs are present in wild-type plants, as well as both the *srp30-1* and *sr45-1* mutants. The *srp30-1* mutant accumulated the same amount of mRNA1 as wild-type plants, while expressing equivalent levels of mRNA2 and mRNA3. Wild-type plants also produced very low levels of these two isoforms. The *sr45-1* mutant produced an equivalent amount of mRNA1 and mRNA3 compared to wild-type plants, but expressed a much higher level of mRNA2. As expected, no *SC35* transcript is produced in any of the three tissues in the *sc35-1* mutants.

Discussion

The findings from this experiment were very diverse regarding how individual SR protein gene mutants affected the alternative splicing of SR protein genes. Six of the 15 alternatively spliced SR protein genes studied here showed an alteration in isoform production in one or more of the mutant lines. There were three genes (*RSp41, SCL30*, and *RSZp32*) which appeared to be completely unaffected by the *srp30-1, sc35-1,* or *sr45-1* mutants. Differences in isoform ratio of three genes, *RSp31, RSZp33*, and *SRp34a*, did not appear to be affected by the individual mutations used in this experiment although these mutations did affect the expression level of the single transcript that encodes the functional protein produced by each of these three SR protein genes.

One of the more interesting expression patterns observed was that of the *SC35* protein gene (Figure 4.5c). The *SC35* gene produces three unique transcripts in roots, mRNA1, in which translation results in a functional protein, and two additional

transcripts which contain intron sequence with premature stop codons located within the RS domain. In wild-type roots, mRNA1 is expressed at very high levels, while mRNA2 and mRNA3 are produced at relatively low levels. Both the *srp30-1* and *sr45-1* mutants affected the isoform ratio and production, but in different ways. The point of most interest is that, in both mutants, the *SC35* functional mRNA is produced at the same or slightly higher level than is seen in wild-type plants. The alternatively spliced isoforms differ the most dramatically. In the *srp30-1* plant, mRNA2 and mRNA3 are expressed at levels near that of mRNA1. In the *sr45-1* plant, mRNA2 is substantially increased while mRNA3 is unchanged compared to that found in wild-type plants.

A possible explanation for this phenomenon is that SC35 may play more of a "helper" role in alternative splicing in *Arabidopsis*. In both the *srp30-1* and *sr45-1* mutants, the total transcript level of *SC35* appears to be expressed at very high levels which is clearly in excess of that seen in wild-type plants as though it is compensating for the loss of either of the other two mutants. The phenotypes associated with these three mutants also suggest that SC35 may not play as significant a role in splicing in *Arabidopsis* as *SRp30* and *SR45* do. Both *srp30-1* and *sr45-1* mutant plant lines display severe phenotypes that affect not only the physical appearance of the plants but also their ability to reproduce. However, the *sc35*-1 mutant, under normal growing conditions, does not display any major mutant phenotype and is, in fact, indistinguishable from wild-type plants at maturity. SC35 may not be as essential in *Arabidopsis* as it is in animals, and it may have more of a "housekeeping" role in alternative splicing in plants.

As previously stated, the advantage of performing this type of analysis in SR protein gene mutants, compared to overexpression experiments, is that it would give a better idea of how a particular SR protein functions *in vivo*. Previous overexpression data from the Barta lab indicates that SRp30 plays a very important role in regulating RSp31 and SRp34/SR1 (Lopato *et al.* 1999). In their report, overexpressing SRp30 results in accumulation of an alternatively spliced isoform of SRp34/SR1 showing intron retention, as well as an alternative splicing isoform of RSp31 mRNA showing exon skipping. However, the RT-PCR analysis of both of these genes in the *srp30-1* mutant plant showed no difference in splicing pattern. The *SRp34/SR1* mRNA in all three tissues of the *srp30-1* mutant was identical to that of wild-type plants. The *RSp31* gene produced only the full-length mRNA in leaves and roots of the *srp30-1* mutant. However, in flowers of the *srp30-1* mutant, the functional-protein coding transcript was induced at a much higher level, roughly three-fold of that seen in wild-type plants. These data significantly differ from that observed in overexpression experiments, and they further stress the need for additional SR protein gene mutants to deduce the true crossregulation ability of *Arabidopsis* SR protein genes.

The overall results from this experiment were surprising when looking at the data from the position of this experiment alone. While we anticipated that an SR protein gene mutant might alter the alternative splicing of another SR protein gene we did not expect to see so many effects. Despite the fact that only three individual mutants were tested, we observed significant alterations in the transcripts of 12 of the 16 alternatively spliced SR protein genes. Taking a step back and analyzing this data with that collected from the ESE assay, it is becoming clear that for an individual

splicing event to occur correctly and efficiently it probably requires multiple SR proteins. While the ESE assay indicated that one of multiple mutations can affect the ability of a target gene to be correctly spliced, this assay also indicates that multiple SR proteins are probably responsible for a single splicing event to take place. This deduction is not only based on the fact that such a large percentage of alternatively spliced genes within this single family are affected by mutations in only three individual SR proteins, but additionally, there were very few instances where a single isoform was completely absent or expressed only in a single mutant line. For example, in the roots of the *srp30-1* mutant, RSp31a mRNA3 is not present while it is expressed at low levels in the other two mutant lines. In the rosette leaves of the *sr45-1* mutant, the SRp30 mRNA2 transcript is present, while it is absent in wild-type, *srp30-1*, and *sc35-1* mutants. Events where an mRNA are absent in a mutant line when present in wild type, are relatively rare in this study and possibly due to another SR protein compensating for the lack of one of the specific mutants used in this study.

The data observed in these two experiments, the ESE assay and the crossregulation study, may hint at why viable mutants in mammals do not exist. While the *Arabidopsis* and human genomes are roughly the same size, *Arabidopsis* has twice as many SR protein genes and roughly three fold fewer targets of alternative splicing (65% of human genes are alternatively spliced compared to 23.5% in *Arabidopsis*). Some splicing events may be affected and regulated by a single SR protein. However, the data from this experiment suggest that the majority of splicing

events are probably regulated by multiple SR proteins, and mutations in multiple genes are needed to completely eliminate or significantly alter splicing.

Like the ESE assay, additional mutants in SR protein genes are necessary to fully deduce the complex interactions of individual SR proteins needed for a specific splicing event to occur. One problem with this particular experiment is that there is still no evidence that an individual SR protein is directly responsible for the processing and splicing of a specific target, in this case, other SR protein genes. It is very reasonable to believe that the effects on splicing seen in the mutants used in this study are indirectly regulated through other affected SR proteins in the same mutant line. Only the combination of this type of experiment, protein-protein interaction assays, and possibly in vitro experiments, will provide a much clearer insight into this very complex question.

A final limitation of this experiment, which is also a point of future progress, was that the experiment was performed using tissue collected under a single growth condition. Ideally, many more samples should be analyzed under various environmental and nutritional conditions. Fortunately, an alternative splicing gene chip is currently being developed in our lab which will alleviate the tedious and time consuming work needed to perform RT-PCR, and many more RNA samples can be analyzed. Finally, the gene chip will also increase the number of gene targets from 16, the number of alternatively spliced SR protein genes, to thousands of candidates of alternative splicing gene regulation.

Materials and Methods

Growth Conditions. Plants that were grown for leaf and flower RNA extraction were grown on soil under 24 hours of light at 20 degrees Celsius and were watered as needed. Plants grown for RNA isolation from roots were grown on MS plates and 1% agar without sucrose or vitamins. Plates were held vertically allowing roots to grow on the surface of the media. Plants grown on plates were also placed under 24 hours of light at 20 degrees Celsius. Petri dishes were wrapped with M3 surgical tape to retain humidity.

RNA Analysis of Homozygous Mutants. RNA was extracted from the 3rd and 4th rosette leaves at the eight rosette leaf stage, flowers from adult plants, and roots of 10 day old plants with a Qiagen RNeasy kit following manufacturer's instructions. cDNA was synthesized from 100 picograms of total RNA using the Taqman™ First Step RT-PCR following the instructions provided by the manufacturer. Ten picograms of cDNA from each sample were subjected to PCR using gene specific primers (Table 4.1). Verification of equal amounts of RNA used in each reverse transcriptase reaction was performed using primers specific for GADPH (Figure 4.6). All primers were designed with a melting temperature of approximately 55 degrees for a PCR program of 35 cycles of 95 degree denaturation for 10 seconds, 55 degree annealing for 10 seconds, and 72 degree elongation for 60 seconds. A 72 degree incubation period for seven minutes was performed after the final PCR cycle. PCR products were run on a 3% agarose gel. RNA analysis was

performed on three unique RNA samples for each specific RT-PCR reaction with

only one is shown in this report.

Figure 4.6 Equal amounts of RNA were used in crossregulation experiments.

RNA samples were standardized with GADPH for equal amounts of RNA to insure

the differences in crossregulation are due to relative differences in specific transcript

concentration and not a difference in overall RNA concentration.

Gene Name	Primer Names	Sequences
RSp40	xz007_RSp40F	AGG TGA CCT GGA ACG ACT ATT CAG G
	xz008_RSp40R	TAG ATC CCG GGT CCT AGT ATT ATC C
RSp41	xz009_RSp41F	GAG TGA TCT TGA GCG GCT TTT CAG A
	xz010_RSp41R	CAA GTC CCG AGT CCT AGT GTT TTG T
RSZ32	xz011_RSZ32F	TAT TAC TTG GAT GGA CGG GAT TTC
	xz012_RSZ32R	GAT ACA GAT CGC TCC ATT GCC TTA
RSZ33	xz013_RSZ33F	GAA CAC TCG TCT TTA CGT TGG CCG A
	xz014_RSZ33R	GCG ATA GAG AGT CAT CCA TCC GCT
SC35	xz015_SC35F	CGA CTG CTG ATG ATC TCT ATC CTC T
	xz016_SC35R	CTG TGG GGA GGA ACT TCT AGG GCT G
SCL30	xz017_SCL30F	GAG GAT ACG GTG GTC GAG GTA GAA GT
	je315_SCL30R	CTGAAGCAACAACCACCGTTATCTC
SCL33	xz018_SCL33F	AGA GGC CGA TAT GGA GGT CGT AG
	xz019_SCL33R	GGA GGA GAA CGA GAG CGA GAG T
SCL33a	xz020_SCL33aF	TAG CCC TCG GGG CCG GTT GGG TGG GA
	xz021_SCL33aR	TGG CTC TTC ACT CTC CCA GGA CTT CC
SRp30	xz023_SRp30F	CTA GTG CTT ATA TAC GGG TGA GGG A
	xz024_SRp30R	CGA TCT TGA TCT TGA TCT TGA TTT T
SRp34a	xz026_SRp34aF	CAA GGA AGG ATC TGA GTA AAT
	xz027_SRp34aR	TAA CAA AAT CAC ACA CTG CCT
SRp34b	xz028_SRp34bF	CGA GGA CAT GAA ATA TGC GAT AA
	xz029_SRp34bR	CAA TGT CTC TCC ACT GTT ACC CAT
SRp31a	xz030_SRp31a(102)F	TCA AGG AAT TCG ATT GCT TTT GCA
	xz031_SRp31a(102)R	CAA AAG TGG TAT TGT CAG TCC TA
SRp34/SR1	xz025_SRp34(SR1)F	GAG GAC ATG AAG TAT GCG CTG A
	je0096 SR1 R	CGATGGACTCCTAGTGTGGA
RSp31	je305_RSp31R	ACGACCACCATTTCGATCATCTCTT
	xz006_RSp31F	CTT CTC CGA TTG TTC AGA GGA AT
SR45	je324_SR45F	GCTGACGCTGAGAAAGATGGTGGTC
	xz022_SR45R	TTG GAG GGG GAG AAG ATG GAG AAC G

Table 4.1 Gene specific primers used to amplify products from cDNA generated by Reverse Transcriptase from RNA isolated from rosette leaves, flowers, and roots of WT plants and plants mutant for SR protein genes SRp30, SC35, and SR45.

Chapter 5: Expression patterns of SR protein genes in the *Arabidopsis thaliana* root

Introduction

      The extent to which individual SR protein genes are redundant or play very specific roles in alternative splicing remains one of the more central questions in SR protein gene study. Molecular interaction experiments, both *in vivo* and *in vitro*, can determine the potential for individual SR proteins to perform a specific task, but do little to definitively identify the exact tissues and points in development at which the respective function naturally occurs. To determine whether or not an individual SR protein gene processes a specific mRNA, or class of mRNA, in a specific organ is only a small part of elucidating the exact role of a specific SR protein gene. Individual organs such as flowers and roots are composed of multiple tissue types,

each of which may express a completely different set of SR protein genes depending upon the unique tissue. Identifying the exact organs, tissues, and cells in which individual SR proteins are expressed will play an important role in helping resolve the exact function of individual SR proteins.

The Barta lab created *SRp30* and *SRp34/SR1* promoter-*GUS* constructs to identify expression patterns of individual SR proteins (Lopato *et al.* 1999). In this study, they found that the two promoters were active simultaneously in some tissues while having completely opposite expression patterns in others (Lopato *et al.* 1999). For instance, in two day old seedlings, the *SRp30* promoter is active in cotyledons but not in the root, while *SRp34/SR1* was expressed in root but not cotyledons. Additionally, the *SRp34/SR1* promoter was active in the entire root but not expressed in any cells in cotyledons. In adult roots, the *SRp30* promoter-*GUS* transgene was expressed in a very specific manner, active only at the lateral root junctions while the *SRp34/SR1* promoter-*GUS* transgene was more broadly expressed in root tips and vascular tissue. However, both promoters were active in the same tissues in flowers at the postanthesis stage (Lopato *et al.* 1999). This experiment shows that while individual SR proteins can have distinct expression patterns, they are also capable of being expressed in the same tissues at different developmental stages. However, as we now know, both of these SR proteins are regulated through alternative splicing which means that while the promoter may be active that does not necessarily indicate that functional native protein is being produced in the same tissues. Additionally, the promoter-*GUS* fusion constructs were very broad in their staining capabilities. In the

SR-GUS pictures from the Barta lab, organs are deeply stained but cell specificity within organs can not easily be deduced.

The Spector lab (Fang *et al.* 2004) used the knowledge that promoters of individual SR proteins can be active in a tissue-specific manner and aimed to identify cell type specificity in a high-resolution manner. The new approach, once again, consisted of fusion proteins of which the native promoter was used to express a transgene *in planta*. In addition to the native promoter, the transgene consisted of the cDNA coding for the functional SR protein being studied with a C-terminal Yellow Fluorescent Protein (YFP) fusion (Fang *et al.* 2004). This approach would allow for more of a cell specific investigation into where the promoter of SR protein genes *SR33* (referred to as SCL33 in this thesis and in all other published works cited in this paper)*, SR1*, and *SRp30* are active. Using confocal imagery, specific cell types were, in fact, identified with a higher degree of specificity than seen with the promoter-*GUS* assay used in the Barta lab. Distinct nuclei can be easily visualized in the YFP fusion transgenic lines and it is therefore easier to deduce the exact tissues that contain those fluorescent nuclei (Fang *et al.* 2004). For example, in the Lapoto paper, the entire root of the *SRp34/SR1* promoter-*GUS* fusion seedlings expresses *GUS* and no cellular specificity is observed (Lopato *et al.* 1999). In contrast, the *SRp34/SR1-YFP* transgenic lines clearly display expression in the epidermis of cells closer to the root meristem and switch to expression mainly in vascular tissue and endodermis of cells that contain root hairs (Fang *et al.* 2004).

While the Fang paper demonstrated a higher level of specificity, the experiments were still under the same constraint as the Lopato *et al.* (1999)

expression pattern experiments: the regulation of alternative splicing was not taken into consideration when the constructs were created. Once again, these transgenic lines are expressing the transgene only where the promoter is active, not necessarily where the functional protein is being produced. Additionally, there was one other anomaly between the two different studies. While both labs investigated the expression patterns of both *SRp34/SR1* and *SRp30*, which was dependent entirely on the active promoter, the data they obtained were remarkably dissimilar, even when taking into consideration the difference in resolution. For example, in the Lopato *et al.* (1999) paper, it is clear that *SRp30* is active in cells only at the junction of the primary root and lateral root as well as root primordia. However, the *Fang et al.* (2004) paper clearly shows that the *SRp30* promoter is active in the vascular tissue of the root in the developmental stage prior to where root hair emerges, and then is more broadly expressed in epidermis and cortex of the root containing root hair. This difference could be the result of several factors including the length of promoter sequence used, the cDNA-*YFP* fusion protein compared to the GUS proteins, environmental conditions the plants were grown in, or nutrient differences in the growth medium among other potential environmental factors.

Another interesting collection of SR protein gene expression data came from a global analysis of gene expression in roots using a microarray gene chip (Birnbaum. *et al.* 2003). The microarray data incorporated making protoplasts from the *Arabidopsis* root and using cell-specific fluorescent cell sorting. Five individual cell types were defined in addition to three unique developmental stages. This resulted in dividing the root into a total of 15 different cell types from which expression data was

collected (Figure 5.1). Of the three expression pattern methods outlined in this

introduction, the microarray method is clearly the most quantitative and the least

subjective in analyzing the data. However, once again, this approach will also

potentially identify all mRNA transcripts of each gene and will not take into

consideration alternative splicing regulation. Like the previous reports, this

experiment will not discriminate against mRNAs that do not produce functional

proteins due to alternative splicing. Additionally, this approach will also miss any

minor nuances in expression that may be present. For instance Lopato *et al.* (1999)

found that *SRp30* is very specifically expressed at lateral and primary root junctions.

This subtlety will not be picked up in the microarray data and will only be recorded as

A Endo+cortex
Endo
Stele
Epi
Lrc

Birnbaum 2003

3
2
1
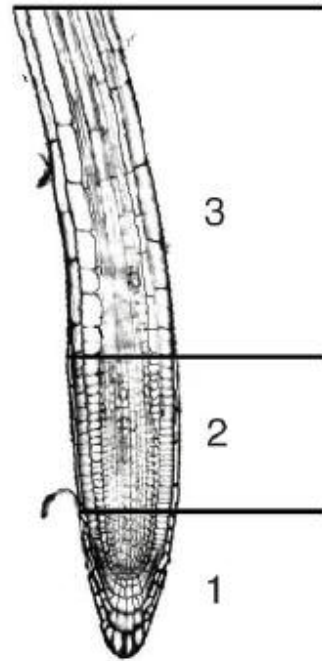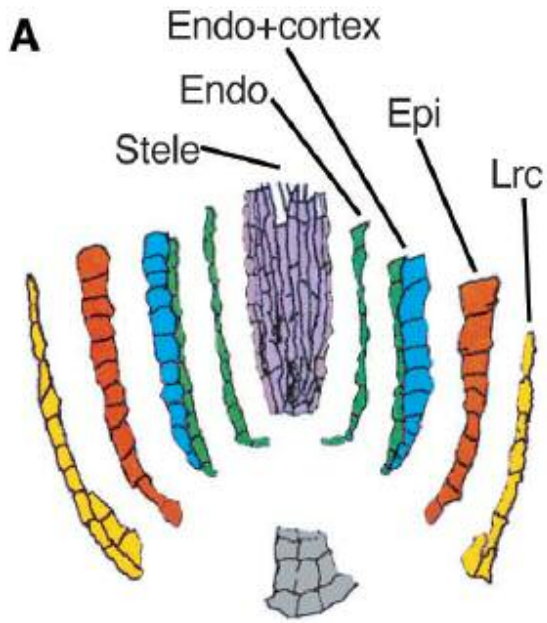
Figure 5.1 Expression profiling cells and tissues at three developmental stages and validation of the technique. Reproduced from Birnbaum *et al* 2003. (A) Radial zone expression profiles were generated by protoplasting roots with cellulase and pectolyase and sorting GFP marker lines. From inner stele to outer lateral root cap (left), GFP marker lines used to isolate cell types and tissues were, respectively, pWOODEN LEG::GFP (Mohonen *et al.* 2000; Bonke *et al.* 2003), pSCARECROW::GFP (Wysocka-Diller. *et al.* 2000), J0571(Haseloff Laboratory), pGLABRA2::GFP(Lin *et al.* 2001), and J3411 (Haseloff Laboratory). Abbreviations are endo, endodermis; endocortex, endodermis and cortex; epi, epidermal atrichoblasts; and lrc, lateral root cap. For developmental stage profiles (right), numbers indicate developmental stages profiled, which were dissected with the use of the following landmarks as upper borders: stage 1, where the root tip reached its full diameter (about 0.15 mm from the root tip); stage 2, where cells transition from being optically dense to a more transparent appearance as they begin longitudinal expansion (about 0.30 mm from the root tip); and stage 3, where root hairs were fully elongated (about 0.45 to 2 mm from the root tip). All zone and stage profiles were repeated in triplicate.

present in certain tissue types of stage three roots as opposed to the much more specific expression pattern that is observed visually. The actual expression pattern can be determined using a combination of microarray data already available, the alternative splicing microarray chip being developed in our lab, and additional transgene experiments.

The Birnbaum microarray gene set included all 20 SR protein genes and demonstrated that SR proteins are both temporally and spatially expressed in Arabidopsis root (Table 5.1). However, as mentioned earlier, these data are only reliable in assaying the active promoter and not necessarily locations of mRNA that codes for functional protein. With this being said, there are some genes that are expressed in very specific tissues and developmental stages. For example, *SCL28* does not appear to be expressed at all in any stage of the root while *RSZp33*, a plant specific SR protein gene, is expressed at very high levels in all tissues at all three developmental stages defined by Birnbaum *et al*. Another SR protein gene, *RSZp21*, is expressed in all tissues of the root but is expressed 3 fold higher in stage one than in stage two or three.

Improvement on the visual SR protein expression experiments required developing a transgene that is sensitive to the regulation, alternative splicing, and other transcriptional and translational regulatory factors that native SR protein genes would be subject to *in planta*. To accomplish this feat, 16 SR protein genes were cloned from genomic DNA consisting of sequence 2Kb upstream of the start codon to the last amino acid before the stop codon of each respective gene. Four SR protein genes, *SR45, RSp41, SCL30*, and *SCL33a* were never successfully cloned from either

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1		stele stage1	stele stage2	stele stage3	endo stage1	endo stage2	endo stage3	cortex-endo stage1	cortex-endo stage2	cortex-endo stage3	epidermis stage1	epidermis stage2	epidermis stage3	lat root cap stage1	lat root cap stage2	lat root cap stage3	
2	SCL28	3	4	3	6	9	7	5	7	6	5	7	6	6	8	7	
3	SRp34b	17	13	14	13	10	10	12	9	9	10	7	8	10	8	8	
4	SRp007	90	36	28	101	40	32	96	38	30	89	35	28	72	29	22	
5	SRp31a	63	52	36	74	61	42	81	67	46	69	57	39	95	79	54	
6	SR45	126	79	56	160	100	71	181	113	81	144	90	64	159	99	71	
7	SCL33	99	89	78	107	97	84	168	152	132	111	101	88	145	131	114	
8	SRp30	118	145	195	102	125	168	99	122	164	92	113	152	89	110	148	
9	RSZp21	299	141	117	389	183	152	324	153	127	252	119	99	234	110	92	
10	SRp41	327	148	104	299	135	95	389	176	124	443	201	141	272	123	87	
11	SCL33a	321	185	81	401	231	101	407	235	102	313	181	79	294	170	74	
12	SRp34a	344	195	110	309	175	99	462	261	148	251	142	80	346	196	111	
13	SC35	372	159	133	300	128	107	446	190	159	374	159	134	334	142	119	
14	RSZp22a	343	223	149	311	203	135	332	216	144	373	243	162	268	174	116	
15	SRp40	223	173	90	493	383	199	343	266	138	249	193	101	332	257	134	
16	RSp31	396	228	140	438	252	155	597	344	211	470	271	166	431	248	152	
17	RSZp32	393	208	135	579	307	199	569	301	195	454	240	156	533	282	183	
18	SCL30	383	390	180	417	425	196	471	480	221	374	381	176	354	361	166	
19	SRp34/SR1	535	285	182	743	396	252	684	364	232	599	319	203	600	320	204	
20	RSZp22	787	533	347	541	366	239	796	539	351	901	610	397	652	441	288	
21	RSZp33	897	519	280	1844	1066	575	1499	867	468	777	449	242	1255	726	391	

Table 5.1 Expression levels of SR protein genes in root of Arabidopsis. The Arabidopsis root was divided into 15 sections based on developmental stage and radial tissue specificity (Figure 5.1). Relative amounts of transcript accumulated for each of the genes was based on the overall expression of the over 22,000 gene specific probes used on the Affymetrix microarray chip described in Birbnbaum *et al* 2003.. Work was performed by Birnbaum et al 2003.

genomic DNA or from BACs despite multiple attempts with various primer pairs. The 16 SR protein genes that were successfully cloned were inserted into an expression vector that provides a C-terminal GFP fusion protein in frame with the SR protein. The resulting transgenes therefore carry the native promoter and introns of each SR protein in fram with GFP. In theory, transgenes created by this method would capture all regulatory elements including upstream regulatory sequences in addition to being regulated by alternative splicing as well as other transcriptional and translational factors. The transgene mRNA should be processed identically as the native genes and thus the GFP signal will be visible only where the functional protein is produced. The majority of minor alternatively spliced isoforms of *Arabidopsis* SR protein genes are truncated and encode premature stop codons. Assuming that the transgene is processed in the same fashion as the native gene in instances in which functional protein is not being produced, *GFP* will not be observed even though the promoter is active. Mutants for *srp30-1, sc35-1*, and *sr45-1* were transformed with the same constructs as wild-type plants.

Using the method described above, I have identified the expression patterns of 16 of the 20 *Arabidopsis* SR protein genes in the root in wild-type plants. Additionally, I found that individual mutants do, in fact, alter the expression pattern of individual SR protein genes. Comparing expression patterns of the transgenes in wild-type plants and mutant plants has identified SR protein genes that are expressed differently in mutant tissues compared to that seen in wild-type plants. The most significant difference in this experiment, compared to similar experiments performed previously in other labs, was that my constructs would be regulated and processed as

a native full-length SR protein. Comparing the data collected from my experiment to that of previously published data (Lopato *et al.* 1999; Birnbaum. *et al.* 2003; Fang *et al.* 2004), I found significant differences in reported expression patterns. This indicates that my transgenes are potentially under the same regulation that native genes are subjected to.

Results

In order to fully describe the expression pattern of *Arabidopsis thaliana* SR protein genes, full genomic sequences, including introns and approximately 2Kb of genomic sequence upstream of the start codon, were PCR amplified from genomic DNA. Through a series of cloning steps, the PCR products were then cloned into a vector (pGlobug) that would result in a Carboxyl-terminal GFP fusion protein and later used for plant transformation. Constructs were designed for the cloning of all 20 *Arabidopsis* SR protein genes but only 15 were successfully cloned from genomic DNA, and *SRp30* could only be cloned from a BAC. Unfortunately, four SR protein genes, *SR45, RSp41, SCL33a*, and *SCL30* were never successfully cloned and thus, *GFP* fusion transgenes were never created for these four SR protein genes. Of the 16 SR protein-*GFP* fusion transgenes created, different success rates of transformation in *Arabidopsis* were observed, which varied from gene to gene and from plant line to plant line. Unfortunately, not every plant line was transformed with all 16 SR-GFP transgenes (Table 5.2). The transgenic plant lines that where never generated failed to show resistance to Finale herbicide on at least 4 occasions. Those that were selected for herbicide resistance were propagated with selection until T2 seeds were

GFP Transgene	WT	*sc35-1*	*sr45-1*	*srp30-1*
SCL28	T2	T2	T2	T2
SRp102	T2	T2	T2	T2
RSZp32	T2	T2	T2	T2
SC35	T2	T2	T2	T2
SRp34b	T2	T2	!	T2
RSZp21	T2	T2	!	T2
RSp31	T2	T2	!	T2
SRp34/SR1	T2	T2	T2	!
SCL33	T2	!	T2	!
SRp30	T2	!	!	T2
SRp007	T2	T2	!	T2
SRp34a	T2	T2	!	!
RSZp22a	T2	T2	T2	!
RSp40	T2	!	!	!
RSZp22	T2	!	!	!
RSZp33	T2	!	!	!
SR45 *	N/A	N/A	N/A	N/A
SRp41*	N/A	N/A	N/A	N/A
SCL33a*	N/A	N/A	N/A	N/A
SCL30*	N/A	N/A	N/A	N/A

Table 5.2 SR protein:GFP transgene constructs created for SR protein gene expression pattern analysis. Attempts to clone all 20 SR protein genes from genomic DNA were successful for 16 individual SR protein genes. Those genes that were not cloned are marked with an asterisk. Constructs that were created but not successfully transformed into a specific plant line are marked with an exclamation point. Constructs that were successfully transformed into a specific plant line and examined for transgene expression in plant roots are designated with a "T2" which represents the generation that was used for this experiment.

collected and this generation was used to grow the plants for the microscopy experiments.

Using confocal microscopy, identifying expression patterns and locations of individual SR-*GFP* fusion transgenes was remarkably simple and straightforward in roots. Very little laser power was needed to observe individual nuclei that were clearly expressing the SR-*GFP* transgenes. This is important because cell walls autofluorescence at roughly the same wavelength as *GFP*. However, it requires a lot more energy from the laser to detect autofluorescence from the cell wall. The difference in intensity of the GFP and autofluorescence allows an unhindered and detailed look at individual nuclei of individual cells located in roots of *Arabidopsis*. Additionally, through Z-series sectioning, confocal microscopy allowed a very detailed look at the radial tissue organization of the root and confirmation of the exact cells and tissue type showing expression of each transgene.

A more detailed analysis of expression patterns will be discussed in the following paragraphs, but there were some general observations that can summarize expression patterns of individual SR protein genes in roots. In general, SR protein genes are very highly expressed in all tissue types of Stage 1 and root tips. If a particular SR protein gene was only expressed in one stage, which occurred frequently, it was typically expressed in all tissues of stage 1 and the root tip. Additionally, there was a great deal of expression pattern variation with respect to age of roots. All data reported in this thesis is based on 10 day old roots. However, at the beginning of this experiment, I observed that the expression pattern of an individual SR protein gene could change dramatically between 10 day old roots and 14 day old

roots. One additional general phenomenon was that expression levels of stage three tissues typically varied in a gradient fashion coinciding with cell age. Older stage three cells, those closer to the stem of the plant, while expressing the respective GFP fusion transgene in the same manner as young stage three cells, typically did so at a much lower level in comparison to the younger tissue. Finally, there was also variation in expression pattern when comparing primary roots to lateral roots. While both types of roots theoretically contain the same tissue type, the expression pattern between the two root types was significantly different in some transgenic lines.

SR45, RSp41, SCL33a, and SCL30. These four genes were never successfully cloned from either genomic or BAC DNA preps. SR45 was successfully amplified from genomic DNA twice but for some unknown reason was never successfully cloned into the TA cloning vector. The remaining three genes were never PCR-amplified, even after multiple attempts with different primer pairs.

RSZp33. The RSZp33-GFP fusion transgene construct was only successfully transformed into wild-type plants. For unknown reasons, the transformation into mutant plants was unsuccessful on several occasions as progeny from T0 plants never survived selection for herbicide resistance. In wild-type plants, GFP expression was observed in all cells of the root tip and stage one tissue of primary roots. Additionally, GFP was observed in the same tissue type of older and longer lateral roots containing root hairs but was not observed in young lateral roots lacking root hair or root primordia. (Figure 5.2)
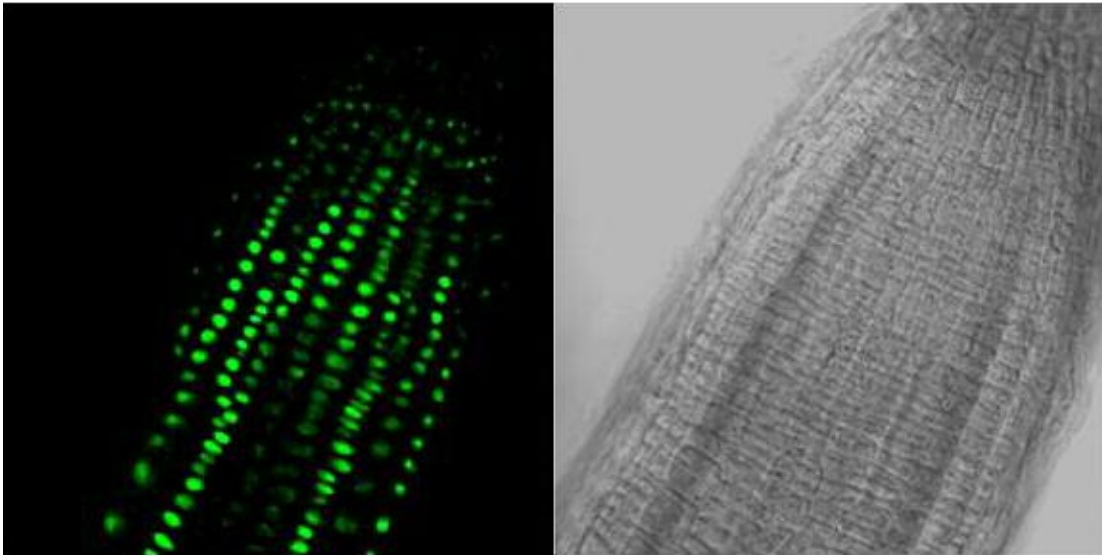
Figure 5.2 *BGlowRSZp33-GFP* transgene expressed in wild-type roots. A single

optical section of a 10 day old primary root tip of a wild-type plant is shown.

Transgene expression was limited to the root tip and stage one cells only of both

primary and lateral roots. No expression was observed in stage two or stage three

tissues.

RSZp22. Like the RSZp33 transgene construct, RSZp22 was only

transformed into wild-type plants although several attempts to transform mutant

plants were carried out. In primary roots, all cells of root tips and stage one tissue

(Figure 5.3a) expressed the fusion transgene, as did vascular tissue throughout all

three stages (Figure 5.3b). Additionally, a really interesting observation was made in

lateral roots. Lateral root primordia where additional epidermal cells had developed,

as opposed to bulges in the primary root where additional tissue will be formed,

expressing GFP in all cells (Figure 5.3c). However, as the lateral root elongated, the

expression pattern became much more specific. While no expression was observed in

stages two and three, stage one lateral root expressed GFP in all cells other than

epidermal and vascular tissue.

RSp40. Expression of the RSp40-GFP fusion transgene was observed only in

root tips and stage one tissues of primary roots only. No expression of the transgene

was observed in lateral roots of any age.

RSZp22a. The construct containing the genomic sequence for the RSZp22a-

GFP fusion protein was successfully inserted into wild-type, *sc35-1,* and *sr45-1*

mutant plant lines. In wild-type plants the GFP transgene was expressed in the

primary root at the root tip and all cells of stage one (Figure 5.4a). Additionally, in

stage two and stage three of the primary root of wild-type plants, the *RSZp22a*

transgene is expressed in the endodermis and vascular tissue and the surrounding
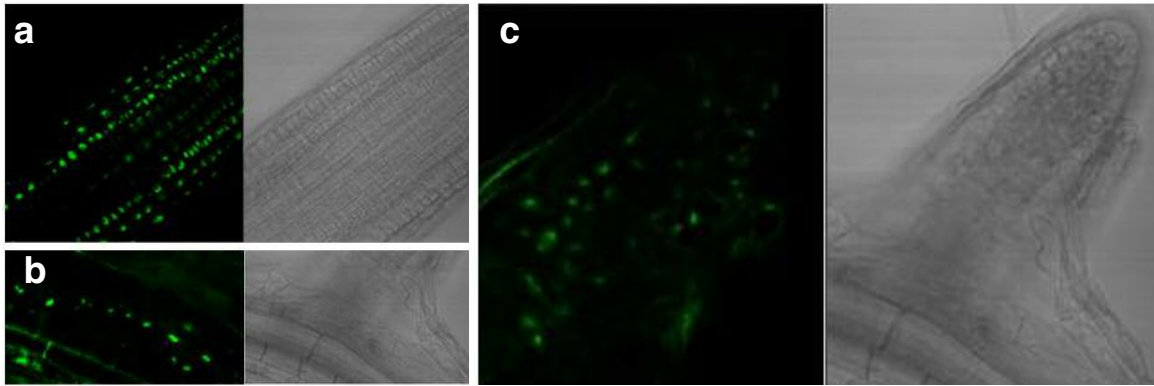
Figure 5.3 *BGlowRSZp22-GFP* transgene expressed in wild-type roots. Single

optical sections were photographed of 10 day old wild-type roots at three different

developmental stages. Stage one (a) and root tip of primary and lateral roots

expressed the transgene in all cells. Additionally, certain cells within the vascular

tissue expressed the *RSZp22-GFP* transgene at stage three (b). Lateral root primordia

emerging from the primary root (c) expressed the transgene in all cells.
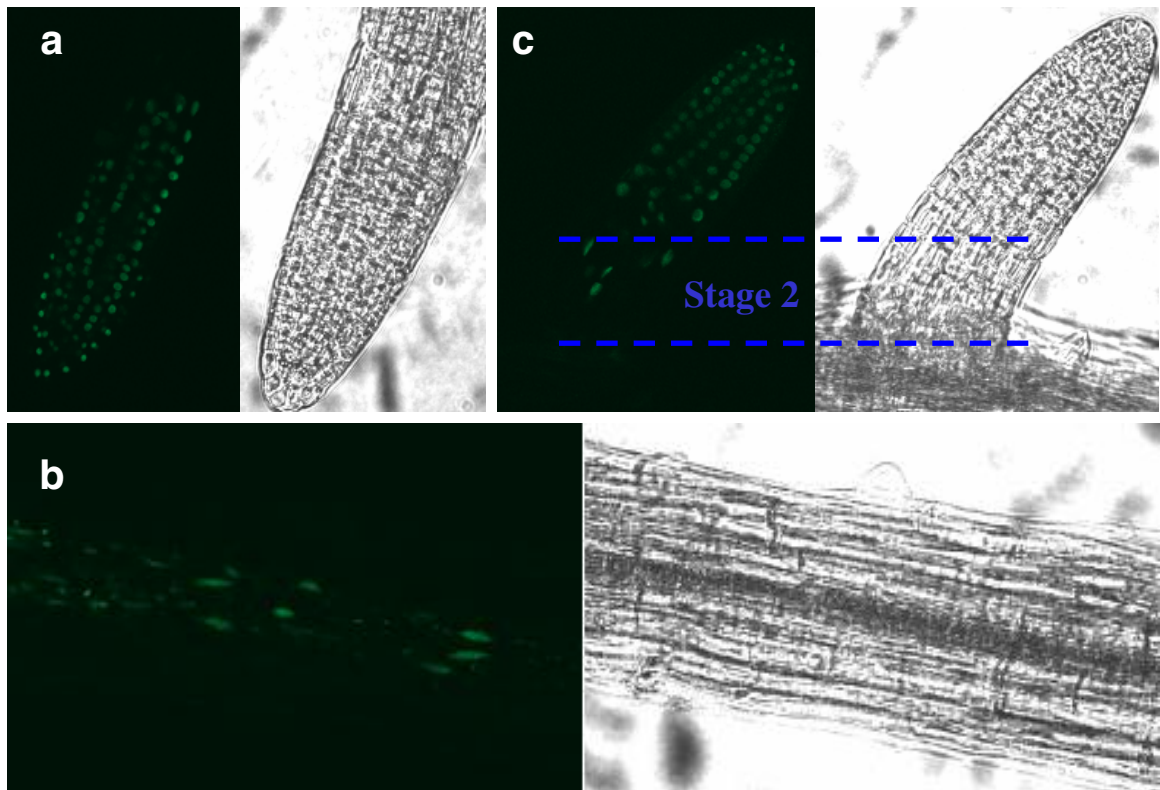
Figure 5.4 *BGlowRSZp22a-GFP* transgene expressed in wild-type roots. Optical sections of wild type roots taken of 10 day old roots at various developmental stages. In wild-type plants, the RSZp22a-GFP transgene is expressed in root tips and stage one cells of both primary roots (a). In stage three of the primary root (b), the transgene is expressed only in the endodermis and vascular tissue. Older lateral roots are expressed in a similar fashion to that seen in primary roots, while developing lateral roots (c) express the transgene in the epidermis of stage two cells.

endodermis but is not present in the epidermis or cortex of either of these two stages (Figure 5.4b). In long and older lateral roots, the exact same expression pattern is observed in all three stages. However, in shorter lateral roots, there is some very light expression of the *RSZp22a* transgene in the epidermis (Figure 5.4c).

In the *sc35-1* mutant background, the *RSZp22a-GFP* fusion protein gene was expressed in a slightly different fashion than seen in that of wild-type plants. Like wild-type plants, all cells of stage one of primary roots expressed GFP at equally high levels as seen in wild-type plants (Figure 5.5a). Additionally, the transgene was expressed in stage three, but not stage two primary root cortex (Figure 5.5b). No vascular tissue appeared showed expression in either stage one or stage two of the primary root but did so in stage three where lateral roots were emerging. While long and older lateral roots mimicked the expression pattern observed in the primary root, young lateral roots had a slightly different expression pattern. Stage one cells expressed the *RSZp22a-GFP* transgene at visibly lower levels while stage two also expressed the transgene in the epidermis and cortex at very low levels (Figure 5.5c).

sr45-1 mutants also displayed a different expression pattern of the *RSZp22a-GFP* transgene compared to that seen in wild-type plants. In primary roots and older lateral roots, the *sr45-1* mutant expressed the transgene in all stage one cells and root tip (Figure 5.6b). In stages two and three, the expression is observed in the cortex only (Figure 5.6a). Younger lateral roots, those consisting of only stage one and two cells, express the transgene in all cells except for vascular tissue (Figure 5.6a). While the differences in expression patterns between the wild-type, *sc35-1,* and *sr45-1*

114
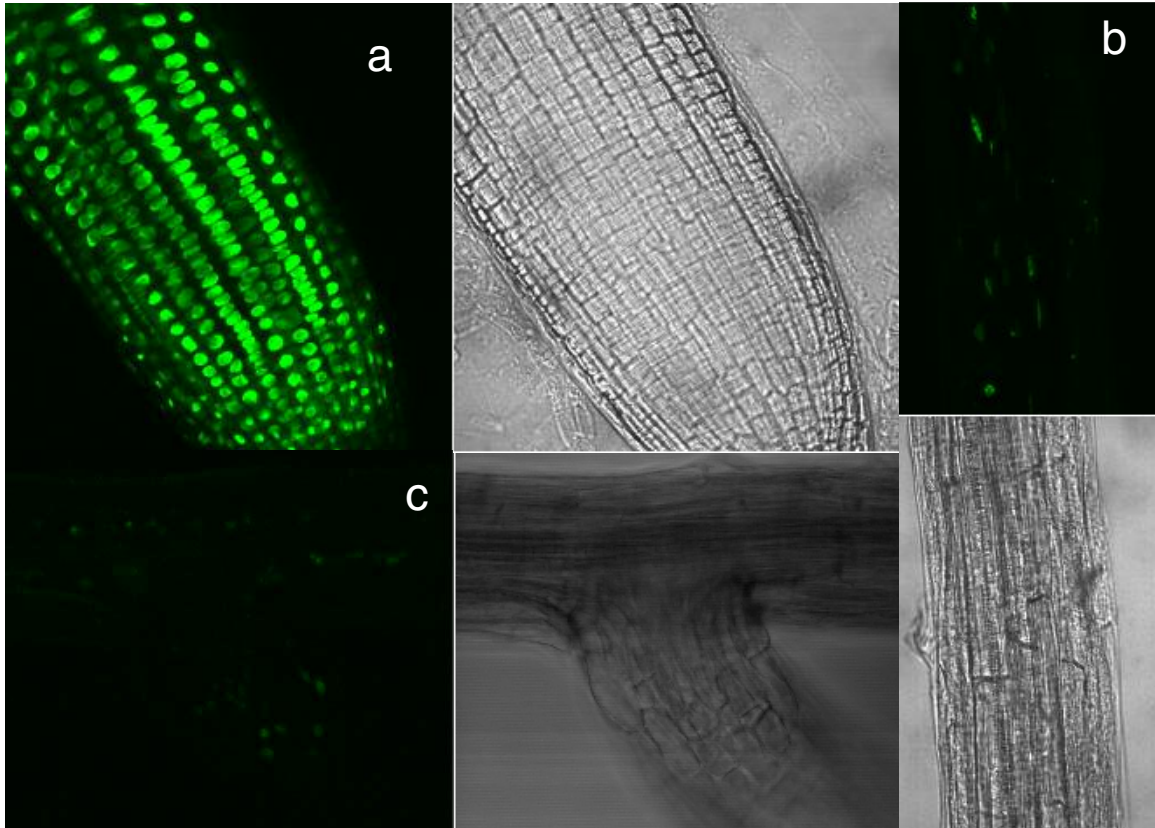
Figure 5.5 *RSZp22a-GFP* expressed in *sc35-1* mutant lines. Single optical

sections were taken of the *sc35-1* mutant line of 10 day old plants at three unique

developmental stages. The primary root and older lateral roots expressed the

RSZp22a-GFP transgene in all cells of stage one and root tip (a). While no

expression was observed in stage two of primary and older lateral roots, the cortex of

stage three (b) did express the transgene. In areas of emerging lateral roots (c), some

vascular tissue was observed to contain the GFP fusion protein in stage three at very

low levels.
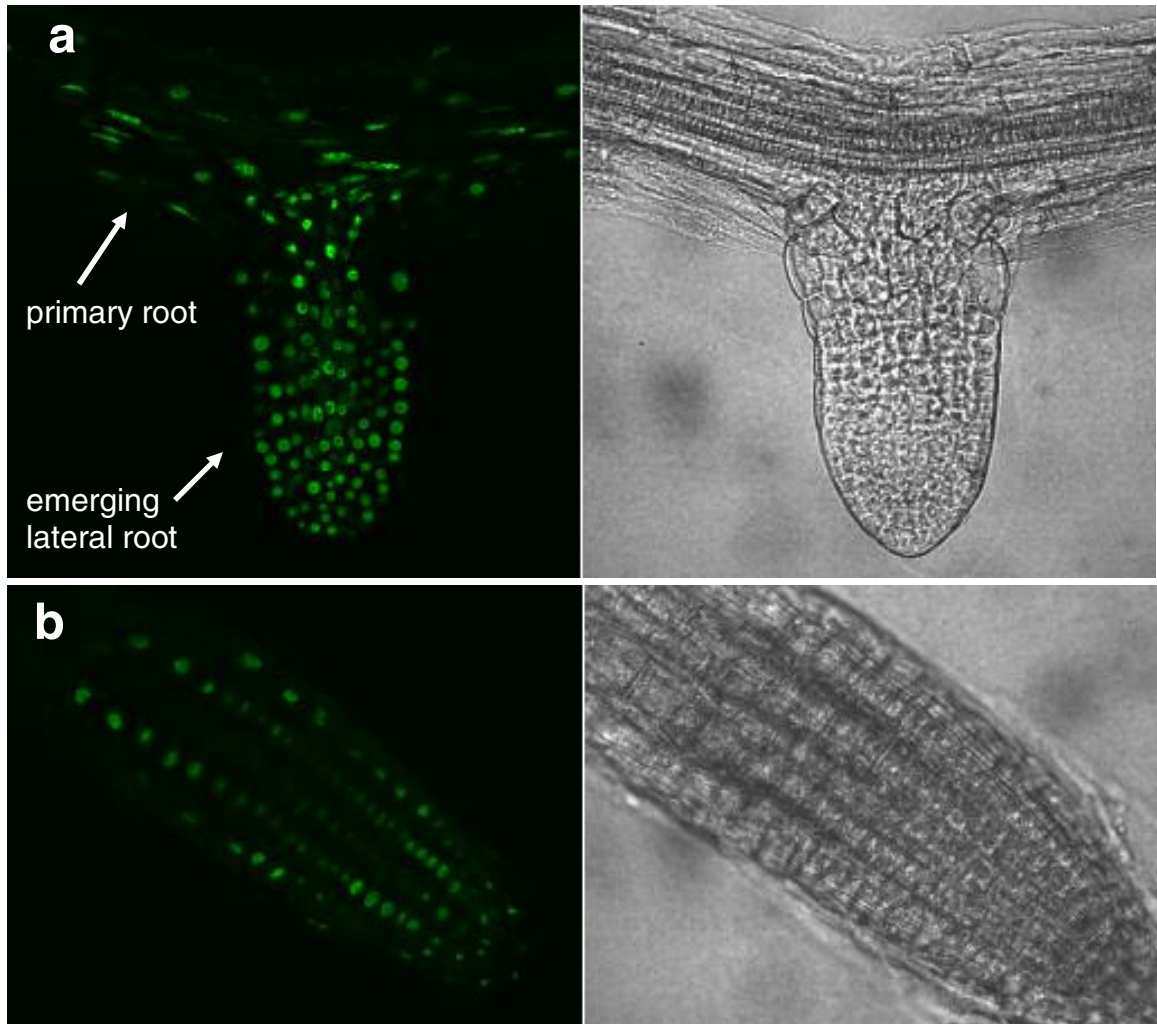
Figure 5.6 *RSZp22a-GFP* expressed in *sr45-1* mutant lines. Optical sections were taken of primary and lateral root tips of *sr45-1* mutant lines at 10 days post germination. *sr45-1* mutants transformed with the *RSZp22a-GFP* transgene display fluorescence in stage one cells and root tip of the primary root and older lateral root (a). In stages two of the same root (b), transgene expression is observed in the cortex. In young lateral roots (a), GFP expression is observed in all cells except vascular tissue.

mutants may be subtle in some tissues, while dramatically different in others, the fact that there are any differences between the two lines at all is very interesting. *RSZp22a* is not one of the known SR protein genes that are subject to alternative splicing. This observation will be expanded on further in the discussion and possible explanations will be offered.

SRp34a. The *SRp34a-GFP* fusion transgene was expressed in an identical pattern in both the WT (Figure 5.7a) and *sc35-1* mutant lines (Figure 6.7b), the only plant lines that were successfully transformed with the construct containing this particular transgene. The GFP fusion protein was seen in the primary root tip and stage one cells only. No other cell type in either developmental stage two or stage three of the primary root was observed. Additionally, no expression was observed in lateral roots in either genetic background.

SRp007. The construct containing the *SRp007-GFP* fusion transgene was transformed into wild-type, *sc35-1*, and *srp30-1* mutant plants. The expression pattern of this transgene was identical in both the wild-type and *srp30-1* plant lines. In primary roots (Figure 5.8a), the GFP fusion protein is observed in all cells of stage one, being highly expressed in all tissue types other than vascular tissue. As cells progress to stage two, only epidermal cells express the *SRp007-GFP* fusion protein gene, and by stage three, no *GFP* expression is observed. Lateral roots display a similar expression pattern to that seen in primary roots. While there is no expression
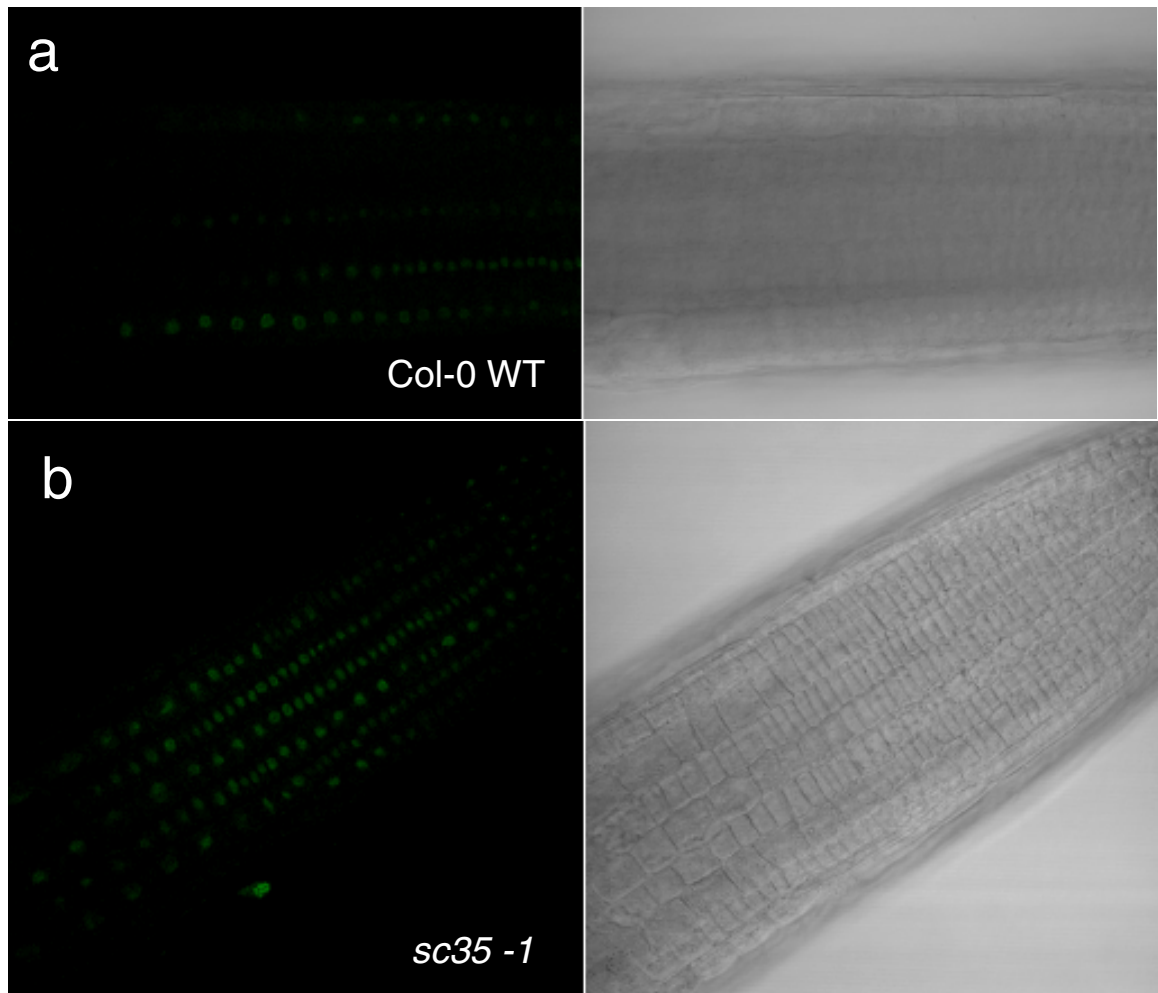
Figure 5.7 *SRp34a-GFP* **expressed in WT and** *sc35-1* **mutant lines**. In both the wild-type and *sc35-1* mutant plant lines, the *SRp34a-GFP* fusion protein gene was expressed in stage one cells of the primary root only. No other tissues in the primary root expressed the transgene nor did any tissue in any stage of lateral root. Wild-type plants expressed the *SRp34a-GFP* transgene at slightly lower levels than seen in *sc35-1* mutant. Expression patterns of both genetic lines were documented at 10 days old.
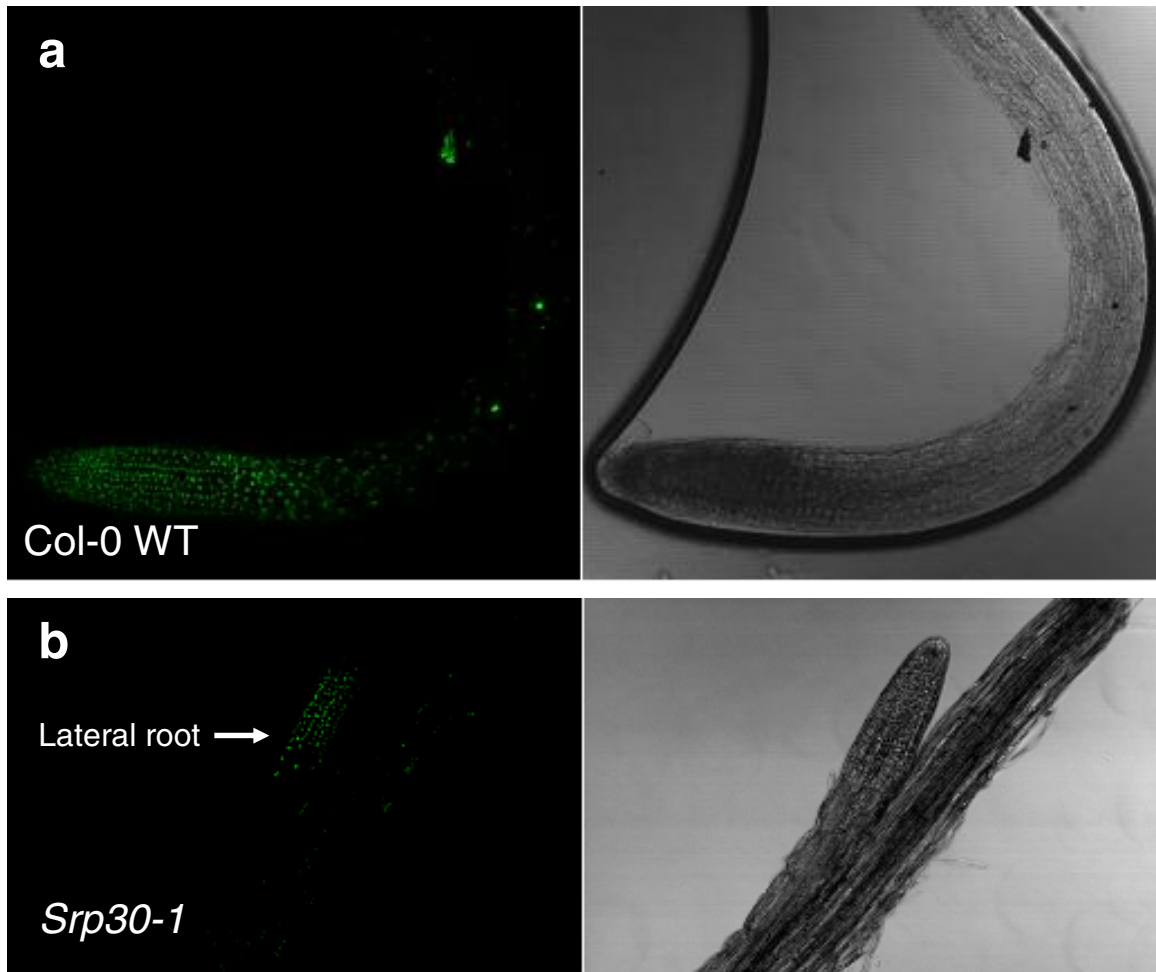
Figure 5.8 *SRp007-GFP* **expressed in WT and** *srp30-1* **mutant lines**. In the primary root of wild-type plants (a) *GFP* expression is observed in all cells of stage one and the root tip except for vascular tissue, and continue that expression pattern into stage two tissue. However, in stage two, the expression pattern, in a gradient manner, stops expressing the transgene in all cells other than the epidermis. At stage three, no *GFP* expression is observed. In lateral roots (b), all cells express the GFP transgene in stage one tissue and no GFP is observed in stages two and three. Photographs were taken of 10 day of wild-type and *sc35-1* mutant roots.

in stages two and three (Figure 5.8b), in all tissues of stage one other than vascular tissue show expression in the lateral roots of wild-type and *srp30-1* mutant lines.

The *sc35-1* mutant plants displayed a very interesting *SRp007-GFP* expression pattern with respect to the lateral roots. Like the wild type and *srp30-1* mutants, expression was observed in stage one lateral root. However, GFP is not visualized at the very early portion of stage one tissue but rather begins expression approximately halfway through stage one (Figure 5.9a). This expression is not observed to be regulated in a gradient manner but rather suddenly is turned on at what appears to be a very specific point in stage one tissue. No expression is observed in stages two and three of lateral roots. The primary roots of the *sc35-1* mutant express the *SRp007-GFP* transgene in all cells of stage one and root tip only (Figure 5.9b).

SCL33. The *SCL33-GFP* fusion gene construct was successfully transformed into WT and *sr45-1* protein gene mutant plant lines. The expression pattern of the *GFP* fusion protein gene was identical between the two plant genetic backgrounds. Expression was limited to the primary root tip (Figure 5.10a) and stage one tissue. Similarly, long and older lateral roots (Figure 5.10b) expressed the GFP fusion gene in the root tip and stage one tissue. However, no expression was seen in younger and shorter lateral roots.

SRp34/SR1. The construct containing the *SRp34/SR1-GFP* fusion transgene was transformed in wild-type, *sc35-1*, and *sr45-1* mutant plants. Wild-type plants expressed the *SRp34/SR1-GFP* fusion protein gene in all cell types in all three stages
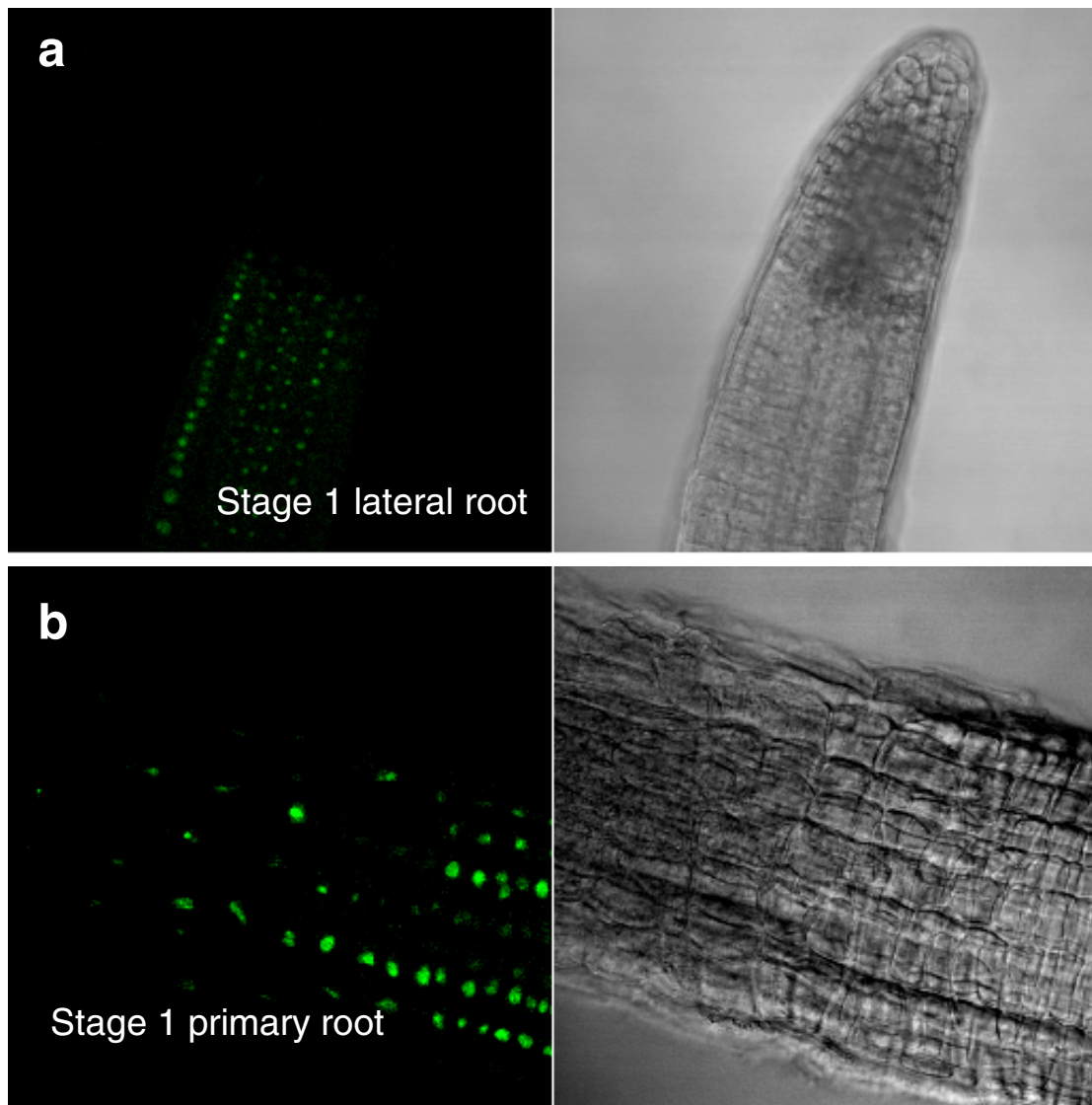
Figure 5.9 *SRp007-GFP* **expressed in** *sc35-1* **mutant lines**. Roots of the *sc35-1* mutants were taken at 10 days old. The *SRp007-GFP* fusion protein gene is expressed in all cells of stage one primary root except for vascular tissue (b). Similarly to primary roots, lateral roots only express the transgene in all cell types of stage one tissue (a) but do so in a very different manner. No expression is observed in the root tip and tissue cells very close to the root tip. However, expression does, in fact, suddenly appear approximately half way through stage one tissue.
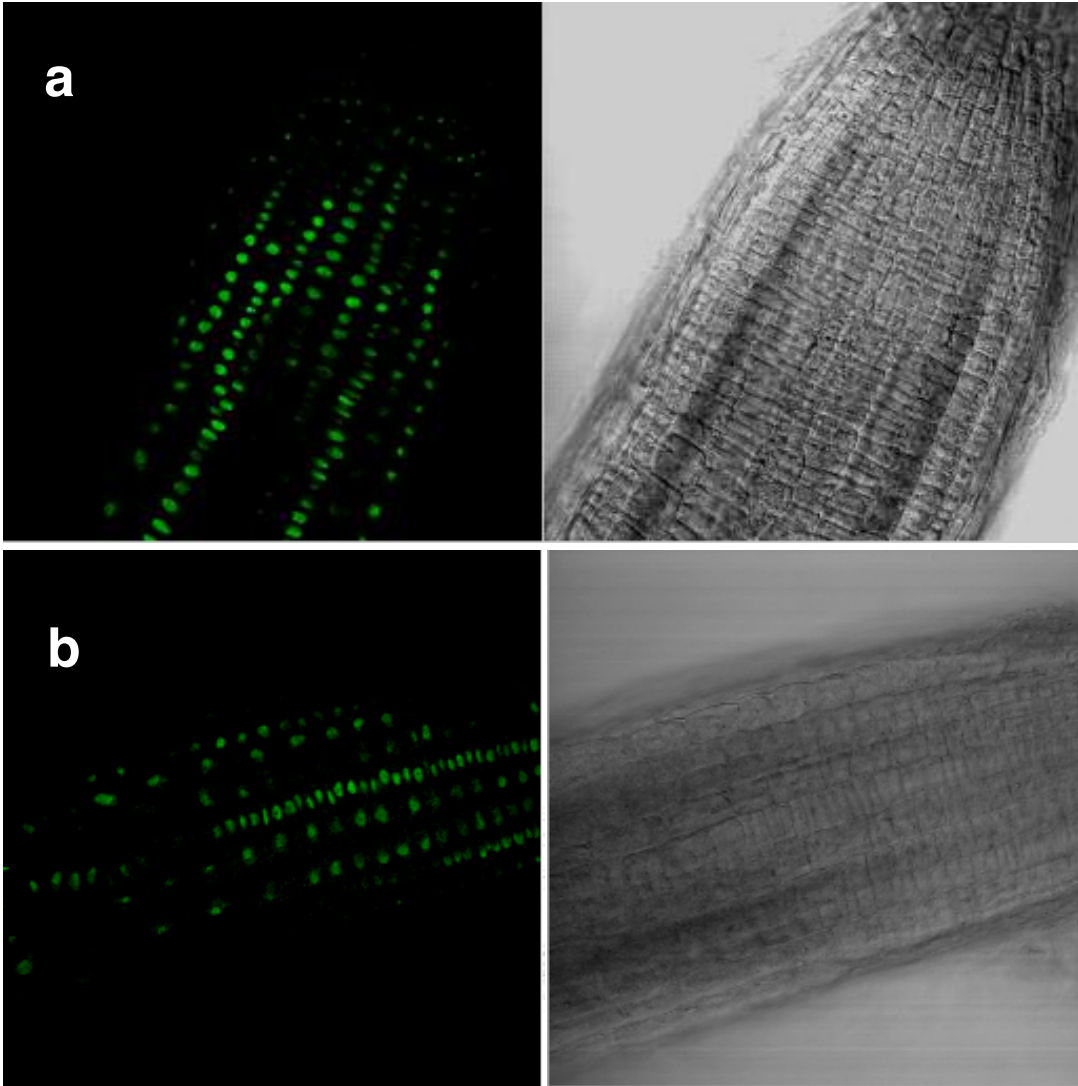
Figure 5.10 *SCL33-GFP* **expressed in WT and** *srp30-1* **mutant lines**. The *SCL33-GFP* fusion protein gene is expressed in the wild-type and *sr45-1* mutant lines in an identical manner. In all cells except vascular tissue of both stage one primary (a) and older lateral (b) roots, GFP expression is observed. No transgene expression is seen in young lateral roots or in stages two and three of primary roots or long lateral roots. Photographs of single optical sections of primary and lateral root tips were taken at 10 days old.
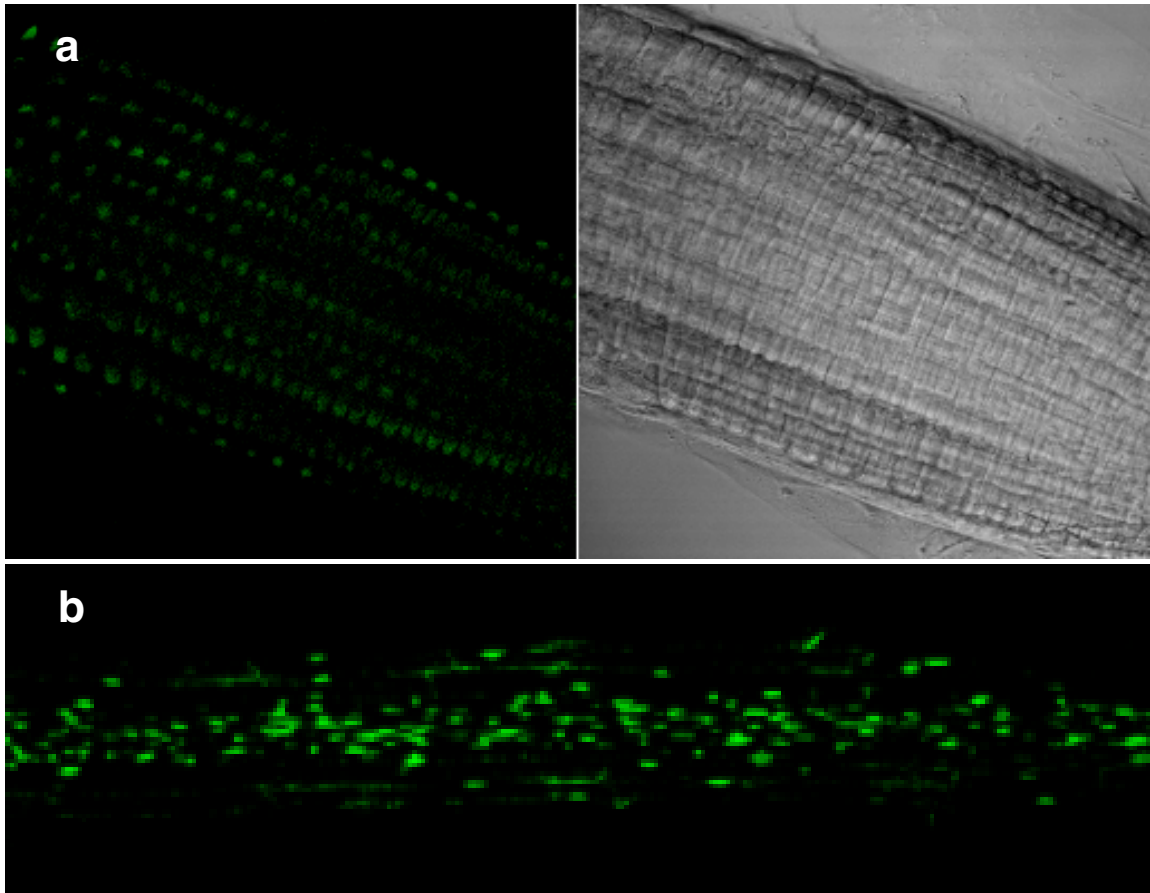
Figure 5.11 *SRp34/SR1-GFP* **expressed in WT roots**. Documentation of the

transgene was recorded in optical slices taken 10 days after germination. The

SRp34/SR1-GFP construct is expressed in all cells in the primary and lateral roots of

wild-type plants. All cell types in the root tip (a) and stages 2 and 3 (b) express this

transgene.

the epidermis and endodermis express the GFP transgene (Figure 5.12b). No

expression is observed in the cortex or vascular tissue or surrounding cells. In stage

three primary roots, both the epidermis and vascular tissue was expressing the fusion

protein (Figure 5.12a). However, the cortex and endodermis was not an active site

for GFP expression. Lateral roots in the *sc35-1* mutant plant expressed the

SRp34/SR1-GFP fusion protein gene in the same manner observed in primary roots

with the exception that no GFP is observed in stage three vascular tissue (Figure

5.12d).

Interestingly, the *sr45-1* mutant plant line expressed the *SRp34/SR1-GFP*

fusion protein gene in a fashion that was an intermediate between wild-type plants

and that observed in the *sc35-1* mutant plant line. Stage one tissue of both primary

and lateral roots (Figure 5.13b) expressed the GFP fusion protein in all cells which is

also observed in wild type and *sc35-1* mutants. An interesting change in expression

pattern is observed in stages two and three. Like wild-type plants, both primary and

lateral roots are expressed in the same fashion. However, the GFP fusion protein

gene is not constitutively expressed like in wild type but rather is limited to the

epidermis, endodermis and some very light expression in vascular tissue of both

stages two and three of primary and lateral roots (Figure 5.13a).

RSp31. The construct containing the *RSp31-GFP* fusion transgene was

transformed in wild-type, *srp30-1,* and *sc35-1* mutant plants. In wild-type plants, the

transgene was consistently expressed in both primary and lateral roots . In all three

stages, one, two, (figure 5.14a) and stage three (Figure 5.14b), the epidermal and
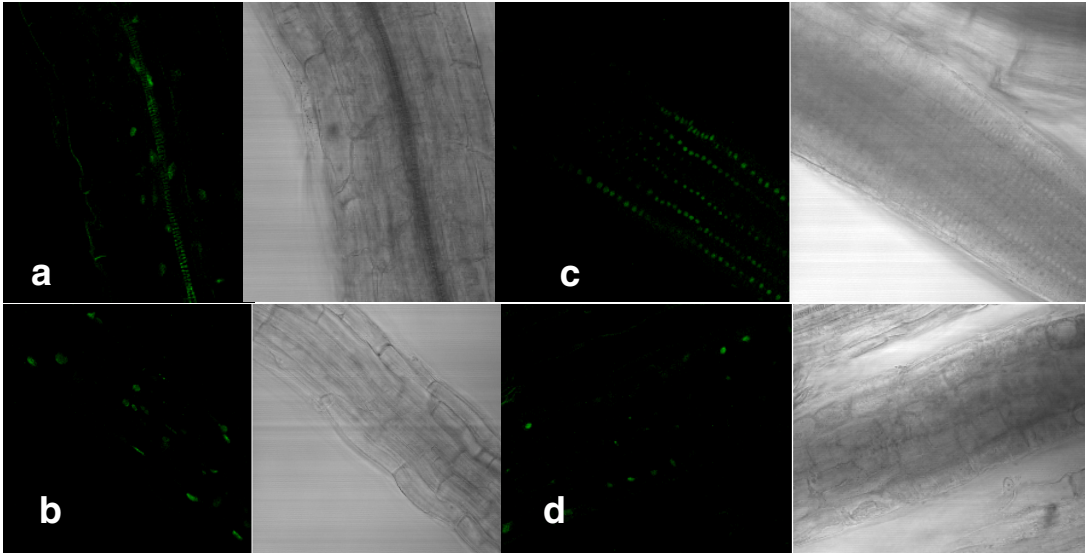
Figure 5.12 *SRp34/SR1-GFP* **expressed in** *sc35-1* **plant lines**. The

SRp34/SR1-GFP fusion protein gene was expressed in *sc35-1* mutant plants in a

tissue and stage dependent manner as seen by the optical sections photographed at 10

days post-germination. In both primary and lateral roots, all cells with the exception

of vascular tissue expressed the *GFP* transgene (c). Likewise, in stage two,

fluorescence was observed in the epidermis and endodermis tissues but lacking in the

cortex and vascular tissues (b). In stage three of primary roots, the *SRp34/SR1-GFP*

transgene is detectible in vascular tissue, endodermis, and very faintly detectable in

the epidermis of the *sc35-1* mutants (a). However, in stage three of lateral roots, *GFP*

expression is limited to the epidermis (d).
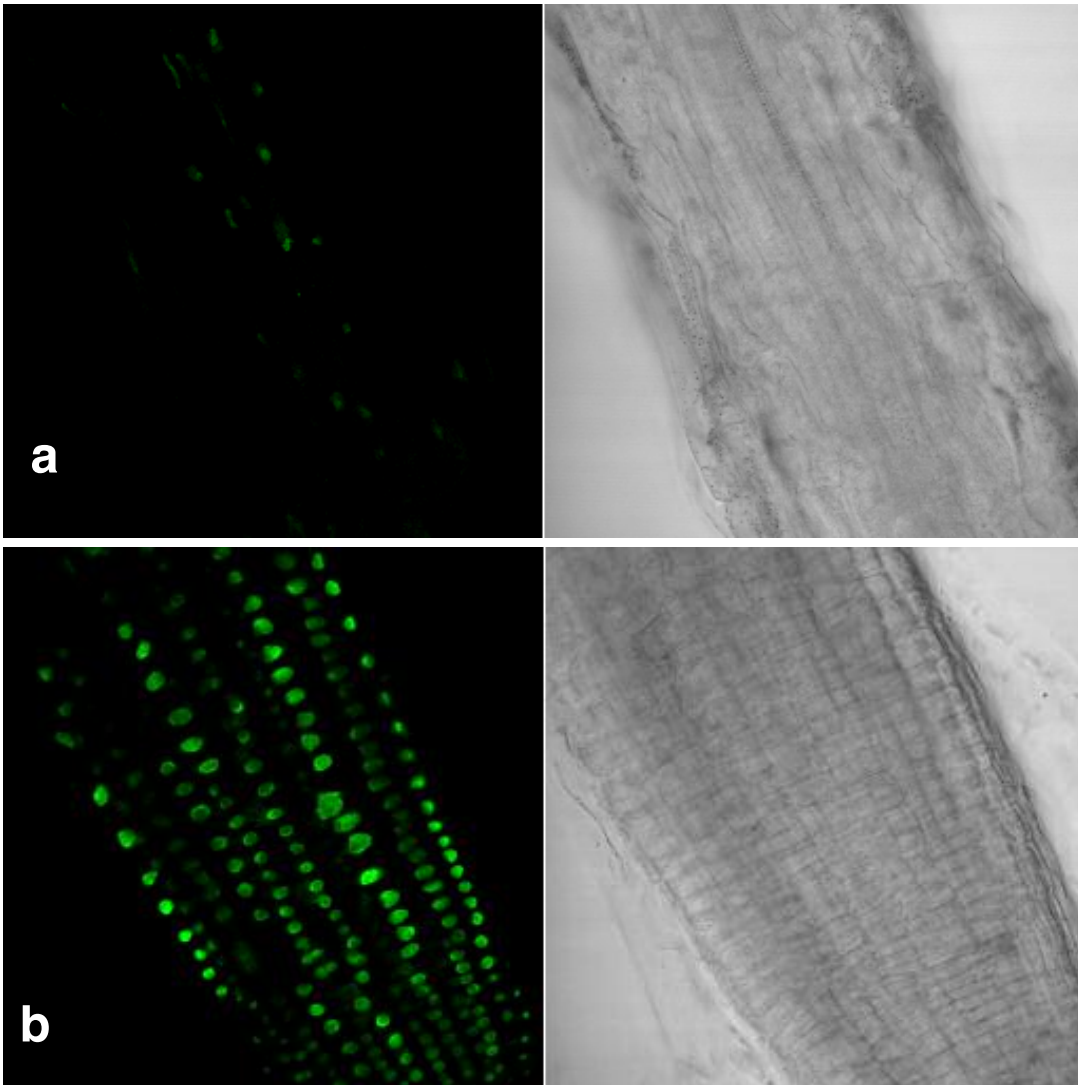
Figure 5.13 *SRp34/SR1-GFP* **expressed in** *sr45-1* **plant lines**. In stage one tissue,

the transgene expression in *sr45-1* mutant plant lines was present in all tissues except

for vascular tissue (b) in both primary and lateral roots. However, in stages two and

three (a), *GFP* was expressed at low levels in the endodermis and epidermis and even

lower levels in some vascular tissue that could not be documented well in pictures

due to autofluorescence. Single optical sections were photographed 10 days post-

germination.

endodermic tissues clearly expressed the transgene while some vascular tissue did as well. The pattern was quite striking and consistent throughout the entire length of the root as well as the entirety of the lateral roots both young and old.

The expression pattern of the *RSp31-GFP* transgene in the *sc35-1* mutant background differed from that observed in wild-type plants. In primary roots, the root tip and all cells of stage one expressed the *GFP* fusion protein gene (Figure 5.15a). Additionally, in primary roots, the only other place in which the fusion protein was observed was in stage three epidermis and vascular tissue but in a very limited number of cells (Figure 5.15b). No other tissue type was observed to express the transgene. Stage two completely lacked active GFP in all cells. Young lateral roots of the *sc35-1* mutant have their own expression pattern which differed from that observed in wild-type plants as well as primary roots of the *sc35-1* mutant line. All tissue of young lateral roots expresses the transgene with the exception of vascular tissue (Figure 5.15c). However, as lateral roots aged and grew longer, a limited amount of vascular tissue also expresses the transgene as seen in primary roots.

Finally, the *srp30-1* mutant line displayed yet another very unique expression pattern. In this mutant line, the only tissue observed to express the *RSp31-GFP* fusion transgene was in the primary root tip and stage one cells. And even then, the expression was very low to the point that a photograph could not be taken. In order to view the individual fluorescing nuclei, the laser needed to be at full power. However, at that point, the autofluorescence from the cell wall would make it difficult to distinguish between background and contribution from the GFP and pictures could not be obtained.
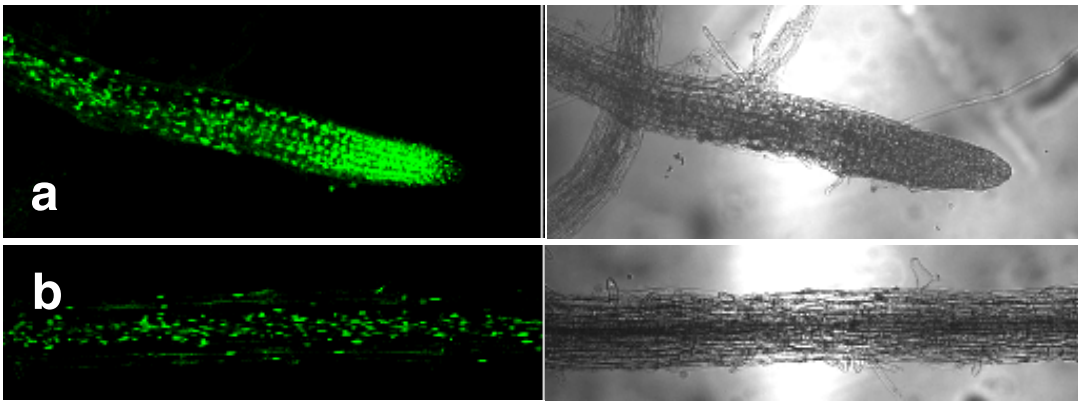
Figure 5.14 *RSp31-GFP* **expressed in WT roots**. In both primary and lateral roots, the epidermis, endodermis, and some vascular tissue is observed in stages one and two (a) as well as in stage three tissue (b). The expression pattern is documented in whole mounts of 10 day old plants.
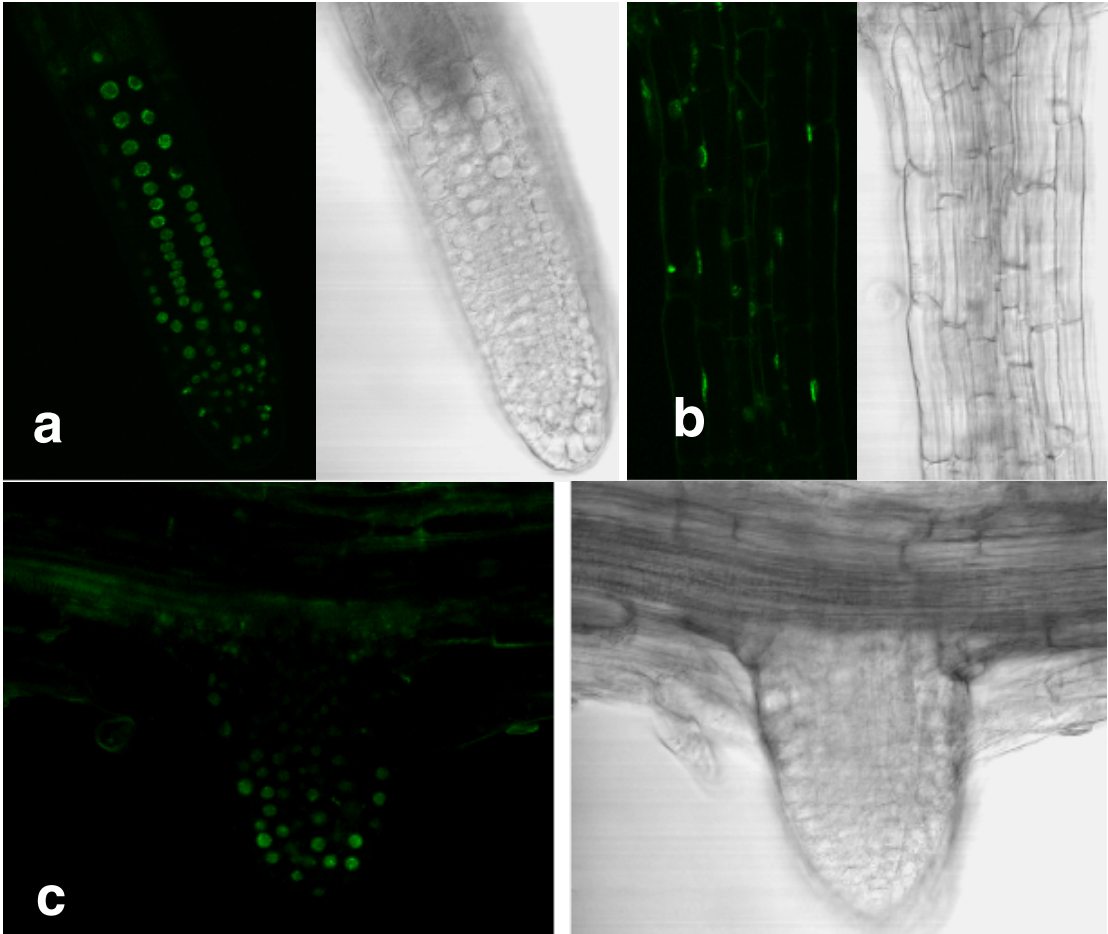
Figure 5.15 *RSp31-GFP* **expressed in** *sc35-1* **mutant roots**. Older lateral roots of the sc35-1 mutant, in which root hairs are present, express the RSp31-GFP transgene in all cells except vascular tissue at stage one and the root tip (a). While there is no expression in stage two of these roots, stage three expresses the GFP fusion gene in the epidermis, endodermis, and some very select vascular tissue (b). In young emerging lateral roots, all cells express the GFP fusion transgene (c). Expression patterns were documented by optical slices of 10 day old roots.

RSZp21. The construct containing the *RSZp21-GFP* fusion transgene was transformed into wild type, *sc35-1*, and *srp30-1* mutant plants. The *RSZp21-GFP* fusion protein gene was expressed in wild-type plants in a very interesting fashion not seen by any of the other fusion proteins.   In of developmental stages two and stage three, in both primary and lateral roots, the GFP fusion protein gene was expressed only in the epidermis (Figure 5.16a). However, in stage one and the root tip, GFP was visible in only endodermic cells and was absent in all other tissues (Figure 5.16b).

In the two mutant lines, *sc35-1* (Figure 5.16d) and *srp30-1* (Figure 5.16c), the expression of the *RSZp21* was limited to all cells of the primary root tip and stage one tissue. These results are very different from those observed in wild-type plants. No other GFP fusion protein was observed at any other developmental stage in the primary roots and no fluorescence was observed in lateral roots at all.

SRp34b. Constructs generated to express the *SRp34b-GFP* fusion protein gene were successfully transformed into wild-type, *sc35-1,* and *srp30-1* mutants plant lines. In wild-type and *srp30-1* mutant lines, no expression was observed in any root tissue. It was absent in both primary and lateral roots. Multiple lines were generated that conferred herbicide resistance but yet never expressed the GFP fusion protein. At first this result was very discouraging in that all of the other SR proteins are expressed somewhere in roots of wild-type plants. However, after comparing this visual observation to the data (Table 5.1) generated by the microarray chip in the Birnbaum paper (Birnbaum. *et al.* 2003) where they did not detect SRp34b transcript
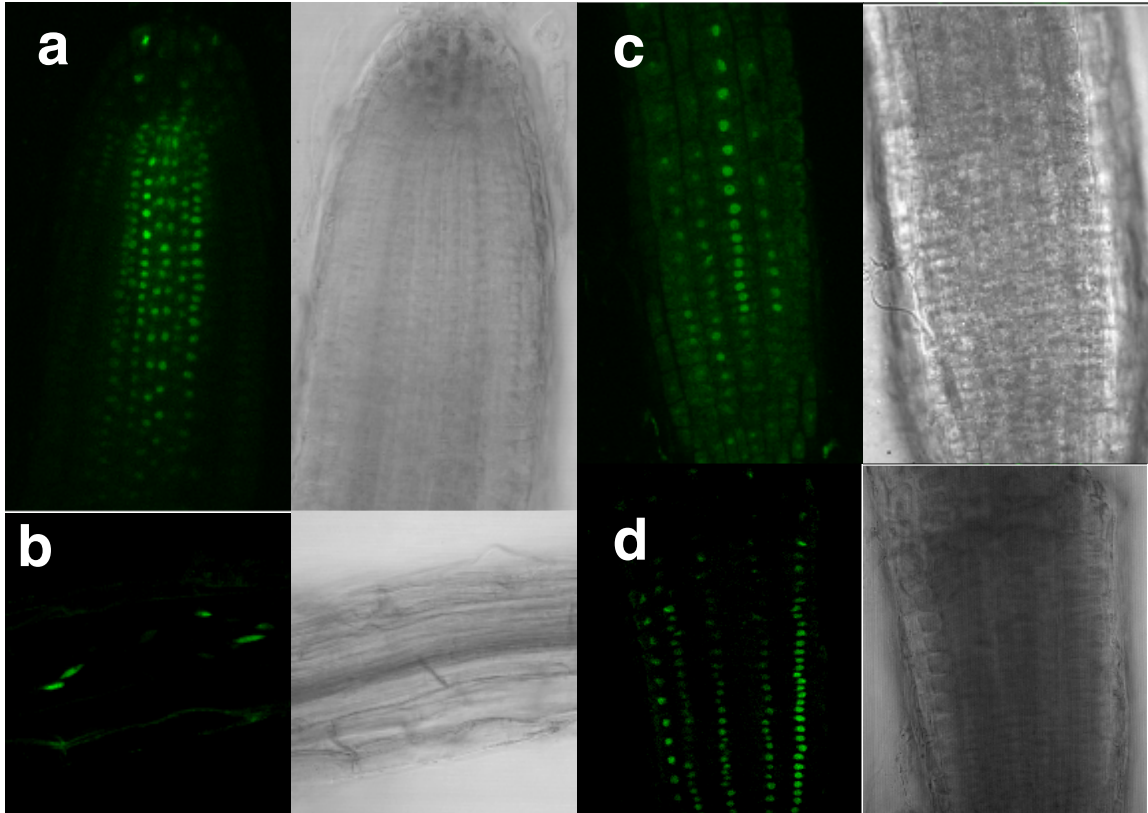
Figure 5.16 *RSZp21-GFP* expressed in WT, *srp30-1*, and *sc35-1* mutant plants. Plants transformed with the RSZp21-GFP fusion gene displayed a very interesting expression pattern. In stage one and root tip of wild-type roots, fluorescence was detectable only the endodermis and no other tissue (a). Additionally, in wild-type plants transformed with the GFP fusion gene, the epidermis of stage three expressed the transgene (b). In plants mutant for either srp30-1 (c) or sc35-1 (d), all cells of stage one primary root expressed the GFP fusion transgene. Photographs are optical slices of 10 day old roots.

in roots, it is very possible that this particular SR protein gene is simply not expressed in wild-type *Arabidopsis* roots.

While the confirmation of my results with the Birnbaum microarray data was exciting enough, the expression pattern of the SRp34b-GFP protein gene in the *sc35-1* mutant background showed a change from that seen in wild-type and *srp30-1* mutant plants. In lateral roots both young and old, there is very strong expression in stage one cells and the root tip. However, in the primary root only, the expression pattern was much more dramatic. The root tip and stage one cells highly expressed the GFP transgene (Figure 5.17a) while vascular tissue in stages two and three expressed the transgene at what appeared to be high levels (Figure 5.17b). The cortex also expressed the *SRp34b-GFP* fusion protein gene in all of stage two. Younger cells of stage three expressed the transgene anywhere from a few cells to a little over ten cells adjacent to the stage two/three boundary (Figure 5.17b). The range of expression in the cortex of stage three cells varied from plant to plant and not between different transgenic lines.

SC35. The construct containing the *SC35-GFP* fusion protein gene was one of only four that was successfully transformed into all four available plant lines, wild-type, *sc35-1, sr45-1*, and *srp30-1* mutant plants. The expression of this fusion gene was identical in wild-type and *sc35-1* mutant lines. In stage one of the primary root, all cells displayed high levels of the GFP fusion protein gene (data not shown). While no tissue in stage two was observed, stage three expressed the fusion gene in the endodermis and vascular tissue only (data not shown). While long lateral roots

132
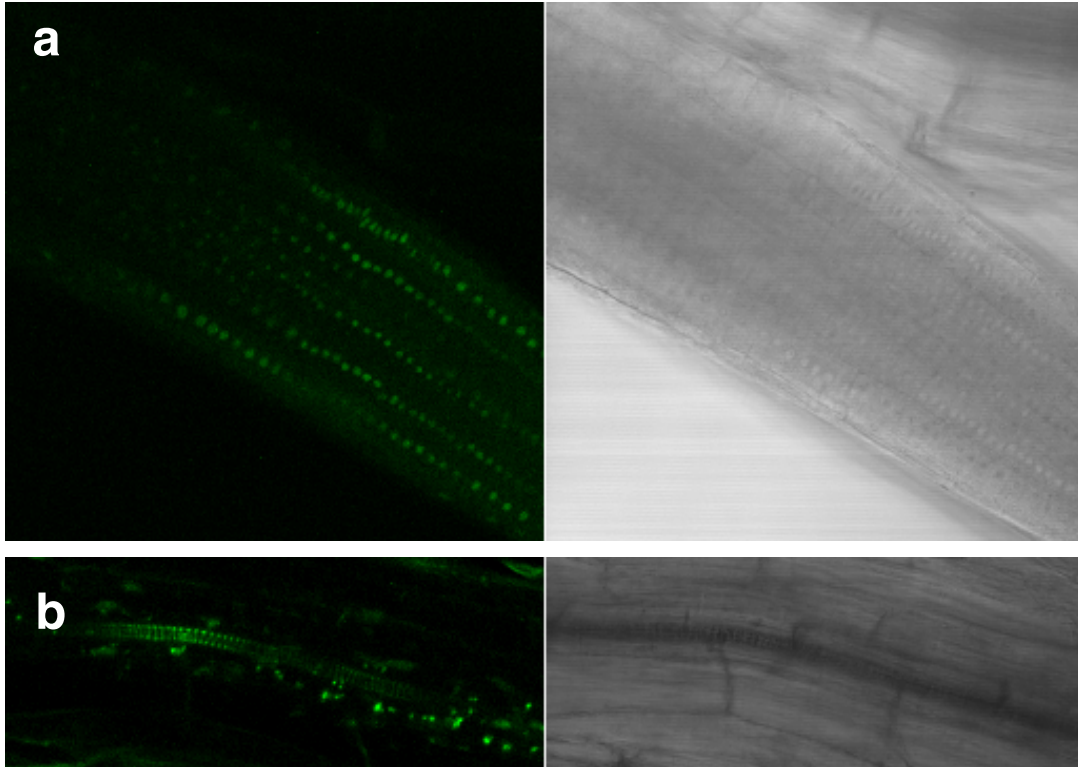
Figure 5.17 Expression of *SRp34b-GFP* in primary root of *sc35-1* mutant plants.
The *SRp34b-GFP* transgene, while not expressed at all in wild-type plants or *srp30-1* plants, is expressed at relatively high levels in stage one of the lateral roots (a). In stages two and three of the primary root only (b), the vascular tissue is very highly expressed. These results were documented in optical slices of 10 day old roots.

were observed to not express GFP, short lateral roots displayed the same exact expression pattern as is found in primary roots. High expression levels were observed in stage one and the root tip and no expression in stage two at all. However, in stage three, the endodermis and vascular tissue did express the transgene.

Both the *srp30-1* and *sr45-1* mutant plant lines affected the expression pattern of the *SC35-GFP* fusion protein gene in different ways. In the *srp30-1* mutant, the only GFP observed was in the root tip and stage one of the primary root. Additionally, the fluorescence was so low that a picture could not be taken due to the high level of laser power needed to view the nuclei which would, in turn, increase the autofluorescence to the point that *bona fide* GFP expression could not be differentiated from the autofluorescence.

While visibly expressed at higher levels than that seen in the *srp30-1* mutants, the expression level of the GFP construct was still expressed in a much more limited fashion in *sr45-1* mutants than observed in wild-type plants and the *sc35-1* mutant plant lines. In the primary root, the only location of expression of the transgene was in the root tip and stage one (Figure 5.18b). In lateral roots, the location of expression was limited even further, only being observed in epidermis, cortex, and endodermis of the stage one/two junction (Figure 5.18a).

RSZp32. The RSZp32-GFP fusion protein construct was successfully transformed into all four plant lines. The fusion protein gene was expressed in primary roots of wild-type and *srp30-1* mutant lines in the same manner (Figure
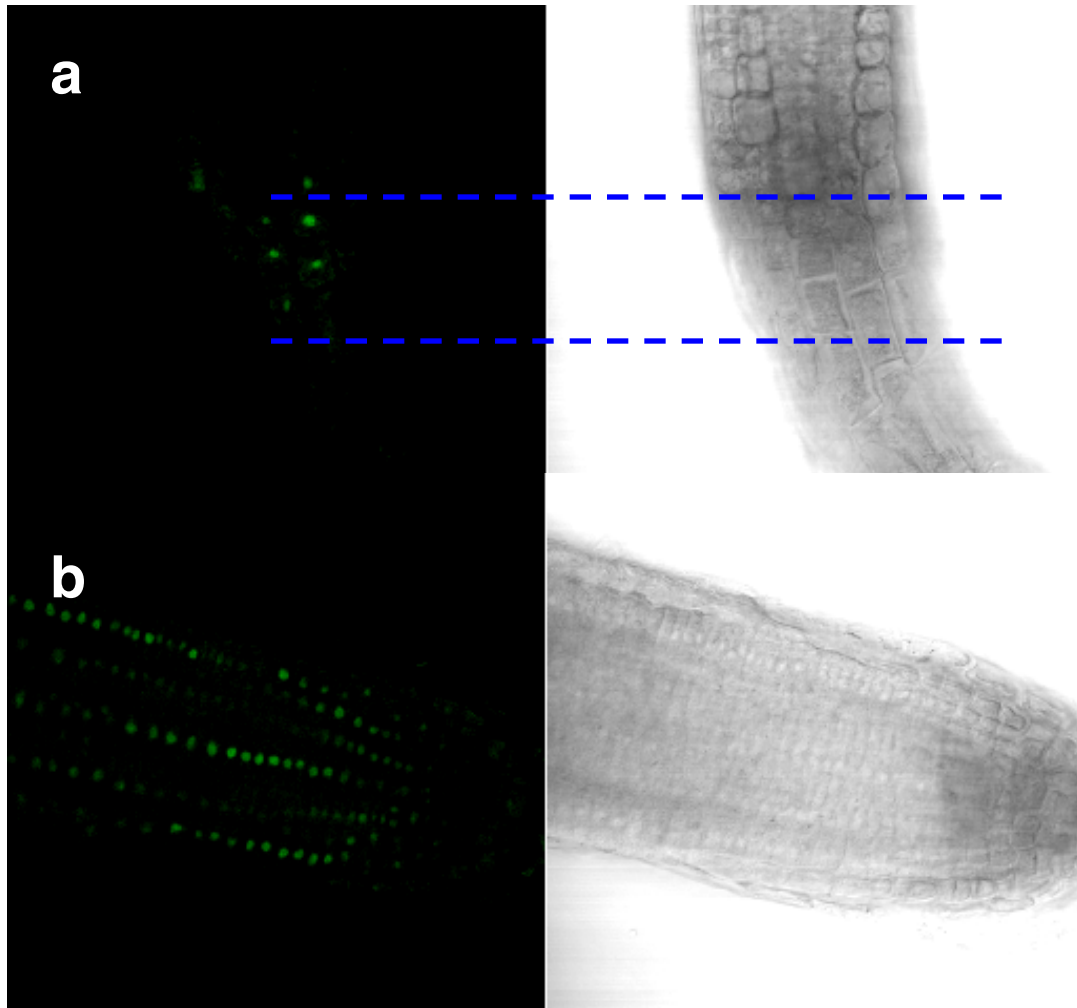
Figure 5.18 Expression of *SC35-GFP* in primary root of *sr45-1* mutant plants.

The expression in lateral roots of the sr45-1 mutant is observed in the endodermis,

cortex, and epidermis but only in a few cells at the boundary between stages one and

two (a). In the primary root of sr45-1 mutants, the *sc35-GFP* fusion protein is

expressed in all cells of stage one primary roots only (b). Expression patterns were

documented in optical slices of 10 day old roots.

5.19a). In both primary and lateral roots, the root tip and all stage one cells expressed the transgene. Additionally, the epidermis and cortex of stage two and three of the primary and lateral roots expressed the GFP fusion transgene (Figure 5.19b). The *sc55-1* mutant line expressed this fusion protein gene in the same fashion with the exception of stage two where only the epidermis was observed to be expressing the *GFP* transgene (Figure 5.20).

In the *sr45-1* mutant plant line, the *RSZp32-GFP* fusion protein gene was expressed in all cells of the root tip (Figure 5.21a), and stage one and two with the exception of the vascular tissue where *GFP* expression was absent (Figure 5.21b). No expression was observed in stage three of the primary root (Figure 5.21b). Lateral roots display a completely different expression pattern than seen in primary roots of the *45-1* mutant plant line. The *GFP* transgene is only visible in all cells of stage one and all cells other than vascular tissue at the stage one/two junction (figure 5.21c).

RSp31a. The *RSp31a-GFP* fusion protein gene was inserted successfully into all four plant lines, wild type, *sc35-1, sr45-1*, and *srp30-1* mutants. The wild-type, *sr45-1*, and *srp30-1* roots all expressed the *RSp31a-GFP* fusion protein transgene in the same manner with no variation between any of these three lines. Primary roots expressed the transgene in root tips and all cells of stage one tissues (Figure 5.22a). The GFP fusion protein was also observed in lateral roots but again, on a very limited basis. Low levels of expression were viewed in lateral root tips and all stage one cells
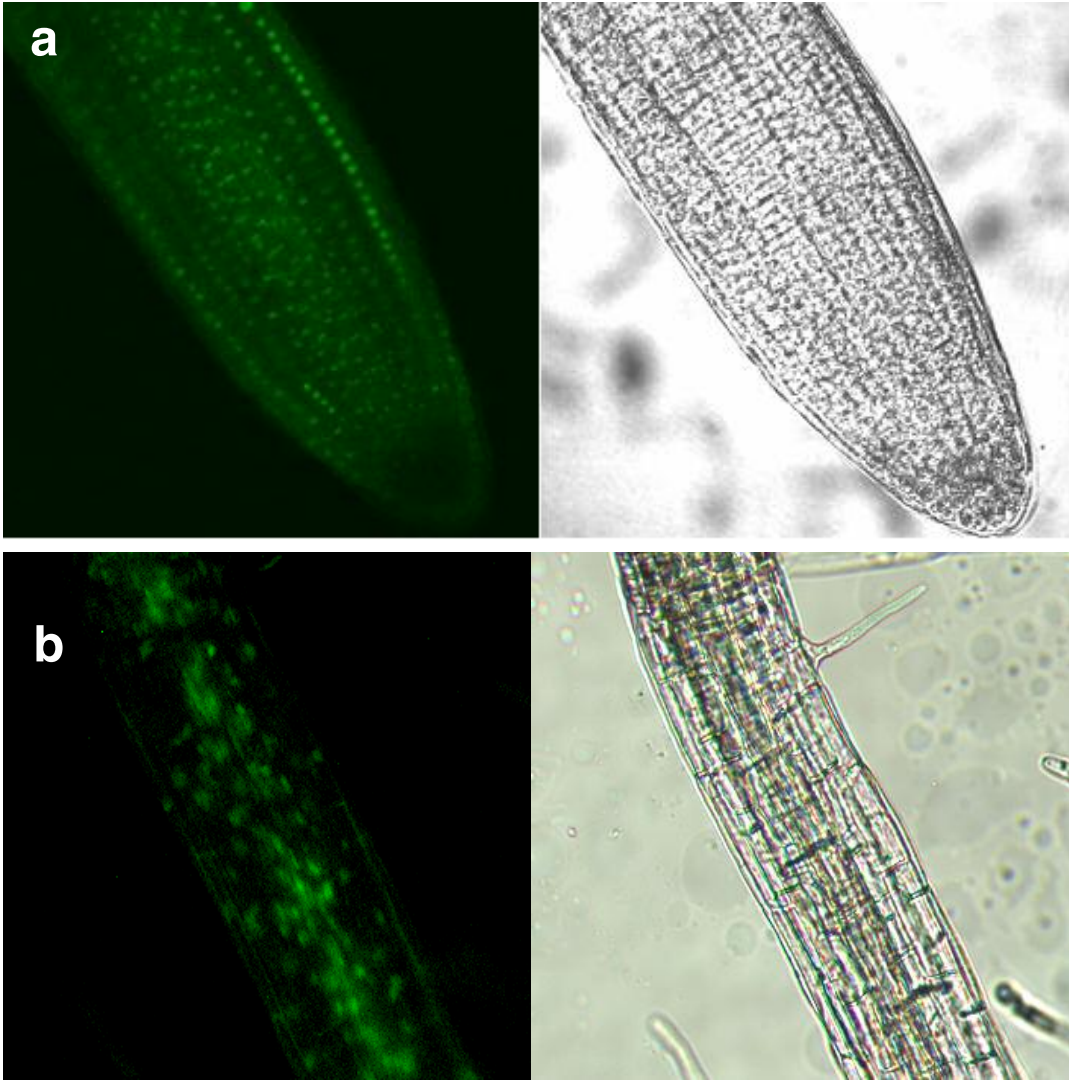
Figure 5.19 Expression of *RSZp32-GFP* in primary root of WT and *srp30-1*

mutant plants. In stage one cells of both primary roots of any age, the transgene is

expressed in all cells at relatively high levels (a). In stages two and three, the

epidermis and cortex express the GFP transgene at high levels lateral roots (b).

Expression patterns were documented in whole mounts of 10 day old roots.
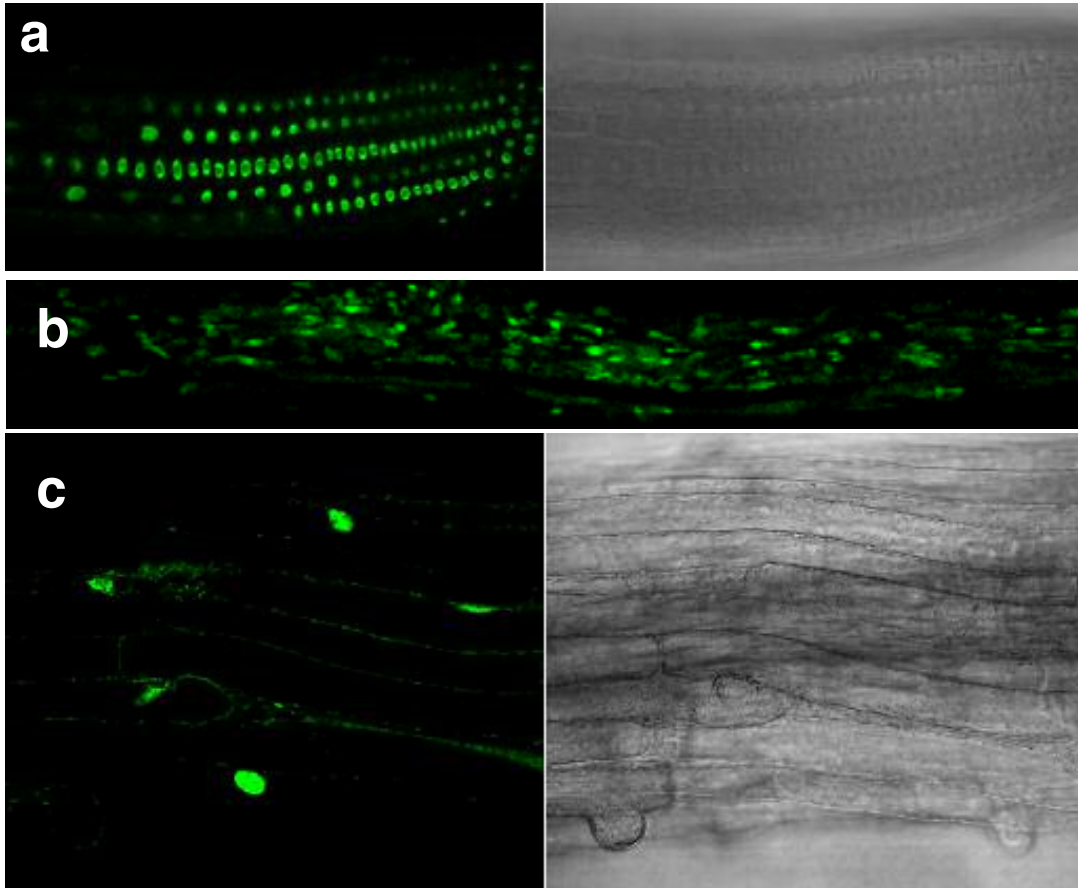
Figure 5.20 Expression of *RSZp32-GFP* in primary root of *sc35-1* mutant plants.

Optical slices of 10 day old roots of various stages were taken to document the

expression pattern of the *RSZp32-GFP* transgene in the *sc35-1* background. In the

sc35-1 mutant, the *RSZp32-GFP* transgene was expressed in a stage specific manner.

In root tips and stage one tissue (a) of both primary roots, the transgene was

expressed in all cells. In stage two (c) only the epidermis displayed fluorescent nuclei

while stage three (b) expressed the transgene in both the epidermis and cortex.
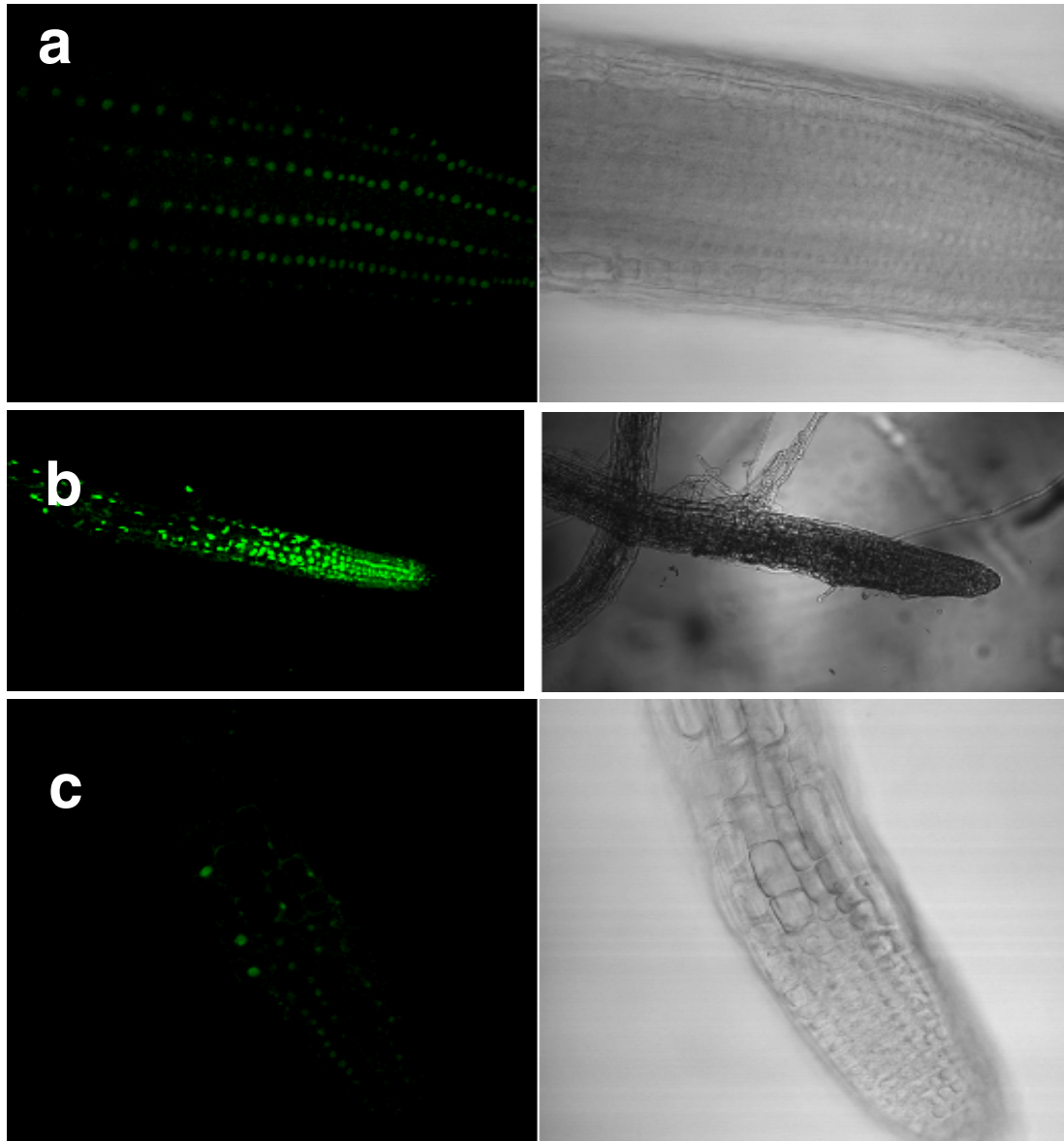
Figure 5.21 Expression of *RSZp32-GFP* in primary root of *sr45-1* mutant plants. In sr45-1 mutant plants, the GFP transgene is expressed in the root tip and stage one cells in both the primary roots (a). In stage two of the primary root (b) the RSZp32-GFP transgene is expressed in all cells except for vascular tissue while in the lateral roots, only the first few stage two cells express the transgene (3). Photographs were taken of optical slices of 10 day old plants.
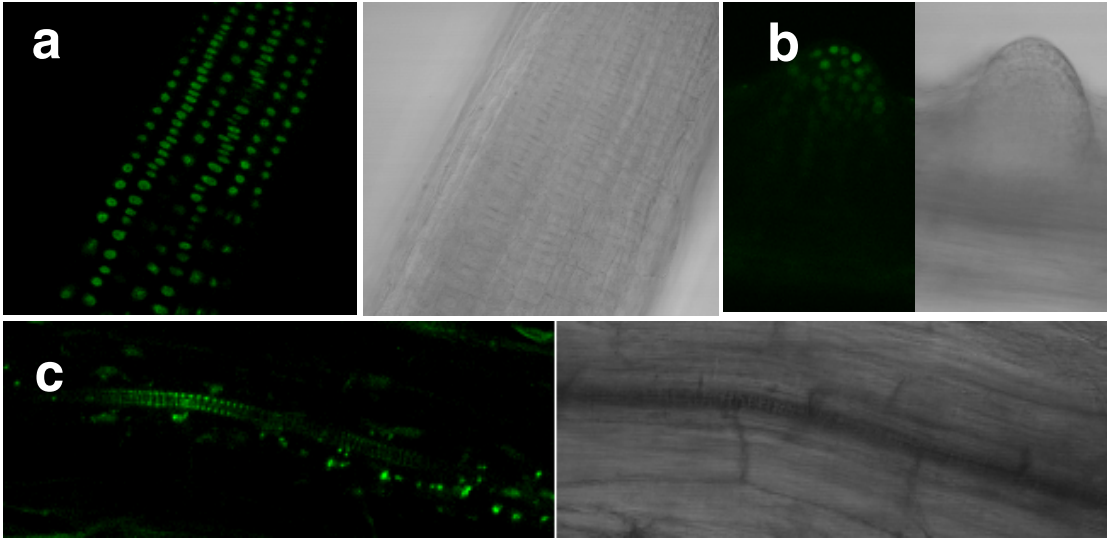
Figure 5.22 Expression of *RSZp31a-GFP* in primary root of WT, *srp30-1*, *sr45-1*, and *sc35-1* mutant plants. All four plant lines expressed the *GFP* transgene in stage one cells and root tip of primary roots (a) as well as in root primordia (b). Wild-type is limited in its expression to this single tissue type and stage, while *sc35-1* mutants express the *RSZp31a-GFP* fusion gene in stages two and three vascular, endodermis, and epidermis tissue (c). Optical slices were obtained from 10 day old roots.

in addition to lateral root primordia (Figure 5.22b). No other tissue of the lateral root expressed the transgene in any of these three lines.

The expression pattern of the *RSp31a-GFP* fusion protein gene in *sc35-1* mutant lines differed greatly from the other three plant lines. While no transgene expression was observed in any cells in lateral roots, GFP was visualized in all three stages of primary roots (Figure 5.22c). In stage one and the root tip, all cells expressed *GFP*. In stages two and three, GFP was visualized in vascular, endodermic, and epidermal cells although expression in epidermal cells was very lightly expressed in comparison to expression in the endodermis and vascular cells. No GFP was visualized in cells making up the cortex.

SCL28. Finally, all four plant genetic backgrounds were successfully transformed with the *SCL28-GFP* fusion protein gene construct. In all four lines, the expression pattern of the transgene was identical. In the primary root, GFP was observed in the stage one cells only and at very low levels (Figure 5.23a) which were barely detectable by the microscope. However, in lateral roots, the expression pattern and level was increased substantially. While no expression was observed in older lateral root tips and stage one cells, GFP was present in epidermal cells (Figure 5.23c). Tissue in stage three nearing stage two expressed GFP in the epidermis at lower levels until no expression was observed at stage two. However, at this junction, vascular tissue actively expressed GFP (Figure 5.23b) and did so throughout the remainder of stage two tissue.
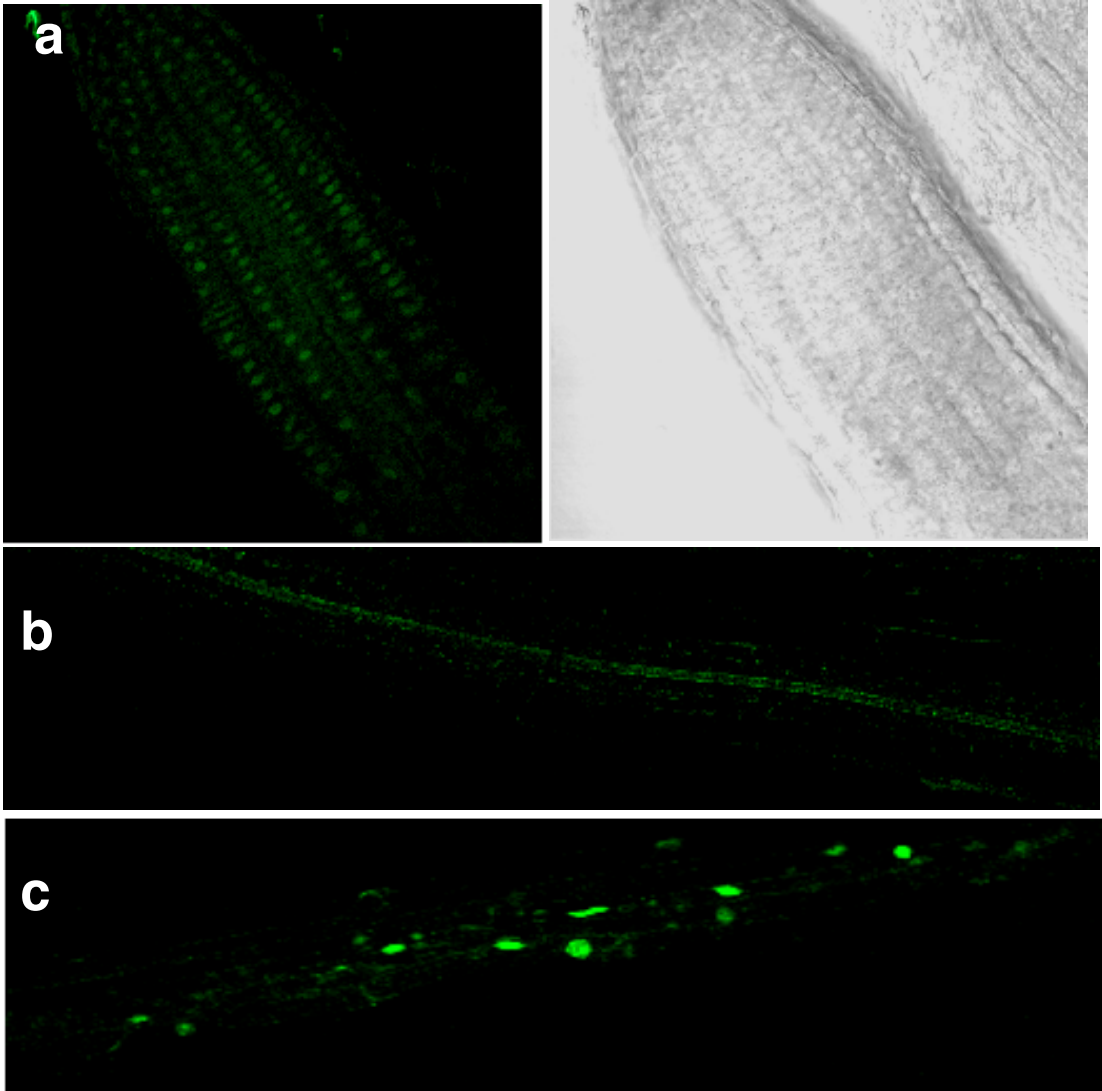
Figure 5.23 Expression of *SCL28-GFP* in primary root of WT, *srp30-1, sr45-1*,

and *sc35-1* mutant plants. In primary roots, the only expression of the *GFP*

transgene is in the root tip and stage one cells (a). In younger lateral roots, not only is

GFP expressed in stage one cells, but it is also observed in vascular tissue of stage

two cells (b) and the epidermis of stage three cells (c).

Discussion

The GFP fusion proteins were clearly visible when observed under confocal microscopy. A small amount of quantitative information could be inferred . In some instances, the power of the laser needed to be turned up to much higher settings or the pinhole opening widened to view the fluorescence. This observation did not vary from root to root of the same genetic plant line or line to line of the same construct, but varied between SR protein transgene or between different plant genetic lines. To me this indicated that not only was the transgene successfully being expressed in the plant, but it was most certainly under more of a specific regulation than simply being expressed or not expressed. A potential problem with generating these transgenic plants is that T-DNA insertions are random and more than one insertion can take place in each line. A concern is that the construct could fall into a location in which it would be under the control of native regulatory sequences. However, these same transgenes were used to rescue the *srp30-1, sc35-1, scl28-1, and rsp31-1* mutants which suggest that the T-DNA is being inserted into locations which allow the transgenes to be expressed in a manner in which functional protein is being produced.

The motivation for creating these constructs in the fashion that I did, was to create a fusion protein gene that would be expressed in the exact same manner and be regulated under the same constraints as the native SR protein gene. While I cannot say for certain that the transgene is expressed exactly as the native gene, the evidence does support that this is a realistic possibility. The constructs engineered in previous experiments performed in the Barta lab or the Spector lab lack this degree of

sensitivity that my constructs are capable of capturing. However, neither approach alone will tell the whole story of how and where SR protein genes are regulated. For a complete understanding, both types of experiments are essential.

The difference between expressing the genomic sequence and cDNA of an SR protein gene in wild-type plants can be best seen in looking at the *SCL33* expression pattern reported in the Fang *et al.* paper in comparison to my observations. Both sets of data confirm that *SCL33* expression is limited to the root tip and stage one tissue. However, there is a major difference in the expression pattern in the primary root. The Fang *et al.* paper, which used native promoter-cDNA-*YFP* fusion constructs, observed their YFP fusion protein in the vascular bundle, endodermis, cortex, and epidermis of the root. This particular paper fails to differentiate between the three different stages but the photograph appears to show expression in stages one and two quite clearly. Their cDNA expression data differs quite dramatically to what was observed in my expression experiments using the entire genomic sequence. My transgene, which contains introns and is subject to alternative splicing, is expressed only in the root tip and stage one tissue.

The Birnbaum microarray data supports the expression pattern observed by Fang *et al.* While this may appear discouraging at first glance, it is important to note that the microarray data used cDNA generated from protoplast cells that could have had any number of unknown effects on the expression level of any number of the SR protein genes. Additionally, like the Fang *et al.* cDNA expression experiment, the microarray experiment did not take into account for alternative splicing gene regulation. The microarray data does not differentiate between an RNA encoding for

a functional protein and one that does not. In fact, *SCL33* has a total of six verified

alternatively spliced isoforms (Palusa *et al.* 2007) of which six are due to full or

partial retention of intron three. None of the six alternative isoforms codes for a full

length protein. Neither the Birnbaum nor the Spector paper differentiates between the

protein coding isoform and any of the six alternatively spliced forms.

The expression pattern of the *SR1/SRp34-GFP* transgene was also of interest

due to the similarities in expression pattern. Both my observations and those of the

Fang *et al.* experiment show that both respective *SR1/SRp34-GFP* transgenes were

expressed in all cells in all stages of the primary root. However, while my fusion

protein gene was expressed in the same pattern in the lateral roots as was observed in

the primary roots, the Fang *et al.* paper reported expression only in the root cap of

lateral roots. The microarray data also reported incredibly high levels of transcript in

all tissues of all developmental stages. The promoter-GUS fusion protein developed

in the Barta lab was only observed in some vascular tissue of older lateral roots as

well as lateral root tips.

The variation between the different sets of experiments was probably due to a

combination of many factors. First and foremost is how the construct was made and

the genomic components of which it was comprised. Additionally, the Palusa *et al.*

(2007) paper looked at alternative splicing of SR protein genes under various

nutritional and environmental conditions and found that those differences can have

dramatic affects on which isoforms are produced (Palusa *et al.* 2007). Clearly the

biological material used in the Birnbaum experiment would have been drastically

different than the plants used in the Fang *et al.*, Lopato *et al.*, or my experiments.

However, there were sure to be differences in temperature, light, humidity, and any number of other environmental conditions between the three labs that could have also caused some slight differences in the expression patterns observed.

In addition to environmental factors, the age of the root played a major role in expression pattern and, in addition to the construct components, could have contributed to the variation seen in the overlapping transgenes between labs. In my own observation, I saw dramatic changes in expression patterns in plants that differed in age from 10 days old to 14 days old. Looking at an individual root also gives support to the notion that SR proteins are very highly developmentally regulated. Only two SR-GFP fusion transgenes, *SRp34b* which did not appear anywhere in the root, and *RSZp21* where expression in the root tip and stage one tissues was limited to the endodermal cells, were not highly expressed in the rapidly dividing root cells of stage one. However, the other 14 transgenes where highly expressed in stage one and the root tip. Additionally, comparing lateral roots on a single plant tells the best story and provides the most evidence of age specific activity. Typically, lateral roots that were longer, older, and with root hair and developing root primordia themselves, expressed the *GFP* transgene in a manner similar to that seen in primary roots. Lateral root primordia, or young lateral roots which had yet to develop root hair, normally displayed a different expression pattern altogether. While all lateral roots consist of the same tissue types and organization of tissues, the key distinguishing factor is age and developmental stage.

One additional observation I would like to expand on is the expression pattern of *RSZp22a*. This was very interesting in that RSZp22a is not one of the 16 SR

146

protein genes that are known to be alternatively spliced. However, the expression pattern was unexpectedly different in *sc35-1* mutant in comparison to wild-type plants. This result could be explained in two ways. First, *RSZp22a* is, in fact, alternatively spliced but the alternative isoforms are very unstable and are degraded at a rate in which their accumulation in tissue is so minute that RT-PCR is not able to detect their presence. Also, as I have alluded to in the previous chapters, there are potentially roles SR proteins play in pre-mRNA processing other than facilitating alternative splicing. The regulation of expression of *RSZp22a* could very well be a product of another post-transcriptional function of SC35.

A very interesting experiment to elucidate and answer for the phenomena, would be to cross the SR protein gene mutants, in this case the *sc35-1* mutant specifically, with plants mutant for genes required for NMD to take place. A family of three genes, the *UPF* family, are known to be required for NMD to occur, and mutants in *Arabidopsis* have been identified in both our lab by Payam Sajedi, and in other labs (Hori *et al.* 2005; Arciga-Reyes *et al.* 2006). The result of the crosses between mutants of SR protein genes and mutants in the *UPF* gene family may result in identifying additional alternatively spliced transcripts of SR protein genes that have not been reported to date. In the case of RSZp22a, if additional alternatively spliced transcripts were found in a *sc35-1* and *upf* double mutant line, it would explain why the *RSZp22a-GFP* fusion protein gene is expressed differently in *sc35-1* mutants than in wild-type. However, a lack of additional alternatively spliced transcripts would possibly indicate that SC35 is, in fact, performing a function other than facilitating alternative splicing in gene regulation. SR proteins are known to shuttle RNAs in and

out of the nucleus and therefore could potentially have post-transcriptional effects on target genes.

This chapter is unique in that the easy solution to further the knowledge gained from the experiment is not necessarily identifying more mutants as is the case for the previous chapters. The sheer magnitude of studying the expression pattern of 20 SR protein genes in a single genetic background of the root alone at different ages, under different environmental conditions, and grown with different nutrients, is daunting enough. However, to be completely thorough, one would need to look at other tissues as well such as leaves, meristem, flowers, and pollen just to name a few. Add into this process additional genetic backgrounds in the form of SR protein gene mutants, a lifetime of work is imminent and creating a global analysis of the expression patterns of all SR protein genes in all mutant plant lines is nearly impossible. However, much interesting information has been obtained in this study and can be applied to future experiments using these same fluorescent lines. I have learned that to say a particular SR protein gene is expressed in a specific tissue in a specific developmental stage may not be a very accurate description at all. Instead, the actual age of the cell as well as the environmental and growth conditions may have just as important of a role in expression as the tissue type itself.

Using confocal microscopy and fluorescent fusion protein genes can and should be a very powerful tool in studying specific SR protein gene mutants. For example, the *srp30-1* mutant displays a *leunig*-like floral phenotype. There is no evidence suggesting that SRp30 directly regulates any of the genes in this pathway to produce such a phenotype. The phenotype could be a result of the cross regulation of

another SR protein gene that is physically interacting with *LEUNIG*. Performing these same GFP fusion experiments and looking at the flower at various time and developmental stages may give insight into not only which SR protein gene or genes is involved in processing the *LEUNIG* pre-mRNA, but at what age, developmental stage, and specific tissues this process is occurring.

The future directions of the previous chapters can be summed up by saying that brute force and sheer volume of additional broad experimentation is needed to significantly advance what was found in the experiments I have performed. However, the future of this experiment and chapter requires more of a finesse approach and demonstrates how all of the experiments in this thesis are intertwined. Identifying a plant mutant for an individual SR protein gene leads to studying a phenotype. A well studied phenotype in turn leads to a molecular characterization of possible SR protein genes involved in producing the phenotype. Observing the expression patterns of SR protein genes identified as candidate factors in causing the phenotype can lead to pinpointing the exact location and point in time at which the fate of the organism is determined.

Gene		stage 1				stage 2				stage 3			
		E	C	N	V	E	C	N	V	E	C	N	V
RSZp33	Primary	+	+	+	+	-	-	-	-	-	-	-	-
	Lateral	+	+	+	+	-	-	-	-	-	-	-	-
RSZp22	Primary	+	+	+	+	-	-	-	-	-	-	-	-
	Lateral	-	+	+	-	-	-	-	-	-	-	-	-
RSp40	Primary	+	+	+	+	-	-	-	-	-	-	-	-
	Lateral	-	-	-	-	-	-	-	-	-	-	-	-
RSZp22a	Primary	+	+	+	+	-	-	+	+	-	-	+	+
	Lateral	+	+	+	+	+	-	+	+	+	-	+	+
RSp34a	Primary	+	+	+	+	-	-	-	-	-	-	-	-
	Lateral	-	-	-	-	-	-	-	-	-	-	-	-
SRp007	Primary	+	+	+	+	-	-	-	-	-	-	-	-
	Lateral	+	+	+	+	-	-	-	-	-	-	-	-
SCL33	Primary	+	+	+	+	-	-	-	-	-	-	-	-
	Lateral	+	+	+	+	-	-	-	-	-	-	-	-
SRp34/SR1	Primary	+	+	+	+	+	+	+	+	+	+	+	+
	Lateral	+	+	+	+	+	+	+	+	+	+	+	+
RSp31	Primary	+	-	+	-	+	-	+	+	+	-	+	+
	Lateral	+	-	+	-	+	-	+	+	+	-	+	+
RSZp21	Primary	-	-	+	-	+	-	-	-	+	-	-	-
	Lateral	-	-	+	-	+	-	-	-	+	-	-	-
SRp34b	Primary	-	-	-	-	-	-	-	-	-	-	-	-
	Lateral	-	-	-	-	-	-	-	-	-	-	-	-
SC35	Primary	+	+	+	+	-	-	-	-	-	-	-	-
	Lateral	+	+	+	+	-	-	-	-	-	-	-	-
RSZp32	Primary	+	+	+	+	+	+	-	-	+	+	-	-
	Lateral	+	+	+	+	+	+	-	-	+	+	-	-
RSp31a	Primary	+	+	+	+	-	-	-	-	-	-	-	-
	Lateral	+	+	+	+	-	-	-	-	-	-	-	-
SCL28	Primary	+	+	+	+	-	-	-	-	-	-	-	-
	Lateral	+	+	+	+	-	-	-	-	-	-	-	-

Table 5.3 Individual SR protein genes are expressed in wild-type plants in a tissue specific manner. Expression patterns of specific *SR-GFP* fusion transgenes in 10 day-old wild-type plants were analyzed using confocal microscopy. The expression pattern was recorded in three stages, one, two, and three as described earlier in this chapter. The presence of nuclear fluorescence was recorded and documented in four categories: epidermis (E), cortex (C), Endodermis (N), and Vascular tissue (V). The presence of visible SR-GFP fusion protein is indicated with a "+" while the absence of visible fusion protein is indicated with a "-".

Gene		stage 1				stage 2				stage 3			
		E	C	N	V	E	C	N	V	E	C	N	V
RSZp33	Primary	#	#	#	#	#	#	#	#	#	#	#	#
	Lateral	#	#	#	#	#	#	#	#	#	#	#	#
RSZp22	Primary	#	#	#	#	#	#	#	#	#	#	#	#
	Lateral	#	#	#	#	#	#	#	#	#	#	#	#
RSp40	Primary	#	#	#	#	#	#	#	#	#	#	#	#
	Lateral	#	#	#	#	#	#	#	#	#	#	#	#
RSZp22a	**Primary**	+	+	+	+	-	-	-	-	-	+	-	+
	Lateral	+	+	+	+	-	-	-	-	+	+	-	-
RSp34a	Primary	+	+	+	+	-	-	-	-	-	-	-	-
	Lateral	-	-	-	-	-	-	-	-	-	-	-	-
SRp007	Primary	+	+	+	+	+	-	-	-	-	-	-	-
	Lateral	+	+	+	+	-	-	-	-	-	-	-	-
SCL33	Primary	#	#	#	#	#	#	#	#	#	#	#	#
	Lateral	#	#	#	#	#	#	#	#	#	#	#	#
SRp34/SR1	**Primary**	+	+	+	+	+	-	+	-	+	-	-	+
	Lateral	+	+	+	+	+	-	+	-	+	+	+	+
RSp31	**Primary**	+	-	+	-	-	-	-	-	+	-	-	+
	Lateral	+	+	+	+	+	+	+	+	+	+	+	+
RSZp21	**Primary**	+	+	+	+	-	-	-	-	-	-	-	-
	Lateral	+	+	+	+	-	-	-	-	-	-	-	-
SRp34b	Primary	-	-	-	-	-	-	-	-	-	-	-	-
	Lateral	-	-	-	-	-	-	-	-	-	-	-	-
SC35	Primary	+	+	+	+	-	-	-	-	-	-	-	-
	Lateral	+	+	+	+	-	-	-	-	-	-	-	-
RSZp32	**Primary**	+	+	+	+	+	-	-	-	+	+	-	-
	Lateral	+	+	+	+	+	-	-	-	+	+	-	-
RSp31a	**Primary**	+	+	+	+	+	-	+	+	+	-	+	+
	Lateral	-	-	-	-	-	-	-	-	-	-	-	-
SCL28	Primary	+	+	+	+	-	-	-	-	-	-	-	-
	Lateral	+	+	+	+	-	-	-	-	-	-	-	-

Table 5.4 Individual SR protein genes are expressed in *sc35-1* roots in a tissue specific manner. Expression patterns of specific *SR-GFP* fusion transgenes in 10 day-old *sc35-1* plants were analyzed using confocal microscopy. The expression pattern was recorded in three stages, one, two, and three as described earlier in this chapter. The presence of visible GFP was recorded and documented in four categories: epidermis (E), cortex (C), Endodermis (N), and Vascular tissue (V). The presence of nuclear fluorescence is indicated with a "+" while the absence of visible fusion protein is indicated with a "-". Transgenes which were not inserted into the *sc35-1* mutant plant line is indicated with a "#" and bold type indicates transgenes with an expression pattern which differs from that observed in wild-type plants.

Gene		stage 1				stage 2				stage 3			
		E	C	N	V	E	C	N	V	E	C	N	V
RSZp33	Primary	#	#	#	#	#	#	#	#	#	#	#	#
	Lateral	#	#	#	#	#	#	#	#	#	#	#	#
RSZp22	Primary	#	#	#	#	#	#	#	#	#	#	#	#
	Lateral	#	#	#	#	#	#	#	#	#	#	#	#
RSp40	Primary	#	#	#	#	#	#	#	#	#	#	#	#
	Lateral	#	#	#	#	#	#	#	#	#	#	#	#
RSZp22a	**Primary**	+	+	+	+	-	-	-	-	-	-	-	-
	Lateral	+	+	+	+	+	+	+	+	-	+	-	-
RSp34a	Primary	#	#	#	#	#	#	#	#	#	#	#	#
	Lateral	#	#	#	#	#	#	#	#	#	#	#	#
SRp007	Primary	#	#	#	#	#	#	#	#	#	#	#	#
	Lateral	#	#	#	#	#	#	#	#	#	#	#	#
SCL33	Primary	+	+	+	+	-	-	-	-	-	-	-	-
	Lateral	+	+	+	+	-	-	-	-	-	-	-	-
SRp34/SR1	**Primary**	+	+	+	+	+	-	+	+	+	-	+	+
	Lateral	+	+	+	+	+	-	+	+	+	-	+	+
RSp31	Primary	#	#	#	#	#	#	#	#	#	#	#	#
	Lateral	#	#	#	#	#	#	#	#	#	#	#	#
RSZp21	Primary	#	#	#	#	#	#	#	#	#	#	#	#
	Lateral	#	#	#	#	#	#	#	#	#	#	#	#
SRp34b	Primary	#	#	#	#	#	#	#	#	#	#	#	#
	Lateral	#	#	#	#	#	#	#	#	#	#	#	#
SC35	**Primary**	+	+	+	+	-	-	-	-	-	-	-	-
	Lateral	+	+	+	-	-	-	-	-	-	-	-	-
RSZp32	**Primary**	+	+	+	+	+	+	+	-	-	-	-	-
	Lateral	+	+	+	+	-	-	-	-	-	-	-	-
RSp31a	Primary	+	+	+	+	-	-	-	-	-	-	-	-
	Lateral	+	+	+	+	-	-	-	-	-	-	-	-
SCL28	Primary	+	+	+	+	-	-	-	-	-	-	-	-
	Lateral	+	+	+	+	-	-	-	-	-	-	-	-

Table 5.5 Individual SR protein genes are expressed in *sr45-1* roots in a tissue specific manner. Expression patterns of specific *SR-GFP* fusion transgenes in 10 day-old *sr45-1* plants were analyzed using confocal microscopy. The expression pattern was recorded in three stages, one, two, and three as described earlier in this chapter. The presence of visible GFP was recorded and documented in four categories: epidermis (E), cortex (C), Endodermis (N), and Vascular tissue (V). The presence of nuclear fluorescence is indicated with a "+" while the absence of visible fusion protein is indicated with a "-". Transgenes which were not inserted into the *sr45-1* mutant plant line is indicated with a "#" and bold type indicates transgenes with an expression pattern which differs from that observed in wild-type plants.

Gene		stage 1				stage 2				stage 3			
		E	C	N	V	E	C	N	V	E	C	N	V
RSZp33	Primary	#	#	#	#	#	#	#	#	#	#	#	#
	Lateral	#	#	#	#	#	#	#	#	#	#	#	#
RSZp22	Primary	#	#	#	#	#	#	#	#	#	#	#	#
	Lateral	#	#	#	#	#	#	#	#	#	#	#	#
RSp40	Primary	#	#	#	#	#	#	#	#	#	#	#	#
	Lateral	#	#	#	#	#	#	#	#	#	#	#	#
RSZp22a	Primary	#	#	#	#	#	#	#	#	#	#	#	#
	Lateral	#	#	#	#	#	#	#	#	#	#	#	#
RSp34a	Primary	#	#	#	#	#	#	#	#	#	#	#	#
	Lateral	#	#	#	#	#	#	#	#	#	#	#	#
SRp007	Primary	+	+	+	+	+	-	-	-	-	-	-	-
	Lateral	+	+	+	+	-	-	-	-	-	-	-	-
SCL33	Primary	#	#	#	#	#	#	#	#	#	#	#	#
	Lateral	#	#	#	#	#	#	#	#	#	#	#	#
SRp34/SR1	Primary	#	#	#	#	#	#	#	#	#	#	#	#
	Lateral	#	#	#	#	#	#	#	#	#	#	#	#
RSp31	**Primary**	+	+	+	+	-	-	-	-	+	-	-	+
	Lateral	+	+	+	+	+	+	+	+	+	+	+	+
RSZp21	**Primary**	+	+	+	+	-	-	-	-	-	-	-	-
	Lateral	+	+	+	+	-	-	-	-	-	-	-	-
SRp34b	Primary	#	#	#	#	#	#	#	#	#	#	#	#
	Lateral	#	#	#	#	#	#	#	#	#	#	#	#
SC35	**Primary**	+	+	+	+	-	-	-	-	-	-	-	-
	Lateral	+	+	+	+	-	-	-	-	-	-	-	-
RSZp32	Primary	+	+	+	+	+	+	-	-	+	+	-	-
	Lateral	+	+	+	+	+	+	-	-	+	+	-	-
RSp31a	Primary	+	+	+	+	-	-	-	-	-	-	-	-
	Lateral	+	+	+	+	-	-	-	-	-	-	-	-
SCL28	Primary	+	+	+	+	-	-	-	-	-	-	-	-
	Lateral	+	+	+	+	-	-	-	-	-	-	-	-

Table 5.6 Individual SR protein genes are expressed in *srp30-1* roots in a tissue specific manner. Expression patterns of specific *SR-GFP* fusion transgenes in 10 day-old *srp30-1* plants were analyzed using confocal microscopy. The expression pattern was recorded in three stages, one, two, and three as described earlier in this chapter. The presence of nuclear fluorescence was recorded and documented in four categories: epidermis (E), cortex (C), Endodermis (N), and Vascular tissue (V). The presence of visible SR-GFP fusion protein is indicated with a "+" while the absence of visible fusion protein is indicated with a "-". Transgenes which were not inserted into the *srp30-1* mutant plant line is indicated with a "#" and bold type indicates transgenes with an expression pattern which differs from that observed in wild-type plants.

Materials and Methods

Transgene construction. Individual SR proteins were PCR amplified from genomic DNA using Invitrogen Elongase™ DNA polymerase using the manufacturer's directions. Gene specific primers (Table 5.7) were designed to include sequences approximately 2KB upstream of the start codon through the last codon prior to the stop codon. For cloning purposes, reverse primers were engineered to contain an NcoI restriction site followed by a guanine, to maintain frame, at the 5' end of the reverse primer with the exception of the reverse primer for SRp30 in which an ApaI site and a cytosine was engineered at that same position. PCR products from

this amplification were run on a 0.45% agarose gel and purified using Quiagen's

Quiaquick™ spin columns per manufacturer's directions.

Once purified, PCR products were ligated overnight into Invitrogen™ cloning

vector 2.1 following directions provided in the Invitrogen™ TA cloning kit.

Resulting ligate was transformed into DH5 alpha chemical competent cells and plated

onto LB plates containing 100ug/ul ampicillin and 0.2% X-galactose (Xgal) and

grown overnight at 37 degrees. Colonies that survived ampicillin selection and were

white in color were verified for containing the cloned gene as well as correct

orientation in that the 5' end of the clone was adjacent to the NotI restriction site.

Verification was performed through a gene specific restriction digestion. Colonies

that were verified for containing their respective cloned SR protein gene were then

digested with NotI and NcoI which produced fragments sizes 1.55Kb, 2.4Kb, and a

gene specific fragment ranging from 4Kb to 6Kb depending on the gene.

The cloned gene isolated from the TA vector was then ligated into cloning

vector pGlowbug that I created in the lab from pND1 (created by Natalie Dye, a

former undergraduate in our lab) and pDN393 which contains the GFP gene. When

partially digested by NotI and a full digest with NcoI, pGlowbug produces four bands

of sizes 0.9Kb, 1.4Kb, 2.9Kb, and 3.8Kb. The 3.8Kb is the pGlowbug fragment that

is used for ligation with the SR protein gene clones isolated from the TA vector.

When ligated, into the 3.8Kb pGlowbug fragment, the genomic clone and GFP are in

frame and are separated by two amino acids, a glycine and a proline. This fusion

protein gene is then followed by the Nos terminator.

Colonies verified to contain their respective SR-GFP fusion genes were then digested with restriction enzyme NotI, gel purified as described earlier, and ligated into cloning vector pMLBart that had been cut to completion with NotI. The resulting ligate was transformed into chemical competent cells and plated onto LB plates containing 50ug/ul spectinomycin and 0.2% Xgal. Colonies that displayed the white phenotype were analyzed by PCR for containing the SR-GFP fusion construct.

DNA preparations were performed on colonies which were confirmed to contain their respective SR-GFP construct and transformed into *Agrobacterium tumefaciens* strain GB3101 via electroporation cell transformation. The transformation was then plated on LB plates containing 50ug/ul spectinomycin, 100ug/ul gentamycin, and 100ug/ul rifampicin. Colonies surviving antibiotic selection were suspended in 12% glycerol and placed at -90 degrees Celsius for long term storage until needed for plant transformation.

Plant transformation. Plants to be used for transformation were allowed to grow at 20 degrees and 16 hour light. *sc35-1* and *sr45-1* homozygous mutants were used for transformation while plants that genotyped as heterozygous for the *srp30-1* mutation were used for transformation. Primary shoots were cut to induce auxiliary shoot formation and growth. Plants were then prepared for transformation via agro bacterium using the floral dip method reviewed in Plant Physiol, December 2000, Vol. 124, pp. 1540-1547. After transformation, T0 plants were put into growth chambers under 24 hours of light at 20 degrees Celsius and watered as needed. Seeds were collected after plants completely senesced and dried for no less than 2 weeks.

Transformant selection. Seeds were collected from plants that underwent transformation and approximately 250 T1 seeds were planted on soil. After 2 days of 4 degrees Celsius under 24 hours dark, pots were then moved to a growth chamber and were allowed to germinate at 20 degrees Celsius and 24 hour light. At the 4 leaf stage, seedlings were sprayed with 1000X dilution of Finale™ herbicide. Wild-type plants, *sc35-1* and *sr45-1* homozygous mutants, and plants which genotyped as heterozygous for the *srp30-1* mutation and survived the herbicide selection, were thinned down to 5 plants per line and were allowed to self-fertilize for T2 seed stock collection.

Growth conditions. Seeds grown in preparation for microscopy of roots were plated on MS plates and 1% agar without sucrose or vitamins. Plates were held vertically allowing roots to grow straight on the surface of the media. Plated plants were grown under 24 hours of light at 20 degrees Celsius. Petri dishes were wrapped with M3 surgical tape to retain humidity.

Gene Name	Forward	Reverse
SCL30	CCAAGTGGTTCGAGCCT	CCATGGGTCTTGGAGATACCTCCACA
SCL28	TATAAGGAGCGGAACACTCAG	CCATGGGACGACTGCAACAAAACACC
SCL33a	CTTTCAGGCAATTACTAGCATAATC	CCATGGGCTGGCTTGGAGAACGGT
SRp30	ACAACTAGGGTCAGATGGTTTG	CCGGGCCAGTTGGTGTTGATTGAAC
SRp31	CACCACCATGAACACAAAATGT	CCATGGGAGGTCTTCCTCTTGGACT
SRp34a	TATGTGTCCAATTGCAAACAA	CCATGGGATTATCCGAAAACAGCAAA
RSZp21	ACTCTTTGGGGACAGGCT	CCATGGGGACCCTCGAGAAACGC
RSp41	CTGTTGCTGACCGCCC	CCATGGGTTCCTCTGCTGGCGG
SR45	ATTTCTGCTCCAAAATTCTATATAAGG	CCGGGCCCAAGTTTTACGAGGTGGA
RSZp33	AGGATTCTTCGCATTCGC	CCATGGGAGGAGACTCACTTCCTCTA
SR1	GTTTGGAGAAATTGGAATTGG	CCATGGGCCTGCAAACAAGAGCTCAG
RSp40	GTTGTTACAATCGTTTTTAAAATTCC	CCATGGGCTCGTCAGCTGGTGGC
RSZp22a	TCAAGTGGTAAAAGTTTTTATTTGAA	CCATGGGGCTCCGGCTTCTGCG
SC35	GACGATATGGATGCAGAGCG	CCATGGGCTTTAAACAAAAACATGTTA
RSZp32	TCTTATCCGCCGCATTC	CCATGGGAGGTGACTCACTGCCTTTA
SCL33	AATATGCAAATGGATTTCTAGACAA	CCATGGGTGCTTCTTCTAGGGCTGGG
RSZp22	TCTCGTTCAGTGTAAAGTTCTTTAAT	CCATGGGGCTCCTGCTTCTGCGTC
SRp007	CACATTGCTTCAGTATTGAGCA	CCATGGGCCAACGTTCATATGATGAA
SRp102	AACGGATCTGAAGAACAAAGTC	CCATGGGACCTCTTGCTCTTTGAATC
SRp34b	CCCATTGGATCAACTCCATTA	CCATGGGTCGAGCTTTTTTATCTGGA

Table 5.7. Gene specific primers for SR protein gene cloning. Primers were designed for all 20 SR protein genes. Reverse primer contains CCATGG, the restriction site for NCOI or, in the case of SRp30 and SR45, a CCGGGC, the restriction site for ApaI. The restriction sites are used for future cloning. Each restriction site is followed by a glycine (G) and then begins at the last amino acid coding codon prior to the stop codon. The forward primer is located approximately 2Kb upstream of the coding region of each respective gene.

Chapter 6: Final thoughts and future directions

<u>Summary of findings</u>

Prior to the initiation of this project, there was already considerable literature on known and proposed functions of SR proteins and regulatory sequences. Additionally reported were experiments studying the crossregulation and autoregulation of overexpressed SR protein gene cDNAs in both plants and animals. All of those individual experiments inspired me to attempt one large and comprehensive long-term project. Doing this would bring together many unique questions into one story. In order to draw any kind of "big picture" conclusions, several questions first had to be answered.

Mutations in SR protein genes in mammals are embryonic lethal (Wang *et al.* 2001; Xu *et al.* 2005) suggesting that SR proteins are necessary for the viability of the organism. The SR protein gene family in plants consists of more than twice the number found in animals. Are SR proteins in plants required for viability as seen in animals, or is there some redundancy in function of the additional members? While multiple labs had identified, characterized, and studied SR proteins in Arabidopsis (Lazar *et al.* 1995; Lopato *et al.* 1996; Lopato *et al.* 1996b; Golovkin *et al.* 1999; Lopato *et al.* 1999; Lazar *et al.* 2000; Lopato *et al.* 2002; Ali *et al.* 2003; Birnbaum. *et al.* 2003; Fang *et al.* 2004), none had isolated a mutant, characterized phenotypes, or performed molecular studies with a mutant. Instead, proposed functions were based on *in vitro* experiments, cell cultures, and over expression studies in animal cell lines as well as overexpression in wild-type plants.

I have successfully identified, from among 45 possible T-DNA insertion lines, five individual mutations that appear to eliminate the expression of *Arabidopsis* SR protein genes. Whether or not they are essential or redundant in plants depends on the SR protein gene. Mutations in three SR protein genes, *SC35, SRp30*, and *SR45* are viable with developmental phenotypes ranging in severity. Mutations in two additional genes, *RSp31* and *SCL28* are embryonic lethal. This outcome was very surprising as I felt my best chance at obtaining a viable mutant would have been to identify a mutant in the SCL subfamily due to there being four members, all of which are very closely related to each other. It was therefore surprising to me to find that *scl28-1* is lethal, while *sc35-1* is viable with a very mild phenotype. Identifying an additional homolog of ASF/SF2, *Arabidopsis SRp30*, and a plant specific SR protein gene, *SR45*, was also a very encouraging find which would strengthen my experiments.

At the beginning of this project, it was known that SR proteins are essential for both constitutive and alternative splicing and that they function by binding to the pre-mRNA through sequence specific factors called ESEs (Blencowe *et al.* 2000). Once bound to pre-mRNA, SR proteins recruit U2AF and the U1 snRNP to initiate spliceosome formation. In the ESE assay performed in my research, I found that in five of the 15 sequences tested, multiple SR protein gene mutants could significantly affect whether or not an individual sequence is processed as an ESE. Unfortunately, I was unable to identify whether or not the individual SR protein played a role in ESE recognition, or if the mutants were responsible for an important protein-protein interaction that facilitates the particular splicing event that the ESE regulates.

However, this question encourages further investigation and future projects resulting from this result and is discussed later in this chapter.

In addition to the question of which SR proteins play an important role in identifying and processing sequences acting as ESEs, very few known genetic targets of individual SR proteins have been reported, and relatively few publications are available describing attempts to identify them. This phenomena is probably more due to the fact that genes that undergo alternative splicing are probably regulated in a very specific manner with precise environmental and tissue specific requirements. To simply select a handful of alternatively spliced genes and hope that mutations in an individual SR protein gene, or set of genes, will affect how the target is spliced, will probably be disappointing in that valuable data will not be generated from this approach. To avoid this problem, the Barta lab used SR protein genes as target genes for alternative splicing (Lopato *et al.* 1999). At the time they performed their experiments, SR protein genes were known to be alternatively spliced. By using the SR protein gene family as their targets, they could collectively study targets of SR proteins as well as describe the crossregulation relationship between individual SR proteins. However, they were limited to overexpressing SR protein genes which, in itself, is a problem. While they demonstrated that overexpressing individual SR proteins altered isoform production of other SR protein gene transcripts, it does not necessarily give a good indication of what occurs naturally.

Both the experiments performed in the Barta lab and those that I performed myself, resulted in some very interesting consistencies. In both experiments, overexpressing an individual SR protein gene, or looking at the effects of

crossregulation in mutants, indicated that an individual gene can crossregulate many other SR protein genes. Likewise, a single gene can be regulated by multiple SR protein genes, essentially undergoing a multitude of checks and balances insuring that the gene is being expressed in a manner which benefits the organism best. This result, in addition to the ESE assay, was my first indication that several SR proteins probably work in concert to efficiently recognize, bind, and facilitate splicing of any one particular splicing event. The determining factor in which SR protein genes participate in any splicing event may be the physical location of where they are expressed, which is discussed in chapters four and five.

One attempt other labs have made to determine if plant SR protein genes could have redundant features or if they are all essential as mammalian SR protein genes appear to be, was to create fusion proteins with visual tags such as GUS and YFP and study the expression patterns. While the Spector and Barta labs showed evidence of both overlapping and differential expression patterns, neither lab included whole genomic data into their expression vectors, relying only on the promoter. The significance of this, once again, is that SR protein genes are crossregulated and the expression patterns observed in the two labs may not be an accurate representation of where those respective SR protein genes are expressed *in planta*. However, using whole genomic sequence including introns, the constructs I generated would be regulated and expressed in a fashion which more closely resembles that of the native genes. In my experiments, I did, in fact, observe that mutants did alter the expression of the *GFP* transgenes. In some cases, an individual *SR-GFP* fusion gene was expressed in mutant tissues when they were not in wild-type. Likewise, there were

165

instances in which tissues of wild-type plants expressed the transgene and did not do so in mutants showing a very specific expression pattern.

Splicing by committee

As I stated at the beginning of this chapter, my goal was to use a number of approaches to determine an overall story of how SR proteins are capable of regulating the expression of genomic targets. Taking all of the data into consideration, the interaction of SR proteins between pre-mRNA, themselves, and other splicing factors is probably more complex than previously thought. The data presented here suggests that a single SR protein does not simply recognize an ESE and recruit other non-SR protein components of the splicing complex individually. Rather, multiple SR proteins are involved in a single splicing event. Whether it is by recognition of the splice site by multiple SR proteins, or if one protein recognizes the splice site and then recruits other SR proteins to perform an unknown job, is still to be confirmed. Additionally, the GFP transgene experiments clearly demonstrated that individual SR proteins are regulated, both temporally and spatially. Mutants in mammals suggest that SR protein genes are not redundant and that the resulting proteins perform unique functions that result in lethality when one SR protein gene is absent. However, as the cross regulation and GFP experiments suggest, it may not be whether an SR protein has the capability to compensate for the loss of another, but whether or not the active and functional isoform of an SR protein that is capable of compensating for the loss an individual SR protein gene, is produced in the same tissue and at the same developmental time.

166

When looking at all the experiments performed with the mutations, it appears that SR proteins do, in plants, have overlapping and redundant functions. However, the functions are not necessarily due to a single SR protein compensating for the loss of another, which is something I expected to see within subfamilies. For instance, *SRp30* is one of four *Arabidopsis* genes in the *ASF/SF2* subfamily of SR protein genes. In the crossregulation experiments of *srp30-1* mutants, it was not only the other members of the subfamily that displayed an alteration of isoform. Rather, non-family members such as *RSZ33, SCL33*, and *RSp40* displayed an alteration of isoform accumulation in at least one of the three tissues lacking functional *SRp30*.

Additional evidence for multiple SR proteins to be involved in a single splicing event comes from the ESE assay. For two separate lines, ESE34 and ESE57, a mutation in more than one SR protein gene leads to a dramatic shift in splicing efficiency. In ESE34, a mutation in either *SC35* or *SRp30* leads to a dramatic shift from exon inclusion to exon skipping. Unfortunately, from this experiment alone there is no way to know if those two SR proteins are performing the same function of binding to the RNA, performing essential protein-protein interactions, or each performing a separate role in splicing. What is quite clear is that both SR proteins play a vital role in processing the same exact ESE sequence.

Another more interesting example is ESE57, where all three mutant lines affected the ability of the ESE57 sequence to direct exon inclusion. Plants mutant for *SRp30* and *SR45* showed reduced inclusion of exons containing ESE57, whereas a *sc35-1* mutant actually improved the efficiency of exon inclusion of the same construct. In this case, wild-type SC35 may be negatively regulating another splicing

factor, possibly an SR protein, that recognizes the ESE57 sequence as an ESE. An interesting observation is that these real-time RT-PCR assays were carried out in leaves from the 8 leaf stage of the plant, which is also the same tissue used in the crossregulation studies. When looking at the crossregulation data of the same tissue type, the *sc35-1* mutant does not alter the isoform production of either *SRp30* or *SR45* which would be expected if *SC35* is a negative regulator of *SRp30* or *SR45* in rosette leaves. This data suggests that yet a fourth splicing factor, possibly another SR protein, is involved in recognition of the ESE57 sequence.

In addition to the ESE assay, the SR-GFP fusion gene experiment also demonstrated that multiple SR proteins have different splicing patterns in mutants. At the beginning of this particular project, my goal was to identify a single SR protein gene that would display an alteration of expression pattern in roots in one mutant. My expectations were exceeded in that each of my mutant plant lines, *srp30-1*, *sc35-1*, and *sr45-1*, had profound effects on the splicing pattern of multiple *SR-GFP* transgenes. Whether these individual genes are compensating for the mutant in each respective line, or are regulated by the individual SR protein gene cannot be confirmed with this study.

While the exact role that each individual SR protein plays in any splicing event is unclear, it has become quite obvious that, at least in plants, SR protein genes work together, by committee, in splicing. While identifying the ability of an individual SR protein gene to influence the processing of a specific ESE is important and a necessary area of research, it is not the only determining factor of SR protein participation of splicing. Rather, a multitude of potential factors are necessary for

efficient splicing to take place. The expression of the SR protein in a specific location at a specific point in time is possibly just as important as the sequence to which it can bind to. Additionally, other non-RNA binding functions may also be the determining factor of an individual splicing event to occur. A combination of many experiments with loss of function mutations is the future for this type of research and only then will a thorough description of splicing be possible.

Future direction and experiments

The data obtained in the experiments of this thesis indicate that SR proteins probably play multiple roles in mRNA processing. However, the most far-reaching observation is that more specific and precise functions could easily be obtained with additional work. None of this would be possible without *bona fide* mutants which, up to this point, did not exist in any model organism. One of the most necessary future goals would be to collect additional SR protein gene mutants including those in *RSp31* and *SCL28* that were found to be lethal in this study. Identifying additional SR protein gene mutants will not only provide an invaluable resource for information as to how SR proteins function, specificities of crossregulation, where they are expressed, and what sequences they target, but they also open the possibility to study other splicing phenomena not addressed in this set of experiments.

A very interesting topic of mRNA research is nonsense-mediated decay (NMD) and how it is used to both regulate gene expression, as well as protect the organism from undesirable proteins resulting from aberrant splicing or mutation. Analysis of plants carrying mutations in one or more SR protein genes as well as

plants mutant for one of three *UPF* genes, all of which are required for NMD (Hori *et al.* 2005; Arciga-Reyes *et al.* 2006; Banihashemi *et al.* 2006; Kertesz *et al.* 2006; Metzstein *et al.* 2006), may possibly lead to further discoveries in other topics of RNA processing or even other fields of research.

One accepted function of SR proteins is that they bind to ESEs to facilitate splicing and a point mutation within the ESE sequence of a target gene could have a dramatic effect on the quantity of functional product being produced. Identifying specific sequences that individual SR proteins recognize will help directly link specific SR protein genes into a more precise splicing mechanism. Using the *in vivo* splicing assay developed in our lab by Mount et al. (in preparation) on more sequences and more mutants, both individual mutant lines as well as lines containing multiple mutations, will greatly aid in accomplishing this task. In a small sample of only 15 potential ESE sequences and in three mutant lines, I have already demonstrated that two individual mutants can drastically affect the ability of the same sequence to facilitate splicing. The reason for this phenomena is unknown but could be explored using a plant line mutant for both SR protein genes. In short, a high throughput of more sequences in as many mutant plant lines as possible that can be identified, will greatly aid in elucidating the exact splicing mechanism in plants which can then be used to develop experiments with the same goals in animals.

Chapter 4 described the ESE assay in mutant plant lines and had one significant result in that *srp30-1* or *sc35-1* mutants reduced the ability of ESE34 to promote exon inclusion. However, there is no apparent explanation for exactly what function either of these two SR proteins are performing to lead to this result. This

result is a significant factor in my belief that multiple SR proteins are involved in a single splicing event. In the discussion section I suggested performing the same experiment using a *srp30-1 sc35-1* double mutant to determine whether or not the two proteins were performing a redundant function or if they had independent functions in processing that particular sequence as an ESE. Another very interesting experiment may be to create chimeric genes by swapping domains between SRp30, SC35, and that of another protein which does not affect the processing of that sequence such as SR45, and attempt to rescue wild-type function. For example, the *sc35-1* mutants could be transformed separately with a chimeric gene composed of the SC35 RRM and SR45 RS domain and the SR45 RRM and the SC35 RS domain. The results of this experiment help to deduce the role of either of those two proteins, SC35 and SR45 in processing that particular ESE.

Additionally, it would be very interesting to see if the same inclusion to skipping ratio is consistent throughout the entire plant in mutants or if there are any differences observed. In viable mutants, there is good indication that some SR proteins are able to compensate for the loss of others. If this is the case, it would be interesting to know exactly which ones do this and if the compensation is both SR protein and tissue specific or if it is dependent on just a single one of those factors. Using ESEs which cause a dramatic change in the inclusion to skipping ratio and performing the same real-time assay in different tissues and with additional mutants, would be the appropriate solution to finding an answer to this question.

In addition to understanding how SR proteins regulate splicing through the processing of ESEs, it is equally important to identify and understand the

crossregulation that SR protein genes undergo and more importantly, the specific requirement and conditions for this type of regulation to take place. Obviously, more SR protein gene mutants need to be identified and when they have, RT-PCR can be performed on these as well. Additionally, my experiments consisted of analyzing crossregulation in only three tissue types, rosette leaves, roots, and flowers. A more comprehensive analysis of additional tissues such as stems, cauline leaves, and sepals. In addition to whole organs more specific tissues within an organ, like pollen of a flower for instance, need to be analyzed. To get a more complete global picture of how and when individual SR proteins are autoregulated, more environmental conditions should be tested. This analysis can be performed under extreme temperatures, various nutrient supplements, atmospheric and chemical stresses, and any other factor desired. It is quite likely that different SR proteins are active in different conditions and may have different roles depending on the relative health and stress on the individual plant. Creating a large database of data obtained from numerous growth conditions can very well lead to an understanding of exactly how these proteins interact with each other and possibly are capable of compensating for the loss of another SR protein gene.

The *SR-GFP* fusion protein genes worked exceptionally well in this experiment in that very specific expression patterns were identified in different tissues of the root and expression was limited to the nucleus, where SR proteins congregate (Fang *et al.* 2004). In addition to tissue specificity, these same constructs were regulated by age and developmental stage of the specific tissue. In some lines, the relative age of the lateral root dramatically affected the splicing pattern of an

172

individual *GFP* transgene. Also, primary roots which differed by only four days could also show a dramatic difference in expression pattern. One of the most important factors in concluding this experiment was a success, is that the expression pattern of some of the genes in wild-type plants differed from that observed in previous research using cDNA expressed by native promoter (Fang *et al.* 2004), GUS driven by native promoter (Lopato *et al.* 1999), or microarray expression analysis (Birnbaum. *et al.* 2003). This is significant in that none of the three previous experiments took into consideration crossregulation at the level of splicing and essentially were identifying specific sites where the promoter was active, not necessarily where functional protein was being produced. My experiments, however, were engineered to be affected by any regulatory factors a native gene would be subject to, regardless of how subtle.

Another obvious direction for this experiment would be to expand on the same experiment but with additional mutants, constructs which I was unable to engineer, and generating transgenic lines that I could not identify through Finale™ herbicide resistance. In addition to roots, other tissues such as leaves, stems, flowers, pollen, etc., can also be analyzed and expression patterns of individual SR proteins documented. While significantly different than that above mentioned experiments, this assay could be improved upon in one very significant way; co-transformation of constructs sensitive to splicing crossregulation and constructs sensitive to active promoter only. Using the Fang *et al.*, and the Lopato *et al*. experiments as a guide, transgenes can be generated using the exact same promoter used in my *GFP* experiments and fused to cDNA encoding for the respective functional SR protein

gene transcript and tagged with another fluorescent reporter gene such as *YFP* or

RFP. Another option is to create a transgene consisting of a fluorescent reporter gene

under the control of the same genomic SR protein gene promoters used in my study.

Co-transformation of the full genomic *GFP* constructs and a construct not under the

regulation of alternative splicing of the same respective gene, will negate any

difference in environmental or nutritional contribution of expression pattern between

two individual plants. A single plant expressing both constructs will guarantee exact

growth conditions for proper comparison between the two respective constructs.

Conclusion

I believe a good Ph.D thesis accomplishes two goals; makes a contribution to

the field and science as a whole, and provides enough data and thorough research to

induce more creative questions than answered. I believe that this research has

successfully accomplished both goals. In my graduate career, I have successfully

identified viable SR protein gene mutants, maintained stable lines, and described

phenotypes associated with the mutations. When this data is published, it will be the

first of its kind to report viable SR protein gene mutants in any model organism.

Using these mutants, I have demonstrated that SR protein gene mutants do, in fact,

alter the ability of unique sequences to function as ESEs.

In my opinion, the greatest outcome from my work is suggesting that the complex

interactions among SR proteins may be, in fact, every bit as complex or more

complicated than previously suspected on the basis of yeast two-hybrid assays and *in*

vitro studies. In reviews, it is binding to ESEs and recruitment of splicing factors that

receive the bulk of the attention. However, additional functions have been proposed

based on more circumstantial evidence such as sub-cellular localization, yeast two-

hybrid assays, and *in vitro* assays in which SR proteins are thought to participate in

nearly all steps of splicing. Understandably, these hypotheses are given little

attention in the same reviews (Graveley *et al.* 2000; Reddy *et al.* 2004). After

incorporating all of my data together and looking at the big picture, I believe the SR

protein gene family is a multi-functional family of genes possibly involved in many

steps of RNA processing and work together as opposed to individually in a target

specific manner. Knocking out two individual SR protein genes out of a possibility

of 20 produced a great reduction in the ability of one specific sequence, ESE34, to

function as an ESE. Both mutants had roughly the same effect suggesting that

possibly both proteins are involved in two different steps of processing that particular

RNA. The crossregulation data was similarly surprising in that I have viable mutants

for only three of the 20 SR protein genes. However, all three altered the transcript

production of several other SR protein genes in the three tissue types I studied. Such

a large effect by a relatively small fraction of the family members indicates that SR

proteins are likely to be involved in many steps of splicing.

Bibliography

Ali, G., Golovkin, M., Reddy, AS. (2003). "Nuclear localization and in vivo dynamics of a plant-specific serine/arginine-rich protein." <u>Plant Journal</u> **36**: 883-893.

Alonso J.M, E. J. R., et. al (2003). "Genome-Wide Insertional Mutagenesis of *Arabidopsis thaliana*." <u>Science</u> **301**: 653-657.

Amrein, H., Hedley, ML., Maniatis, T. (1994). "The role of specific protein-RNA and protein-protein interactions in positive and negative control of pre-mRNA splicing by Transformer 2." <u>Cell</u> **76**: 735-746.

Arciga-Reyes, L., Wootton, L., Kieffer, M., Davies, B. (2006). "UPF1 is required for nonsense-mediated mRNA decay (NMD) and RNAi in Arabidopsis." <u>Plant Journal</u> **47**: 480-489.

Banihashemi, L., Wilson, G.,M., Das, N., Brewer, G. (2006). "Upf1/Upf2 regulation of 3' untranslated region splice variants of AUF1 links nonsense-mediated and A+U-rich element-mediated mRNA decay." <u>Mol Biol Cell</u> **23**: 8743-8754.

Bennett, V. D., Pallante, K.M., Adams, S.L. (1991). "The splicing pattern of fibronectin mRNA changes during chondrogenesis resulting in an unusual form of the mRNA in cartilage." <u>J. Biol. Chem</u> **266**: 5918-5924.

Berget, S., M., Moore, C., Sharp, P.,A. (1977). "Spliced segments at the 5' terminus of adenovirus 2 late mRNA." <u>Pro Natl Acad Sci</u> **74**: 3171-3175.

Birnbaum., K., Shasha, D.,E., Wang, J.,Y., Jung, J.,W., Lambert, G.,M., Galbraith, D.,W., Benfey., P.,N. (2003). "A gene expression map of the Arabidopsis root." <u>Science</u> **Dec 12;302**(5652): 1956-1960.

Blencowe, B., J. (2000). "Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases." <u>Trends Biochem Sci.</u> **25**: 106-110.

Bonke, M., Thitamadee, S., Mahonen, A.,P., Hauser, M.,T., Helariutta., Y. (2003). "APL regulates vascular tissue identity in Arabidopsis." <u>Nature</u> **426**(6963): 181-186.

Bubulya, P., A., Prasanth., K.,V., Deerinck., T.,J., Gerlich., D., Beaudouin., J., Ellisman., M.,H., Ellenberg., J., Spector., D.,L. (2004). "Hypophosphorylated SR splicing factors transiently localize around active nucleolar organizing regions in telophase daughter nuclei." <u>J Cell Biol</u> **167(1)**(Oct 11): 51-63.

Burton-Wurster, N., Lust, G., Macleod, J.,N. (1997). "Cartilage fibronectin isoforms--In search of functions fora special population of matrix glycoproteins." <u>Matrix Biol.</u> **15**: 441-454.

Caceres, J., F, Stamm, S., Helfman, D.,M., Krainer, A.,R. (1994). "Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors." Science **265**: 1706-1709.

Campbell, M., A., Haas, B.,J., Hamilton, J.,P., Mount, S.,M., Buell, C.,R. (2006). "Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis." BMC Genomics **7**(327).

Cartegni, L., Wang, J., Zhu, Z., Zhang, M.,Q., Krainer, A.,R. (2003). "ESEfinder: A web resource to identify exonic splicing enhancers." Nucleic Acids Research **31**: 3568-3571.

Chandler, S., D., Mayeda, A., Yeakley, J.,M., Krainer, A.,R., Fu, X.,D. (1997). "RNA splicing specificity determined by the coordinated action of RNA recognition motifs in SR proteins." Pro Natl Acad Sci
 94: 3596-3601.

Chew, S. L., H.-L. Liu, A. Mayeda, and A. R. Krainer (1999). "Evidence for the function of an exonic splicing enhancer after the first catalytic step of pre-mRNA splicing." Proc. Natl. Acad. Sci. **96**: 10655-10660.

Chow, L., C., Gelinas, R.,E., Broker, T.,R., Roberts, R.,J. (1977). "An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA." Cell **12**: 1-8.

Colwill, K., Pawson, T., Andrews, B., Prasad, J., Manley, J.,L., Bell, J.,C., Duncan, P.,I. (1996). "The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution." EMBO J **15**: 265-275.

Crayton, M., E., 3rd, Powell, B.,C., Vision, T.,J., Giddings, M.,C. (2006). "Tracking the evolution of alternatively spliced exons within the Dscam family." BMC Evol Biol **Feb 16**(6): 16.

Desfeux, C., Clough, S.,J., Bent, A.,F. (2000). "Female reproductive tissues are the primary target of Agrobacterium-mediated transformation by the Arabidopsis floral-dip method." Plant Physiol **123**: 895-904.

Dinesh-Kumar, S., P., Baker, B.,J. (2000). "Alternatively spliced N resistance gene transcripts: their possible role in tobacco mosaic virus resistance." Pro Natl Acad Sci **97**: 392-398.

Ding Jian-Hua, F. X.-D., et. al. (2004). "Dilated cardiomyopathy caused by tissue-specific ablation of SC35 in the heart." EMBO J **23**: 885-896.

Du, C., McGuffin, M., E., Dauwalder, B., Rabinow, L., Mattox, W. (1998). " Protein Phosphorylation Plays an Essential Role in the Regulation of Alternative Splicing and Sex Determination in Drosophila." Mol Cell **2**: 741-750.

Fairbrother, W., G., Holste, D., Burge, D.,B., Sharp, P.,A. (2004b). "Single nucleotide polymorphism-based validation of exonic splicing enhancers." PLoS Biol. **Sep;2**(9): E268.

Fairbrother, W., G., Yeo, G.,W., Yeh, R., Goldstein, P., Mawson, M., Sharp, P.,A., Burge, C.,B. (2004a). "RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons." Nucleic Acids Research **32**: W187-W190.

Fang, Y., Hearn, S., Spector, D.,L. (2004). "Tissue-specific expression and dynamic organization of SR splicing factors in Arabidopsis." Mol Biol Cell **15**: 2664-2673.

Franks, R., G., Liu, Z., Fischer, R.,L. (2006). "SEUSS and LEUNIG regulate cell proliferation, vascular development and organ polarity in Arabidopsis petals." Planta **224**: 801-811.

Fu, X., D., Mayeda, A., Maniatis, T., Krainer, A.,R. (1992). "General splicing factors SF2 and SC35 have equivalent activities in vitro, and both affect alternative 5' and 3' splice site selection." Pro Natl Acad Sci **89**: 11224-11228.

Ge, H., Manley, J.,L. (1990). "A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro." Cell **62**: 25-34.

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

Golovkin, M., Reddy, A.,S.,N. (1999). "An SC35-like protein and a novel serine/arginine-rich protein interact with Arabidopsis U1-70K protein." Journal of Biological Chemistry **274**: 36428-36438.

Graveley, B., R. (2000). "Sorting out the complexity of SR protein functions." RNA **6**: 1197-1211.

Graveley, B., R., Hertel KJ, Maniatis T (1998). "A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers." EMBO J **17**: 6747-6756.

Graveley BR, M. T. (1998). "Arginine/Serine-Rich Domains of SR Proteins Can Function as Activators of Pre-mRNA Splicing." Molecular Cell **1**: 765-771.

Gui, J., F., Lane, W.,S., Fu, X.,D. (1994). "A serine kinase regulates intrcellular localization of splicing factors in the cell cycle." Nature **369**: 678-682.

Haas, B., J., Delcher, A.,L., Mount, S.,M., Wortman, J.,R., Smith, R.,K., Jr, Hannick, L.,I., Maiti, R., Ronning, C.,M., Rusch, D.,B., Town, C.,D., Salzberg, S.,L., White, O. (2003). "Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies." Nucleic Acids Research **31**: 2141-2144.

Hori, K., Watanabe, Y. (2005). "UPF3 suppresses aberrant spliced mRNA in Arabidopsis." Plant Journal **43**(530-540).

Iida, K., Mitiko, G. (2006). "Survey of conserved alternative splicing events of mRNAs encoding SR proteins in land plants." Mol Biol Evol **23**(5): 1085-1094.

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

Johnson, J., M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P.,M., Armour, C.,D., Santos, R., Schadt, E.,E., Stoughton, R., Shoemaker, D.,D. (2003). "Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays." Science **302**: 2141-2144.

Jordan, T., Schornack, S., Lahaye, T. (2002). "Alternative splicing of transcripts encoding Toll-like plant resistance proteins - what's the functional relevance to innate immunity?" Trends Plant Sci **7**: 392-398.

Jumaa H, W. G., Nielsen PJ (1999). "Blastocyst formation is blocked in mouse embryos lacking the splicing factor SRp20." Curr Biol **9**: 899-902.

Jurica, M., S., Moore, M.J. (2003). "Pre-mRNA splicing: awash in a sea of proteins." Mol. Cell(12).

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

Kalyna, M., Lopato, S., Barta, A. (2003). "Ectopic expression of atRSZ33 reveals its function in splicing and causes pleiotropic changes in development." Mol Biol Cell **Sep**(14(9)): 3565-3577.

Kalyna, M., Lopato, S., Voronin, V., Barta, A. (2006). "Evolutionary conservation and regulation of particular alternative splicing events in plant SR proteins." Nucleic Acids Research **September**(34(16)): 4395-4405.

Kan, J., L., Green, M.,R. (1999). "Pre-mRNA splicing of IgM exons M1 and M2 is directed by a juxtaposed splicing enhancer and inhibitor. ." Genes Development **13**: 462-471.

Kertesz, S., Kerenyi, Z., Merai, Z., Bartos, I., Palfy, T., Barta, E., Silhavy, D. (2006). "Both introns and long 3'-UTRs operate as cis-acting elements to trigger nonsense-mediated decay in plants." Nucleic Acids Research **34**: 480-489.

Kohtz, J., D., Jamison, S.,F., Will, C.,L., Zuo, P., Lu¨ hrmann, R., Garcia-Blanco, M.,A., and Manley, J.,L. (1994). "Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors." Nature **368**: 119-124.

Krainer, A., R., Conway, G.,C., Kozak, D. (1990). "The essential pre-mRNAsplicing factor SF2 influences 5' splice site selection by activating proximal sites." Cell **62**: 35-42.

Kraus, M., E., Lis, J.,T. (1994). "The concentration of B52, an essential splicing factor and regulator of splice site choice in vitro, is critical for Drosophila development." Mol Biol Cell **14**: 5360-5370.

Labourier, E., Bourbon, H.,M., Gallouzi, I.,E., Fostier, M., Allemand, E., Tazi, J. (1999). "Antagonism between RSF1 and SR proteins for both splice-site recognition in vitro and Drosophila development." Genes and Development **13**: 740-753.

Lamond, A., I., Spector, D,.L. (2003). "Nuclear speckles: a model for nuclear organelles." Nat. Rev. mol. Cell Biol. **4**: 605-612.

Lazar, G., Goodman, H.,M. (1995). "Identification of a plant serine-argenine-rich proteins similar to the mammalian splicing factor SF2/ASF." Pro Natl Acad Sci **92**: 7672-7676.

Lazar, G., Goodman, H.,M. (2000). "The Arabidopsis splicing factor SR1 is regulated by alternative splicing." Plant Mol. Biol. **42**: 571-581.

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

Le Hir, H., Izaurralde, E., Maquat, L.,E., Moore. M.,J. (2000). "The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions." EMBO J **19**: 6860-6869.

Lemaire, R., Prasad, J., Kashima, T., Gustafson, J., Manley, J.,L., Lafyatis, R. (2002). "Stability of a PKCI-1-related mRNA is controlled by the splicing factor ASF/SF2: a novel function for SR proteins." Genes Development **16**: 594-607.

Liang, H., Norton, P.,A. (2005). "Overexpression of SR proteins and splice variants modulates chondrogenesis." Experimental Cell Research **Sept 1**: 9.

Lin, Y., Schiefelbein, J. (2001). "Embryonic control of epidermal cell patterning in the root and hypocotyl of Arabidopsis." Development **128**(19): 3697-3705.

Liu, H., X., Chew, S., L., Cartegni, L., Zhang, M., Krainer, A.,R. (2000). "Exonic splicing enhancer motif recognized by human SC35 under splicing conditions. ." Mol. Cell. Biol. **20**: 1063-1071.

Liu, H., X., Zhang, M., Krainer, A.,R. (1998). "Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins." Genes Development **12**: 1998-2012.

Liu, Z., Meyerowitz, E.,M. (1995). "LEUNIG regulates AGAMOUS expression in Arabidopsis flowers." Development **121**: 975-991.

Longman, D., Johnstone, I.,L., Caceres, J.,F. (2000). "Functional characterization of SR and SR-related genes in Caenorhabditis elegans." EMBO J. **19**: 1625-1637.

Lopato, S., Barta, A. (2002). "Network of interactions of a novel plant-specific Arg/Ser-rich protein, atRSZ33, with atSC35-like splicing factors." Journal of Biological Chemistry **277**: 39989-39998.

Lopato, S., Gattoni, R., Fabini, G., Stevenin, J., Barta, A. (1999). "A novel family of plant splicing factors with a Zn knuckle motif: examination of RNA binding and splicing activities." Plant Mol. Biol. **39**: 761-773.

Lopato, S., Kalyna, M., Dorner, S., Kobayashi, R., Krainer, A.,R., Barta, A. (1999). "atSRp30, one of two SF2/ASF-like proteins from Arabidopsis thaliana, regulates splicing of specific plant genes." Genes Development **8**(April 15;13): 987-1001.

Lopato, S., Mayeda, A., Krainer, A.,R., Barta, A. (1996). "Pre-mRNA splicing in plants: characterization of Ser/Arg splicing factors." Pro Natl Acad Sci **93**: 3074-3079.

Lopato, S., Mayeda, A., Krainer, A.,R., Barta, A. (1996a). "Pre-mRNA splicing in plants: characterization of Ser/Arg splicing factors." Pro Natl Acad Sci **93**: 3074-3079.

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

Lorkovic, Z., J., Lopato, S., Pexa., M., Lehner, R., Barta, A. (2004). "Interactions of Arabidopsis RS domain containing cyclophilins with SR proteins and U1 and U11 snRNP-specific proteins suggest their involvement in pre-mRNA splicing." Journal of Biological Chemistry **279**: 33890-33898.

Lorkovic, Z., J., Wieczorek, Kirk, D.,A., Lambermon, M.,H., Filipowicz, W. (2000). "Pre-mRNA splicing in higher plants." Trends Plant Sci **5**: 160-167.

Lorson, C., L., Androphy, E.,J. (2000). "An exonic enhancer is required for inclusion of an essential exon in the SMA-determining gene SMN." Hum Mol Genet **9**: 259-265.

Macknight, R., Duroux, M., Laurie, R., Dijkwel, P., Simpson, G., Dean, C. (2002). "Functional significance of the alternative transcript processing of the Arabidopsis floral promoter FCA." Plant Cell **14**: 877-888.

Maniatis, T., Reed, R. (2002). "An extensive network of coupling among gene expression machines." Nature **416**: 499-506.

Manley, J., L., Tacke, R. (1996). "SR proteins and splicing control." Genes Development **10**: 1569-1579.

Metzstein, M., M., Krasnow, M.,A. (2006). "Functions of the nonsense-mediated mRNA decay pathway in Drosophila development." PLoS Genet. **2**: e180.

Misteli, T., Caceres, J.,F., Clement, J.,Q., Krainer, A.,R., Wilkinson, M.,F., Spector, D.,L. (1998b). "Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo." J Cell Biol **143**: 297-307.

Misteli, T., Caceres, J.,F., Spector, D.,L. (1997). "The dynamics of a pre-mRNA splicing factor in living cells." Nature **387**: 523-527.

Misteli, T., Spector, D.,L. (1998a). "The cellular orgainzation of gene expression." Curr Opin Cell Biol **10**: 323-331.

Misteli, T., Spector, D.,L. (1999). "RNA polymerase II targets pre-mRNA splicing factors to transportation sites in vivo." Mol Cell **3**: 697-705.

Mohonen, A. P., Helariutta, Y., et al (2000). "A novel two-component hybrid molecule regulates vascular morphogenesis of the Arabidopsis root." Genes Development **14**(23): 2938-2943.

Moore, M., J., Query, C.,C., Sharp, P.,A. (1993). The RNA World. Plainview, NY, Cold Spring Harbor Lab., Press.

Mount, S., M., Benoit, J., Choi, Y., Dye, N., Pan, A. (in preparation). "Exonic Splicing Enhaner activity in Arabidopsis thaliana."

Nilsen, T., W. (1994). "RNA-RNA interactions in the spliceosome: unraveling the ties that bind." Cell **78**: 1-4.

Nishiyama, T., Fujita, T., Shin-I, T., Seki, M., Nishide, H., Uchiyama, I., Kamiya, A., Carninci, P., Hayashizaki, Y., Shinozaki, K., Kohara, Y., Hasebe, M. (2003). "Comparative genomics of Physcomitrella patens gametophytic transcriptome and Arabidopsis thaliana: implication for land plant evolution." Proc. Natl. Acad. Sci **100**: 8007-8012.

Palusa, S., G., Ali, G.,S., Reddy, A.,S. (2007). "Alternative splicing of pre-mRNAs of Arabidopsis serine/arginine-rich proteins: regulation by hormones and stresses." Plant Journal **49(6)**(March): 1091-1107.

Peng, X., Mount, S.,M. (1995). "Genetic enhancement of RNA-processing defects by a dominant mutation in B52, the Drosophila gene for an SR protein splicing factor." Mol Cell Biol **15**: 6273-6282.

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

Quesada, V., Macknight, R., Dean, C., Simpson, G.,G. (2003). "Autoregulation of FCA pre-mRNA processing controls Arabidopsis flowering time." EMBO J **22**: 3142-3152.

Rappsilber, J., Ryder, U., Lamond, A.,I., Mann, M. (2002). "Large-scale proteomic analysis of the human spliceosome." Genome Res. **12**: 1231-1245.

Reddy, A., S. (2004). "Plant serine/arginine-rich proteins and their role in pre-mRNA splicing." Trends Plant Sci **9**: 541-547.

Reddy, A., S.,N. (2001). "Nuclear pre-mRNA splicing in plants." CRC Crit. Rev. Plant Sci.(20): 523-572.

Reddy, A., S.,N. (2004). "Plant serine/arginine-rich proteins and their role in pre-mRNA splicing." Trends Plant Sci **9**(11): 1360-1385.

Reed, R. (1996). "Initial splice-site recognition and pairing during pre-mRNA splicing." Curr Opin Genet Dev **6**: 215-220.

Reichert, V., L., Le, Hir, H., Jurica, M.,S., Moore, M.,J. (2002). "5' exon interactions within the human spliceosome establish a framework for exon junction complex structure and assembly." Genes Development **16**: 2778-2791.

Rian, H., Kole, R. (1995). "Selection of novel exon recognition elements from a pool of random sequences." Mol. Cell Biol. **15**: 6291-6298.

Ring HZ, L. J. (1994). "The SR protein B52/SRp55 is essential for Drosophila development." Mol Cell Biol **14**: 7499-7506.

SALK (2007). SIGnAL "T-DNA Express" Arabidopsis Gene Mapping Tool.

Sanford J., R., Gray, N.,K., Beckmann, K., Caceres, J.,F. (2004). "A novel role for shuttling SR proteins in mRNA translation." Genes Development **18**: 755-768.

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

Schaal, T., D., Maniatis, T. (1999). "Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences." Mol. Cell Biol. **19**: 1705-1719.

Shi, H., Hoffman, B.,E., Lis, J.,T. (1997). "A specific RNA hairpin loop structure binds the RNA recognition motifs of the Drosophila SR protein B52." Mol Cell Biol **17**: 2649-2657.

Shin, C., Feng ,Y., Manley, J.,L. (2004). "Dephosphorylated SRp38 acts as a splicing repressor in response to heat shock." Nature **427**: 553-558.

Simpson, G., G., Dijkwel, P.,P., Quesada, V., Henderson, I., Dean, C. (2003). "FY is an RNA 3' end-processing factor that interacts with FCA to control the Arabidopsis floral transition." Cell **113**: 777-787.

Tacke, R., Manley, J.,L (1995). "The human splicing factors ASF/SF2 and SC35 possess different, functionally significant RNA binding specificities." EMBO J **14**: 3540-3551.

Tian, M., Maniatis, T. (1994). "A splicing enhancer exhibits both constitutive and regulated activities." Genes Development **8**: 1703-1712.

Tuerk, C., Gold, L. (1990). "Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. ." Science **249**: 505-510.

Wang, B., B., Brendel, V. (2004). "The ASRG database: identification and survey of Arabidopsis thaliana genes involved in pre-mRNA splicing." Genome Biology **5**(R): 102.

Wang, H., Y., Lin, W., Dyck, J.,A., Yeakley, J.,M., Songyang, Z., Cantley, L.,C., Fu, X.,D. (1998). "SRPK2: a differentially expressed SR protein-specific kinase involved in mediating the interaction and localization of pre-mRNA splicing factors in mammalian cells." J Cell Biol **140**: 737-750.

Wang, H., Y., Xu, X., Ding, J.,H., Bermingham, J.,R .Jr, Fu, X.,D. (2001). "SC35 plays a role in T cell development and alternative splicing of CD45." Mol Cell **7**: 331-342.

Wang J, T. Y., Manley JL (1996). "Targeted disruption of an essential vertebrate gene: ASF/SF2 is required for cell viability." Mol Cell **10**: 2588-2599.

Wang, Z., Rolish, M.,E., Yeo, G., Tung, V., Mawson, M., Burge, C.,B. (2004). "Systematic identification and analysis of exonic splicing silencers." Cell **Dec 17;**(119(6)): 831-845.

Wu JY, M. T. (1993). "Speific interactions between proteins implicated in splice site selection and regulated alternative splicing." Cell **75**: 1061-1070.

Wysocka-Diller., J. W., Benfey, P.N., et al (2000). "Molecular analysis of SCARECROW function reveals a radial patterning mechanism common to root and shoot." Development **127**(3): 595-603.

Xiao, S.-H., Manley, J.,L. (1997). "Phosphorylation of the ASF/SF2 RS domain affects both protein-protein interactions and is necessary for splicing." Genes and Development **11**: 334-344.

Xiao, S.-H., Manley, J.,L. (1998). "Phosphorylation-dephosphorylation differentially affects activities of splicing factor ASF/SF2." EMBO J **17**: 6359-6367.

Xu, X., Yang, D., Ding, J.,H., Wang, W., Chu, P.,H., Dalton, N.,D., Wang, H.,Y., Bermingham, J.,R., Jr, Ye, Z., Liu, F., Rosenfeld, M.,G., Manley, J.,L., Ross, J., Jr, Chen, J., Xiao, R.,P., Cheng, H., Fu, X.,D. (2005). "ASF/SF2-regulated CaMKIIdelta alternative splicing temporally reprograms excitation-contraction coupling in cardiac muscle." Cell **120**: 59-72.

Zahler, A., M., Lane, W.,S., Stolk, J.,A., Roth, M.,B. (1992). "SR proteins: a conserved family of pre-mRNA splicing factors." Genes Development **6**: 837-847.

Zhang, X., C., Gassmann, W. (2003). "RPS4-mediated disease resistance requires the combined presence of RPS4 transcripts with full-length and truncated open reading frames." Plant Cell **15**: 2333-2342.

Zhou, Z., Licklider, L.,J., Gygi, S.,P., Reed, R. (2002). "Comprehensive proteomic analysis of the human spliceosome." Nature **419**: 182-185.