SECOND-STEP SPLICING FACTORS AND HEAVY METAL STRESS MANAGEMENT IN *ARABIDOPSIS THALIANA*

BY

AMY SARAH GLEKAS

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Biochemistry
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Doctoral Committee:

Professor Mary A. Schuler, Chair, Director of Research
Professor Susan A. Martinis
Professor Yi Lu
Associate Professor Michel Bellini

ABSTRACT

Pre-mRNA splicing in plants is mechanistically the same as splicing in other eukaryotes, but its mode of intron recognition is unique. Whereas in animals a pyrimidine-tract contributes to the definition of exons, in plants, introns are defined by the boundaries of AU-rich intron and GC-rich exon. Some of these differences may depend on splicing factors mediating recognition of the unusual sequence-dependent transition points and/or factors mediating the first- and second-steps in intron definition and excision.

The second transesterification step of splicing involves many protein factors (second-step splicing factors) that have not been previously characterized in *Arabidopsis*. Among these are numerous genes encoded by multicopy genes in this model plant: PRP16 (1 copy), PRP17 (2 copies), PRP18 (2 copies), PRP22 (3 copies) and SLU7 (3 copies). This work is aimed at defining the structural differences and expression patterns of these multiple second-step splicing factors and their involvements in heavy metal stress response.

The first goal of this project was to determine unique characteristics of the multicopy second-step splicing factors and the tissue and developmental stages in which each of these genes is expressed. My studies have indicated that all of these second-step splicing factors genes are expressed in *Arabidopsis*. While most are constitutively expressed throughout each tissue type and developmental stage, a specialized subset including PRP17-2, PRP18B, PRP22-3 and SLU7-2 are primarily expressed in flower and silique tissue in four-week-old plants and throughout the entire seven-week-old plant. Each of these specialized genes has unique structural features when compared to their homologs, suggesting that they form unique spliceosomal networks. These include variations in sequence identity between each other and their homologs in other organisms. Additionally, homology modeling revealed specific sites in which changes in

residues will likely contribute to their interactions with other proteins. For example, residues important for protein-protein interaction in *Sc*PRP17 are unique on *At*PRP17-2. Basic surface residues on *Sc*PRP18 that contribute to interaction with *Sc*SLU7 are in slightly different locations on both *At*PRP18A and *At*PRP18B and fewer in number on *At*PRP18B.

The second goal of this project was to determine the response of second-step splicing factors to heavy metal stress conditions ($HgCl_2$, $Hg(OAc)_2$, $CdSO_4$, $CuSO_4$ and $ZnSO_4$). My studies have shown that plants subjected to increasing concentrations of $Hg(OAc)_2$ and $CdSO_4$ for three weeks from germination accumulate pre-mRNA transcripts of genes not subject to alternative splicing. This effect is observed for some of the second-step splicing factors and some genes that are involved in other cellular processes like plant defense and transcription regulation. In contrast, pre-mRNA transcripts of genes subject to alternative splicing, such as the Ser/Arg-rich (SR) proteins involved in intron recognition, accumulate varied proportions of alternatively spliced transcripts but not pre-mRNA transcripts.

To determine whether translation-dependent nonsense-mediated decay (NMD) was involved in the accumulation of pre-mRNAs, the effects of chemically inactivating translation were examined. My studies have shown that the patterns of second-step splicing factor transcript accumulation observed when NMD is chemically knocked out most resemble the patterns of transcript accumulation when plants are treated with cadmium. The fact that these patterns do not match exactly suggests that while the metals may affect NMD, they do not abolish it in the same manner as chemically abolishing it by halting translation and other effects are independent of NMD. It was also determined that many of the second-step splicing factor expression levels are regulated by NMD under normal conditions and the loss of NMD causes an accumulation of these transcripts.

To my family
past, present and future

ACKNOWLEDGEMENTS

I would like to deeply thank my advisor, Professor Mary A. Schuler. She is a brilliant scientist and a wonderful mentor; patient, nurturing, understanding and encouraging. She has supported me and taught me so much. I know that if I ever find myself managing a group of scientists, I will model myself after her as much as I can. It is not easy to take a chance on a student who is moving from one lab to another, and I am eternally grateful that Mary accepted me into her lab. I also would like to thank the members of my committee for their support and critiques throughout the years. I have benefited from having each of these people in my scientific life. Professor Susan Martinis has been a guide and supporter for me since I joined the department of Biochemistry. She gave me the opportunity to instruct a class, an invaluable experience that I could not have matched with being a teaching assistant. Professor Michel Bellini taught me many useful techniques in cell biology very early in my graduate career and has been an advisor and confidante since. He was a strong advocate for me in my early time in the department and as a new instructor of an upper level course. Professor Yi Lu gave me founded and helpful criticism and inspired me to work toward understanding material that I had found difficult. My short time as a rotation student in his lab gave me perspective that I have used in my approach to problems since. I would also like to thank my former advisor, Lyne Levesque, who fostered my love for the entire life and death of mRNA.

The past and present members of the Schuler lab have been of great support to me. I learned a great deal from all of the previous graduate students and postdocs and I would like to thank them all. A few have been my closest colleagues and biggest teachers: Hui Duan, Sanjeewa Rupasinghe, Iyoncy Rodrigo, Kara Sanford, Metin Bilgin, Devi Annamalai, Cynthia

McDonnell and Yinghong Pan. I have also had the pleasure of working with many bright and promising undergraduates and high school students. They are as impressive as one could imagine and they inspire me every day.

An added bonus of graduate school was learning how to teach. Of the many people that I interacted with as a teaching assistant, two truly taught me what it means to be a devoted, caring and compassionate teacher: Melissa Reedy and Brad Mehrtens. These two people run the massive class that is MCB 150, and I am so grateful that I was able to be a small part of this course for so many semesters. I learned what it means to truly care about young students and how important it is to make sure that they are getting the absolute best education possible. When I went on to instruct an upper-level course, Melissa coordinated it for me. This was a massive sacrifice on her part, but I would never have been able to succeed without her. I am so lucky that she was my colleague and remains my friend.

My family has been extremely supportive thought my entire life, and my time in graduate school was no exception. I want to thank my parents, Dr. Jeffrey and Sharon Dunlap, and my sister and brother-in-law, Stephanie and Chad Apple. I love them very much; but I think that I am most thankful for the fact that they never asked when I would finish! I have been fortunate to know all four of my grandparents and three of my great-grandparents. My two grandmothers, Elizabeth and Hazel are still here, and I love that every time we talk, they ask about school. I also want to thank my newest family: Dimitrios, Evdokia, Peter, Jen and Eleni Glekas. My in-laws are wonderful people and I am so happy that I have them in my lives. I am especially lucky that two of my best friends in the world are my sisters, Stephanie and Eleni.

Lastly, I have made the most wonderful friends here in Illinois. I cannot list them all, but I have had the pleasure of playing cards, watching football, knitting, curling, fantasy football

drafting, bar-b-qing, drive-in movie attending, cookie exchanging and going out with the best friends that I have ever had. But the greatest friend that I made here at University of Illinois is George Glekas. We met, dated, fell in love, got married and had our beautiful son, Dimitrios, all while I was a student in the Schuler lab. I have never met a person who made me happier and I never knew how wonderful it would be to share my life with a fellow scientist who is so kind and giving. I am eternally grateful for the love and support of my perfect husband, George.

TABLE OF CONTENTS

CHAPTER 1

Introduction

1.1 Pre-mRNA splicing

Virtually all eukaryotes contain primary transcripts that are composed of exons and intervening introns. As it is coded, messenger RNA (mRNA) is a precursor to the mature RNA that will be translated. *Cis*-splicing, or the removal of the non-coding introns and subsequent joining of the coding exons, is paramount to proper gene expression in eukaryotes. Splicing is carried out in the nucleus by the spliceosome, large ribonucleoprotein (RNP) complexes, in two steps (Padgett et al., 1984; Ruskin et al., 1984). The mRNA must also be capped by a guanine which is methylated at position 7 at the 5' end of the transcript and a polyadenosine (poly (A)) tail is added to the 3' end for the mRNA to be mature (reviewed in Shatkin and Manley, 2000). Export of the mature mRNA out of the nucleus for translation is mediated in part by a second complex of proteins called the exon junction complex (EJC). The EJC is present at the exon-exon junction and is important in RNA surveillance. Aberrant splicing may result in a premature stop codon (PTC) prior to an EJC indicating that the gene was improperly spliced and would result in the production of a truncated protein. RNA that fails to be completely processed usually undergoes a form of RNA surveillance, or RNA degradation (reviewed in Bono and Gehring, 2011).

The number and length of introns in a gene can vary. In an example gene transcript (Figure 1.1), there is an untranslated region (UTR) at the 5' end of the gene. This is followed by the start codon (AUG) and coding sequence. All of this described region would be considered exon 1 as UTRs can also contain introns and exons. Following exon 1 is the sole intron in this model. This is the region that will be spliced out of the gene. Following the intron is exon 2, the

second and final coding region in this example. Included in this region are the stop codon and a second stretch of untranslated RNA, the 3' UTR. The intergenic region is found on either side of the UTRs on the chromosomal DNA and is not transcribed (not shown).

The removal of the non-coding intron and joining of the surrounding exons requires to chemical steps. Both of the chemical steps of the splicing reaction involve transesterifications (Figure 1.2) (Padgett et al., 1984; Ruskin et al., 1984). The first step occurs between the 5' end of the intron and a conserved adenosine inside of the intron (Figure 1.3). The hydroxyl group of the conserved adenosine attacks the phosphodiester bond between the exon and the intron of the 5' splice site. The result is a free 3' hydroxyl group on the first (5') exon and a lariat structure in the intron. This is an unusual phosphodiester bond between the 2' hydroxyl of the adenosine and 5' phosphate group at the 5' splice site. The second step occurs between the 3' end of the intron and the free hydroxyl group at the former 5' end of the intron, joining of the two surrounding exons together (Figure 1.4). The free 3' hydroxyl group of exon 1 attaches the phosphodiester bond between the intron and exon of the 3' splice site. The phosphate group at the 5' end of exon 2 is now bound to the 3' end of exon 1 in a typical phosphodiester bond. The lariat structure contains a free hydroxyl group at its 3' end and is released and targeted for degradation (Maschhoff and Padgett, 1992; Moore and Sharp, 1992).

The composition of the spliceosome is dynamic and the assembly of proteins and RNAs is ordered allowing for the factors involved in each step to differ. The major spliceosome (the U2 spliceosome) is comprised of 5 small nuclear RNPs (snRNPs): U1, U2, U4, U5 and U6 along with non-snRNP protein factors. snRNPs consist of the uridine-rich small nuclear RNA, U1, U2, U4/U6 or U5, each of which are complementary to conserved sequences in and around the intron, and associated proteins. A minor spliceosomal complex (the U12 complex) splices a

minority of introns and is composed of U11, U12, U4atac, U5 and U6atac snRNPs. The functions of the snRNPs found in the minor spliceosome are analogous to those of the major spliceosome (*e.g.* U11 functions as U1, U12 as U2, etc.) and U5 is shared between both spliceosome types (Tarn and Steitz, 1997).

In the major spliceosome, the U1 snRNP is complementary to a conserved sequence at the 5' splice site (Lerner et al., 1980; Rogers and Wall, 1980) and the U2 snRNP is associated with the branchpoint region of the intron to expose it for use as a nucleophile in the first transesterification step of splicing forming a pre-catalytic complex (Parker et al., 1987; Query et al., 1994; Ritchie et al., 2009). The U4 and U6 snRNAs are arranged into a single complex. They interact with the U5 snRNP to form the U4/U6·U5 tri-snRNP. The tri-snRNP binds with the complex resulting in a rearrangement (Cheng and Abelson, 1987; Konarska and Sharp, 1987; Pikielny et al., 1986). The U5 RNA directly interacts with the conserved sequence at the 3' splice site of the intron (Umen and Guthrie, 1995). U1 and U4 snRNAs are released from the spliceosome and the U2 and U6 snRNAs form base pairs with each other (Cheng and Abelson, 1987; Lamond et al., 1988; Pikielny et al., 1986) (Figure 1.5). With the assistance of other protein factors within the spliceosome, splicing is catalyzed; the intron is removed and the two surrounding exons are joined together and released from the spliceosome.

There are over 300 factors in the spliceosome. The snRNPs are a small fraction of the molecules entwined in this process. This study focuses on the effects on the SR (serine/arginine-rich) proteins and the second step splicing factors involved in the spliceosome. Many of these second step splicing factors are designated as PRP (precursor RNA processing) proteins, yet the PRPs are diverse in their function and step in involvement in splicing; PRP8 is really a snRNP-associated protein (Pinto and Steitz, 1989).

The chemistry of splicing is the same in plants as that in other eukaryotes. Beyond the mechanism, many variations to splicing, including regulation, protein involvement, organelles involved, etc., occur in plants. This work is focused on the study of splicing in the diploid model organism *Arabidopsis thaliana*. *Arabidopsis* has five chromosomes which were sequenced in 2000 (Arabidopsis Genome Initiative, 2000). Work on *Arabidopsis* splicing not only has implication for the basic understanding of splicing, but is relatable to many agricultural issues including production of genetically modified crops, disease resistance, pollution and pest management.

1.2 Intron features

While the mechanisms of intron splicing between eukaryotes are similar, the frequency and size of introns vary between organisms (Table 1.1). In *Saccharomyces cerevisiae*, 3% of the genes contain an intron and of these genes, almost all of them contain a single intron that is approximately 300 base pairs (Lopez and Séraphin, 1999). Multiple introns are present in most human genes, and the size of those introns can vary from hundreds to thousands of base pairs (Lander et al., 2001; Venter et al., 2001). In *Arabidopsis* genes, introns are present in 80-85% of the genes and range in size from less than 100 base pairs to over 1000 base pairs (Brown et al., 1996; Lorković et al., 2000). *Arabidopsis* genes also tend to carry a varied amount of introns. The majority of *Arabidopsis* genes have 1-3 introns, but a single gene can have over 40 introns. As expected, nearly all mature transcripts in human and *Arabidopsis* genes are spliced, but it is also important to note that in yeast, the 3% of intron-containing genes account for 27% of the total transcripts (Lopez and Séraphin, 1999).

There are surprisingly few characteristics at the intron that render the splice site recognizable (Table 1.1). The 5' splice site maintains the conserved sequence AG/gu (capital

designates exonic nucleotides, slashes designate splice sites, lowercase designates intron nucleotides, underlines designate highly conserved nucleotides) across eukaryotes for splice sites recognized by the major spliceosome. Directly downstream from the 5' splice site is another conserved sequence. In plants and mammals, the sequence is aagu and in yeast, the sequence is augu (Schuler, 2008). The conserved sequence at the 3' splice site is ag/G for introns spliced by the major spliceosome. Immediately upstream of the 3' splice site is a pyrimidine in mammals, a cytosine in yeast or gc in plants (Schuler, 2008). More recently, a few different splice site sequences have been found to be recognized by the major spliceosome. The additional sequences include /gc…ag/ and /au…ac/, yet no /au…ac/ introns were identified (Sheth et al., 2006). The minor spliceosome recognizes a separate subset of introns, those with either a /gu…ac/ or a /gu…ag/ intron junction sequence (Levine and Durbin, 2001; Sheth et al., 2006; Zhu and Brendel, 2003).

In addition to the two splice site sequences that are recognized by the spliceosome, the branchpoint region includes the absolutely conserved adenosine involved in lariat formation and the less strictly conserved surrounding nucleotides. In mammals, the consensus branchpoint sequence is YNYURAC (Nelson and Green, 1989), in yeast it is UACUAAC (Domdey et al., 1984; Rodriguez et al., 1984) and in plants it is CURAY (Brown, 1996; Liu and Filipowicz, 1996; Simpson et al., 1996) (the branchpoint adenosine is underlined, Figure 1.6). The location of the branchpoint adenosine can vary between 19 and 50 nucleotides upstream of the 3' splice site (Schuler, 2008).

A final spliceosomal recognition site is the polypyrimide tract (Py-tract), a stretch of pyrimidines, is located between the branchpoint and the 3' splice site. The Py-tract is recognized by either the U2 snRNP heterodimer U2AF or a subunit of the heterodimer in *S. cerevisiae* and

humans (Abovich et al., 1994; Zamore et al., 1992). Yet the necessity of the of Py-tract is splice site recognition is unclear, as the U2AF recognition of the region between the branchpoint and the 3' splice site does not always depend upon the presence of a strong Py-tract (Sridharan and Singh, 2007). Plant introns, however, have a notable lack of a Py-tract (Wiebauer et al., 1988). Plants have a uridine-rich tract in which adenosines can be substituted (AU-tract) within the intron (Baynton et al., 1996). Interfering with the length of the AU-tract can cause recognition of alternative 5' and 3' splice sites (McCullough and Schuler, 1993; Merritt et al., 1997). The boundary between the AU-rich intron and the relatively AG-rich exon are important for plant splice site recognition (McCullough et al., 1993).

1.3 Plant splicing

While some similarities between plant splicing and splicing in other eukaryotes are evident, the differences are significant. Previously described differences in intron features likely cause mammalian genes expressed in transgenic plants to be processed incorrectly (Barta et al., 1986; van Santen and Spritz, 1987; Wiebauer et al., 1988). Conversely, plant introns can be processed in animal systems (Brown et al., 1986; Hartmuth and Barta, 1986; van Santen and Spritz, 1987; Wiebauer et al., 1988). The Py-tract/AU-tract found in animals or plants, respectively, is likely the reason for this incongruity. The animal spliceosome is able to correctly identify the plant splice sites when a Py-tract is not present (but and AU-tract is) but the plant spliceosome cannot correctly identify the splice sites of an AU-tract-lacking animal gene (McCullough and Schuler, 1993).

The contrast between the AU-rich intron and the exon of the plant gene has led to the theory of the intron definition and intron boundary recognition by the plant spliceosome (Belostotsky and Rose, 2005; McCullough et al., 1993). This is a marked contrast from the

widely accepted idea of exon definition in animals that lead to correct identification of splice sites and introns (Belostotsky and Rose, 2005; Berget, 1995). This contrast suggests a major difference in the fundamentals of intron recognition between plants and other eukaryotes and may be a clue as to why the animal spliceosome can recognize plant introns but not vice versa.

Because major differences in intron recognition have been identified in plants versus other eukaryotes, it is no surprise that the factors that make up the spliceosomes also vary. Seventy-four genes encoding snRNAs and 395 genes encoding homologs to animal splicing-related proteins have been identified in the *Arabidopsis* genome (Wang and Brendel, 2004). 70 of the genes encoding snRNAs are thought to be associated with the major spliceosome while only 4 of those genes are thought to encode minor spliceosomal snRNAs (Wang and Brendel, 2004). The 74 identified snRNAs are spread out over the 5 *Arabidopsis* chromosomes and are generally not adjacent to each other in contrast to the gene repeats found in the hundreds of human U1, U2 and U4 snRNAs (Bark et al., 1986; Lund and Dahlberg, 1984; Van Arsdell and Weiner, 1984; Wang and Brendel, 2004).

1.4 Alternative splicing in plants

Alternative splicing, a mechanism in which alternative mature mRNAs can be produced from a single gene, adds an additional layer of gene regulation in eukaryotes and has been examined through several different methods in plants. Both individual gene examination and bioinformatics tools lend information on the study of different types and outcomes of alternative splicing in *Arabidopsis*. Alternative splicing can be achieved through intron retention, exon skipping, alternative 5' and 3' splice sites and any combination of the above (Figure 1.7).

The human genome has between 20,000 and 30,000 genes, a number that was initially thought to be quite low (Lander et al., 2001; Venter et al., 2001). It is now known that the

number of different mature mRNA transcripts outnumber the number of genes by as much as eight-fold as a result of alternative splicing. Based on exon junction microarrays of the human genome, 74% of human genes are predicted to be alternatively spliced (Johnson et al., 2003). *Arabidopsis* has nearly the same number of genes (25,000) as human (Arabidopsis Genome Initiative, 2000), but only 42% of *Arabidopsis* genes are suspected to undergo alternative splicing as determined with cDNA and EST data (Filichkin et al., 2010). *S. cerevisiae* do not undergo alternative splicing.

The most predominant form of alternative splicing in human is exon skipping whereas the most common form of alternative splicing in *Arabidopsis* is intron retention (Haas, 2008). This is interesting albeit not surprising considering the unique theories for how the splice site is recognized: exon recognition versus intron recognition. Of the alternative splicing events that occur in *Arabidopsis*, the use of an alternative 3' splice site is twice as frequent as the use of an alternative 5' splice site (Wang and Brendel, 2006).

Alternative splicing was initially discovered in *Drosophila melanogaster* developmental genes that were alternatively spliced in sex determination (Baker, 1989). This type of regulation has since been identified in other insects, silkworm and *C. elegans* (Fujii and Shimada, 2007; Kerins et al., 2010; Salz, 2011). Alternative splicing has also been implicated in mammalian and the haploid cucumber sex determination (Guo et al., 2010; Lalli et al., 2003). Sex determination is not the sole developmental purpose for alternative splicing. The Down syndrome cell adhesion molecule (*Dscam*) gene in *Drosophila* has over 38,000 different isoforms as a result of alternative splicing. These protein products are important in the establishing of neural circuits, and at least some of the isoforms are spliced a specific points in development as well as different locations within the organism (Celotto and Graveley, 2001).

In addition to developmental roles, alternative splicing in *Arabidopsis* is implicated in regulation of the jasmonic acid hormone pathways and in circadian responses. Jasmonic acids are involved in plant growth as well as defense and development. The JASMONATE-ZIM DOMAIN (ZIM) proteins regulate this pathway, among others (Gfeller et al., 2010). Three of these JAZ proteins have been confirmed to be alternatively spliced and these alternative splicing events lead to jasmonate-insensitive protein variants or dominant JAZ repressors (Chini et al., 2009; Pauwels and Goossens, 2011; Thines et al., 2007; Yan et al., 2007). Regulation of the circadian clock in *Arabidopsis* is also partially regulated by alternative splicing. Eighteen splice variants have been identified in 5 of the 10 core clock proteins and at least some of these alternative splicing events are a result of the subtle temperature changes that are characteristic of the day-to-night switch and vice versa (James et al., 2012).

Alternative splicing as a means of defense is a valuable tool that may be unique to plants and some invertebrates. Because plants lack an innate immune system, this use of alternative splicing may allow plants to remain viable biotic and abiotic stress conditions such as temperature, light, radiation, osmotic, drought, salt and heavy metals (Ali and Reddy, 2008; Gassmann, 2008; Glekas et al., in preparation).

Conservative predictions suggest that only 42% of *Arabidopsis* genes are alternatively spliced, but the majority of all stress-related genes undergo alternative splicing (Chen et al., 2007; Filichkin et al., 2010). Currently, research of abiotic stress response is limited to the genes that are alternatively spliced and the downstream effects remain to be elucidated. Genes that are alternatively spliced include both specific stress-related genes and that have wider functionality (Ali and Reddy, 2008). Cold stress results in the alternative splicing of β-hydroxyacyl ACP dehydratase, an enzyme involved in fatty acid biosynthesis and involved in temperature stress

response (Tai et al., 2007). Acetyl CoA carboxylase, an enzyme involved in lipid metabolism, utilizes an alternative promoter as a result of heat stress-related alternative splicing (Podkowinski et al., 2003). Transcription factors and splicing factors are also alternatively spliced in temperature stress conditions, events which are duplicated in other abiotic stress conditions (Ali and Reddy, 2008; Egawa et al., 2006; Palusa et al., 2007; Xue and Loveridge, 2004). The effects of heavy metals on both alternative and general splicing have not been examined until this study and these data are important for determination of consequences of environmental contamination.

More progress has been made with plant resistance proteins (R proteins) that are alternatively spliced as an effect from biotic stress or microbial pathogen infection. Some *Arabidopsis* R proteins are alternatively spliced and the ratio of different splice products is important for the defense function (Gassmann, 2008). In particular, some R protein products are alternatively spliced resulting in a truncated protein. This protein must be present at certain levels in combination with the major splice product in order for the stress response to be activated (Dinesh-Kumar and Baker, 2000; Zhang and Gassmann, 2007).

1.5 Splicing-related proteins in plant

In order to understand the intricacies of plant splicing, it is necessary to characterize the *trans*-acting splicing related protein factors that are involved. The *Arabidopsis* genome commonly has multiple loci for genes that are found in single copy in mammalian or yeast genomes. There is no exception for splicing proteins in *Arabidopsis*. The collection of spliceosomal proteins in *Arabidopsis* is quite complex and it is important to examine the function of the proteins and their multiple loci, as their expression levels lead to the notion that they are not likely redundant in function (Glekas and Schuler, in preparation).

Splicing is not limited to two transesterification steps; proteins of the spliceosome can play many roles prior to and after participating in only those two steps. Spliceosomal proteins are involved in splice site recognition, the first and second transesterification steps, fidelity assurance and mRNA export from the nucleus. Some of the spliceosomal proteins fit into certain classes (Sm, SR, PRP, etc.) but these classes can overlap, are general and do not encompass all spliceosomal proteins. And while this list of spliceosomal proteins is not comprehensive, it does highlight many important spliceosomal proteins.

RNA recognition motif (RRM) proteins are a large class of proteins that are not limited to splicing factors. Within the RRM family of proteins are the snRNP proteins, SR proteins and some post splicing proteins such as shuttling proteins. In addition to the RRM proteins, non-RRM proteins, like first- and second-step splicing factors, play a role in the catalyses necessary for splicing. It is clear that intron composition and splice-site recognition in plants are different than those of other eukaryotes. A major focus of this study is to characterize one subset of *Arabidopsis* splicing factors.

1.6 snRNP proteins

SnRNP proteins assist snRNAs in carrying out their previously outlined functions. Ninety-one snRNP-associated proteins have been identified in *Arabidopsis* via alignment data, but few of these proteins have been characterized functionally. Twenty-four of the snRNP protein factors contain an RNA-binding domain called the Sm or Sm-like domain (Achsel et al., 2001). These 24 proteins are homologs of 15 human Sm or Sm-like proteins (Wang and Brendel, 2004), meaning multiple copies of many of the snRNPs are present in *Arabidopsis* as opposed to their human counterparts. For example, there are 22 U1 and U2-spcific proteins in *Arabidopsis* and only 5 in human (Wang and Brendel, 2006). Currently, it is unclear if some of these multiple

loci have compensatory effects, recognize different target or even if all of these genes are expressed.

1.7 SR proteins

SR proteins are a highly conserved, large subclass of RRM proteins that play a role in recognition of splice sites across introns (Reddy, 2004; Wang and Brendel, 2006). These proteins are not considered snRNP-related proteins, yet they do play a role in splice site recognition and recruitment of snRNPs in both constitutive and alternative splicing (Graveley, 2000; Shepard and Hertel, 2009). SR proteins are, in fact, heavily implicated in not only general alternative splicing, but in the alternative splicing of their own transcripts as SR proteins are so commonly themselves alternatively spliced (Reddy, 2004). Arabidopsis possess 19 identified SR proteins, while about half that many are found in the human genome (Barta et al., 2008; Reddy, 2004). Many of these *Arabidopsis* SR proteins are unique to plants alone and do not have homologs in mammalian or yeast organisms.

1.8 Second step splicing factors

Second step splicing factors are among the non-RRM class of proteins, although some of the proteins do bind RNA. Most of the second step splicing factors are designated as PRP proteins, although SLU7 is also considered a second step splicing factor. The functions of second step splicing factors range from 3' splice site recognition to release of the fully spliced and proofread RNA (Schwer, 2008; Schwer and Guthrie, 1991). As with other examples of splieceosomal proteins, there are multiple copies of all but one second step splicing factor, 11 in *Arabidopsis* and 1 in human and yeast (Wang and Brendel, 2006).

Table 1.1 Comparison of intron characteristics between organisms. Capital letters designate exonic nucleotides, slashes designate splice sites, lowercase letters designate intronic nucleotides, underlined letters designate highly conserved nucleotides. Major and minor splice site consensus sequences are recognized by the major or minor spliceosome.

	S. cerevisiae	Human	*Arabidopsis*
Major 5' splice site consensus	AG/guaagu	AG/guaagu	AG/guaugu
Minor 5' splice site consensus	/gu	/gu	/gu
Major 3' splice site consensus	yag/G	cag/G	gcag/G
Minor 3' splice site consensus	ac/ or ag/	ac/ or ag/	ac/ or ag/
Branchpoint consensus	UACUA<u>A</u>C	YNYUR<u>A</u>C	CUR<u>A</u>Y
Approximate length (bp)	300	100s – 1000s	60 – 1000s
Percent of genes	3%	90%	80 – 85%
Number of introns per gene	1	multiple	1 - 40

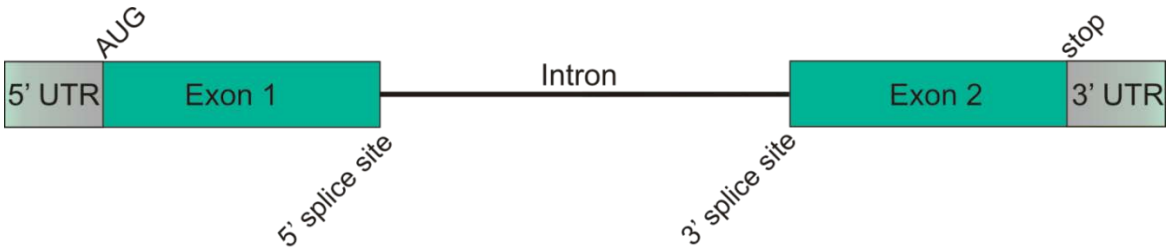


Figure 1.1 Single intron-containing gene model. The 5' and 3' untranslated regions (UTRs) are gray. Exons 1 and 2 are the open boxes and the single intron is a black bar. The start codon is labeled as AUG and the stop codon is indicated. The splice sites, the regions where the exons and introns meet are labeled on either side of the intron.

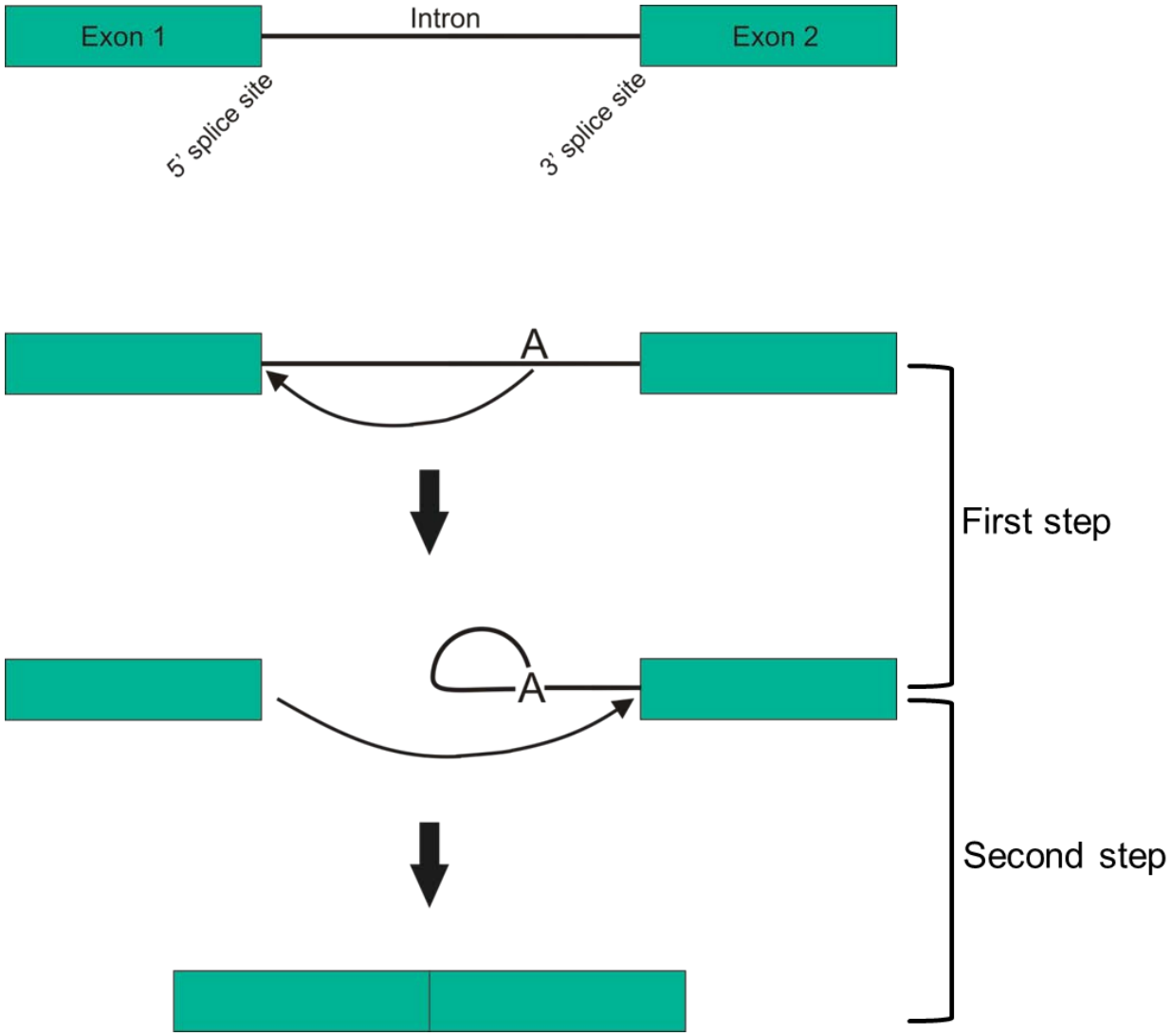


Figure 1.2 Basic steps of intron splicing. Two exons (green boxes) are joined together by the splicing out of a single intron (black bar). The reaction involves two transesterifications, represented by curved arrows. The A represents the conserved adenosine in the intron involved in nucleophilic attack on the 5' splice site and lariat formation. The two separate transesterification steps are bracketed.

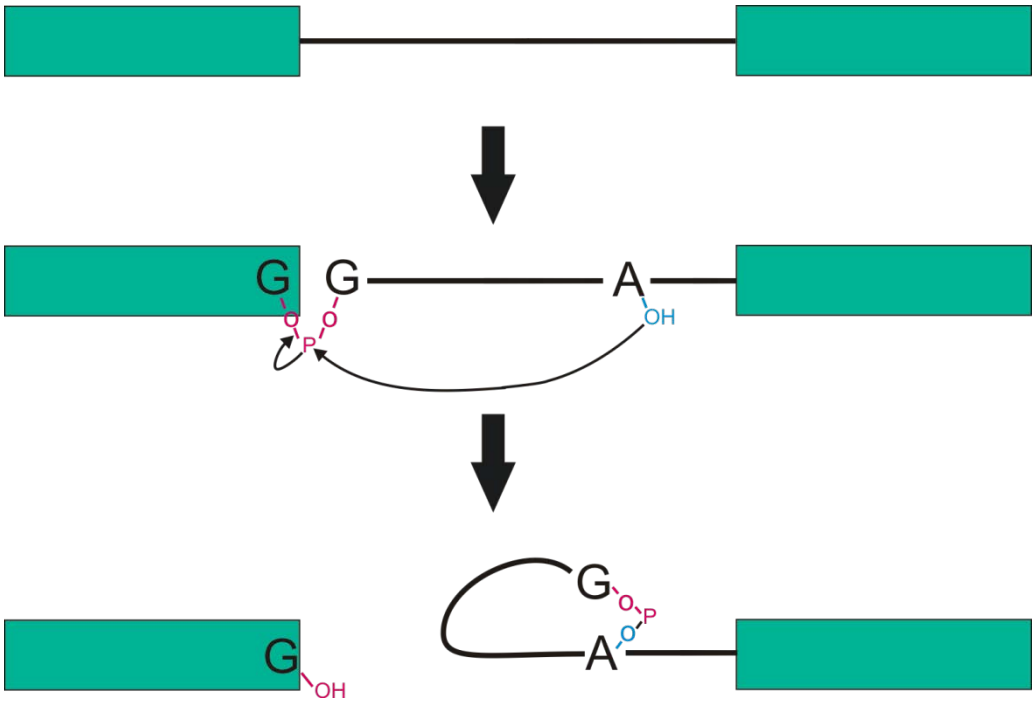


Figure 1.3 First transesterification of splicing. The 5’ splice site contains a guanosine in the exon (green boxes) and a guanosine in the intron (black bar), both labeled as G. The conserved adenosine necessary for the first transesterification and lariat formation is labeled as A in the intron. Typical phosphodiester bonds are shown in red and the 2’ hydroxyl is shown in blue. Curved, black arrows indicate the two nucleophilic attacks of the first transesterification step.

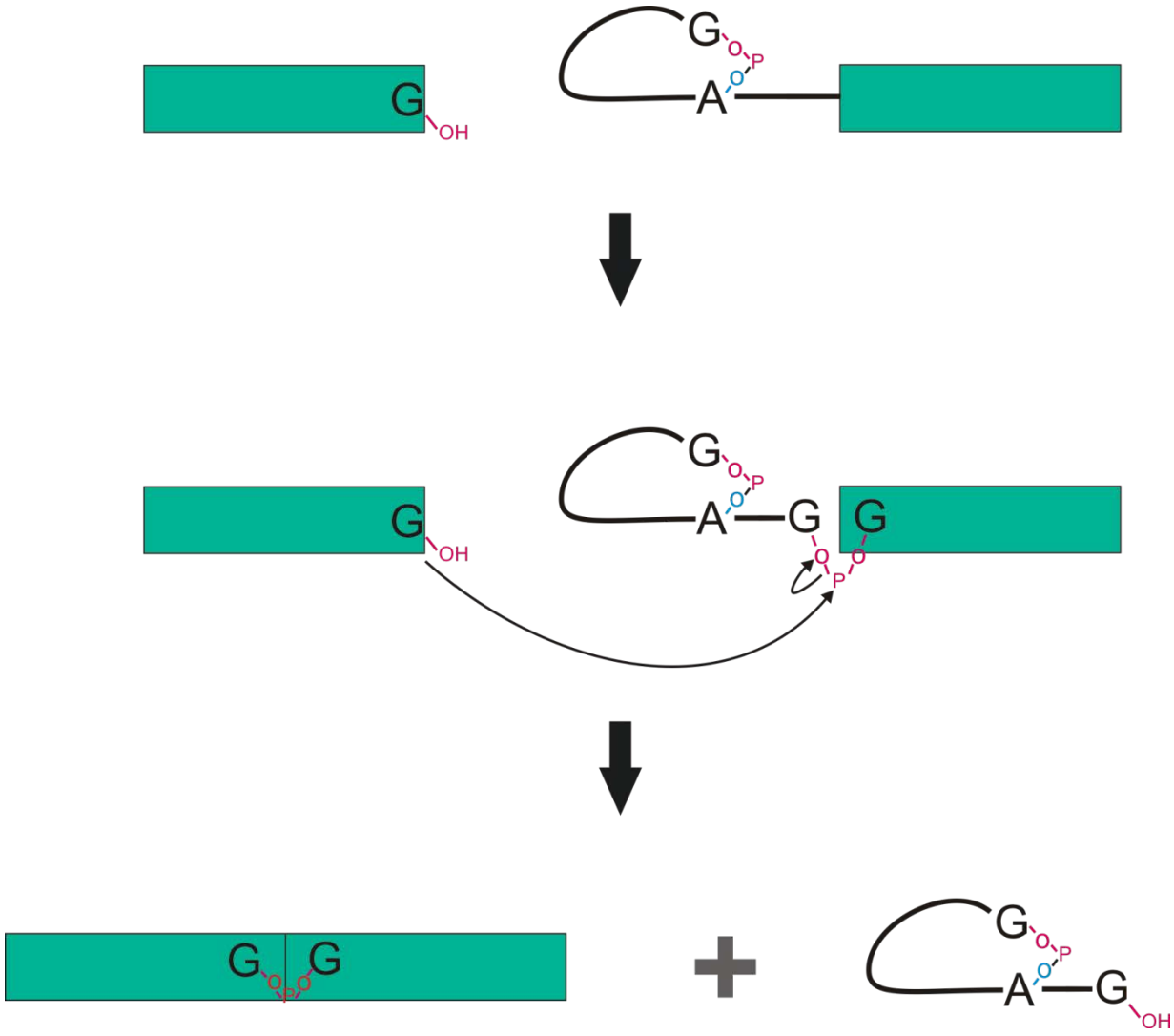


Figure 1.4 Second transesterification of splicing. The 5' splice site contains a guanosine in the exon (green boxes) and the 3' splice site contains a guanosine in the exon and in the intron (black bar), all labeled as G. The conserved adenosine necessary for the first transesterification and lariat formation is labeled as A in the already formed lariat. Typical phosphodiester bonds are shown in red and the 2' hydroxyl is shown in blue. Black arrows indicate the two nucleophilic attacks of the second transesterification step.

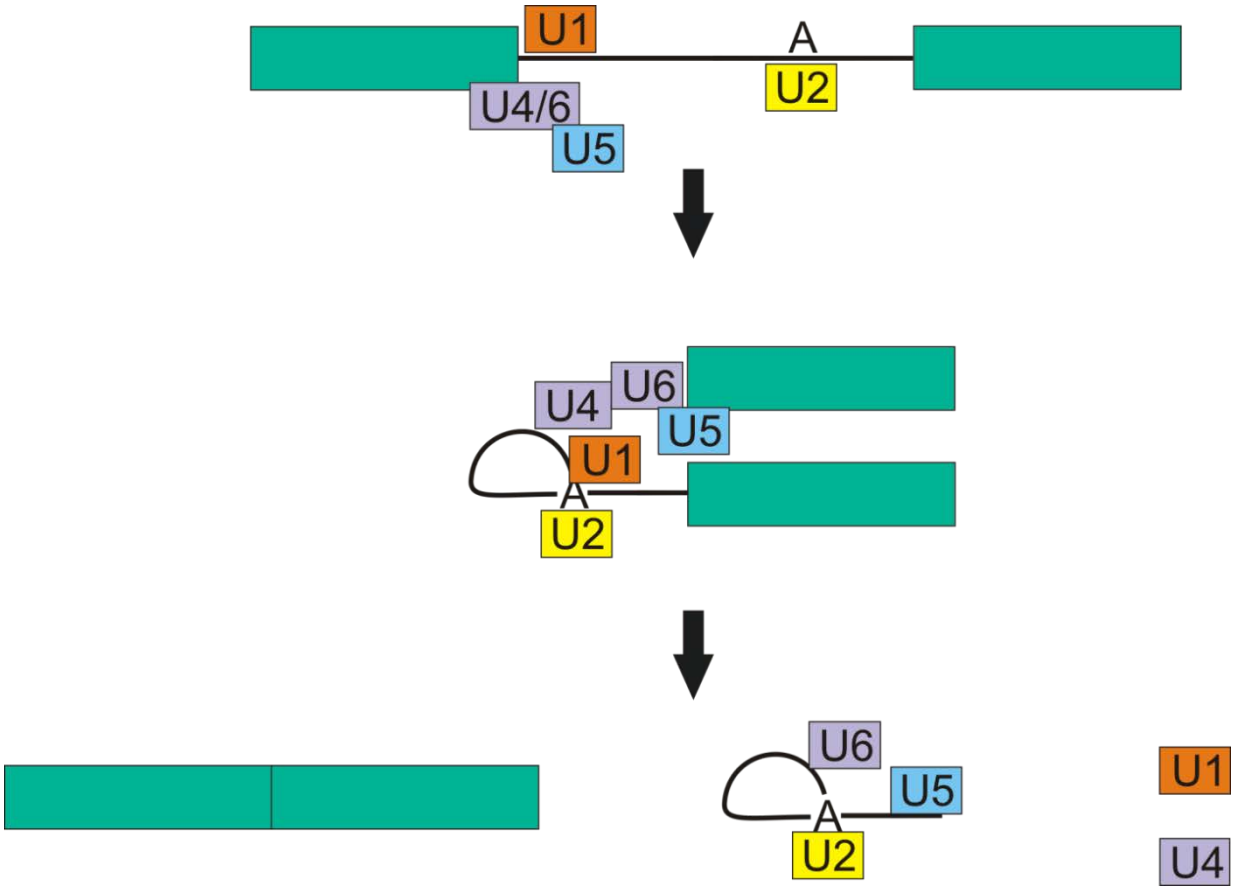


Figure 1.5 Assembly of snRNPs on the major spliceosome. snRNPs are labeled based on the snRNA that is at the core of each complex. U1 anneals to the conserved sequence at the 5' splice site. U2 binds the branchpoint. The U4/U6·U5 tri-snRNP rearranges the exons; the U5 RNA directly interacts with the conserved sequence at the 3' end of the intron. U1 and U4 snRNAs are released from the spliceosome. The remaining snRNPs are released with the lariat structure.

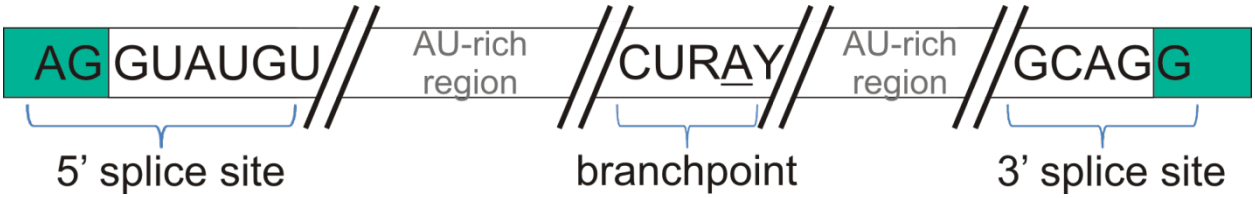


Figure 1.6 *Arabidopsis* intron sequence model. Consensus sequences for the 5’ splice site, branchpoint and 3’ splice site in *Arabidopsis* are labeled. The diagonal slashes indicate a break in the intron sequence and the AU-rich regions within the intron are labeled.

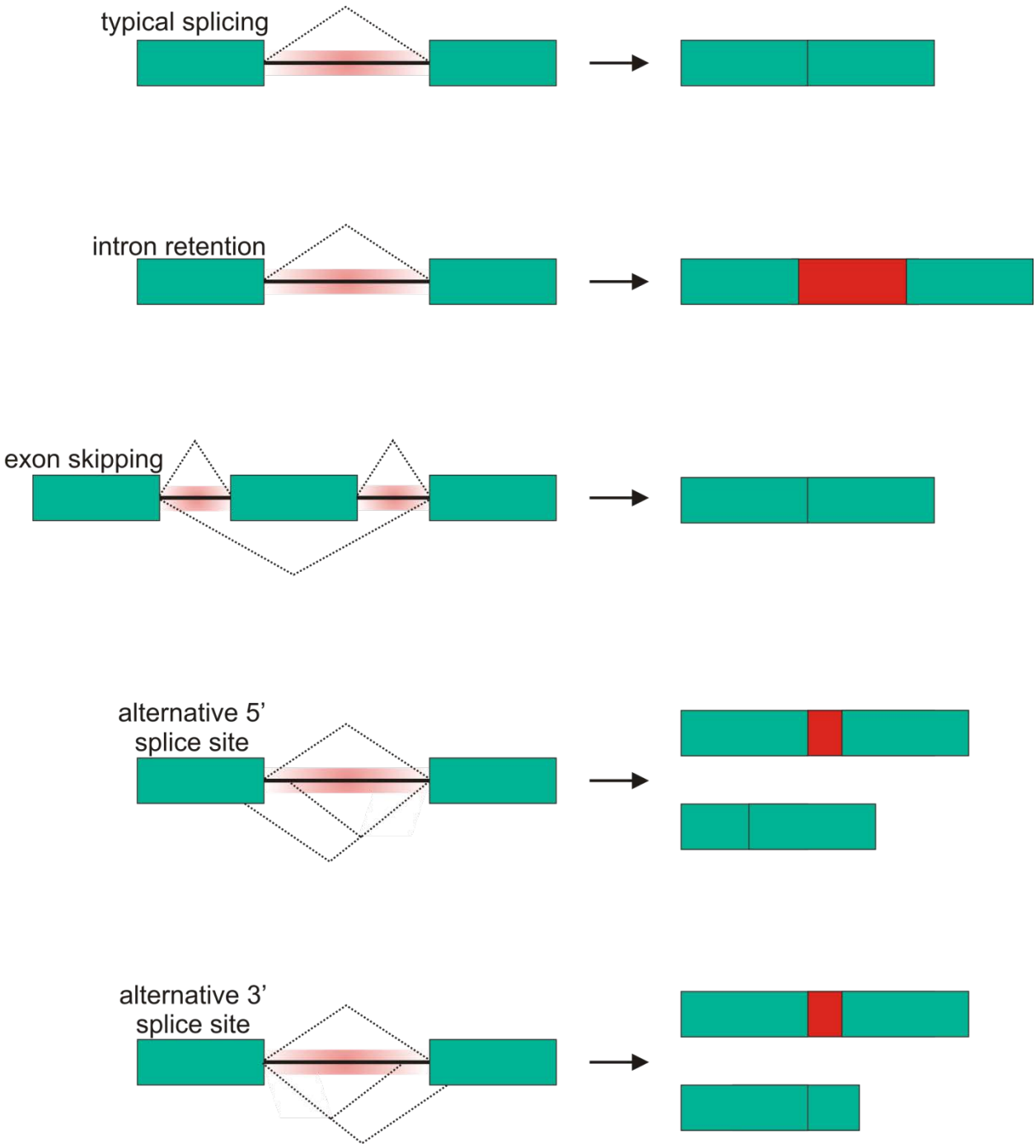


Figure 1.7 Types of alternative splicing. Exons are green boxes and introns are black bars with transparent red boxes. Alternatively retained introns or intron fragments are red boxes.

1.9 References

Abovich, N., Liao, X. C. and Rosbash, M. (1994). The yeast MUD2 protein: an interaction with PRP11 defines a bridge between commitment complexes and U2 snRNP addition. *Genes and Development* **8**, 843-854.

Achsel, T., Stark, H. and Lührmann, R. (2001). The Sm domain is an ancient RNA-binding motif with oligo(U) specificity. *Proceedings National Academy of Sciences USA* **98**, 3685-3689.

Ali, G. S. and Reddy, A. S. (2008). Regulation of alternative splicing of pre-mRNAs by stresses. *Current Topics Microbiology and Immunology* **326**, 257-275.

Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796-815.

Baker, B. S. (1989). Sex in flies: the splice of life. *Nature* **340**, 521-524.

Bark, C., Weller, P., Zabielski, J. and Pettersson, U. (1986). Genes for human U4 small nuclear RNA. *Gene* **50**, 333-344.

Barta, A., Kalyna, M. and Lorkovic, Z. J. (2008). Plant SR proteins and their functions. *Current Topics in Microbiology and Immunology* **326**, 83-102.

Barta, A., Sommergruber, K., Thompson, D., Hartmuth, K., Matzke, M. A. and Matzke, A. J. M. (1986). The expression of a nopaline synthase - human growth hormone chimaeric gene in transformed tobacco and sunflower callus tissue. *Plant Molecular Biology* **6**, 347-357.

Baynton, C. E., Potthoff, S. J., McCullough, A. J. and Schuler, M. A. (1996). U-rich tracts enhance 3′ splice site recognition in plant nuclei. *Plant Journal* **10**, 703-711.

Belostotsky, D. A. and Rose, A. B. (2005). Plant gene expression in the age of systems biology: integrating transcriptional and post-transcriptional events. *Trends Plant Science* **10**, 347-353.

Berget, S. M. (1995). Exon recognition in vertebrate splicing. *Journal Biological Chemistry* **270**, 2411-2414.

Bono, F. and Gehring, N. H. (2011). Assembly, disassembly and recycling: The dynamics of exon junction complexes. *RNA Biology* **8**, 24-29.

Brown, J. W., Feix, G. and Frendewey, D. (1986). Accurate in vitro splicing of two pre-mRNA plant introns in a HeLa cell nuclear extract. *EMBO Journal* **5**, 2749-2758.

Brown, J. W., Smith, P. and Simpson, C. G. (1996). *Arabidopsis* consensus intron sequences. *Plant Molecular Biology* **32**, 531-535.

Brown, J. W. S. (1996). *Arabidopsis* intron mutations and pre-mRNA splicing. *Plant Journal* **10**, 771-780.

Celotto, A. M. and Graveley, B. R. (2001). Alternative splicing of the *Drosophila* Dscam pre-mRNA is both temporally and spatially regulated. *Genetics* **159**, 599-608.

Chen, F. C., Wang, S. S., Chaw, S. M., Huang, Y. T. and Chuang, T. J. (2007). Plant gene and alternatively spliced variant annotator. A plant genome annotation pipeline for rice gene and alternatively spliced variant identification with cross-species expressed sequence tag conservation from seven plant species. *Plant Physiology* **143**, 1086-1095.

Cheng, S. C. and Abelson, J. (1987). Spliceosome assembly in yeast. *Genes and Development* **1**, 1014-1027.

Chini, A., Fonseca, S., Chico, J. M., Fernández-Calvo, P. and Solano, R. (2009). The ZIM domain mediates homo- and heteromeric interactions between *Arabidopsis* JAZ proteins. *Plant Journal* **59**, 77-87.

Dinesh-Kumar, S. P. and Baker, B. J. (2000). Alternatively spliced N resistance gene transcripts: their possible role in tobacco mosaic virus resistance. *Proceedings of the Naional Academy of Sciences USA* **97**, 1908-1913.

Domdey, H., Apostol, B., Lin, R.-J., Newman, A., Brody, E. and Abelson, J. (1984). Lariat structures are in vivo intermediates in yeast pre-mRNA splicing. *Cell* **39**, 611-621.

Egawa, C., Kobayashi, F., Ishibashi, M., Nakamura, T., Nakamura, C. and Takumi, S. (2006). Differential regulation of transcript accumulation and alternative splicing of a DREB2 homolog under abiotic stress conditions in common wheat. *Genes and Genetic Systems* **81**, 77-91.

Filichkin, S. A., Priest, H. D., Givan, S. A., Shen, R., Bryant, D. W., Fox, S. E., Wong, W. K. and Mockler, T. C. (2010). Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome Research* **20**, 45-58.

Fujii, T. and Shimada, T. (2007). Sex determination in the silkworm, *Bombyx mori*: A female determinant on the W chromosome and the sex-determining gene cascade. *Seminars Cell and Developmental Biology* **18**, 379-388.

Gassmann, W. (2008). Alternative splicing in plant defense. *Current Topics Microbiology and Immunology* **326**, 219-233.

Gfeller, A., Liechti, R. and Farmer, E. E. (2010). *Arabidopsis* jasmonate signaling pathway. *Science Signaling* **3**, cm4.

Glekas, A. S., Annamalai, D., Schuler, M. A. (2013). Differential accumulation of *Arabidopsis* splicing factor transcripts in response to heavy metal stress. In preparation.

Glekas, A. S. and Schuler, M. A. (2013). Differential and specialized expression of second-step splicing factors in Arabidopsis thaliana. In preparation.

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

Guo, S., Zheng, Y., Joung, J. G., Liu, S., Zhang, Z., Crasta, O. R., Sobral, B. W., Xu, Y., Huang, S. and Fei, Z. (2010). Transcriptome sequencing and comparative analysis of cucumber flowers with different sex types. *BMC Genomics* **11**, 384.

Haas, B. J. (2008). Analysis of alternative splicing in plants with bioinformatics tools. *Current Topics Microbiology and Immunology* **326**, 17-37.

Hartmuth, K. and Barta, A. (1986). *In vitro* processing of a plant pre-mRNA in a HeLa cell nuclear extract. *Nucleic Acids Research* **14**, 7513-7528.

James, A. B., Syed, N. H., Bordage, S., Marshall, J., Nimmo, G. A., Jenkins, G. I., Herzyk, P., Brown, J. W. S. and Nimmo, H. G. (2012). Alternative splicing mediates responses of the *Arabidopsis* circadian clock to temperature changes. *Plant Cell* **24**, 961-981.

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

Kerins, J. A., Hanazawa, M., Dorsett, M. and Schedl, T. (2010). PRP-17 and the pre-mRNA splicing pathway are preferentially required for the proliferation versus meiotic development decision and germline sex determination in *Caenorhabditis elegans*. *Developmental Dynamics* **239**, 1555-1572.

Konarska, M. M. and Sharp, P. A. (1987). Interactions between small nuclear ribonucleoprotein particles in formation of spliceosomes. *Cell* **49**, 763-774.

Lalli, E., Ohe, K., Latorre, E., Bianchi, M. E. and Sassone-Corsi, P. (2003). Sexy splicing: regulatory interplays governing sex determination from *Drosophila* to mammals. *Journal Cell Science* **116**, 441-445.

Lamond, A. I., Konarska, M. M., Grabowski, P. J. and Sharp, P. A. (1988). Spliceosome assembly involves the binding and release of U4 small nuclear ribonucleoprotein. *Proceedings National Academy of Sciences USA* **85**, 411-415.

Lander, E. S. Linton, L. M. Birren, B. Nusbaum, C. Zody, M. C. Baldwin, J. Devon, K. Dewar, K. Doyle, M. FitzHugh, W. et al. (2001). Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921.

Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. and Steitz, J. A. (1980). Are snRNPs involved in splicing? *Nature* **283**, 220-224.

Levine, A. and Durbin, R. (2001). A computational scan for U12-dependent introns in the human genome sequence. *Nucleic Acids Research* **29**, 4006-4013.

Liu, H.-X. and Filipowicz, W. (1996). Mapping of branchpoint nucleotides in mutant pre-mRNAs expressed in plant cells. *Plant Journal* **9**, 381-389.

Lopez, P. J. and Séraphin, B. (1999). Genomic-scale quantitative analysis of yeast pre-mRNA splicing: implications for splice-site recognition. *RNA* **5**, 1135-1137.

Lorković, Z. J., Wieczorek Kirk, D. A., Lambermon, M. H. L. and Filipowicz, W. (2000). Pre-mRNA splicing in higher plants. *Trends Plant Science* **5**, 160-167.

Lund, E. and Dahlberg, J. E. (1984). True genes for human U1 small nuclear RNA. Copy number, polymorphism, and methylation. *Journal Biological Chemistry* **259**, 2013-2021.

Maschhoff, K. L. and Padgett, R. A. (1992). Phosphorothioate substitution identifies phosphate groups important for pre-mRNA splicing. *Nucleic Acids Research* **20**, 1949-1957.

McCullough, A. J., Lou, H. and Schuler, M. A. (1993). Factors affecting authentic 5' splice site selection in plant nuclei. *Molecular Cellular Biology* **13**, 1323-1331.

McCullough, A. J. and Schuler, M. A. (1993). AU-rich intronic elements affect pre-mRNA 5' splice site selection in Drosophila melanogaster. *Molecular Cellular Biology* **13**, 7689-7697.

Merritt, H., McCullough, A. J. and Schuler, M. A. (1997). Internal AU-rich elements modulate activity of two competing 3′ splice sites in plant nuclei. *Plant Journal* **12**, 937-943.

Moore, M. J. and Sharp, P. A. (1992). Site-specific modification of pre-mRNA: the 2'-hydroxyl groups at the splice sites. *Science* **256**, 992-997.

Nelson, K. K. and Green, M. R. (1989). Mammalian U2 snRNP has a sequence-specific RNA-binding activity. *Genes and Development* **3**, 1562-1571.

Padgett, R. A., Konarska, M. M., Grabowski, P. J., Hardy, S. F. and Sharp, P. A. (1984). Lariat RNAs as intermediates and products in the splicing of messenger RNA precursors. *Science* **225**, 898-903.

Palusa, S. G., Ali, G. S. and Reddy, A. S. (2007). Alternative splicing of pre-mRNAs of *Arabidopsis* serine/arginine-rich proteins: regulation by hormones and stresses. *Plant Journal* **49**, 1091-1107.

Parker, R., Siliciano, P. G. and Guthrie, C. (1987). Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell* **49**, 229-239.

Pauwels, L. and Goossens, A. (2011). The JAZ proteins: a crucial interface in the jasmonate signaling cascade. *Plant Cell* **23**, 3089-100.

Pikielny, C. W., Rymond, B. C. and Rosbash, M. (1986). Electrophoresis of ribonucleoproteins reveals an ordered assembly pathway of yeast splicing complexes. *Nature* **324**, 341-345.

Pinto, A. L. and Steitz, J. A. (1989). The mammalian analogue of the yeast PRP8 splicing protein is present in the U4/5/6 small nuclear ribonucleoprotein particle and the spliceosome. *Proceedings National Academy of Science USA* **86**, 8742-8746.

Podkowinski, J., Jelenska, J., Sirikhachornkit, A., Zuther, E., Haselkorn, R. and Gornicki, P. (2003). Expression of cytosolic and plastid acetyl-coenzyme A carboxylase genes in young wheat plants. *Plant Physiology* **131**, 763-772.

Query, C. C., Moore, M. J. and Sharp, P. A. (1994). Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes and Development* **8**, 587-597.

Reddy, A. S. N. (2004). Plant serine/arginine-rich proteins and their role in pre-mRNA splicing. *Trends Plant Science* **9**, 541-547.

Ritchie, D. B., Schellenberg, M. J. and MacMillan, A. M. (2009). Spliceosome structure: Piece by piece. *Biochimica Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1789**, 624-633.

Rodriguez, J. R., Pikielny, C. W. and Rosbash, M. (1984). *In vivo* characterization of yeast mRNA processing intermediates. *Cell* **39**, 603-610.

Rogers, J. and Wall, R. (1980). A mechanism for RNA splicing. *Proceedings National Academy of Sciences USA* **77**, 1877-1879.

Ruskin, B., Krainer, A. R., Maniatis, T. and Green, M. R. (1984). Excision of an intact intron as a novel lariat structure during pre-mRNA splicing *in vitro*. *Cell* **38**, 317-331.

Salz, H. K. (2011). Sex determination in insects: a binary decision based on alternative splicing. *Current Opinion in Genetics and Development* **21**, 395-400.

Schuler, M. A. (2008). Splice site requirements and switches in plants. *Current Topics Microbiology and Immunology* **326**, 39-59.

Schwer, B. (2008). A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. *Molecular Cell* **30**, 743-754.

Schwer, B. and Guthrie, C. (1991). PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. *Nature* **349**, 494-499.

Shatkin, A. J. and Manley, J. L. (2000). The ends of the affair: capping and polyadenylation. *Nature Structural Molecular Biology* **7**, 838-842.

Shepard, P. J. and Hertel, K. J. (2009). The SR protein family. *Genome Biology* **10**, 242.

Sheth, N., Roca, X., Hastings, M. L., Roeder, T., Krainer, A. R. and Sachidanandam, R. (2006). Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Research* **34**, 3955-3967.

Simpson, C. G., Clark, G., Davidson, D., Smith, P. and Brown, J. W. S. (1996). Mutation of putative branchpoint consensus sequences in plant introns reduces splicing efficiency. *Plant Journal* **9**, 369-380.

Sridharan, V. and Singh, R. (2007). A conditional role of U2AF in splicing of introns with unconventional polypyrimidine tracts. *Molecular Cellular Biology* **27**, 7334-7344.

Tai, H. H., Williams, M., Iyengar, A., Yeates, J. and Beardmore, T. (2007). Regulation of the beta-hydroxyacyl ACP dehydratase gene of *Picea mariana* by alternative splicing. *Plant Cell Reports* **26**, 105-113.

Tarn, W.-Y. and Steitz, J. A. (1997). Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. *Trends Biochemical Sciences* **22**, 132-137.

Thines, B., Katsir, L., Melotto, M., Niu, Y., Mandaokar, A., Liu, G., Nomura, K., He, S. Y., Howe, G. A. and Browse, J. (2007). JAZ repressor proteins are targets of the SCFCOI1 complex during jasmonate signalling. *Nature* **448**, 661-665.

Umen, J. G. and Guthrie, C. (1995). A novel role for a U5 snRNP protein in 3' splice site selection. *Genes and Development* **9**, 855-868.

Van Arsdell, S. W. and Weiner, A. M. (1984). Human genes for U2 small nuclear RNA are tandemly repeated. *Molecular Cellular Biology* **4**, 492-499.

van Santen, V. L. and Spritz, R. A. (1987). Splicing of plant pre-mRN As in animal systems and vice versa. *Gene* **56**, 253-265.

Venter, J. C. Adams, M. D. Myers, E. W. Li, P. W. Mural, R. J. Sutton, G. G. Smith, H. O. Yandell, M. Evans, C. A. Holt, R. A. et al. (2001). The sequence of the human genome. *Science* **291**, 1304-1351.

Wang, B.-B. and Brendel, V. (2006). Genomewide comparative analysis of alternative splicing in plants. *Proceedings National Academy of Sciences USA* **103**, 7175-7180.

Wang, B. B. and Brendel, V. (2004). The ASRG database: identification and survey of *Arabidopsis thaliana* genes involved in pre-mRNA splicing. *Genome Biology* **5**, R102.

Wiebauer, K., Herrero, J. J. and Filipowicz, W. (1988). Nuclear pre-mRNA processing in plants: distinct modes of 3'-splice-site selection in plants and animals. *Molecular Cellular Biology* **8**, 2042-2051.

Xue, G. P. and Loveridge, C. W. (2004). HvDRF1 is involved in abscisic acid-mediated gene regulation in barley and produces two forms of AP2 transcriptional activators, interacting preferably with a CT-rich element. *Plant Journal* **37**, 326-339.

Yan, Y., Stolz, S., Chételat, A., Reymond, P., Pagni, M., Dubugnon, L. and Farmer, E. E. (2007). A downstream mediator in the growth repression limb of the jasmonate pathway. *Plant Cell* **19**, 2470-2483.

Zamore, P. D., Patton, J. G. and Green, M. R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* **355**, 609-614.

Zhang, X. C. and Gassmann, W. (2007). Alternative splicing and mRNA levels of the disease resistance gene RPS4 are induced during defense responses. *Plant Physiology* **145**, 1577-1587.

Zhu, W. and Brendel, V. (2003). Identification, characterization and molecular phylogeny of U12-dependent introns in the *Arabidopsis thaliana* genome. *Nucleic Acids Research* **31**, 4561-4572.

CHAPTER 2

Differential expression of second-step splicing factors in *Arabidopsis*

1.1 Introduction

Pre-mRNA splicing is a well-studied process, yet in plants, there are a large number of uncharacterized factors acting in the spliceosome. While some identities exist at the splice sites of plant, animal and yeast introns, there are significant differences in other features of their introns. Some of these cause mammalian genes to be processed incorrectly in transgenic plants (Barta et al., 1986; van Santen and Spritz, 1987; Wiebauer et al., 1988) and plant introns to be processed incorrectly in animal systems (Brown et al., 1986; Hartmuth and Barta, 1986; van Santen and Spritz, 1987; Wiebauer et al., 1988).

PRP (pre-mRNA processing proteins) proteins involved in the first- and second-steps of splicing have been well characterized in *Saccharomyces cerevisiae*, *Homo sapiens* and, in some cases, *Drosophila melanogaster* and *Caenorhabditis elegans* but not in plants. Among these are the second-step splicing factors designated PRP16, PRP17, PRP18, PRP22 and SLU7 that are encoded by single copy genes in these other eukaryotes and, typically, by multicopy genes in plants. Functions defined for these are summarized below.

PRP16

In *S. cerevisiae*, the ATPase PRP16 is an essential DExD/H box helicase that interacts with snRNAs at the spliceosome (Burgess et al., 1990; Madhani and Guthrie, 1994; Schwer and Guthrie, 1992). PRP16 binds after 5' splice site cleavage and lariat formation and proofreads proper 5' splice site cleavage and lariat formation (Burgess and Guthrie, 1993; Koodathingal et al., 2010; Schwer and Guthrie, 1991). PRP16 helicase activity provokes a structural rearrangement in the spliceosome at the second transesterification step, although the exact nature

of the rearrangement is poorly understood (Schwer and Guthrie, 1992). *In vitro*, it has been shown that PRP16 can indiscriminately unwind or travel along single-stranded or double-stranded RNA, respectively, which presented evidence, along with genetic data, that PRP16 can displace secondary structure complexes of RNA *in vivo* (Madhani and Guthrie, 1994; Mefford and Staley, 2009; Wang et al., 1998). In 2011, it was shown that PRP16 can displace spliceosomal protein factors necessary for the first transesterification of splicing allowing the second-step to proceed (Tseng et al., 2011).

PRP17

S. cerevisiae PRP17 is a splicing factor that functions during the second transesterification step of splicing and is thought to bind PRP16 (Frank et al., 1992; Jones et al., 1995; Xu et al., 1998). In addition, a PRP17 null mutant in *S. cerevisiae* is blocked from entering S-phase of mitosis because there is a lack of splicing of the *Anc1* transcript, linking cell-cycle regulation with pre-mRNA splicing (Ben-Yehuda et al., 2000; Boger-Nadjar et al., 1998; Chawla et al., 2003). It has been noted that the *Anc1* intron contains non-consensus thymidine (uridine in RNA) residues in the branchpoint to 3' exon border needed for its recognition by PRP17 (Dahan and Kupiec, 2004) suggesting that PRP17 plays a role in branchpoint recognition. There is no evidence, however, that PRP17 interacts directly with RNA. Additionally, PRP17 in yeast is necessary for the splicing of long introns (>200 nucleotides) and introns with long distances between the branchpoint and the 3' splice site (>13 nucleotides) (Sapra et al., 2004). PRP17 in *S. cerevisiae* as well as its homologs in other organisms have highly conserved WD40 domains that are predicted to form a 7-blade propeller (Lindsey-Boltz et al., 2000). The WD40 motif contains exposed loops that are often important for protein-protein interactions and scaffolding-roles. This would match with the importance of PRP17 in splicing of long introns or introns that contain a

lengthy branchpoint to 3' splice site distance in yeast as it is likely acting as a scaffold to bring components at different regions of the spliceosome together.

PRP18

In *S. cerevisiae*, PRP18 is a second-step splicing factor implicated in splice site recognition (Crotti and Horowitz, 2009; Horowitz and Abelson, 1993; Vijayraghavan and Abelson, 1990). Like PRP17, PRP18 function is also important in the splicing of introns with a lengthy distance between the branchpoint and the 3' splice site (Zhang and Schwer, 1997). Mutational analysis demonstrates interaction between PRP18 and another second-step splicing factor, SLU7, as well as PRP18's ability to stabilize the interaction of adjoining exons with U5 snRNA (Bacikova and Horowitz, 2005; Crotti et al., 2007).

PRP22

S. cerevisiae PRP22 (and its human homolog HRH1) is a DExD/H box helicase that functions after the second-step splicing factor helicase, PRP16. It plays several roles late in the second-step of splicing. PRP22 is responsible for the completion of the second-step of splicing, the release of the mature mRNA from the spliceosome and in that capacity it maintains a proofreading function to ensure quality control (Mayas et al., 2006; Ohno and Shimura, 1996; Sawa and Shimura, 1991; Schneider and Schwer, 2001; Wagner et al., 1998). The function of PRP22 in the completion of the second-step of splicing does not require its ATPase activity (and in turn its helicase activity), but the ATPase activity is necessary for the disassembly of the spliceosome, which includes the release of the mature RNA (Campodonico and Schwer, 2002; Schneider and Schwer, 2001). PRP22's proofreading activity functions two-fold. It stalls the release of a spliced transcript *in vivo* that contains a mutation at the 3' splice site (Mayas et al., 2006). *In vitro*, it can also halt the exon ligation of a splicing intermediate that contains a

mutated 5’ splice site or an aberrant branchpoint structure (Mayas et al., 2006). Utilizing 5’ to 3’ directionality, the helicase disrupts the interaction between the U5 snRNA and the mature mRNA to release the mRNA from the spliceosome (Schwer, 2008).

SLU7

SLU7 is essential for 3’ splice site selection (Chua and Reed, 1999b; Frank and Guthrie, 1992). SLU7 associates with the spliceosome after the first step of splicing but before the identification of the 3’ splice site, acts in concert with PRP16 and continues to act after PRP16 activity is finished (Ansari and Schwer, 1995; Chua and Reed, 1999a). Of the studied second-step splicing factors, SLU7 is the first to be identified as playing a role in alternative splicing. Human cells under UV stress cause the subcellular localization of SLU7 to alter, resulting in alternative splicing (Janowicz et al., 2011; Shomron et al., 2005). Its expression is also regulated by two opposing transcription factors which, when either is depleted, results in alternative splicing of other gene products (Alberstein et al., 2007).

Although there is no evidence indicating that the second-step of splicing mechanistically differs in plants and other eukaryotes, the splicing factors involved still need to be functionally characterized since the model plant *Arabidopsis thaliana* contains multiple loci for four of the five second-step splicing factors and a single locus for PRP16. Comparisons of the amino acid sequence identities between *A. thaliana*, *S. cerevisiae* and *H. sapiens* homologues (Table 2.1) indicate that the single-copy *At*PRP16 shares 33-47% amino acid sequence identity with its human and yeast counterparts. In contrast, *At*PRP17 is present in two copies (*At*PRP17-1, *At*PRP17-2) that share only 25% identity with each other and 24-48% identity with human and yeast. *At*PRP18 is present in two copies (*At*PRP18A, *At*PRP18B) that share 47% identity with each other and 16-38% identity with human and yeast. *At*PRP22 is present in three copies

(*At*PRP22-1, *At*PRP22-2, *At*PRP22-3) and *At*SLU7 is present in three copies (*At*SLU7-1A, *At*SLU7-1B, *At*SLU7-2). While *At*PRP22-2 and *At*PRP22-3 are relatively similar to one another (60% identity), each only shares 30-31% sequence identity with *At*PRP22-1. Likewise, while *At*SLU7-1A and *At*SLU7-1B are very similar (91% identity), each only shares 58-59% identity with *At*SLU7-2.

It is likely that the multiple loci for PRP17, PRP18, PRP22 and SLU7 are the result of several genome duplication events, at least one of which occurred prior to differentiation from ancient angiosperm ancestors and at least two of which occurred after this differentiation (Bowers et al., 2003; Maere et al., 2005). While duplicated genes are commonly associated with functional redundancy, the facts that none of the duplicated copies are adjacent to their closest relatives (Figure 2.1) and that many have substantially diverged suggest that this is not the case for the *Arabidopsis* second-step splicing factors. This suggestion is further supported by the fact that most attempts aimed at knocking out the second-step splicing factors have generated embryo lethal mutants and a few phenotypically abnormal mutants (unpublished data).

This study, aimed at characterizing the second-step splicing factors, initially examined sequence alignments for sequence and structure anomalies and subsequently produced homology models for evaluation of important domains and residues. Expression profiles of each second-step splicing factor determined in multiple tissue types and at multiple ages identified subsets of genes co-expressed in aging plants.

2.2 Results

Multiple sequence alignments, domain maps and homology models of the *Arabidopsis* second-step splicing factors

*At*PRP16, the only single isoform second-step splicing factor in this study, is a DExD/H superfamily RNA helicase with a Walker A motif in motif I of the DEAD box superfamily domain and a DEAH sequence in motif II of the DEAD box superfamily domain (Figure 2.2A). Like other DExD/H superfamily proteins, *At*PRP16 contains DEAH, Helicase-C, HA2 and C-terminal helicase domains but their lengths vary in comparison to *Sc*PRP16 (Figure 2.2B). Preceding the region shown in Figure 2A, the first 572, 546 and 364 amino acids of the *Arabidopsis*, human and yeast proteins, respectively, are not alignable even though the first 288 N-terminal amino acids of *Sc*PRP16 are known to be necessary for nuclear localization and function *in vivo* (Wang and Guthrie, 1998). The unusually long N-terminal region of *At*PRP16 is similar in length to other plant PRP16 proteins (soy, grass and grapevine) and shares 53-60% identity with these; this region does not contain any discernible motifs.

Within the more conserved PRP16 sequences shown in Figure 2.2A, the Walker A motif (GXXXXGKT) contains GKT (at 378-380 in *Sc*PRP16 and 585-587 in *At*PRP16, Figure 2.2B) that has been shown to bind ATP and be essential for function in yeast (Hotz and Schwer, 1998; Linder et al., 1989). The sequence of this motif is absolutely conserved even at its four variable positions in the *Arabidopsis*, yeast and human PRP16 proteins. Several other helicase motifs and their surrounding regions are also absolutely conserved in the *Arabidopsis*, yeast and human PRP16 proteins. These include motif Ia (TQPRRVAA at 580-586 in *At*PRP16) that has been associated with single-stranded DNA binding by the DNA helicase UL9 of HSV-9 (Tuteja and Tuteja, 2004); motif II, (DEAH at 650-654 in *At*PRP16) that has the DE necessary for ATPase

activity (Schneider et al., 2002); motif III (SAT at 681-683 in *At*PRP16) that has been shown to couple ATPase and helicase activities in the DNA helicase ULV5 or HSV-I (Tuteja and Tuteja, 2004); motif V (ATNIAETSLT at 836-845) that is important for helicase function in the DEAH helicases PRP43 and PRP22 (Schneider et al., 2004; Tanaka and Schwer, 2006) While the remaining motifs IV and VI each have one variant position, motif VI contains conserved Q854 and R858 that are necessary for ATPase activity (Schneider et al., 2002) as well as R855, G857 and R861 that are necessary for *S. cerevisiae* viability (Hotz and Schwer, 1998).

Molecular modeling of the helicase domain (amino acids 450-1214) in *At*PRP16 against the crystal structure of the *S. cerevisiae* PRP43 (PDB 2XAU, Walbott et al., 2010), a spliceosome disassembly factor (Arenas and Abelson, 1997; Walbott et al., 2010) positions the critical GKT side-chains (green stick-format in Figure 2C), DEAH side-chains (blue stick-format), SAT side-chains (yellow stick-format) and QR side-chains (orange stick-format) in close proximity in the ATPase catalytic pocket (Figure 2.2C). Other important motifs in this helicase, including motif Ia that mediates single-stranded nucleic acid binding, Helicase-C domain found in four helicase superfamilies and is critical for helicase activity in many RNA helicases, and originally analyzed in RNA helicase II/Gu (Valdez et al., 1997) and the HA2 domain with unknown function (assumed to bind nucleic acids). Motif Ia and the Helicase-C domain are situated opposite of each other on the exterior of the protein, while the HA2 domain consists of three α-helices nestled in the interior of the protein.

PRP17 in *S. cerevisiae* and its homologs in other organisms have a highly conserved WD40 domain that is thought to form a 7-blade propeller (Lindsey-Boltz et al., 2000) with exposed loops implicated in protein-protein interactions (Xu and Min, 2011). Preceding the WD40 domain is an N-terminal region of varying lengths, which in *At*PRP17-1 and *Sc*PRP17 is

about 275 amino acids long and in *At*PRP17-2 and *Hs*PRP17 is about 153 amino acids long (Figure 2.3A). This region is not well-defined and contains almost no discernible domain. The single exception that is the region between amino acids 42-61 in *Sc*PRP17. These amino acids are required for localization to the nucleus (Lindsey-Boltz et al., 2000), but they are not conserved in any of the other examined PRP17s (Figure 2.3B). The first half of both *At*PRP17-1 and *At*PRP17-2 do not contain any identified domains, nor do they share any significant sequence similarity with each other. But, the N-terminal half of *At*PRP17-1 does share strong sequence identity with other predicted splicing factors while the N-terminal half of *At*PRP17-2 shares sequence identity with other plant PRP17 proteins. There are multiple homologs of PRP17 in many different plant organisms and among these, one has stronger identity to *At*PRP17-1 and the other has stronger identity to *At*PRP17-2.

Mutational analysis on yeast PRP17 residues show that the TETG at 235-238 is likely to be involved in protein-protein interactions (Lindsey-Boltz et al., 2000). *At*PRP17-1 contains this exact sequence but *At*PRP17-2 has V and R replacing the two T residues, respectively (Figure 2.3B). This, in effect, replaces the polar side-chains with hydrophobic and charged residues, likely altering the way in which *At*PRP17-2 interacts with its binding partners.

Not all of the WD repeats in the *At*PRP17 proteins (vertical hash marks in Figure 2.3B) necessarily end in WD residues, but many end in XD. Molecular modeling of the WD repeats in the *At*PRP17 proteins against the WDR5 template (PDB 2CO0, Ruthenburg et al., 2006) shows that the typical seven-blade propeller is predicted for both proteins (Figure 2.3C) and that this domain follows an N-terminal region of variable length. The TETG (*At*PRP17-1) and VERG (*At*PRP17-2) residues predicted to be involved in protein-protein interactions are highlighted in yellow stick format in Figure 2.3C with their side-chains protruding from the protein. Based on

these models, it is likely that *Sc*PRP17 and *At*PRP17-1 maintain similar interactions with their pairing partners and that *At*PRP17-2 has alternate pairing partners that can accommodate the extended R side-chain.

Both *At*PRP18 proteins differ in their domain arrangements from that of *S. cerevisiae* (Figure 2.4A). *At*PRP18A and *Hs*PRP18 contain N-terminal splicing factor motifs (SMF) (highlighted in yellow in Figure 2.4A) that were originally identified in PRP4 (He et al., 2012) and subsequently shown to be necessary for its interaction with cyclophilin H (cypH) in the first step of splicing (Ayadi et al., 1997). While molecular modeling revealed that *Hs*PRP18 and cypH might interact similarly to *Hs*PRP4 and cypH (He et al., 2012), there is no evidence that *Hs*PRP18 and cypH interact *in vivo* in the second-step of splicing. This domain is not present in the truncated *At*PRP18B or *Sc*PRP18, but it is found in other plant species (He et al., 2012). Both *At*PRP18A and *At*PRP18B also contain the PRP18 superfamily domain that is conserved among all PRP18 proteins (Figure 2.4B). Quite different from their yeast and human counterparts, the C-terminal regions of *At*PRP18A and *At*PRP18B are found in many plant species including soybean, grape, poplar, *Ricinus communis* (castor oil plant), *Brachypodium distachyin* (grass), maize, rice, barley and moss. The crystal structure of *Sc*PRP18 solved by Jiang and colleagues (2000, PDB 1DVK) indicates a positively charged surface region consistent with previous indications that PRP18 interacts with SLU7. This region located in helices 1 and 2 has H118, K125, K140, K141, R151 and R152 (indicated by arrows in Figure 2.4B) contributing to the surface charge. Of these, amino acids H118, R151 and R152 are essential for SLU7 interaction and amino acids R80, K140 and K141 are important for SLU7 interaction (Bacikova and Horowitz, 2002). Asterisks above *At*PRP18A indicate the basic residues predicted to be surface amino acids and likely contribute to interactions with SLU7 (Figure 2.4B). Each of these

residues (visualized in stick format in Figure 2.4C) are outward-facing and also likely candidates for SLU7 interactions in *Arabidopsis*.

The three *At*PRP22 proteins contain the expected DEAH, Helicase-C, HA2 and C-terminal helicase domains found in *At*PRP16 and *Sc*PRP22 (Figure 2.5A). *At*PRP22-1 also contains a S1 RNA binding domain (originally identified in the ribosomal protein S1) that exists in *Sc*PRP22; *At*PRP22-2 and *At*PRP22-3 do not contain this domain. The additional SR (serine/arginine-rich) domain is found in *At*PRP22-1 and human HRH1 (Ono et al., 1994) (Figure 2.5B). While analysis of HRH1 has indicated that this region is responsible for the localization of HRH1 to the nuclear speckles (Ohno and Shimura, 1996), *Sc*PRP22 that lacks this region is also capable of localizing to these same regions.

All *At*PRP22 proteins also contain the common DExD/H superfamily domain. In motif I of the DEAH region, *At*PRP22-1 and *At*PRP22-2 contain a Walker A motif (GKT) that is necessary for the ATPase activity of *Sc*PRP22 (Schneider et al., 2004; Wagner et al., 1998). At this position *At*PRP22-3 has an alternate GKS (Figure 2.5B), a replacement found in a number of other Walker A motifs. In addition, all *At*PRP22 proteins contain DEAH at 603-606 (numbered as in *Sc*PRP22) and Q804 and R808 that are necessary for the ATPase activity of *Sc*PRP22 (Schneider et al., 2002; Schneider et al., 2004). In motif III of the DEAH domain, the SAT at 635-637 in ScPRP22 are necessary for helicase activity and release of mature mRNA from the spliceosome (Schwer and Meszaros, 2000). The substitution of T for S in *At*PRP22-2 is unlikely to affect its activity. Motif V is important for linking the ATPase activity of PRP22 to the RNA substrate binding (Schneider et al., 2004).

Each of the three *At*PRP22 homology models (PDB 3I4U, Kudlinzki et al., 2012) exhibit the typical arrangement of a DEAH superfamily domain: a Helicase-C domain, an HA2 domain

and an OB (oligonucleotide-binding) domain (Figure 2.5C). The two regions that are shown to be necessary for ATPase activity: GKT in motif I and SAT in motif III are highlighted in stick format along with the DEAH in motif II.

SLU7 is a zinc knuckle protein encoded by three loci in *Arabidopsis*. The zinc knuckle domain occurs in the N-terminal half of each of the three *Arabidopsis* SLU7 proteins (Figure 2.6A). Overlapping with this region are four putative NLS regions that are important for localization of *Hs*SLU7 to the nucleus (Shomron et al., 2004). Not conserved in the *At*SLU7 proteins is the PRP18 binding region from amino acids 215-224 in *Sc*SLU7 (Figure 2.6B). The glutamic acid at 217 is absolutely required for interaction with PRP18 and proper splicing of pre-mRNA (James et al., 2002) and is present in human SLU7. In contrast, *At*SLU7-1A and *At*SLU7-1B contain a valine and *At*SLU7-2 contains an isoleucine at this position; both *At*SLU7 proteins and *Hs*SLU7 have substantial sequence divergence in the region surrounding this position. *At*SLU7-1A, *At*SLU7-1B and *Hs*SLU7 share a C-terminal insert that is not found in *Sc*SLU7 or *At*SLU7-2 (Figure 2.6B). The 144 and 145 amino acid C-terminal inserts in *At*SLU7-1A and *At*SLU7-1B are nearly identical and both contain a lysine-rich region not found in any of the other SLU7 proteins examined in this study. Homology models for the SLU7s of *Arabidopsis* cannot be generated because there is no structural data for any SLU7 protein or homolog.

Analysis of expression levels of the second-step splicing factors in development

In order to further differentiate between the multiple loci of the second-step splicing factors and to determine if subsets of the splicing factors act in concert, transcript levels of the second-step splicing factors were analyzed in multiple tissue types. To investigate the possibility that expression levels of the second-step splicing factors varied in different tissue types, with focus on rosette, stem, cauline, flower and silique at both 4 weeks and 7 weeks as compared to

the entire plant (seedling) at 1 week. Genevestigator (Hruz et al., 2008) was used for preliminary expression analysis (Figure 2.7). The types of tissue analyzed are classified as germinated seed, seedling, young rosette, developed rosette, bolting, young flower, developed flower, flowers and siliques, mature siliques and senescence. Trends of expression do vary, which warranted further investigation in order to narrow the results to more specific tissue types and time points.

Subsets of splicing factors in four-week-old plants

To clarify the expression patterns of each second-step splicing factor, their transcript levels in four- and seven-week-old plants were evaluated by RT-PCR analysis and compared to the transcript levels in one-week-old seedling after normalization to the transcript levels of constitutive EF1α. As shown in Figure 2.8, different subsets of second-step splicing factors accumulate in each tissue-type of four-week-old plants. In rosette tissue, *At*PRP16, *At*PRP18B, *At*PRP22-2, *At*SLU7-1A and *At*SLU7-1B are expressed at levels at least twice their levels in one-week-old seedlings. In stem tissue, *At*PRP16 and *At*PRP22-2 are also expressed at higher levels. In cauline tissue, *At*PRP16, *At*PRP17-1, *At*PRP22-2, *At*PRP22-3 and *At*SLU7-1B are expressed at higher levels with *At*PRP22-2 expressed four-fold higher times that of in one-week-old seedlings, a more drastic change than what is found in analysis of other genes. In flower tissue, *At*PRP16, *At*PRP17-1, *At*PRP17-2, *At*PRP22-1, *At*PRP22-2 and *At*PRP22-3 expressions are expressed at high levels with *At*PRP17-2 having 13-fold higher expression than in one-week-old seedlings and *At*PRP22-2 having seven-fold higher expression than in one-week-old seedlings. In silique tissue, *At*PRP17-1, *At*PRP17-2, *At*PRP22-1, *At*PRP22-2, *At*PRP22-3, *At*SLU7-1A, *At*SLU7-1B and *At*SLU7-2 are expressed at high levels with. *At*PRP17-2 has eight-fold higher levels than in one-week-old seedlings and *At*SLU7-2, which is not expressed in any other mature tissues, having nearly nine-fold higher levels than in one-week-old seedling. *At*PRP18A

expression is consistent throughout the four-week-old plant except in stem tissue where it, along with *At*PRP17-2, *At*PRP18B and *At*SLU7-2, is downregulated.

Subsets of splicing factors in seven-week-old plants

In all tissues of seven-week-old plants, *At*PRP16, *At*PRP17-1, *At*PRP17-2, *At*PRP22-1, *At*PRP22-2 and *At*PRP22-3 transcript levels are consistently at least two-fold higher than in one-week-old seedlings (Figure 2.9). *At*PRP18A expression does not undergo significant increases or decreases any of the seven-week-old tissues. Both *At*PRP18B and *At*SLU7-2 expressions are significantly higher in flower and silique tissues of seven-week-old plants. Notably, seven-week-old silique tissue expresses every second-step splicing factor examined, seven-week-old rosette tissue expresses many of the same second-step splicing factors as in four-week-old rosette tissue, four-week-old stem tissue expresses very few of these transcripts at higher levels than in one-week-old seedlings and seven-week-old stem tissue expressed nearly all of these transcripts at higher levels.

As compared to the same tissues in a 4-week-old, plant, there is less downregulation of most genes at 7 weeks of age (Figure 2.10). *At*PRP22-2 is upregulated in every tissue at both the four week age and the seven week age, but when comparing tissue at seven weeks versus tissue at four weeks, the level of expression is higher in each of the older tissue. *At*PRP17-1, *At*PRP22-1, *At*PRP22-3, *At*SLU7-1A and *At*SLU7-1B levels tend to increase in most (but not all) tissues over time. Variation over time is more pronounced in *At*PRP17-2, *At*PRP18B and *At*SLU7-2, and in some genes, expression levels decrease.

2.3 Discussion

Multiple sequence alignments and homology models were utilized to identify similarities and differences between the *Arabidopsis* second-step splicing factors and the better-characterized

yeast and human second-step splicing factors. *At*PRP16 contains an additional 450 amino acid region at its N-terminus that is not present in yeast PRP16. The stretch of SR-rich residues present in AtPRP16 could play an additional role in the localization of PRP16 to the nucleus as it does in HsPRP16 where it has been shown to be necessary for localization to the spliceosome (Zhou and Reed, 1998). The expression of *At*PRP16 remains relatively steady across most tissue types with expression levels never dipping below those of one-week-old seedlings except in seven-week-old silique tissue. This suggests that this protein might be necessary for splicing in all tissue types.

*At*PRP17-1 has a TETG motif on the outside of its WD40 propeller blade that closely aligns with the TETG motif important for interactions of yeast PRP17 with the spliceosome. *At*PRP17-2 lacks this TETG motif and instead, has nonpolar and charged side-chain replacements that could allow for recruitment of alternate factors to *At*PRP17-2-containing spliceosomes. Like *At*PRP16, *At*PRP17-1 expression is steady in most tissues and developmental stages with increases noted only in seven-week-old silique tissue. In contrast, *At*PRP17-2 expression is increased in four-week-old flower and silique tissues and all seven-week-old tissues. This suggests that *At*PRP17-2 is important in splicing in reproductive tissues and developing embryos.

*At*PRP18A contains an SMF domain-containing insert in N-terminal region that is also present in other plant PRP18 proteins, but not in *At*PRP18B or any other PRP18 proteins that have been functionally characterized. This domain could facilitate the recruitment of *At*PRP18A to specific spliceosomes or could recruit other proteins to *At*PRP18A-containing spliceosomes. Both *At*PRP18A and *At*PRP18B contain additional plant-specific C-terminal domains that have yet to be characterized for function.

*At*PRP22-1 contains SR and S1 domains in its N-terminal region that are not present in *At*PRP22-2 and *At*PRP22-3. As in the case of the PRP16 proteins, this SR domain is not present in yeast PRP22 but is present in human PRP22, where it has been shown to be important for the localization of PRP22 to nuclear speckles (Ohno and Shimura, 1996). Based on the differences of these three *Arabidopsis* PRP22 proteins, it will be important to determine if they localize to nuclear speckles and if they colocalize to the same nuclear speckles. *At*PRP22-2 also has several substitutions at the catalytic regions of the protein that are likely to be inconsequential because of their occurrence in other DEAD-box helicases. Like *At*PRP16, *At*PRP22-1 is expressed in every tissue type without much variation in expression levels. *At*PRP22-2, however, exhibits increased expression in four-week-old flower and in all mature tissue types. *At*PRP22-3 is upregulated in only seven-week-old cauline, flower and silique tissues.

It should be noted, that none of the *Arabidopsis* SLU7 proteins contain a discernible PRP18-binding domain. *At*SLU7-1A and *At*SLU7-1B expression is fairly steady across all tissue types at varying ages except *At*SLU7-1B is upregulated in seven-week-old silique tissue. *At*SLU7-2 is downregulated in all tissues except four-week-old silique and seven-week-old cauline, flower and silique tissues. In silique tissues of both ages, *At*SLU7-2 is upregulated.

From these data, it seems that some second-step splicing factors are expressed at steady levels and likely responsible for the splicing of introns in a wide number of tissues. Another subset of second-step splicing factors is expressed in specialized tissues and potentially responsible for the splicing of unusual introns in these tissues. The specialized splicing factors *At*PRP17-2 and *At*PRP22-2 are simultaneously upregulated in four-week-old flower and silique tissues and all seven-week-old tissues. The specialized splicing factors *At*PRP18B and *At*SLU7-2

exhibit increased expression in both four- and seven-week-old silique tissues. Determination of direct binding between each of these sets of genes will be a natural next step.

It will be interesting to determine if each of these second-step splicing factors function within the same spliceosome or it they are more likely to function in a gene-specific manner. That is, does the splicing of certain pre-mRNAs recruit specific second-step splicing factors? In addition to the possibility of the multiple loci of the second-step splicing factors acting in a gene-specific manner, it will be prudent to examine if certain stresses imparted on the plant cause an alteration in the expression and function of the second-step splicing factors. We predict that the actions of the second-step splicing factors can influence alternative splicing and that the specialized splicing factors act on certain pre-mRNA transcripts.

2.4 Materials and Methods

Multiple sequence alignments

Alignments were generated using the AlignX feature of the Vector NTI Suite 7 (Invitrogen, Carlsbad, CA) (Lu and Moriyama, 2004). The following parameters were employed: gap opening penalty of 10, gap extension penalty of 0.05, gap separation penalty range of 8. The score matrix that was employed was blosum62 scoring matrix (Eddy, 2004). Visualization and adjustments were made when deemed necessary upon inspection. Accession numbers for *Arabidopsis* sequences used are in Table 2.2 and accession numbers for the non-*Arabidopsis* sequences used are in Table 2.3.

Homology models

Models were generated using MOE: Molecular Operating Environment (Chemical Computing Group, Montreal, Canada). Templates were identified using the ALIGN function in MOE which employs the blosum62 scoring matrix (Eddy, 2004). Ten intermediate models were

generated for each target sequence using the respective crystal or NMR structure chosen for that particular model. The best intermediate models chosen based on their packing scores and rotamer outliers were subjected to further energy minimization using the CHARMM22 force field (MacKerell et al., 1998). The models were further rendered using PyMol (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC).

Plant growth conditions and material

Wild type *Arabidopsis thaliana* (Columbia ecotype) seeds were sterilized by submerging in 70% alcohol for 30 seconds, 20% bleach and a drop of Tween20 for 15 minutes, and washed three times with sterile water. Seeds were then plated on ½ MS agar plates (pH 5.7). Seedlings were incubated in the dark at 4°C for 24 hours then grown under long day conditions. At the age of 14 days, plants were transferred to soil 1 week seedling samples. One-week-old seedlings were collected from plates and rosette, stem, cauline, flower and silique tissues were collected from 4- and 7-week-old plants. All tissue was snap-frozen in liquid nitrogen and stored at -80°C.

RNA isolation and cDNA synthesis

RNA used in tissue characterizing reverse transcriptase PCR was isolated using the RNeasy Plant minikit (QIAGEN), following the recommended procedure including on-column DNase treatment. In the case of metal stress experiments, RNA was extracted by the beadbeater method described in Thimmapuram et al. (2005) including DNase treatment. cDNA was synthesized by Superscript III (Invitrogen) according to the user manual, 1 µg of total RNA was used in a 20 µl reaction followed by dilution to 40 µl in water.

Analysis of transcript levels

Semi-quantitative reverse transcriptase PCR (RT-PCR) primers are listed in Table 2.4. cDNA was amplified by 25 cycles of PCR, with the exception of EF1a, which was amplified for 18 cycles. Each cycle consists of 94°C for 45 seconds, 58°C for 45 seconds and 72°C for 2 minutes. The first cycle is preceded by a 3 minute 95°C denaturation period and followed by a 10 minute 72°C elongation period. PCR products were run on 1.3% agarose gels, transferred to Hybond-N (Amersham-Pharmacia Biotech), and probed with ^{32}P-labeled DNA probe made with random primers (Invitrogen) and gene-specific cDNA template. Blots were hybridized and washed according to Lou et al (1993). Band strength was determined by ImageJ (NIH) (Rasband, W. S. ImageJ, US National Institutes of Health, Bethesda, MD, USA) after background signal was reduced with a rolling ball radius of 50 – 70 pixels. Transcript levels were normalized to the corresponding EF1α measurements and compared to untreated samples. These performed in biological triplicate and fold-change is plotted on a logarithmic scale. The p-value is less than 0.05.

Table 2.1 Percent sequence identities of PRP16, PRP17, PRP18, PRP22 and SLU7 genes.

	***At*PRP16**	***Hs*PRP16**	***Sc*PRP16**		
***At*PRP16**	100	33	47		
***Hs*PRP16**		100	34		
***Sc*PRP16**			100		
	***At*PRP17-1**	***At*PRP17-2**	***Hs*PRP17**	***Sc*PRP17**	
***At*PRP17-1**	100	25	48	29	
***At*PRP17-2**		100	24	37	
***Hs*PRP17**			100	31	
***Sc*PRP17**				100	
	***At*PRP18A**	***At*PRP18B**	***Hs*PRP18**	***Sc*PRP18**	
***At*PRP18A**	100	47	31	16	
***At*PRP18B**		100	31	38	
***Hs*PRP18**			100	30	
***Sc*PRP18**				100	
	***At*PRP22-1**	***At*PRP22-2**	***At*PRP22-3**	***Hs*HRH1**	***Sc*PRP22**
***At*PRP22-1**	100	31	30	60	45
***At*PRP22-2**		100	60	28	32
***At*PRP22-3**			100	27	31
***Hs*HRH1**				100	43
***Sc*PRP22**					100
	***At*SLU7-1A**	***At*SLU7-1B**	***At*SLU7-2**	***Hs*SLU7**	***Sc*SLU7**
***At*SLU7-1A**	100	91	59	45	19
***At*SLU7-1B**		100	58	44	19
***At*SLU7-2**			100	33	43
***Hs*SLU7**				100	12
***Sc*SLU7**					100

Table 2.2 *Arabidopsis* second-step splicing factors. The second-step splicing factors' common names and locus in *Arabidopsis*. In the locus, the At represents the organism: *Arabidopsis* thaliana, the following number represents the chromosome on which it is located, and the final number represents where the gene is located on the chromosome.

Common Name	Locus	Accession number
PRP16	At5g13010	AED91841
PRP17-1	At1g10580	ABO38749
PRP17-2	At5g54520	AED96505
PRP18A	At1g03140	AEE27535
PRP18B	At1g54590	AEE33122
PRP22-1	At3g26560	AEE77178
PRP22-2	At1g26370	AEE30683
PRP22-3	At1g27900	AEE30889
SLU7-1A	At1g65660	AEE34408
SLU7-1B	At4g37120	AEE86756
SLU7-2	At3g45950	AEE78093

Table 2.3 Homologs of the second-step splicing factors in *Arabidopsis*. These proteins were used in the generation of multiple sequence alignments of the *Arabidopsis* second-step splicing factors. *Sc = S. cerevisiae*, *Hs = Homo sapiens*.

Common Name	Accession number
*Sc*PRP16	CAA82165
*Hs*PRP16	NP_054722
*Sc*PRP17	NP_010652
*Hs*PRP17	NP_056975
*Sc*PRP18	AAA34915
*Hs*PRP18	NP_003666
*Sc*PRP22	CAA41530
*Hs*HRH1 (PRP22)	NP_004932
*Sc*SLU7	CAA98908
*Hs*SLU7	NP_006416

Table 2.3 Primers used in RT-PCR for tissue characterization.

Gene	Forward Primer	Reverse Primer
PRP16	gagaaaaccaggtgatagtg	ctgcctcaatctcatcttgc
PRP17-1	gcttgctgcacagagtttgg	cgagattgatacaacactaaag
PRP17-2	ggtcaaagcaagggttatcg	ggtgacgtttgatgcaagg
PRP18A	ggaaacaggagcttgatgc	ccatagtcactccaatagg
PRP18B	ggtttcgttattcatgatgtgag	ggtctcatgaatgataggaag
PRP22-1	cgacgtgctcaagatgtgagg	ggtcatacaacggctcaatcc
PRP22-2	ggagagttgcattcgcttgg	cgtttggaaatggtgaggag
PRP22-3	catgatgaagattctaatcttcc	gagacatcaaatttatttctgtcc
SLU7-1A	gttcaaacagctaaacatcc	ggaaatctttcattggatcc
SLU7-1B	ccatcccaagctgaattgc	ggaaatttctcattgggtcc

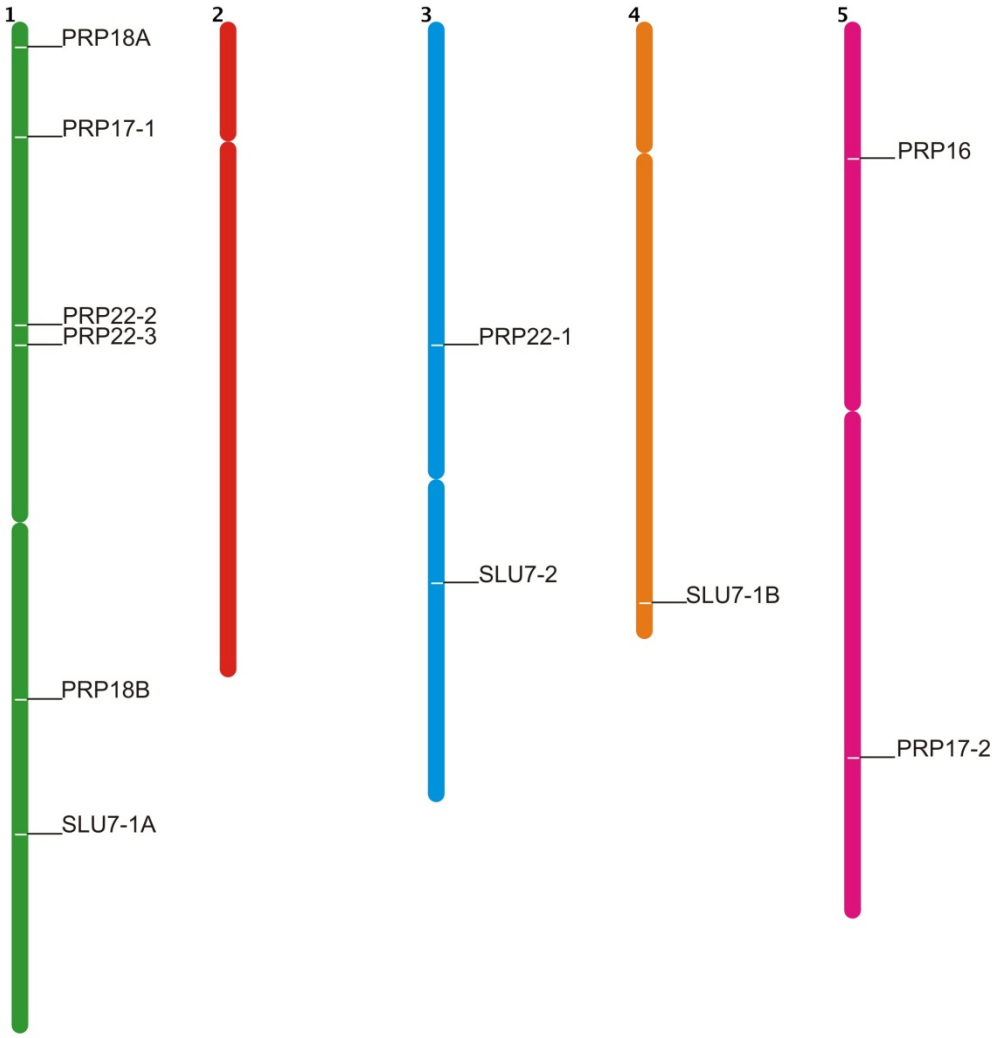
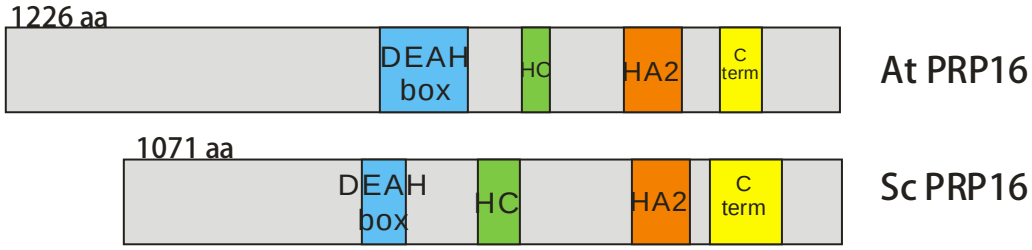


Figure 2.1 Arrangement of the second-step splicing factors on the *Arabidopsis* chromosomes. The five *Arabidopsis* chromosomes, depicted to scale. Location of each second-step splicing factor is indicated on the chromosomes.

A



B

```
AtPRP16       1  MGVDPFKTTETLEADKETNGGVPVKDKLTFKAPERKSRLGLDARAIEKKD
HsPRP16       1  -----------MGDTSEDASIHRLEGTDLDCQVGGLICKSKSAASEQHV
ScPRP16       1  --------------------------------------------------

AtPRP16      51  NAKTEGEFKVPKKSAISVTSSLDEEDKSDVSGLDFGTENTRPVHSSRRYR
HsPRP16      39  FKAPAPRPSLLGLDLLASLKRREREEKDDGEDKKKSKVSSYKDWEESKDD
ScPRP16       1  --------------------------------------------------

AtPRP16     101  EKSSRSQSAQESTVTTENAGTSDISITPRTLSCTSSYERGGSNRHREEHR
HsPRP16      89  QKDAEEEGGDQAGQNIRKDRHYRSARVETPSHPGGVSEEFWERSRQRERE
ScPRP16       1  -----------------------------MGHSGREERIKDIFKELTSK

AtPRP16     151  RDRSETPRSRQRNTYDEMDHYRRRESYRQSDRDYHGEKRRRYNSDWRTPG
HsPRP16     139  RREHGVYASSKEEKDWKKEKSRDRDYDRKRDRDERDRSRHSSRSERDGGS
ScPRP16      21  ELTPGLLLTLQKLAQKPN------TNLEQFIASCKALTKLSSNNPIIFNE

AtPRP16     201  RSDWDDGQDEWERSPHGDRGSSYSRRPQPSPSPMLAAASPDARLASPWLD
HsPRP16     189  ERSSRRNEPESPRHRPKDAATPSRSTWEEEDSGYGSSRRSQWESPSPTPS
ScPRP16      65  LLELLKNKSEEDSTGPKKIAPSIN--KRKKFKIQLDLDDNEDELDSPVQK
```

Figure 2.2 (A) Domain maps of PRP16 proteins. The putative scale domain map of *At*PRP16 is shown in comparison to the domain map of *Sc*PRP16. **(B) Sequence alignments of PRP16**. Shown are amino acid sequences of *At*, *Hs* and *Sc* PRP16s. Identical residues are red, conserved residues are royal blue and similar residues are light blue. The DEAD box superfamily domain is underlined with a solid line with brackets indicating motifs I-VI. HA2 domain is underlined with a dotted line, Helicase-C domain is underlined with a dashed lined. Essential amino acids are indicated with an asterisk above each one including the Walker A motif: GKT located in motif I. **(C) Homology model of *At*PRP16**. Top panel: PRP16 protein encompassing residues 450 through 1214 modeled against *Sc*PRP43. The DEAD box superfamily domain is colored green, the Helicase-C domain is colored blue and the HA2 domain is colored yellow. Bottom panel: The residues necessary for ATPase activity are highlighted in stick format. The 585GKT587 of motif I is green, the 678DEAH281 of motif II is blue, the 710SAT712 of motif III is yellow. Q883 and R886 in motif VI are orange.

B (cont.)

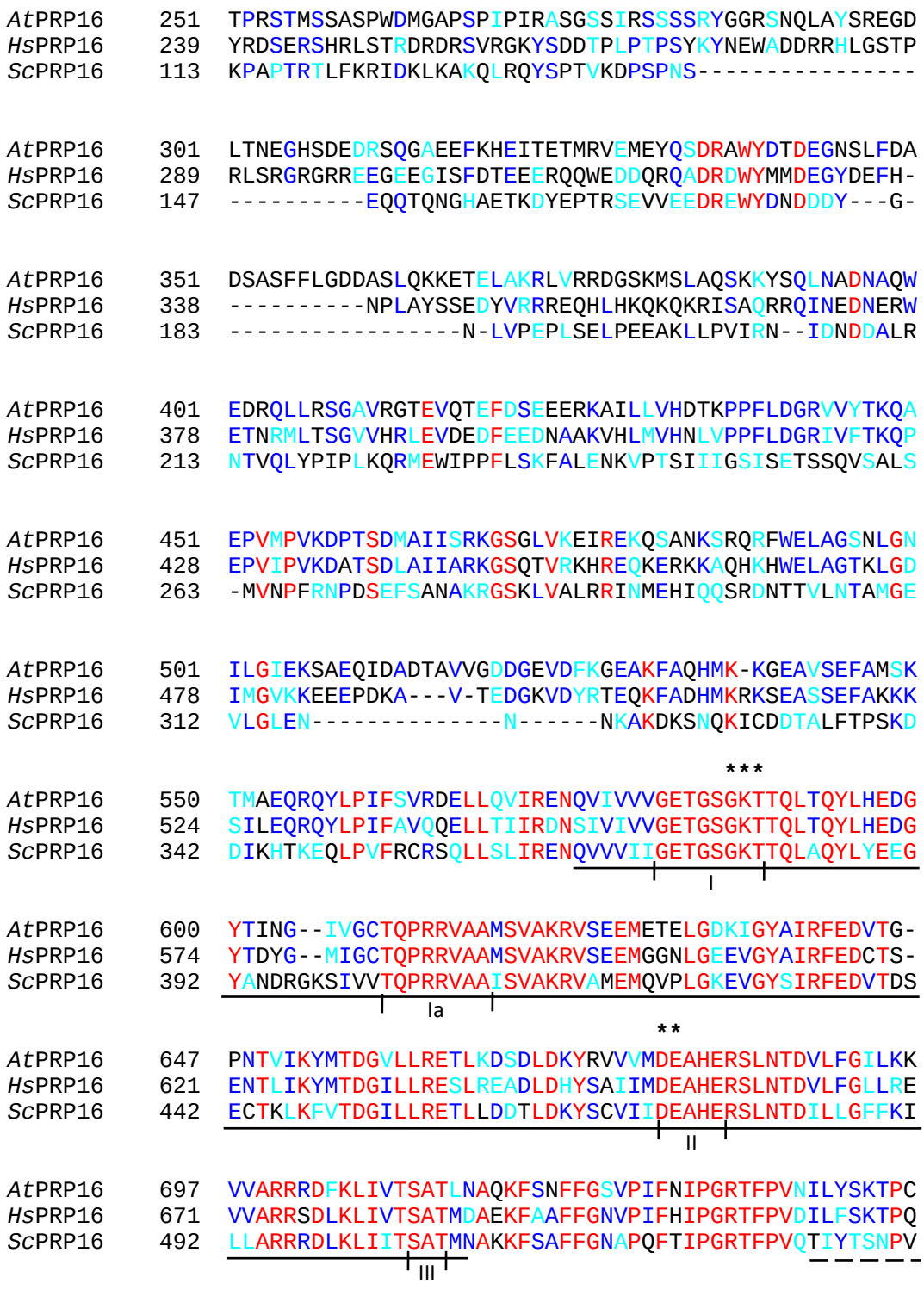


Figure 2.2 (cont.)

B (cont.)

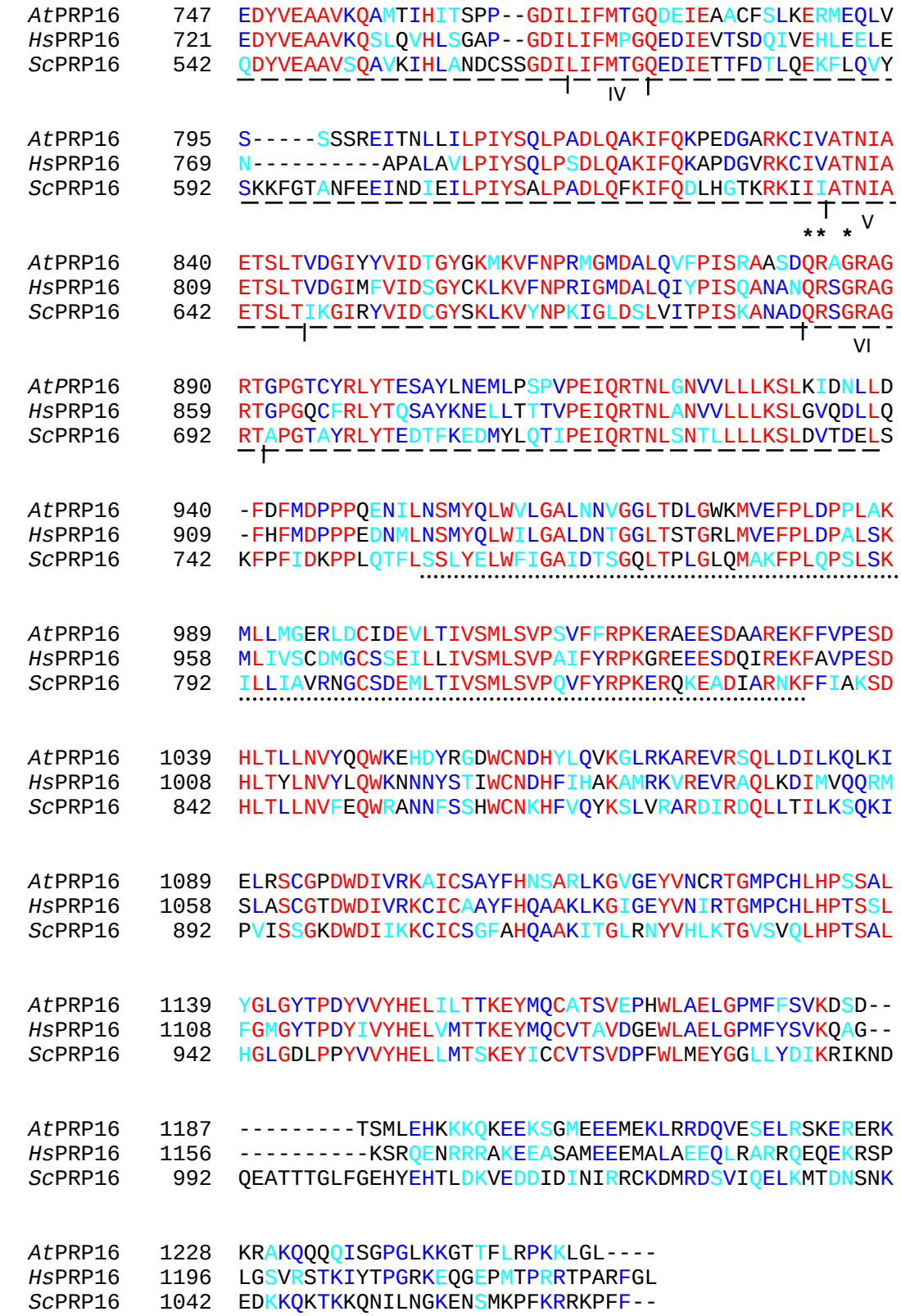


Figure 2.2 (cont.)

C

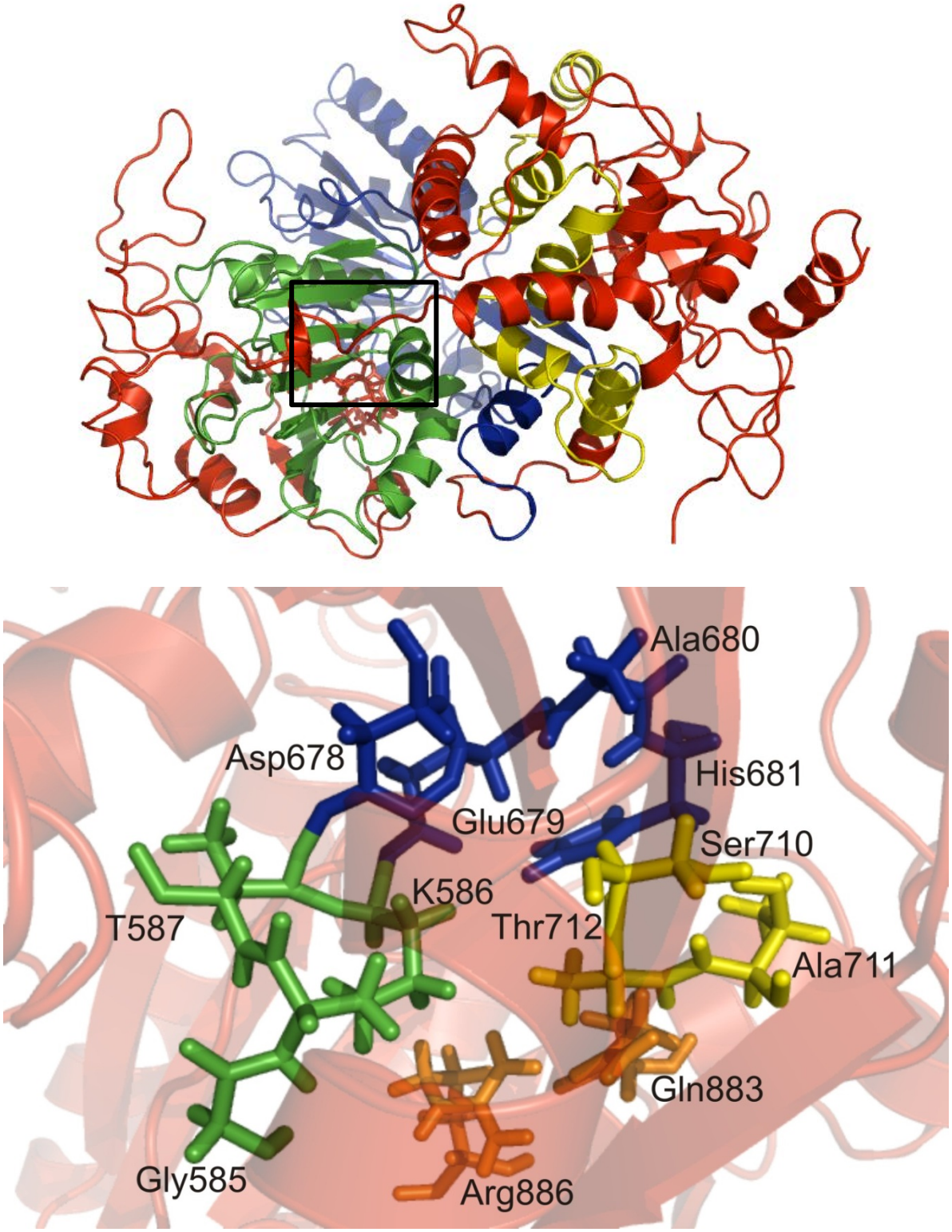
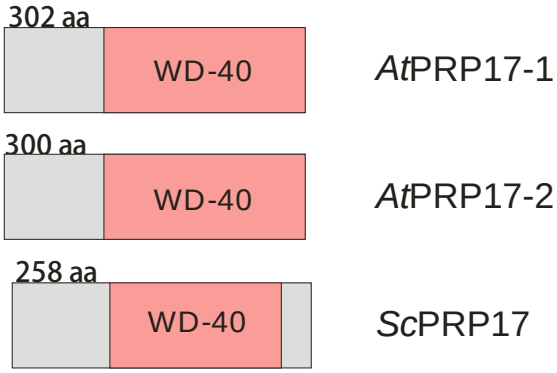


Figure 2.2 (cont.)

A



B

```
AtPRP17-1      1  -----MDLIQSYEEEDEAVASSPESSP----------PRMLKAKSSAPEV
AtPRP17-2      1  ----MDLICNSYANDSDDEPEPVTNER----------------------
  HsPRP17      1  MSAAIAALAASYGSGSGSESDSDSESSRCPLPAADSLMHLTKSPSSKPSL
  ScPRP17      1  -----MGLVDGYDTSSDSDLNFDEGKS----------------------

AtPRP17-1     36  DDTALALTVANVNQSKSK--PIDPTQHVVFYNPTHDQLWAPMFGPAHPYA
AtPRP17-2     24  ----LTSKIAAAPSIPSK--RPYPVPEERQYKP----------PRRP--
  HsPRP17     51  AVAVDSAPEVAVKEDLETGVHLDPAVKEVQYNPTYETMFAPEFGPENPFR
  ScPRP17     23  ----------VHEKKNG-------NLHEDTS----------YEPSS---

AtPRP17-1     84  KDGIAQGMRNHKLGSVEDASIGSFGFEEQYHTFHKCGYAADPSGMN----
AtPRP17-2     55  ------------------------------------------------
  HsPRP17    101  TQQMAA-PRNMLSGYAEPAHINDFMFEQQRRTFATYGYALDPSLDNHQVS
  ScPRP17     42  ----------------------------------------------N----
```

Figure 2.3 (A) Domain maps of PRP17 proteins. The putative scale domain map of *At*PRP17s are shown in comparison to the domain map of *Sc*PRP16. **(B) Sequence alignments of PRP17**. Shown are amino acid sequences of *At*PRP17-1 and *At*PRP17-2, human and *Sc*PRP17s. Identical residues are red, conserved residues are royal blue and similar residues are light blue. WD40 domain is underlined with a solid line. Seven WD40 repeats are present, indicated by the vertical hash marks. The *Sc* 235TETG238 region that is required for splicing is highlighted in yellow. In PRP17-2 of *At*, the two Ts are substituted with V and R, respectively and are highlighted in green (Lindsey-Boltz et al., 2000). The region between amino acids 42 and 61 are required for localization to the nucleus in yeast (highlighted in pink) and these amino acids are in the least conserved N-terminal region of PRP17. **(C) Homology model of *At*PRP17-1 and *At*PRP17-2**. Top panel: *At*PRP17-1 protein encompassing residues 261through the C-terminus of the protein modeled against human WDR5. Bottom panel: *At*PRP17-2 protein encompassing residues 144 through the C-terminus of the protein modeled against human WDR5. In both panels, each WD40 repeat is colored differently. In order from the N-terminus to the C-terminus, the colors are green, red, yellow, blue, brown, magenta and orange. On each model, the residues corresponding to the yeast 235TETG238 region are displayed as sticks, highlighted in yellow and enclosed in a box.

B (cont.)

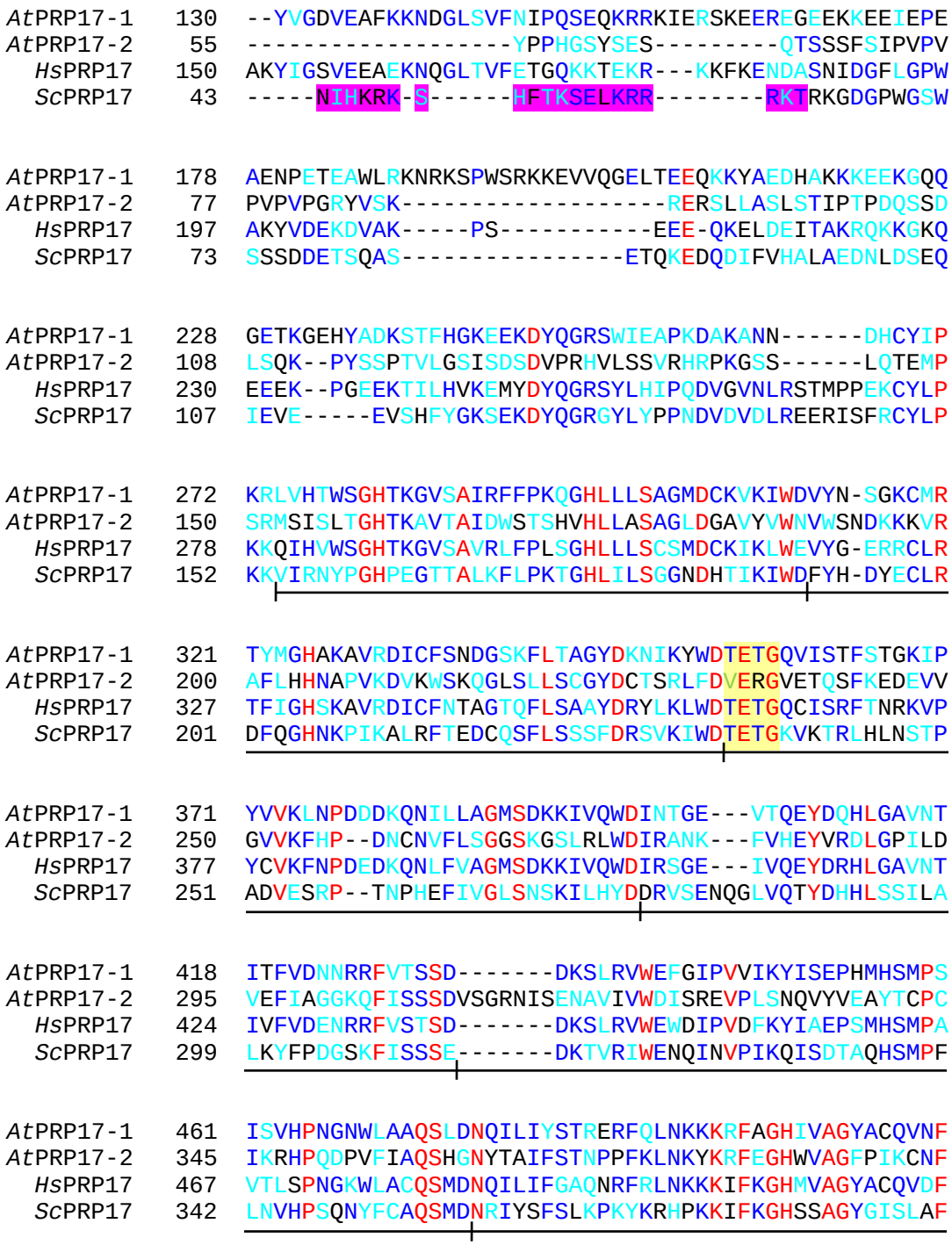


Figure 2.3 (cont.)

B (cont.)

```
AtPRP17-1   511  SPDGRFVMSGDGEGKCWFWDWKSCKVFRTLKCH-NGVCIGAEWHPLEQSK
AtPRP17-2   395  SPDGETLASGSSDGSVYMYDYKSTALIKKLKAY-EQPCVNVSYHPVLPNV
  HsPRP17   517  SPDMSYVISGDGNGKLNIWDWKTTKLYSRFKAH-DKVCIGAVWHPHETSK
  ScPRP17   392  SGDGRYICSGDSKSRLFTWDWNTSRLLNNIKIPGNKPITQVDWHPQETSK

AtPRP17-1   560  VATCGWDGLIKYWD
AtPRP17-2   444  VAACSWNGQVSVFE
  HsPRP17   566  VITCGWDGLIKLWD
  ScPRP17   442  VICSGAAGKIYVCD
```

C

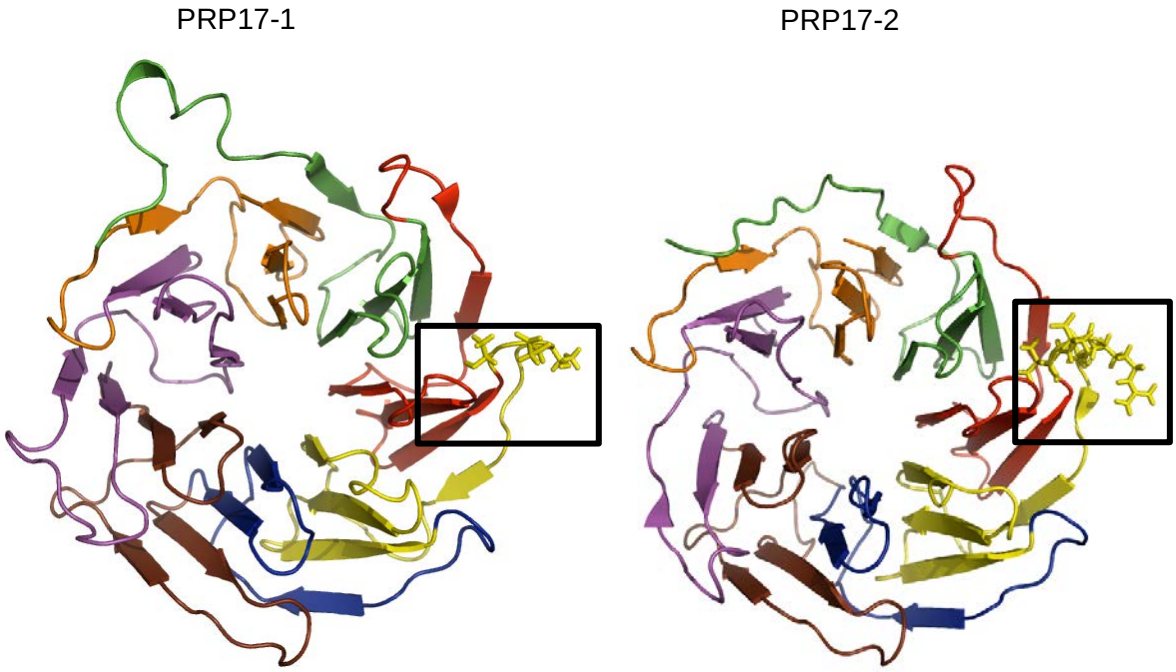
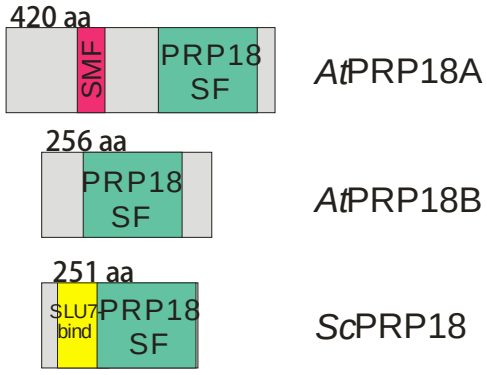


Figure 2.3 (cont.)

A



B

```
AtPRP18A      1  MDLLREEILKKRKSLAEESGGKKFFKRSEIEQKKIQKLREEERREHELKA
AtPRP18B      1  -----MVVSKSTNSRDDNKHLGADENPNEFINSKCFKLAS------HFKT
 HsPRP18      1  MDILKSEILRKRQ-LVEDRNLLVE-NKKYFKRSELAKKEEE----AYFER
 ScPRP18      1  --------------------------------------------------

AtPRP18A     51  QRRAAAAASGGDGKSSGSAPGSSNAATSASSKSSASDAAAIADSKALTDE
AtPRP18B     40  RDFRFGQLSRKDGNNFK-------------------------SQRVTHR
 HsPRP18     45  CGYKIQPKEEDQKPLTSSNP---------------------VLELELAEE
 ScPRP18      1  --MDLDLASILKGEISKK-----------------------KKELANS

AtPRP18A    101  N--LILPRQEVIRRLRFLKQPMTLFGEDDQSRLDRLKYVLKE-GLFEVDS
AtPRP18B     64  RDTQP------N---RGRGSNLTYSWMSFIVRLDRLKYVLKEGLFEEVDN
 HsPRP18     74  KLPMTLSRQEVIRRLRERGEPIRLFGETDYDAFQRLRKIEILTPEV----
 ScPRP18     24  KGVQPP--------CTEKFQPHESANIDETPRQVEQE--------STDEE
```

Figure 2.4 (A) Domain maps of PRP18 proteins. The putative scale domain map of *At*PRP18s is shown in comparison to the domain map of *Sc*PRP18. **(B) Sequence alignments of PRP18**. Shown are amino acid sequences of *At*PRP18A and *At*PRP18B, *Hs* and *Sc*PRP18s. Identical residues are red, conserved residues are royal blue and similar residues are light blue. Five α-helix regions are underlined. Helical regions 1 and 2 are important for SLU7 interaction and the following residues contributing to surface charge: H118, K125, K140, K141, R151 and R152, indicated by arrows. In *At*PRP18A, asterisks indicate the basic residues that are predicted to be surface amino acids and would likely interact with SLU7. These residues do not necessarily align with the identified yeast surface residues that are necessary for SLU7 interaction; H118, R151 and R152 are essential for SLU7 interaction and amino acids R80, K140 and K141. SMF domains in *At*PRP18A and *Hs*PRP18 are highlighted in yellow. The SLU7-binding region in *Sc*PRP18 is highlighted in green. **(C) Homology model of *At*PRP18A and *At*PRP18B**. Top panel: *At*PRP18A protein encompassing residues 195 through 373 of the protein modeled against *Sc*PRP18. Bottom panel: *At*PRP18B protein encompassing residues 144 through 315 of the protein modeled against *Sc*PRP18. In both panels, helix 1 is colored blue and helix 2 is colored green. Residues that are predicted to be involved in SLU7 interaction are displayed in stick format

B (cont.)

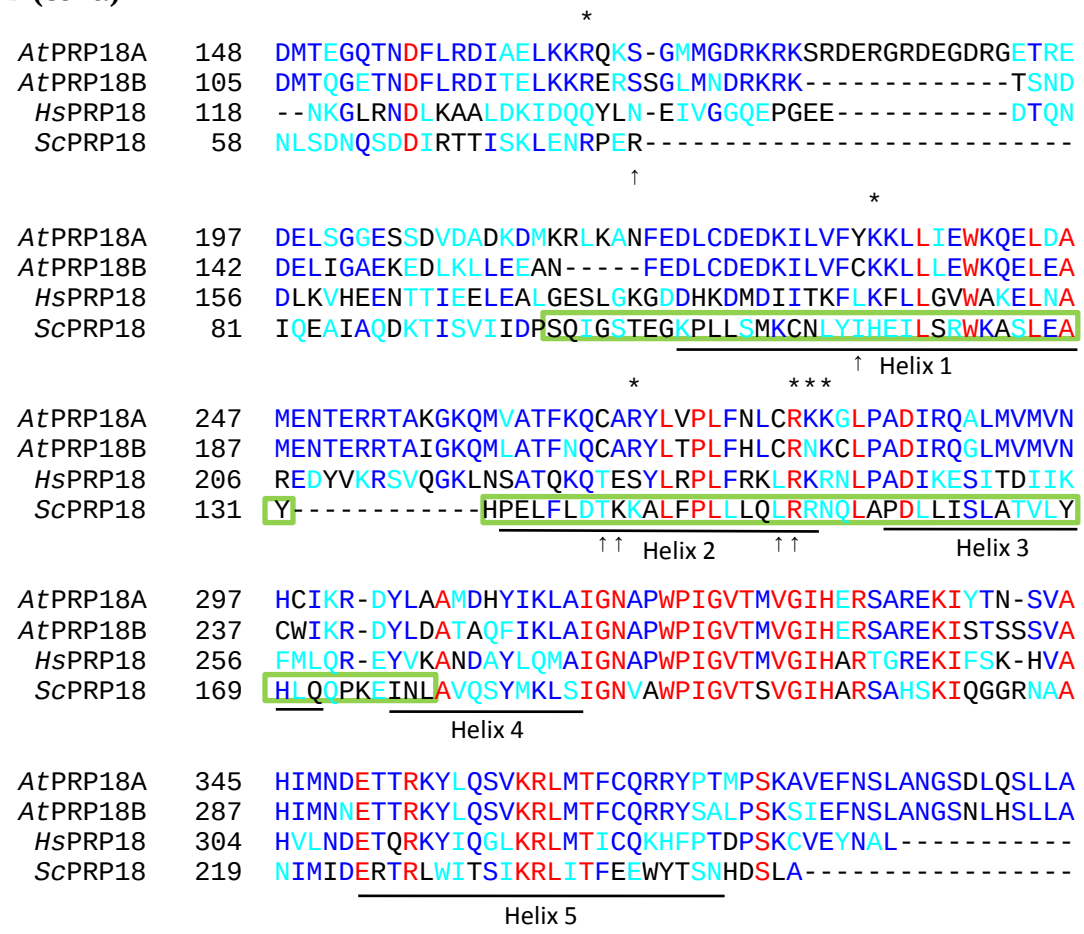


Figure 2.4 (cont.)

C

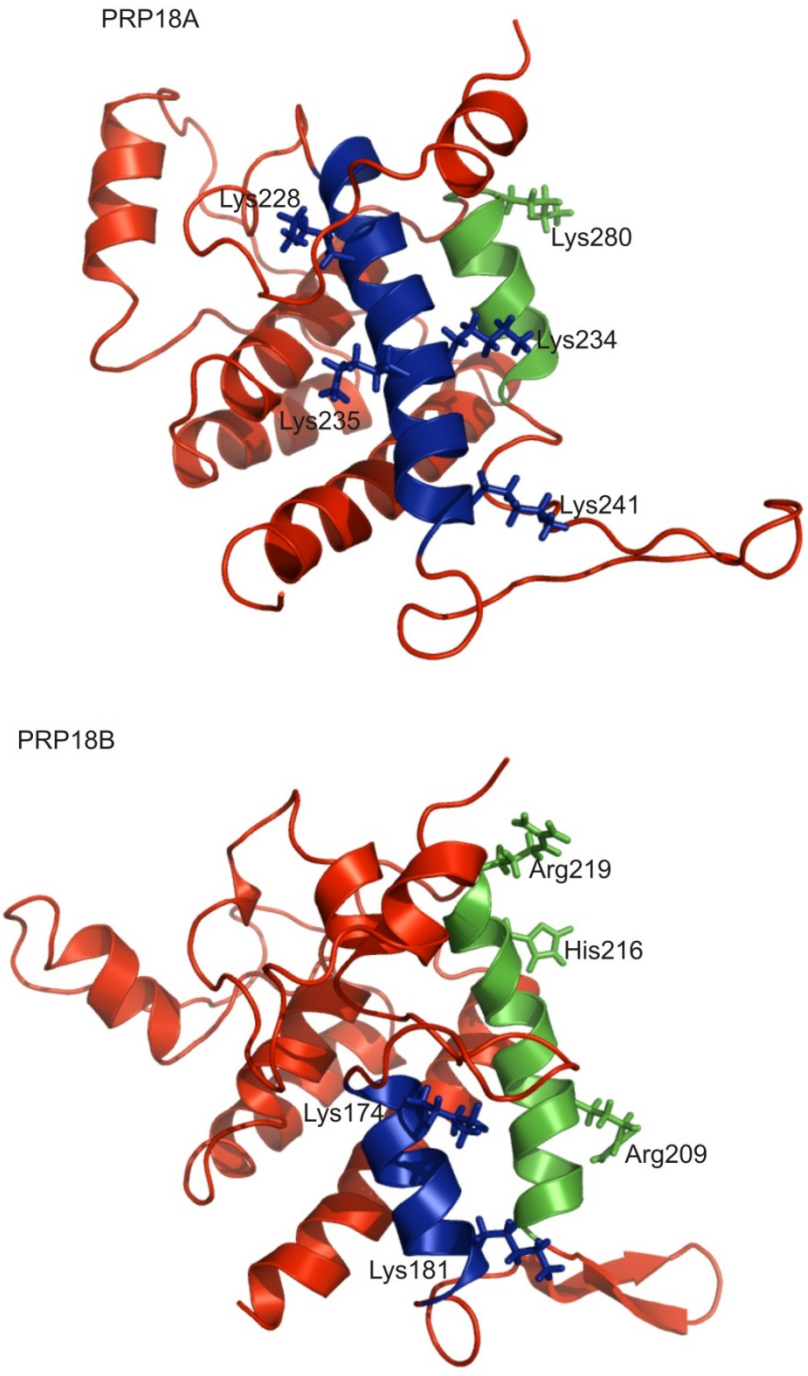


Figure 2.4 (cont.)

A

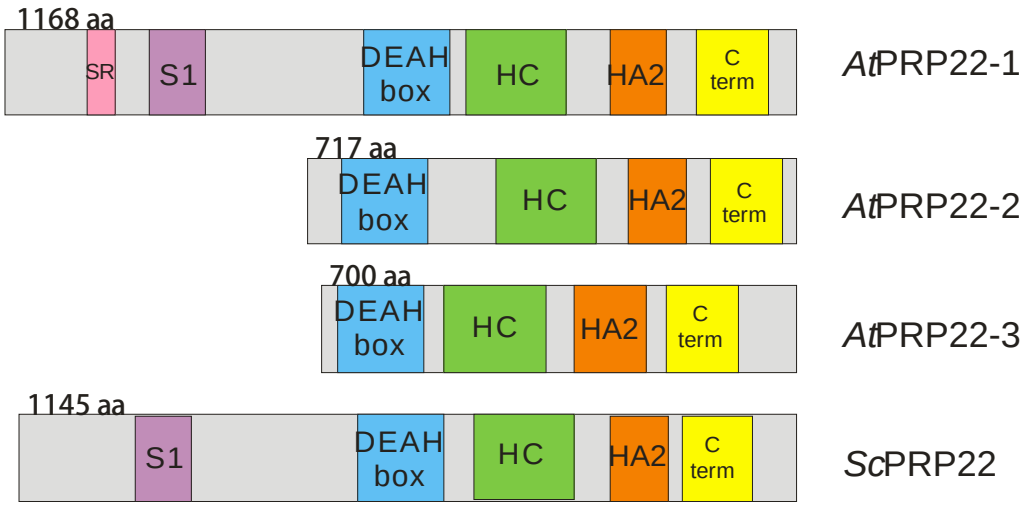


Figure 2.5 (**A**) **Domain maps of PRP22 proteins**. The putative scale domain map of *At*PRP22s is shown in comparison to the domain map of *Sc*PRP22. (**B**) **Sequence alignments of PRP22**. Shown are amino acid sequences of *At*PRP22-1, *At*PRP22-2 and *At*PRP22-3, *Hs*HRH1 and *Sc*PRP22. Identical residues are red, conserved residues are royal blue and similar residues are light blue. DExDc superfamily domain is underlined with a solid line with brackets indicating motifs I-VI. The Helicase-C domain is underlined with a dashed lined and the HA2 domain is underlined with a dotted line. The C-term OB domain is underlined in cyan. A putative RS domain is found only in *Hs*HRH1 and *At*PRP22-1 and is bracketed with a black box. The Walker A motif (GKT) in the helix 1 region is indicated by asterisks. The S1 motif, found only in *At*PRP22-1 and in *Sc*PRP22 bracketed with a yellow box. *Sc* residues S635 and T637 are necessary for helicase activity and are indicated with an asterisk. (**C**) **Homology model of *At*PRP22-1**. Top panel: *At*PRP22-1 protein encompassing residues 427 through1160 modeled against *Sc*PRP43. The DEAD box superfamily domain is colored green, the Helicase-C domain is colored blue, the HA2 domain is colored yellow and the OB domain is colored cyan. Bottom panel: The residues necessary for ATPase activity are highlighted in stick format. The 543GKT545 of motif I is green, the 635DEAH638 of motif II is blue, the 666SAT668 of motif III is yellow. (**D**) **Homology model of *At*PRP22-2**. Top panel: *At*PRP22-2 protein encompassing the entire protein modeled against *Sc*PRP43. The DEAD box superfamily domain is colored green, the Helicase-C domain is colored blue, the HA2 domain is colored yellow and the OB domain is colored cyan. Bottom panel: The residues necessary for ATPase activity are highlighted in stick format. The 69GKT71 of motif I is green, the 635DEAH638 of motif II is blue, the 235SAS237 of motif III is yellow. (**E**) **Homology model of *At*PRP22-3**. Top panel: *At*PRP22-3 protein encompassing the N-terminus through residue 673 modeled against *Sc*PRP43. The DEAD box superfamily domain is colored green, the Helicase-C domain is colored blue, the HA2 domain is colored yellow and the OB domain is colored cyan. Bottom panel: The residues necessary for ATPase activity are highlighted in stick format. The 32GKS34 of motif I is green, the 124DEAH127 of motif II is blue, the 157SAT159 of motif III is yellow.

B

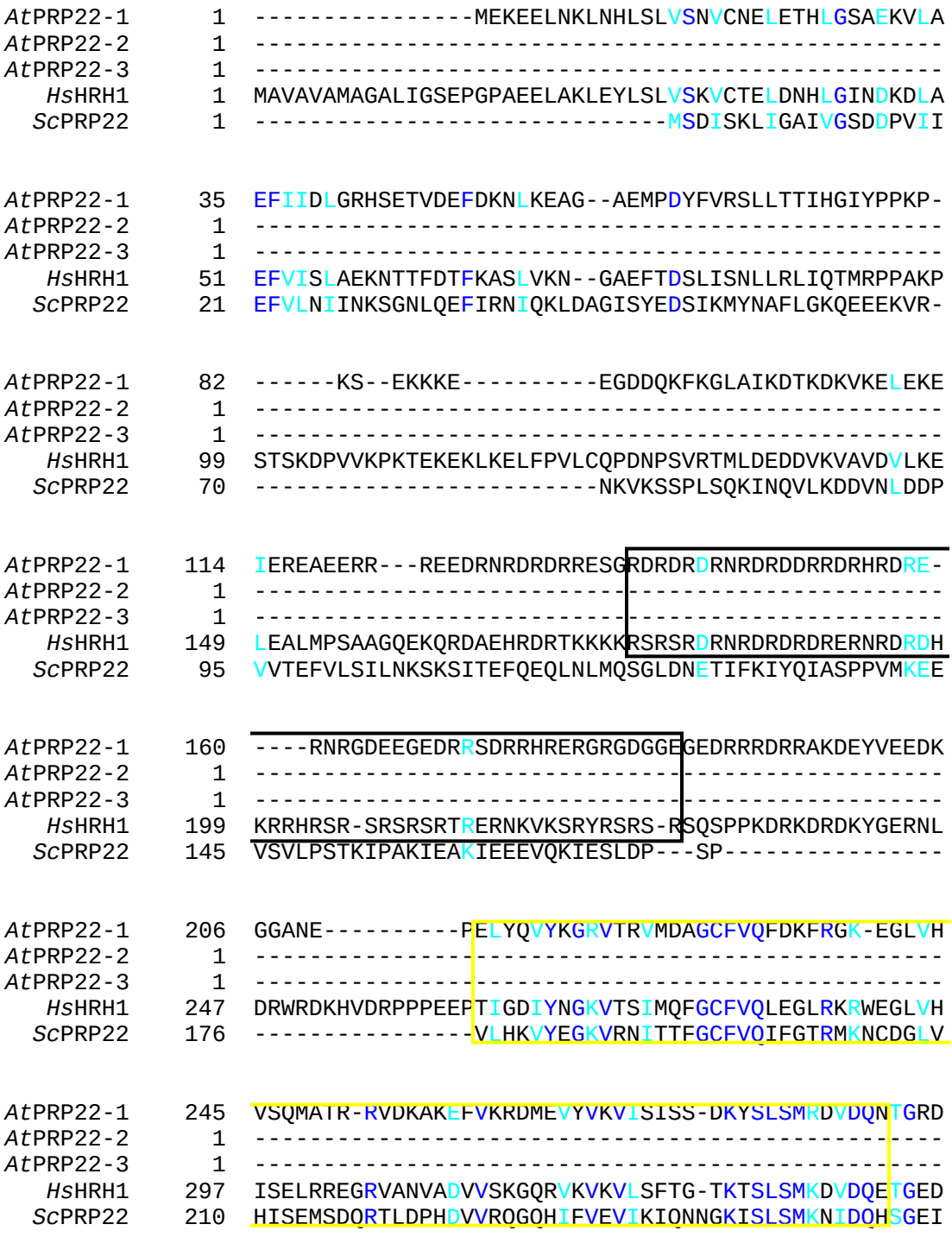


Figure 2.5 (cont.)

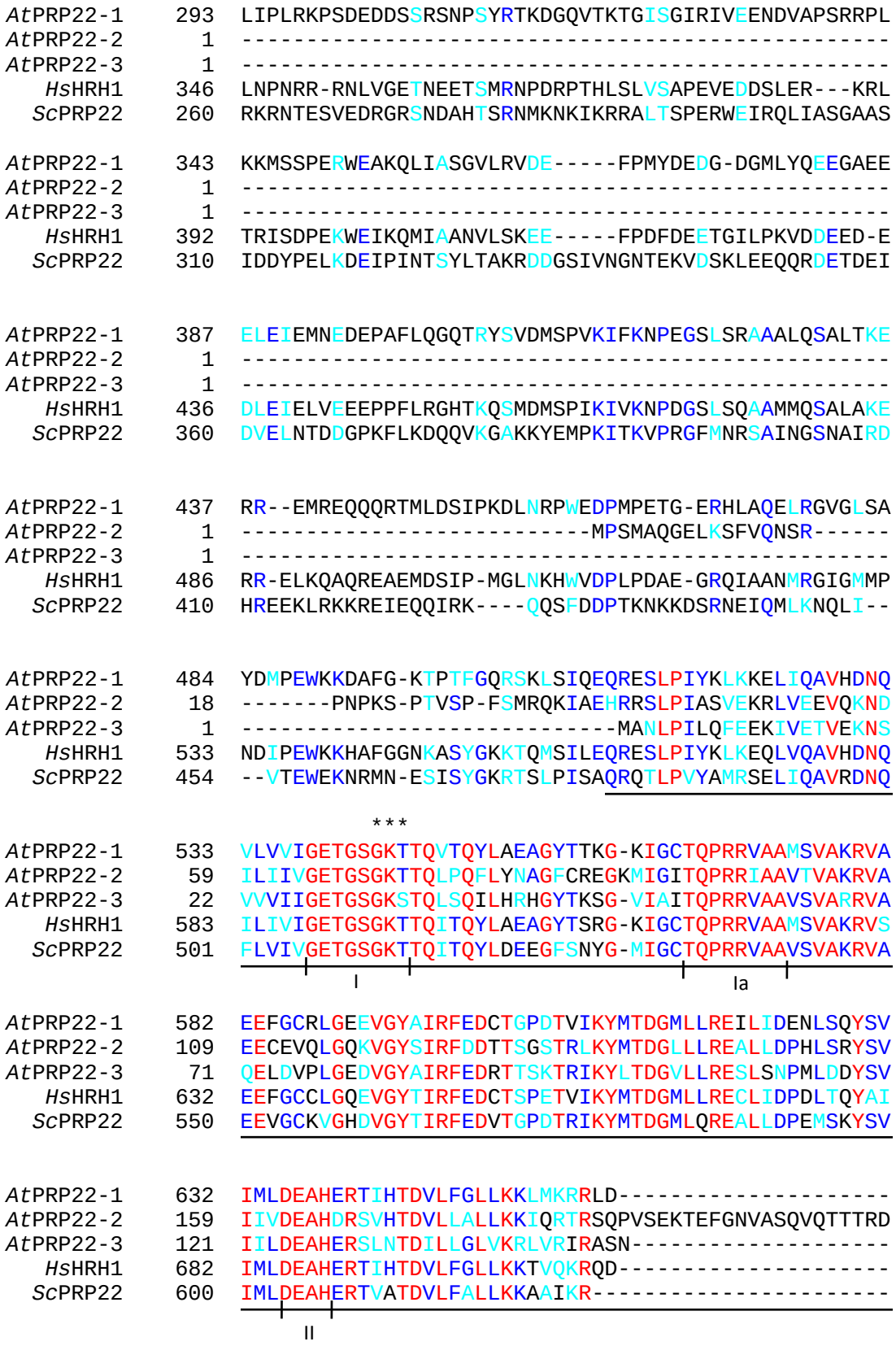


Figure 2.5 (cont.)

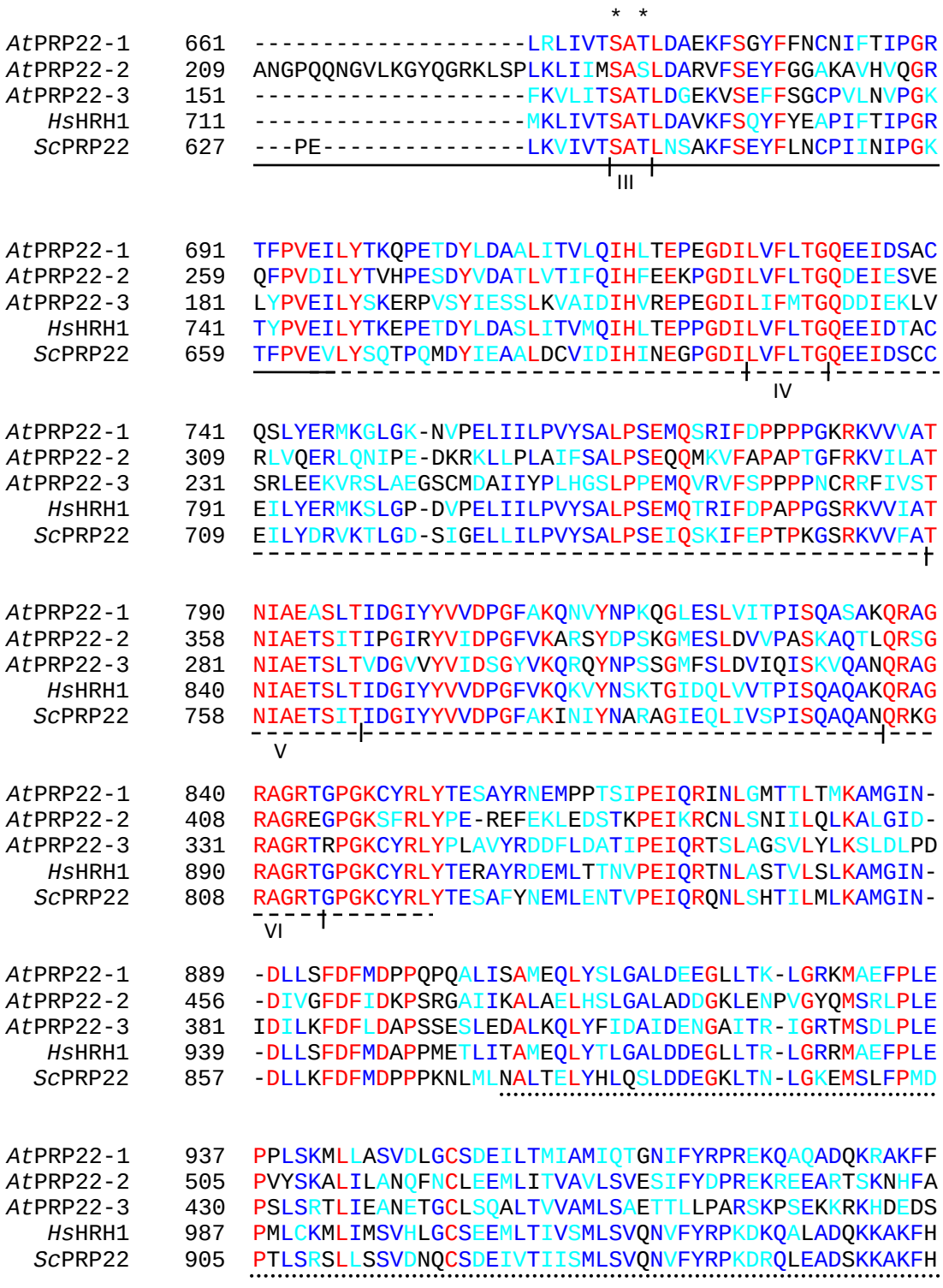


Figure 2.5 (cont.)

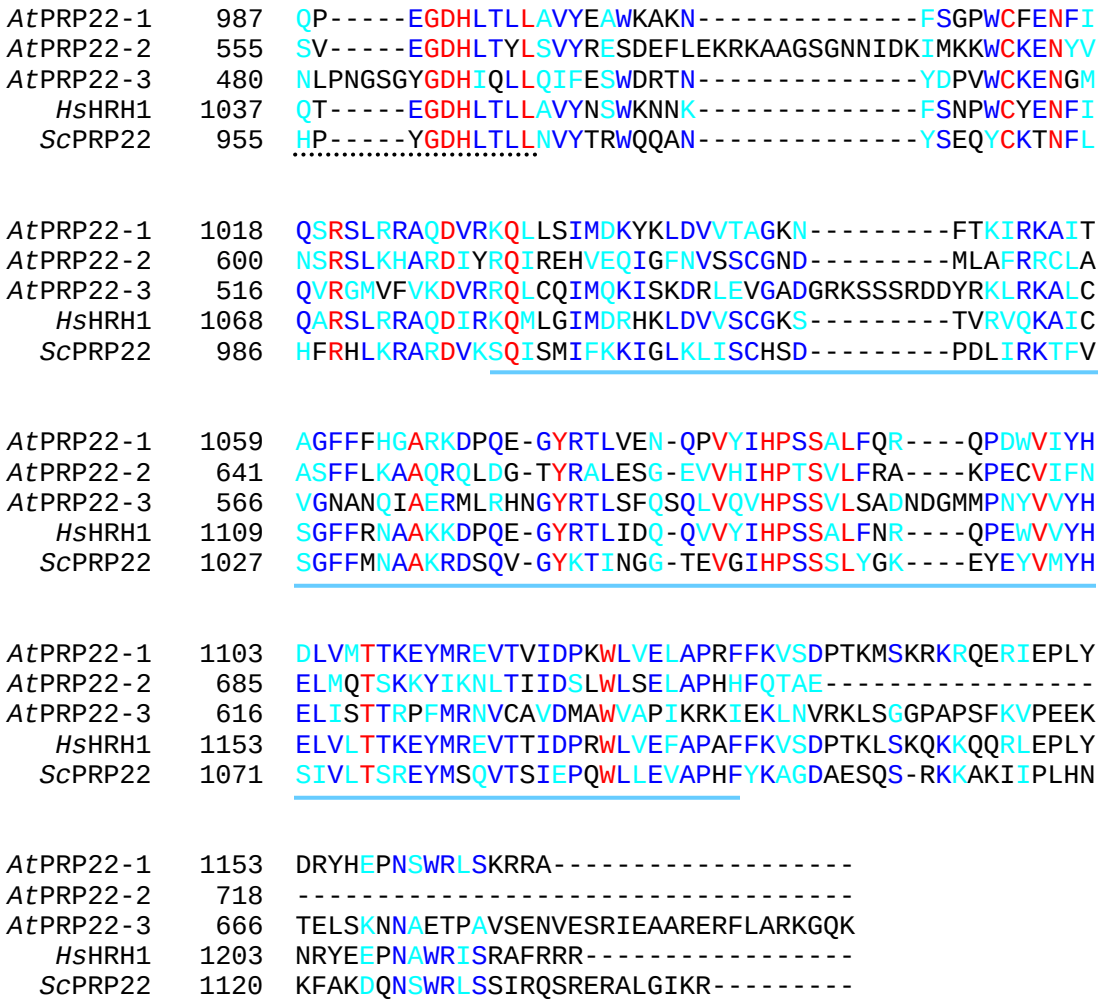


Figure 2.5 (cont.)

C

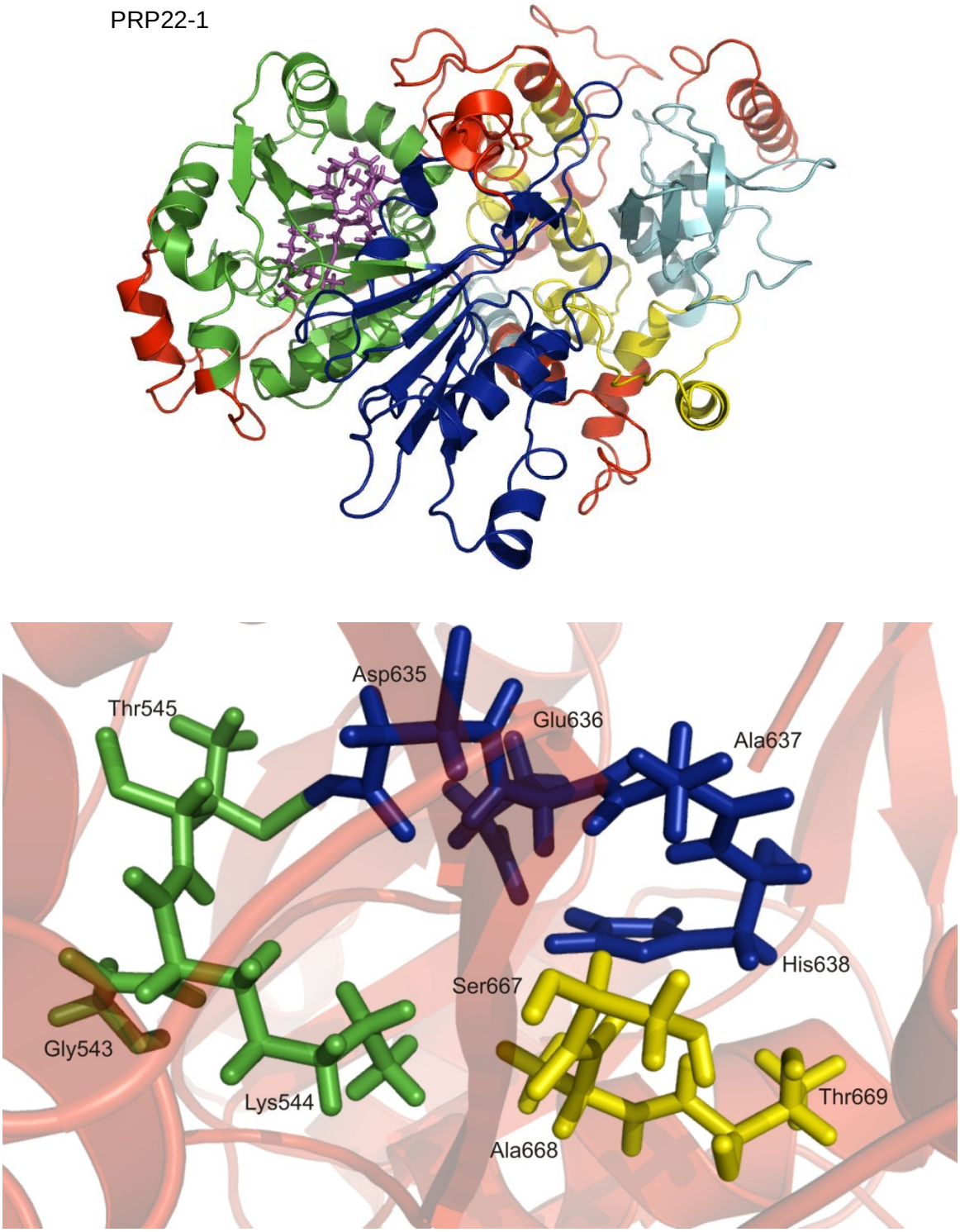


Figure 2.5 (cont.)

D

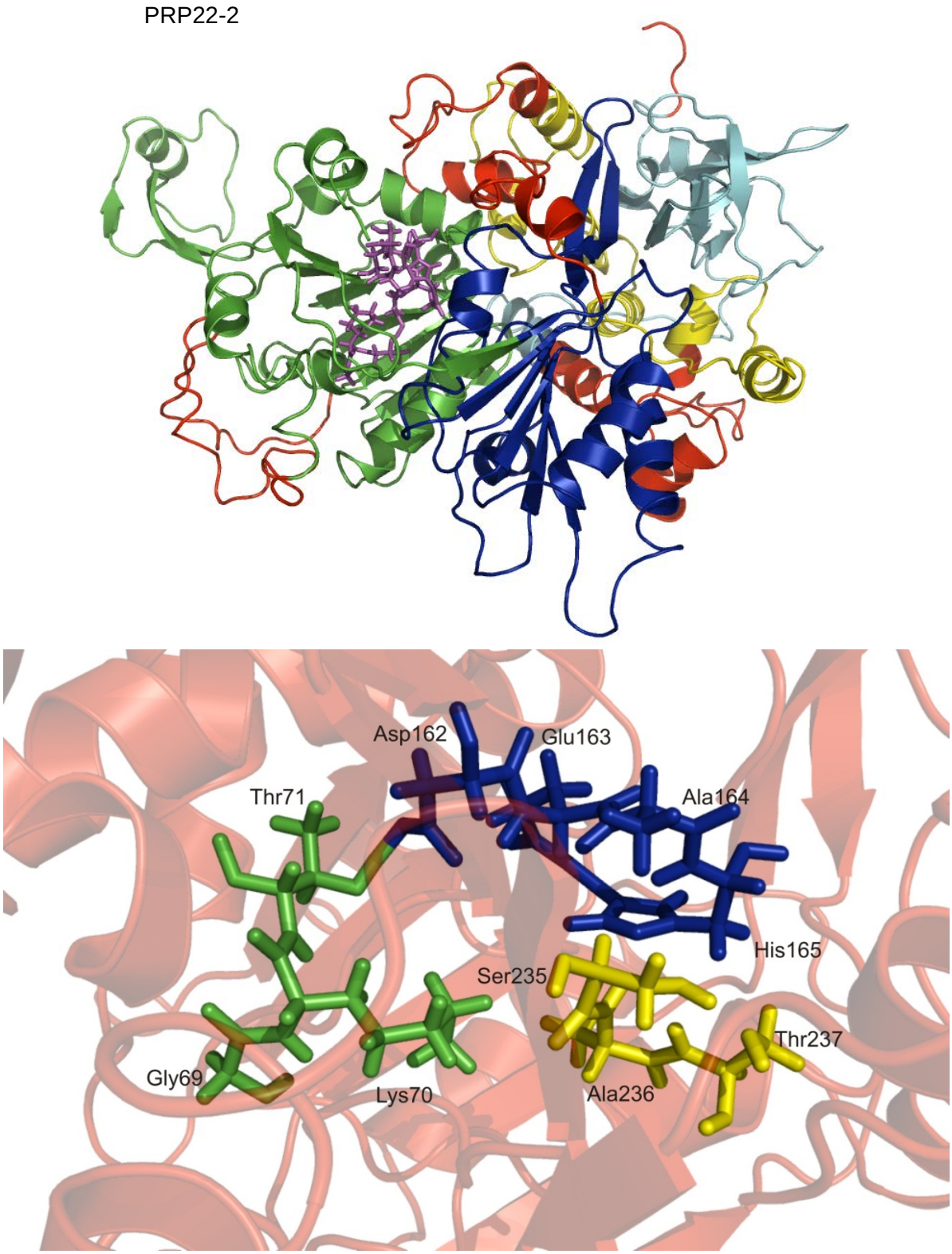


Figure 2.5 (cont.)

E

PRP22-3

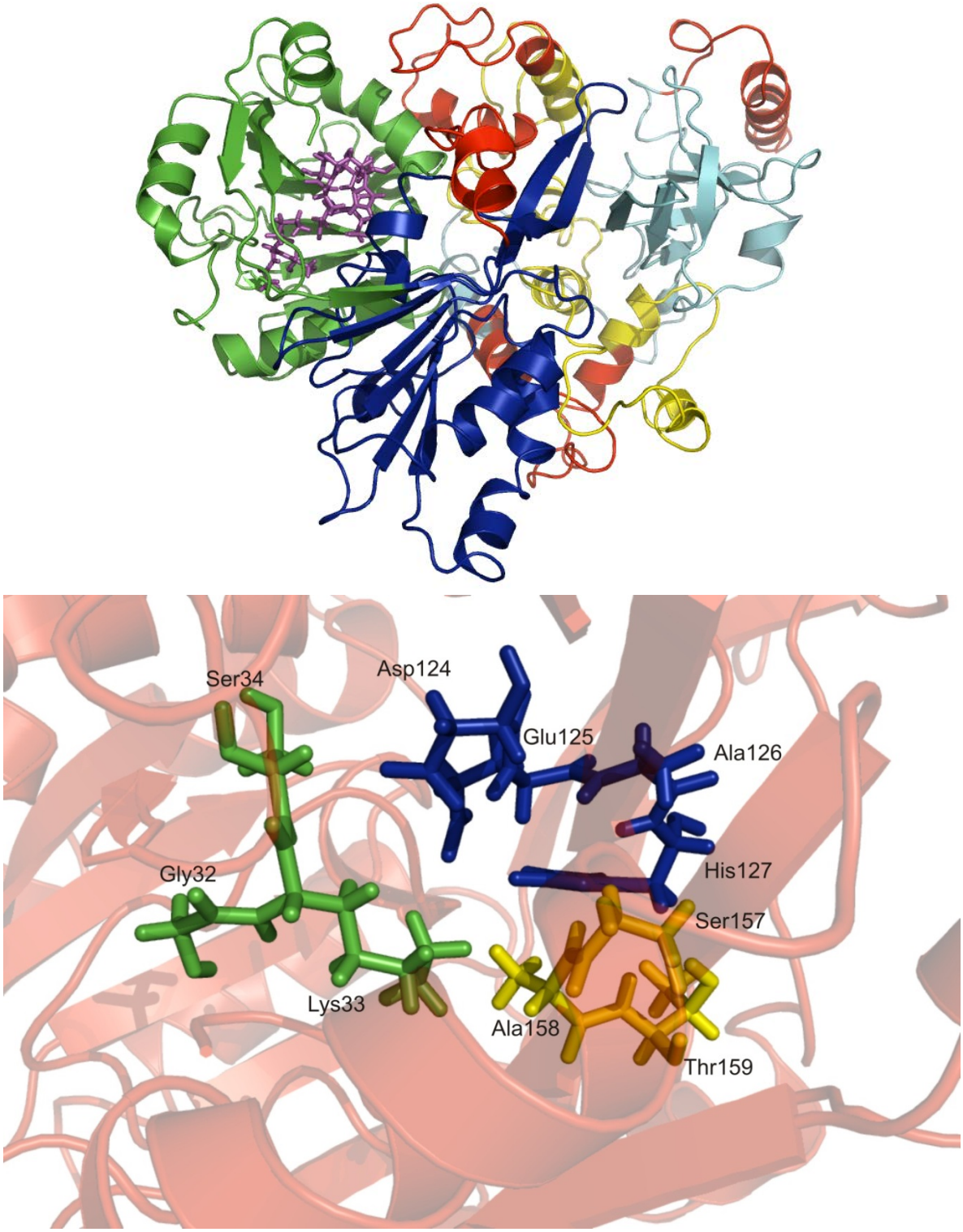
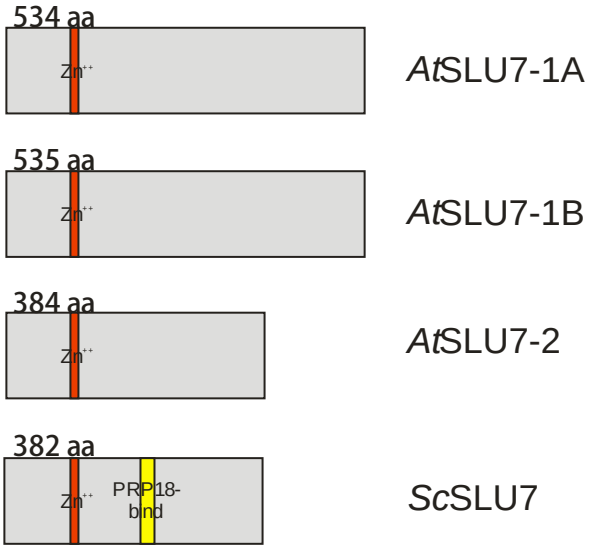


Figure 2.5 (cont.)

A



B

```
AtSLU7-1A      1  ------------------MATASVAFKSR-EDHRKQIELEEARKAGLAP
AtSLU7-1B      1  ------------------MATASVAFKSR-EDHRKKLELEEARKAGLAP
 AtSLU7-2      1  ------------------MATASVAFKSR-KDHRKQKELEEARKAGLAP
   HsSLU7      1  MSATVVDAVNAAPLSGSKEMSLEEPKKMTR-EDWRKKKELEEQRKLGNAP
   ScSLU7      1  ------------------MNNNSRNNENRSTINRNKRQLQQAK------

AtSLU7-1A     31  AEVDEDGKEINPHIPQYMSSAPWYLN-SEKPS-----------------
AtSLU7-1B     31  AEVDEDGKEINPHIPEYMSKAPWYLK-SEQPS-----------------
 AtSLU7-2     31  AEVDEGGKEINLHIPKYLTIPPLYAK-SEKPS-----------------
   HsSLU7     50  AEVDEEGKDINPHIPQYISSVPWYIDPSKRPT-----------------
   ScSLU7     26  -EKNE-----NIHIPRYIRNQPWYYKDTPKEQEGKKPGNDDTSTAEGGEK

AtSLU7-1A     62  ----LKHQRKWKSDPN--------------YTKSWYDRGAK-IFQAEKYR
AtSLU7-1B     62  ----LKHQKNWKIEPE--------------PKKIWYDRGKK-IYQAEQYR
 AtSLU7-2     62  ----LKHQKNWKTKPV--------------STTSYYDRGAK-TYQAEKYR
   HsSLU7     82  ----LKHQRPQPEKQKQFS-----------SSGEWYKRGVKENSIITKYR
   ScSLU7     70  SDYLVHHRQKAKGGALDIDNNSEPKIGMGIKDEFKLIRPQK-MSVRDSHS
```

Figure 2.6 (A) Domain maps of SLU7 proteins. The putative scale domain map of *At*SLU7s are shown in comparison to the domain map of *Sc*SLU7. **(B) Alignment of SLU7**. Zinc knuckle CCHC motif is underlined in black. The zinc knuckle motif overlaps with the blocked putative NLS region. The non-conserved *Sc*PRP18 binding region is highlighted in yellow. A lysine-rich region only found in *Hs*SLU7 is underlined with a dotted line

B (cont.)

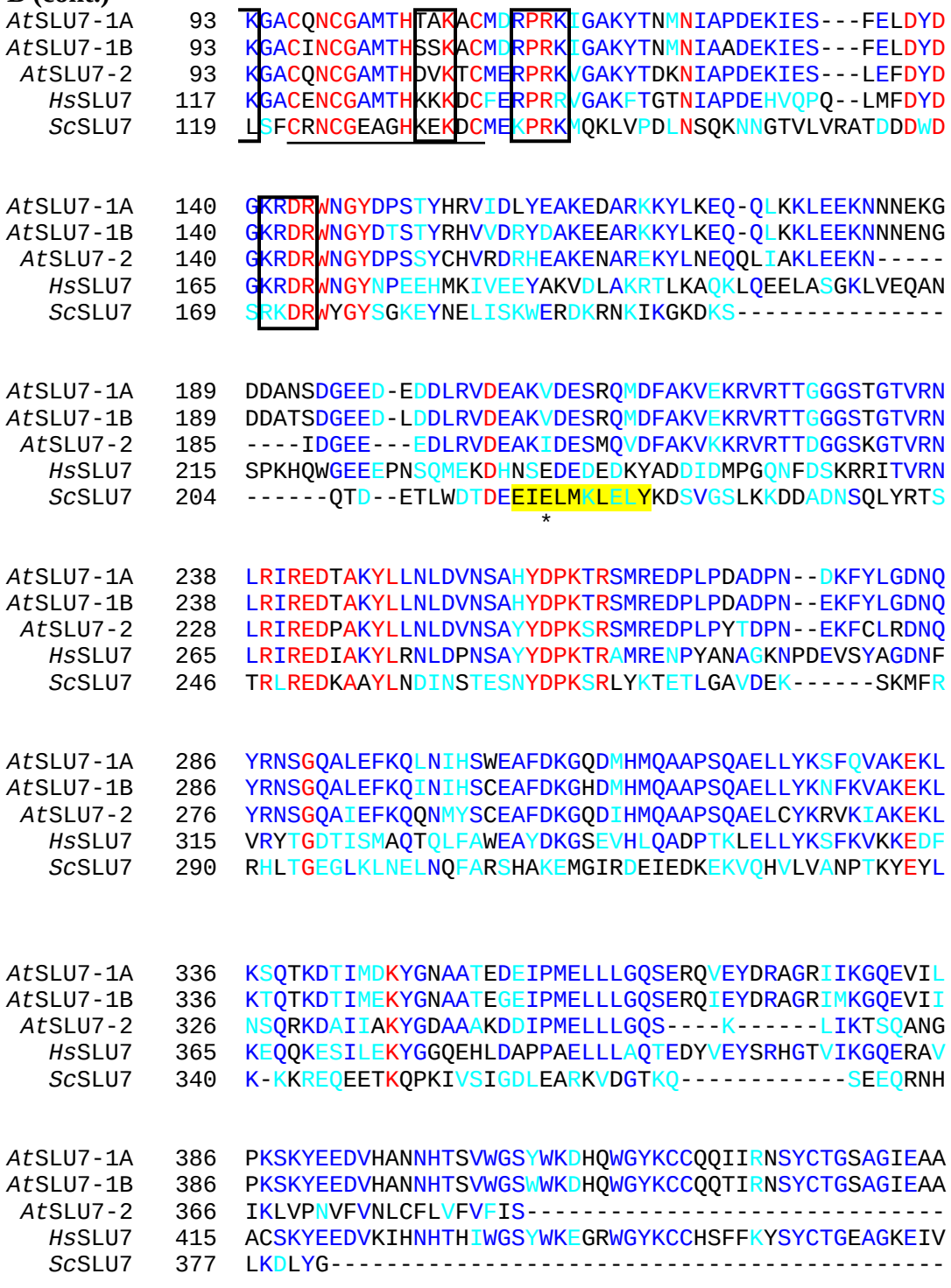


Figure 2.6 (cont.)

```
AtSLU7-1A   436  EAA------------------LDLMKANIARKEATEESPKKVEEKRMAS
AtSLU7-1B   436  EAS------------------IDLMKANIARKEASKESPKKVEEKKMAT
 AtSLU7-2   386  --------------------------------------------------
   HsSLU7   465  NSEECIINEITGEESVKKPQTLMELHQEKLKEEKKKKKKKKKKKHRKSSSD
   ScSLU7   383  --------------------------------------------------


AtSLU7-1A   467  WGTDIPEDLELNEEALANALKKEDLSRREEKDERKRKYNVKYNN-DVTPE
AtSLU7-1B   467  WGTDIPEDLELNEEALANALKKEDLSRREEKDERKRKYNVNYTN-DVTSE
 AtSLU7-2   386  --------------------------------------------------
   HsSLU7   515  SDDEEKKHEKLKKALNAEEARLLHVKETMQIDERKRPYNSMYETREPTEE
   ScSLU7   383  --------------------------------------------------


AtSLU7-1A   516  EMEAYRMKRVHHEDPMKDFL--
AtSLU7-1B   516  EMEAYRMKRVHHEDPMRNFPG-
 AtSLU7-2   386  ----------------------
   HsSLU7   565  EMEAYRMKRQRPDDPMASFLGQ
   ScSLU7   383  ----------------------
```

Figure 2.6 (cont.)

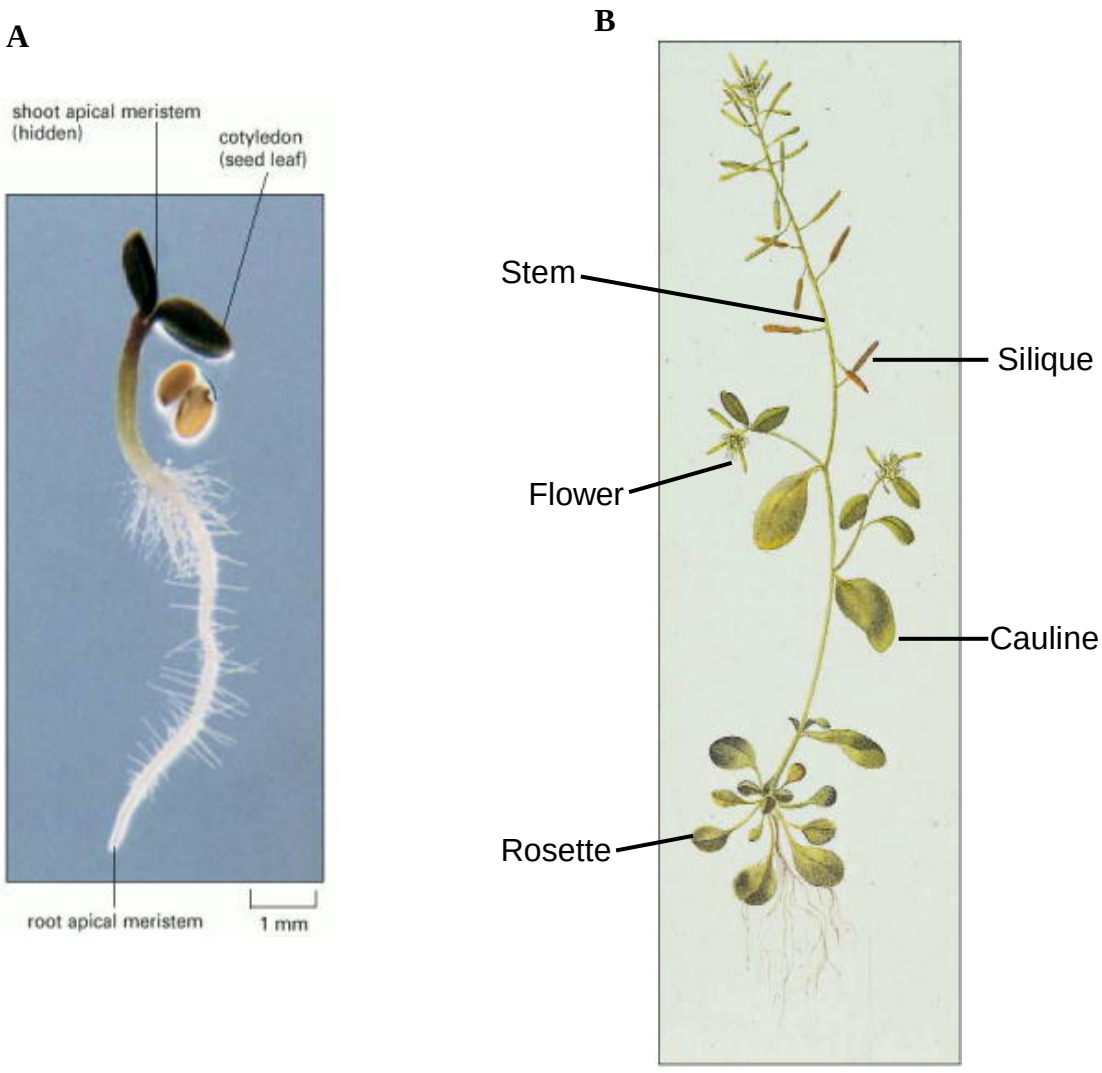


Figure 2.7 Tissue types of *Arabidopsis thaliana*. (A) One-week-old seedling includes the root and the emerging cotyledon. **(B) Mature plant**. These tissues are present at both four and seven weeks (modified from Alberts, 2002).

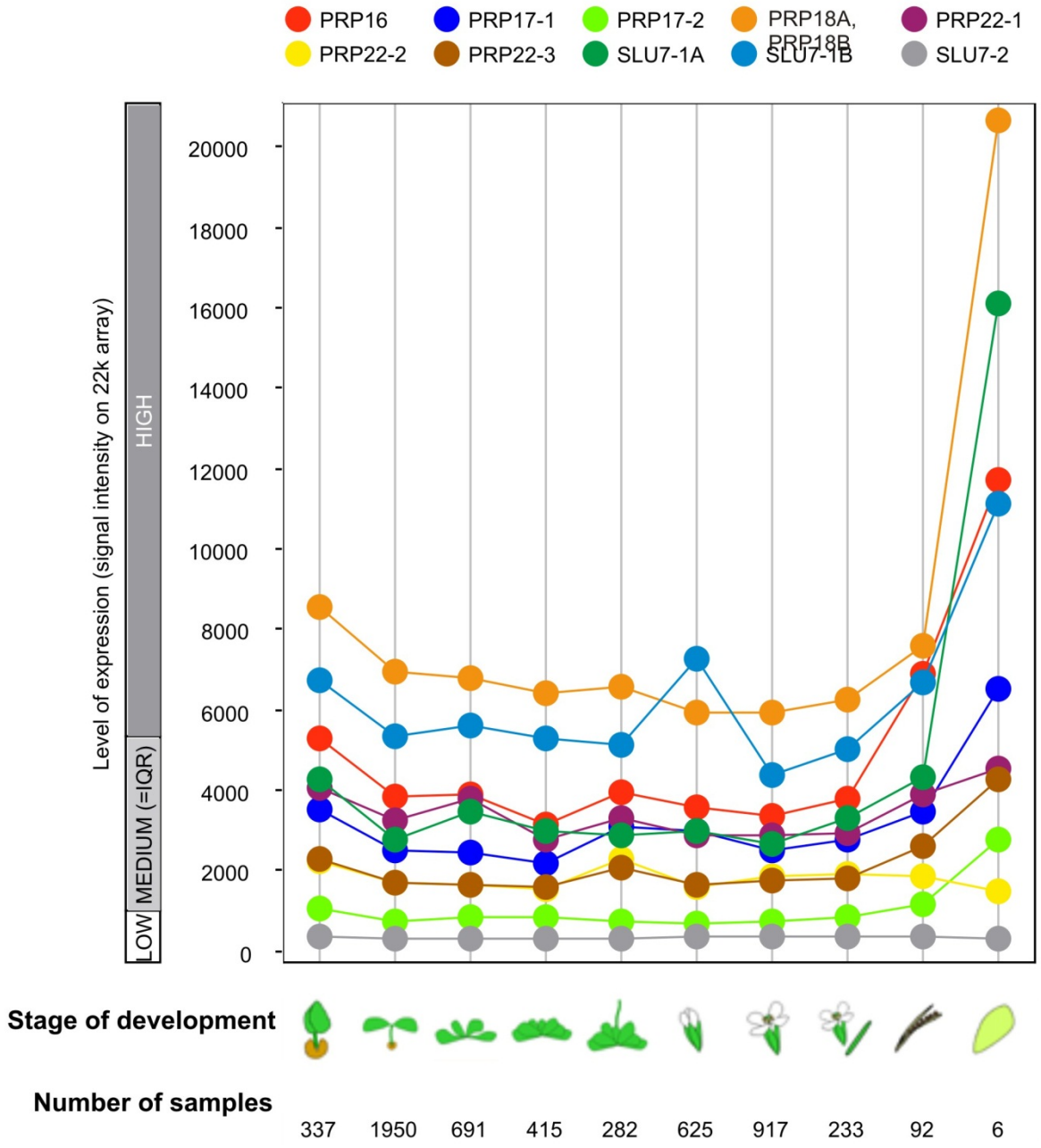
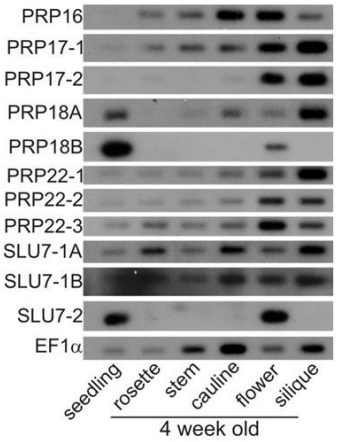
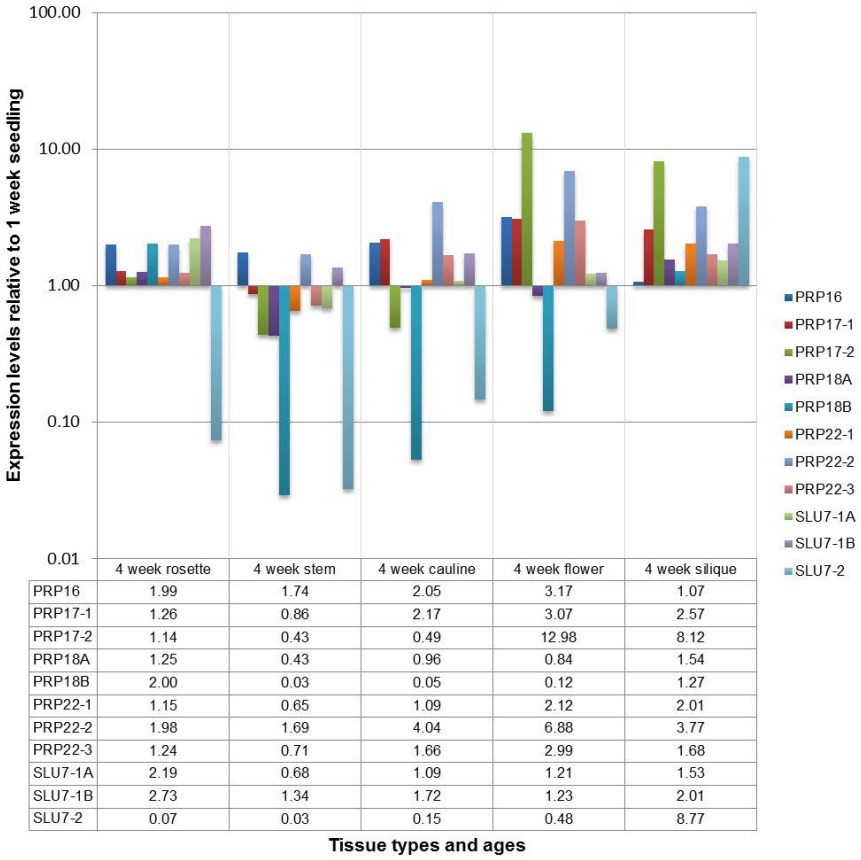


Figure 2.8 Microarray data of expression of second-step splicing factors. Combination of open source 22k microarray data acquired from Genevestigator (Hruz et al., 2008). Raw signal intensity data of each *At* second-step splicing factor is plotted against the stage of development. The number of samples contributing to each data point is indicated at the bottom of the graph. The stages of development from left to right are as follows: germinated seed, seedling, young rosette, developed rosette, bolting, young flower, developed flower, flowers and siliques, mature siliques and senescense. PRP18A and PRP18B probes are unable to distinguish between the two genes.

A





	4 week rosette	4 week stem	4 week cauline	4 week flower	4 week silique
PRP16	1.99	1.74	2.05	3.17	1.07
PRP17-1	1.26	0.86	2.17	3.07	2.57
PRP17-2	1.14	0.43	0.49	12.98	8.12
PRP18A	1.25	0.43	0.96	0.84	1.54
PRP18B	2.00	0.03	0.05	0.12	1.27
PRP22-1	1.15	0.65	1.09	2.12	2.01
PRP22-2	1.98	1.69	4.04	6.88	3.77
PRP22-3	1.24	0.71	1.66	2.99	1.68
SLU7-1A	2.19	0.68	1.09	1.21	1.53
SLU7-1B	2.73	1.34	1.72	1.23	2.01
SLU7-2	0.07	0.03	0.15	0.48	8.77

Figure 2.9 (A) Expression levels of second-step splicing factors in 4-week-old tissue. A representative panel of Southern blot analysis on the left. Experiments were performed in biological triplicate and relative expression is represented on a logarithmic scale. The calculated values are listed below the chart. Performed in biological triplicate; p-value < 0.05. **(B) Model for spliceosomal subsets of second-step splicing factors in 4-week-old tissue**. Heat maps ranging from no expression (brightest red) to the highest observed expression (brightest green) to model subsets of second-step splicing factors in 4 week *At* tissue based on fold change determined by Southern blot analysis.

B

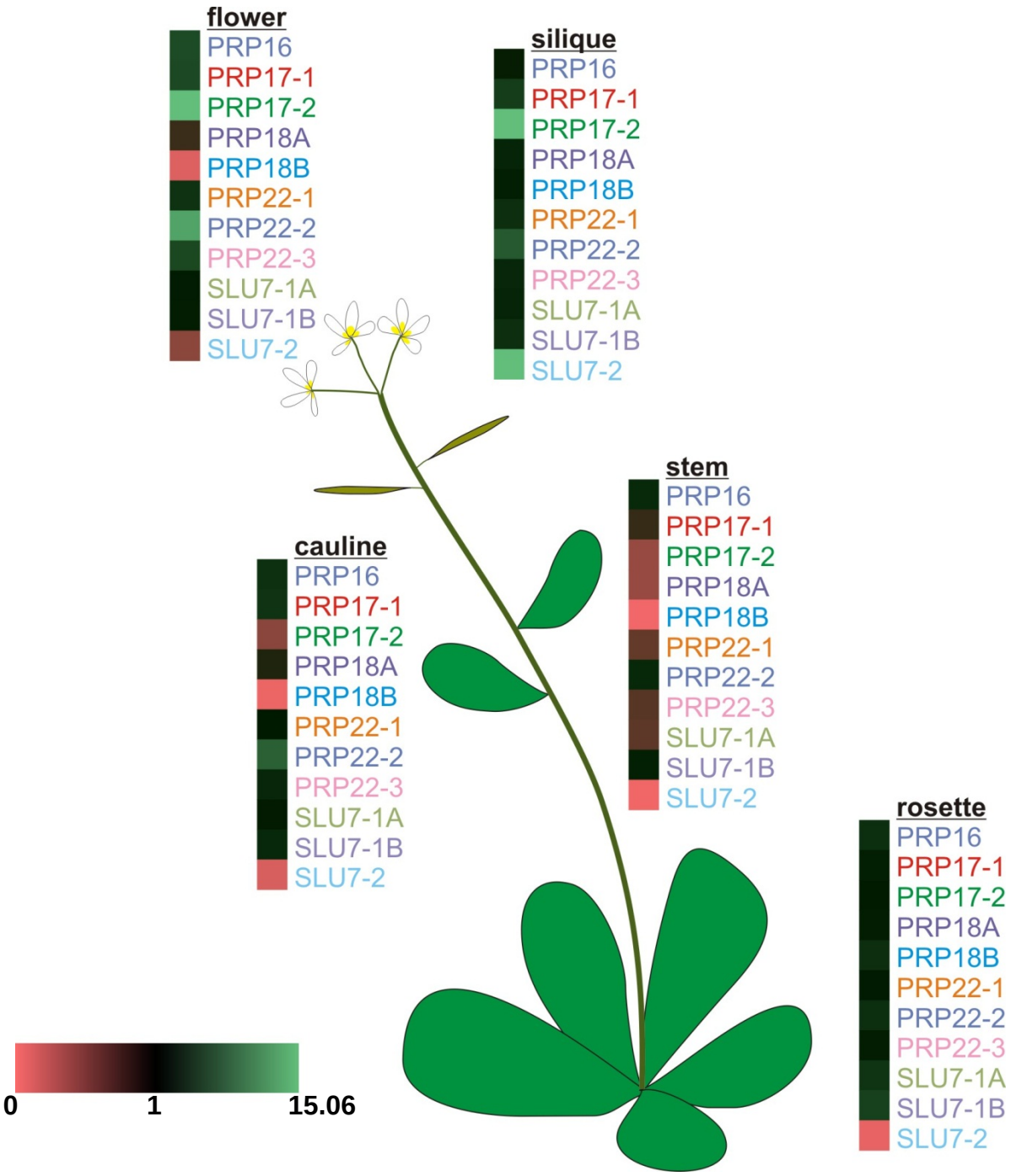
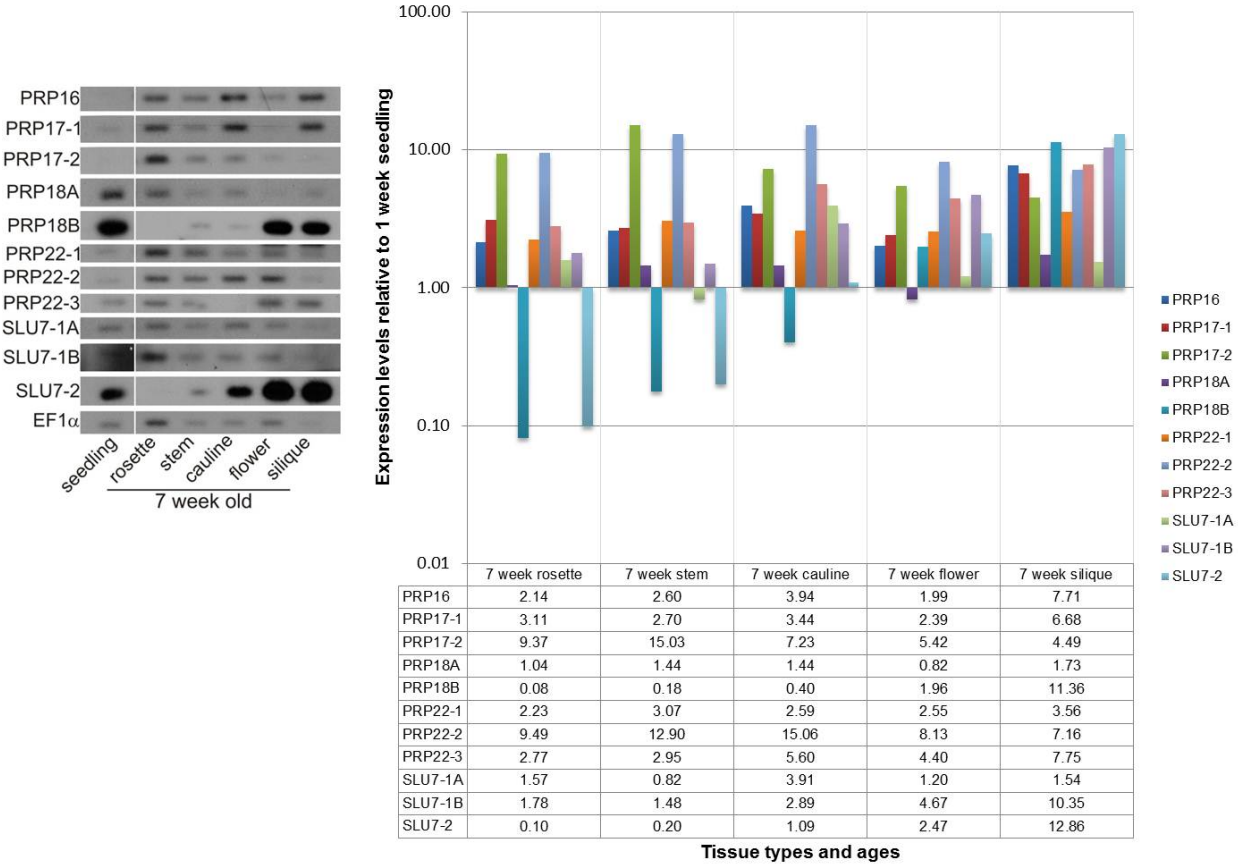


Figure 2.9 (cont.)

A



	7 week rosette	7 week stem	7 week cauline	7 week flower	7 week silique
PRP16	2.14	2.60	3.94	1.99	7.71
PRP17-1	3.11	2.70	3.44	2.39	6.68
PRP17-2	9.37	15.03	7.23	5.42	4.49
PRP18A	1.04	1.44	1.44	0.82	1.73
PRP18B	0.08	0.18	0.40	1.96	11.36
PRP22-1	2.23	3.07	2.59	2.55	3.56
PRP22-2	9.49	12.90	15.06	8.13	7.16
PRP22-3	2.77	2.95	5.60	4.40	7.75
SLU7-1A	1.57	0.82	3.91	1.20	1.54
SLU7-1B	1.78	1.48	2.89	4.67	10.35
SLU7-2	0.10	0.20	1.09	2.47	12.86

Figure 2.10 (A) Expression levels of second-step splicing factors in 7-week-old tissue. A representative panel of Southern blot analysis on the left. Experiments were performed in biological triplicate and relative expression is represented on a logarithmic scale. Performed in biological triplicate; p-value < 0.05. **(B) Model for spliceosomal subsets of second-step splicing factors in 7-week-old tissue**. Heat maps ranging from no expression (brightest red) to the highest observed expression (brightest green) to model subsets of second-step splicing factors in 4 week *At* tissue based on fold change determined by Southern blot analysis.

B

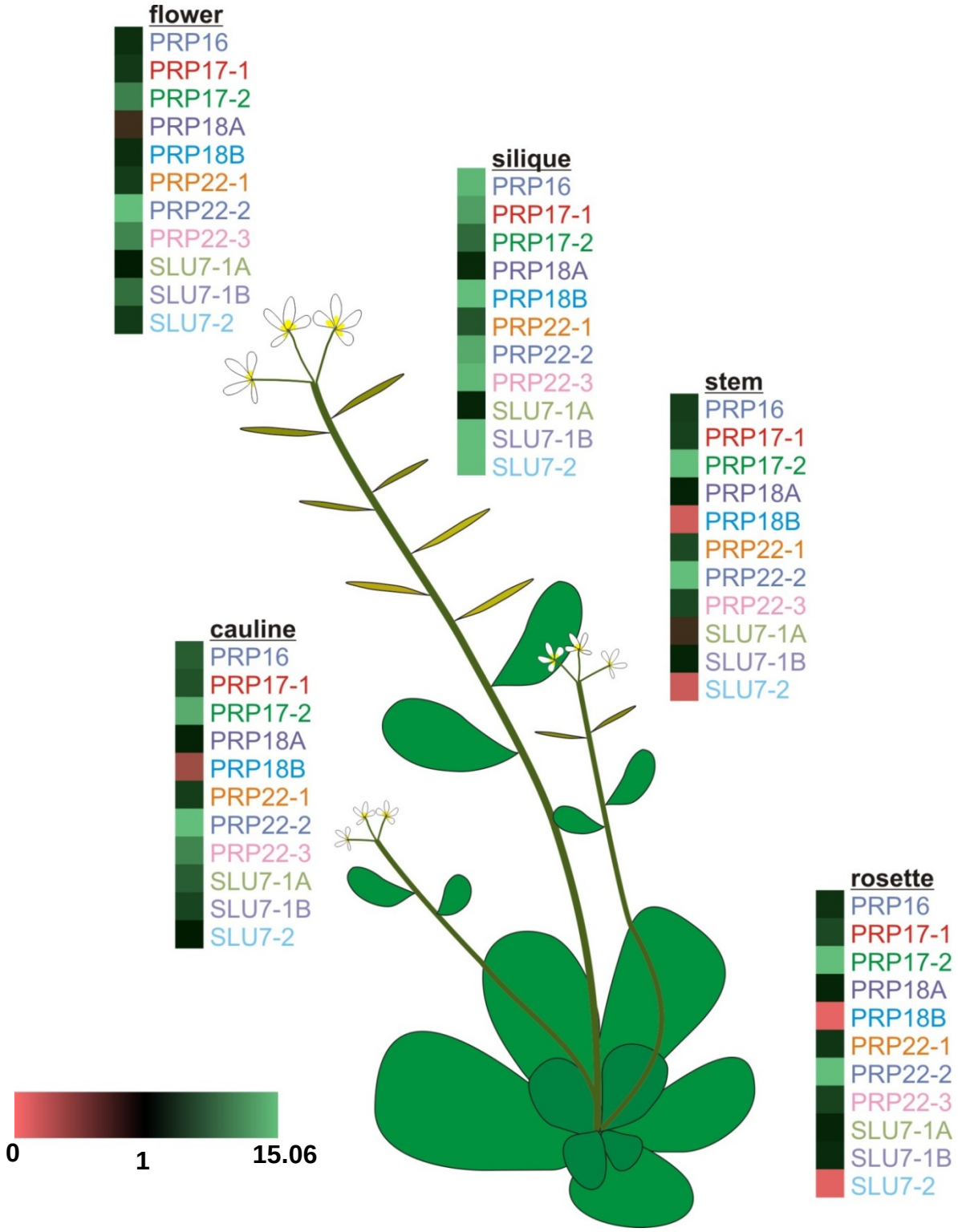


Figure 2.10 (cont.)

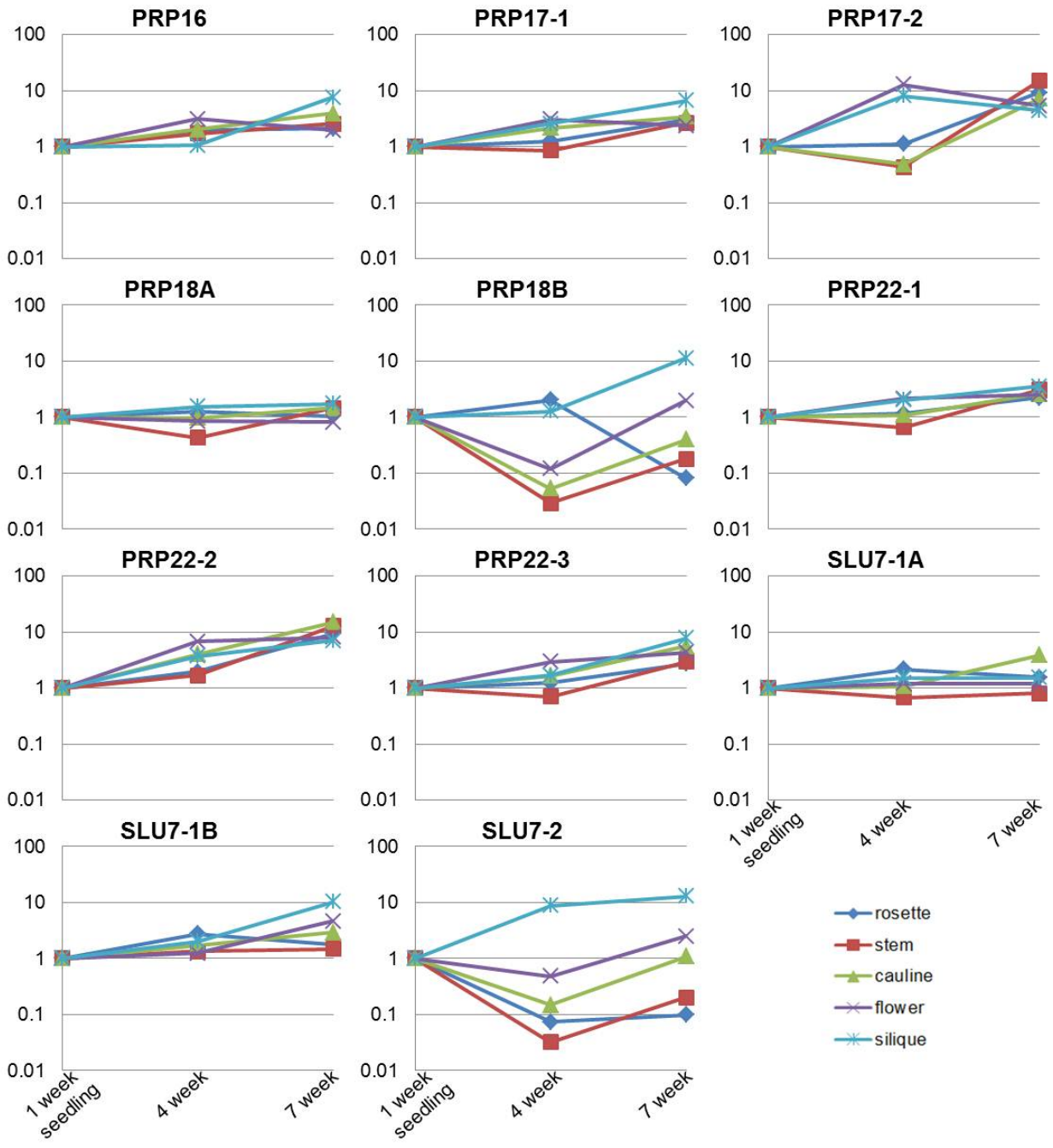


Figure 2.11 Tissue specific changes over time. The changes in expression levels for each gene over time in certain tissue-types. Logarithmic scale of fold change from a base of one week seedling and four week and seven-week-old rosette, stem, cauline, flower or silique. Performed in biological triplicate; p-value < 0.05.

2.5 References

Alberstein, M., Amit, M., Vaknin, K., O'Donnell, A., Farhy, C., Lerenthal, Y., Shomron, N., Shaham, O., Sharrocks, A. D., Ashery-Padan, R. et al. (2007). Regulation of transcription of the RNA splicing factor hSlu7 by Elk-1 and Sp1 affects alternative splicing. *RNA* **13**, 1988-1999.

Alberts, B. (2002). Molecular biology of the cell. New York: Garland Science.

Ansari, A. and Schwer, B. (1995). SLU7 and a novel activity, SSF1, act during the PRP16-dependent step of yeast pre-mRNA splicing. *EMBO Journal* **14**, 4001-4009.

Arenas, J. E. and Abelson, J. N. (1997). Prp43: An RNA helicase-like factor involved in spliceosome disassembly. *Proceedings National Academy of Sciences USA* **94**, 11798-11802.

Ayadi, L., Miller, M. and Banroques, J. (1997). Mutations within the yeast U4/U6 snRNP protein Prp4 affect a late stage of spliceosome assembly. *RNA* **3**, 197-209.

Bacikova, D. and Horowitz, D. S. (2002). Mutational analysis identifies two separable roles of the *Saccharomyces cerevisiae* splicing factor Prp18. *RNA* **8**, 1280-1293.

Bacikova, D. and Horowitz, D. S. (2005). Genetic and functional interaction of evolutionarily conserved regions of the Prp18 protein and the U5 snRNA. *Molecular Cellular Biology* **25**, 2107-2116.

Barta, A., Sommergruber, K., Thompson, D., Hartmuth, K., Matzke, M. A. and Matzke, A. J. M. (1986). The expression of a nopaline synthase — human growth hormone chimaeric gene in transformed tobacco and sunflower callus tissue. *Plant Molecular Biology* **6**, 347-357.

Ben-Yehuda, S., Dix, I., Russell, C. S., McGarvey, M., Beggs, J. D. and Kupiec, M. (2000). Genetic and physical interactions between factors involved in both cell cycle progression and pre-mRNA splicing in *Saccharomyces cerevisiae*. *Genetics* **156**, 1503-1517.

Boger-Nadjar, E., Vaisman, N., Ben-Yehuda, S., Kassir, Y. and Kupiec, M. (1998). Efficient initiation of S-phase in yeast requires Cdc40p, a protein involved in pre-mRNA splicing. *Molecular General Genetics* **260**, 232-241.

Bowers, J. E., Chapman, B. A., Rong, J. and Paterson, A. H. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* **422**, 433-438.

Brown, J. W., Feix, G. and Frendewey, D. (1986). Accurate in vitro splicing of two pre-mRNA plant introns in a HeLa cell nuclear extract. *EMBO Journal* **5**, 2749-2758.

Burgess, S., Couto, J. R. and Guthrie, C. (1990). A putative ATP binding protein influences the fidelity of branchpoint recognition in yeast splicing. *Cell* **60**, 705-717.

Burgess, S. M. and Guthrie, C. (1993). A mechanism to enhance mRNA splicing fidelity: the RNA-dependent ATPase Prp16 governs usage of a discard pathway for aberrant lariat intermediates. *Cell* **73**, 1377-1391.

Campodonico, E. and Schwer, B. (2002). ATP-dependent remodeling of the spliceosome: intragenic suppressors of release-defective mutants of *Saccharomyces cerevisiae* Prp22. *Genetics* **160**, 407-415.

Chawla, G., Sapra, A. K., Surana, U. and Vijayraghavan, U. (2003). Dependence of pre-mRNA introns on PRP17, a non-essential splicing factor: implications for efficient progression through cell cycle transitions. *Nucleic Acids Research* **31**, 2333-2343.

Chua, K. and Reed, R. (1999a). Human step II splicing factor hSlu7 functions in restructuring the spliceosome between the catalytic steps of splicing. *Genes and Development* **13**, 841-850.

Chua, K. and Reed, R. (1999b). The RNA splicing factor hSlu7 is required for correct 3' splice-site choice. *Nature* **402**, 207-210.

Crotti, L. B., Bacikova, D. and Horowitz, D. S. (2007). The Prp18 protein stabilizes the interaction of both exons with the U5 snRNA during the second-step of pre-mRNA splicing. *Genes and Development* **21**, 1204-1216.

Crotti, L. B. and Horowitz, D. S. (2009). Exon sequences at the splice junctions affect splicing fidelity and alternative splicing. *Proceedings Naional Academy of Sciences USA* **106**, 18954-18959.

Dahan, O. and Kupiec, M. (2004). The *Saccharomyces cerevisiae* gene CDC40/PRP17 controls cell cycle progression through splicing of the ANC1 gene. *Nucleic Acids Research* **32**, 2529-2540.

Eddy, S. R. (2004). Where did the BLOSUM62 alignment score matrix come from? *Nature Biotechnology* **22**, 1035-1036.

Frank, D. and Guthrie, C. (1992). An essential splicing factor, SLU7, mediates 3' splice site choice in yeast. *Genes and Development* **6**, 2112-2124.

Frank, D., Patterson, B. and Guthrie, C. (1992). Synthetic lethal mutations suggest interactions between U5 small nuclear RNA and four proteins required for the second-step of splicing. *Molecular Cellular Biology* **12**, 5197-5205.

Hartmuth, K. and Barta, A. (1986). *In vitro* processing of a plant pre-mRNA in a HeLa cell nuclear extract. *Nucleic Acids Research* **14**, 7513-7528.

He, F., Inoue, M., Kigawa, T., Takahashi, M., Kuwasako, K., Tsuda, K., Kobayashi, N., Terada, T., Shirouzu, M., Guntert, P. et al. (2012). Solution structure of the splicing factor motif of the human Prp18 protein. *Proteins* **80**, 968-974.

Horowitz, D. S. and Abelson, J. (1993). A U5 small nuclear ribonucleoprotein particle protein involved only in the second-step of pre-mRNA splicing in *Saccharomyces cerevisiae*. *Molecular Cellular Biology* **13**, 2959-2970.

Hotz, H. R. and Schwer, B. (1998). Mutational analysis of the yeast DEAH-box splicing factor Prp16. *Genetics* **149**, 807-815.

Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., Widmayer, P., Gruissem, W. and Zimmermann, P. (2008). Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Advanced Bioinformatics* **2008**, 420747.

James, S. A., Turner, W. and Schwer, B. (2002) How Slu7 and Prp18 cooperate in the second step of yeast pre-mRNA splicing. *RNA* **8**, 1068-1077.

Janowicz, A., Michalak, M. and Krebs, J. (2011). Stress induced subcellular distribution of ALG-2, RBM22 and hSlu7. *Biochimica Biophysica Acta* **1813**, 1045-1049.

Jones, M. H., Frank, D. N. and Guthrie, C. (1995). Characterization and functional ordering of Slu7p and Prp17p during the second-step of pre-mRNA splicing in yeast. *Proceedings Naional Academy of Sciences USA* **92**, 9687-9691.

Koodathingal, P., Novak, T., Piccirilli, J. A. and Staley, J. P. (2010). The DEAH box ATPases Prp16 and Prp43 cooperate to proofread 5' splice site cleavage during pre-mRNA splicing. *Molecular Cell* **39**, 385-395.

Kudlinzki, D., Schmitt, A., Christian, H. and Ficner, R. (2012). Structural analysis of the C-terminal domain of the spliceosomal helicase Prp22. *Biological Chemistry* **393**, 1131-1140.

Linder, P., Lasko, P. F., Ashburner, M., Leroy, P., Nielsen, P. J., Nishi, K., Schnier, J. and Slonimski, P. P. (1989). Birth of the D-E-A-D box. *Nature* **337**, 121-122.

Lindsey-Boltz, L. A., Chawla, G., Srinivasan, N., Vijayraghavan, U. and Garcia-Blanco, M. A. (2000). The carboxy terminal WD domain of the pre-mRNA splicing factor Prp17p is critical for function. *RNA* **6**, 1289-1305.

Lou, H., McCullough, A. J. and Schuler, M. A. (1993). Expression of maize Adh1 intron mutants in tobacco nuclei. *Plant Journal* **3**, 393-403.

Lu, G. and Moriyama, E. N. (2004). Vector NTI, a balanced all-in-one sequence analysis suite. *Briefings in Bioinformatics* **5**, 378-388.

MacKerell, A. D., Bashford, D., Bellott, Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S. et al. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal Physical Chemistry B* **102**, 3586-3616.

Madhani, H. D. and Guthrie, C. (1994). Genetic interactions between the yeast RNA helicase homolog Prp16 and spliceosomal snRNAs identify candidate ligands for the Prp16 RNA-dependent ATPase. *Genetics* **137**, 677-687.

Maere, S., De Bodt, S., Raes, J., Casneuf, T., Van Montagu, M., Kuiper, M. and Van de Peer, Y. (2005). Modeling gene and genome duplications in eukaryotes. *Proceedings Naional Academy of Sciences USA* **102**, 5454-5459.

Mayas, R. M., Maita, H. and Staley, J. P. (2006). Exon ligation is proofread by the DExD/H-box ATPase Prp22p. *Nature Structural Molecular Biology* **13**, 482-490.

Mefford, M. A. and Staley, J. P. (2009). Evidence that U2/U6 helix I promotes both catalytic steps of pre-mRNA splicing and rearranges in between these steps. *RNA* **15**, 1386-1397.

Ohno, M. and Shimura, Y. (1996). A human RNA helicase-like protein, HRH1, facilitates nuclear export of spliced mRNA by releasing the RNA from the spliceosome. *Genes and Development* **10**, 997-1007.

Ono, Y., Ohno, M. and Shimura, Y. (1994). Identification of a putative RNA helicase (HRH1), a human homolog of yeast Prp22. *Molecular Cellular Biology* **14**, 7611-7620.

Ruthenburg, A. J., Wang, W., Graybosch, D. M., Li, H., Allis, C. D., Patel, D. J. and Verdine, G. L. (2006). Histone H3 recognition and presentation by the WDR5 module of the MLL1 complex. *Nature Structural Molecular Biology* **13**, 704-712.

Sapra, A. K., Arava, Y., Khandelia, P. and Vijayraghavan, U. (2004). Genome-wide analysis of pre-mRNA splicing: intron features govern the requirement for the second-step factor, Prp17 in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *Journal Biological Chemistry* **279**, 52437-52446.

Sawa, H. and Shimura, Y. (1991). Requirement of protein factors and ATP for the disassembly of the spliceosome after mRNA splicing reaction. *Nucleic Acids Research* **19**, 6819-6821.

Schneider, S., Campodonico, E. and Schwer, B. (2004). Motifs IV and V in the DEAH box splicing factor Prp22 are important for RNA unwinding, and helicase-defective Prp22 mutants are suppressed by Prp8. *Journal Biological Chemistry* **279**, 8617-8626.

Schneider, S., Hotz, H. R. and Schwer, B. (2002). Characterization of dominant-negative mutants of the DEAH-box splicing factors Prp22 and Prp16. *Journal Biological Chemistry* **277**, 15452-15458.

Schneider, S. and Schwer, B. (2001). Functional domains of the yeast splicing factor Prp22p. *Journal Biological Chemistry* **276**, 21184-21191.

Schwer, B. (2008). A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. *Molecular Cell* **30**, 743-754.

Schwer, B. and Guthrie, C. (1991). PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. *Nature* **349**, 494-499.

Schwer, B. and Guthrie, C. (1992). A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. *EMBO Journal* **11**, 5033-5039.

Schwer, B. and Meszaros. T. (2000). RNA helicase dynamics in pre-mRNA splicing. *EMBO Journal* **19**, 6582-6591.

Shomron, N., Reznik, M. and Ast, G. (2004). Splicing factor hSlu7 contains a unique functional domain required to retain the protein within the nucleus. *Molecular Biology Cell* **15**, 3782-3795.

Shomron, N., Alberstein, M., Reznik, M. and Ast, G. (2005). Stress alters the subcellular distribution of hSlu7 and thus modulates alternative splicing. *Journal Cell Science* **118**, 1151-1159.

Tanaka, N. and Schwer, B. (2006). Mutations in PRP43 that uncouple RNA-dependent NTPase activity and pre-mRNA splicing function. *Biochemistry* **45**, 6510-6521.

Thimmarpuram, J., Duan, H., Liu, L. and Schuler, M. A. (2005). Bicistronic and fused monocistronic transcripts are derived from adjacent loci in the *Arabidopsis* genome. *RNA* **11**, 128-138.

Tseng, C. K., Liu, H. L. and Cheng, S. C. (2011). DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. *RNA* **17**, 145-154.

Tuteja, N. and Tuteja, R. (2004). Unraveling DNA helicases. *European Journal Biochemistry* **271**, 1849-1863.

Valdez, B. C., Henning, D., Perumal, K. and Busch, H. (1997). RNA-unwinding and RNA-folding activities of RNA Helicase II/Gu - Two activities in separate domains of the same protein. *European Journal Biochemistry* **250**, 800-807.

van Santen, V. L. and Spritz, R. A. (1987). Splicing of plant pre-mRN As in animal systems and vice versa. *Gene* **56**, 253-265.

Vijayraghavan, U. and Abelson, J. (1990). PRP18, a protein required for the second reaction in pre-mRNA splicing. *Molecular Cellular Biology* **10**, 324-332.

Wagner, J. D., Jankowsky, E., Company, M., Pyle, A. M. and Abelson, J. N. (1998). The DEAH-box protein PRP22 is an ATPase that mediates ATP-dependent mRNA release from the spliceosome and unwinds RNA duplexes. *EMBO Journal* **17**, 2926-2937.

Walbott, H., Mouffok, S., Capeyrou, R., Lebaron, S., Humbert, O., van Tilbeurgh, H., Henry, Y. and Leulliot, N. (2010). Prp43p contains a processive helicase structural architecture with a specific regulatory domain. *EMBO Journal* **29**, 2194-2204.

Wang, Y. and Guthrie, C. (1998). PRP16, a DEAH-box RNA helicase, is recruited to the spliceosome primarily via its nonconserved N-terminal domain. *RNA* **4**, 1216-1229.

Wang, Y., Wagner, J. D. and Guthrie, C. (1998). The DEAH-box splicing factor Prp16 unwinds RNA duplexes *in vitro*. *Current Biology* **8**, 441-451.

Wiebauer, K., Herrero, J. J. and Filipowicz, W. (1988). Nuclear pre-mRNA processing in plants: distinct modes of 3'-splice-site selection in plants and animals. *Molecular Cellular Biology* **8**, 2042-2051.

Xu, C. and Min, J. (2011). Structure and function of WD40 domain proteins. *Protein Cell* **2**, 202-214.

Xu, D., Field, D. J., Tang, S. J., Moris, A., Bobechko, B. P. and Friesen, J. D. (1998). Synthetic lethality of yeast *slt* mutations with U2 small nuclear RNA mutations suggests functional interactions between U2 and U5 snRNPs that are important for both steps of pre-mRNA splicing. *Molecular Cellular Biology* **18**, 2055-2066.

Zhang, X. and Schwer, B. (1997). Functional and physical interaction between the yeast splicing factors Slu7 and Prp18. *Nucleic Acids Research* **25**, 2146-2152.

Zhou, Z. and Reed, R. (1998). Human homologs of yeast prp16 and prp17 reveal conservation of the mechanism for catalytic step II of pre-mRNA splicing. *EMBO Journal* **17**, 2095-2106.

CHAPTER 3

Growth of *Arabidopsis* in highly metallic environments affects expression of second-step splicing factors.

3.1 Introduction

Stress management mechanisms in plants differ depending on the type of environmental conditions encountered. Heavy metal stresses caused by mercury (Hg^{2+}), cadmium (Cd^{2+}), copper (Cu^{2+}) and zinc (Zn^{2+}) are especially problematic since several of these divalent cations, along with iron (Fe^{2+}) and magnesium (Mg^{2+}), are key constituents of metabolic proteins and nucleic acid processing complexes. Copper and zinc are essential micronutrients due to the existence of many Cu^{2+}- and Zn^{2+}-dependent proteins. Mercury and cadmium are nonessential micronutrients that have potential to displace other metals in the binding sites of proteins. Cu^{2+} and Zn^{2+} have high potential to disrupt protein activity in this manner (Lippard and Berg, 1994). Additionally, Cd^{2+} can replace Zn^{2+} in Zn^{2+}-dependent reactions and, as a result, impede them (Stohs and Bagchi, 1995).

Hg^{2+} and Cu^{2+} are likely to undergo uptake via metal transporters that are loosely-specific for their intended divalent cation (Clemens, 2001). Specifically, copper uptake is mediated by COPT1, a member of the five-gene copper transporter (CTR) family in *Arabidopsis*, and transported to copper chaperones in the cytoplasm (Kampfenkel et al., 1995; Sancenón et al., 2003). Copper is then delivered to one of its very important targets, Cu/Zn superoxide dismutase (Cu/ZnSOD), by the high affinity copper chaperone CCS (Cu chaperone for SOD), which exists in both the cytosol and chloroplast (Abdel-Ghany et al., 2005; Chu et al., 2005). Additional chaperones are CCH (Calcium CHannel) and ATX1 (AnTioXidant protein 1), both homologs of

the yeast copper chaperone ATX1 (Andrés-Colás et al., 2006; Himelblau et al., 1998; Puig et al., 2007). *Arabidopsis* ATX1 has specifically been shown to bind the P-type ATPase HMA5 (heavy metal ATPase 5) that maintains copper homeostasis through Cu^{2+} detoxification (Andrés-Colás et al., 2006; Puig et al., 2007). The two remaining copper chaperones are related to yeast *COX17* that delivers copper to the mitochondrial cytochrome oxidase (Amaravadi et al., 1997; Balandin and Castresana, 2002; del Pozo et al., 2010; Glerum et al., 1996; Srinivasan et al., 1998; Wintz and Vulpe, 2002). *Arabidopsis* COX17-1 transcripts are upregulated in the presence of Cu^{2+} and other metals and its protein complements yeast *cox17* mutants (Balandin and Castresana, 2002; del Pozo et al., 2010). *At*COX17-2 transcript levels are not upregulated in the presence of Cu^{2+} its protein does not complement yeast *cox17* mutants (Balandin and Castresana, 2002) indicating that its function remains to be elucidated. Cd^{2+} is taken up both by Zn^{2+} transporters (ZNT1, IRT1) and Fe^{2+} transporters (IRT1, *At*Nramp3) (Korshunova et al., 1999; Pence et al., 2000; Thomine et al., 2000). Zinc is utilized as a cofactor for over 300 enzymes (Guerinot, 2000). Zn^{2+} is taken up by the ZIP4 transporter (Grotz et al., 1998; Wintz et al., 2003), and the types and differences between Zn chaperones are numerous.

While most heavy metals do not directly undergo redox variations in plants (Clemens, 2006), they can cause a variety of downstream effects that include oxidative stress. Increased amounts of redox-inactive mercury block water transport (Javot and Maurel, 2002), induce oxidative stress (Cho and Park, 2000; Rellán-Álvarez et al., 2006; Ortega-Villasante et al., 2007), and, at very high levels, cause cell and tissue necrosis (Cho and Park, 2000; Javot and Maurel, 2002; Ortega-Villasante et al., 2007; Rellán-Álvarez et al., 2006). Elevated levels of redox-inactive cadmium alter some of the same cellular processes (*i.e.*, water uptake) as well as nutrient uptake, photosynthesis, redox control and lipid peroxidation (Lagriffoul et al., 1998; Montillet et

al., 2004; Schützendübel and Polle, 2002; Semane et al., 2010; Van Assche and Clijsters, 1990). Increased levels of redox-active copper significantly alter redox potential, reactive oxygen species (ROS) accumulation and also lead to lipid peroxidation (Cuypers et al., 2011; Drążkiewicz et al., 2004; Halliwell and Gutteridge, 1984). Elevated levels of redox-inactive zinc inhibit root growth, photosynthesis and induce chlorosis (Broadley et al., 2007). Plants cope with higher levels of these divalent cations by enhancing their transport, chelation and sequestration systems.

Cadmium stress induces phytochelatin (PC) synthesis from glutathione (GSH), which is mediated by phytochelatin synthase (PCS), and transport of PC-bound metal complexes into vacuoles (Clemens, 2001; DalCorso et al., 2008; Schützendübel and Polle, 2002). Accompanying these changes are early depletions of GSH, inhibitions of antioxidant enzymes and accumulation of hydrogen peroxide. Similar increases in PC synthesis also occur in mercury, zinc and heavy metal stresses (Grill, 1987; Maitani et al., 1996). Analyses of PC-deficient *Arabidopsis cad* mutants have shown that these increases in PC synthesis are important for detoxification of excess Zn^{2+} (Tennstedt et al., 2009) and that, in their absence, *cad* mutants are more sensitive to cadmium and mercury (Ha et al., 1999; Howden et al., 1995; Li et al., 2006b). In addition to these mechanisms for directly sequestering Cd^{2+}, Zn^{2+} and Hg^{2+}, these cations and Cu^{2+} also induce expression of metallothionein (MT) proteins that are direct sinks for these cations (Clemens, 2001; DalCorso et al., 2008). Zinc stress also induces MTP1 (Metal Tolerance Protein 1) that facilitates Zn^{2+} transport into the vacuole (Blaudez et al., 2003; Desbrosses-Fonrouge et al., 2005; Kobae et al., 2004; van der Zaal et al., 1999). In contrast to other plant metal transporters that are promiscuous with respect to their divalent metal transport properties,

Arabidopsis MTP1 is specific for zinc and not capable of moving Cd^{2+} or other divalent cations (Kobae et al., 2004; Krämer, 2005).

While much has been done to look at the downstream effects of these heavy metal stresses on photosynthesis and other biochemical processes (DalCorso et al., 2008; Schützendübel and Polle, 2002), less has been done examining the effects of heavy metals on upstream events affecting transcript abundance. Some studies have globally assessed the range of transcripts induced in 3-week-old *Arabidopsis* shoots and roots in response to Cd^{2+} exposure using microarrays (Herbette et al., 2006) and, in more limited tryptic analysis, assessed the range of proteins varying in 6-week-old hydroponically-grown seedling roots in response to Cd^{2+} exposure (Roth et al., 2006). Other studies have compared the responses to elevated Zn^{2+} exposure in *Arabidopsis thaliana* to *Arabidopsis halleri*, a Zn^{2+} and Cd^{2+} hyperaccumulator (Becher et al., 2004; Chiang et al., 2006; Shanmugam et al., 2011; Talke et al., 2006; Weber et al., 2004). All of these studies have focused on the changes in transcript and protein abundances without analyzing for specific effects on pre-mRNA splicing and/or degradation, post-transcriptional steps that ultimately control mRNA abundance and functionality in all eukaryotic organisms.

With no information available on the effects of heavy metal stress on post-transcriptional splicing events in *Arabidopsis*, this study initiated analyzing metal stress response for eleven second-step splicing factors, *At*PRP16, *At*PRP17-1, *At*PRP17-2, AtPRP18A, *At*PRP18B, *At*PRP22-1, *At*PRP22-2, *At*PRP22-3, *At*SLU7-1A, *At*SLU7-1B and *At*SLU7-2. The data presented here provide evidence that the proportions of *Arabidopsis* second-step splicing factor spliced and unspliced transcripts vary after exposure to several heavy metals. In particular, we show that pre-mRNA transcripts accumulate with Cd^{2+} and Hg^{2+} exposure suggesting that some

of the post-transcriptional effects of cadmium and mercury stress mediate downstream phenotypic changes associated with each of these elements.

The highly conserved serine/arginine-rich (SR) family of proteins, which are non-snRNP proteins involved in both constitutive and alternative splicing, as well as in diverse other functions within the nucleus, has varied localizations within cells (Reddy, 2004). Many of these are plant-specific and most are the products of alternatively spliced transcripts. We chose to examine the alternatively spliced SR proteins RSp31, SRp30, SR33, SRp34/SR1 and U1-70K (**Table 3.1**) in plants that are exposed to heavy metals. These genes in plants also undergo changes in expression of their alternatively spliced transcripts under heavy metal treatment conditions, but not their pre-mRNA transcripts.

Several other constitutively and alternatively spliced plant transcripts (Table 3.2), whose products are not splicing factors, were also examined under heavy metal stress conditions. The effects that heavy metal treatment has on these transcripts offers further evidence that cadmium and mercury can affect the proportions of alternatively spliced transcripts for those with multiple splice site choices and the proportion of total RNA (pre-mRNA plus mature mRNA) for those with only constitutive splice site choices.

3.2 Results

Some metals cause retention of all introns in the splice products of second-step splicing factors

In order to determine the effect of heavy metals on the transcript levels of the second-step splicing factors, Arabidopsis seedling were grown for three weeks on ½ MS plates supplemented with 30 μM 30 $HgCl_2$, 10 μM $Hg(OAc)_2$, 10 μM $CdSO_4$, 20 μM $CuSO_4$ or 60 μM $ZnSO_4$, RNA was isolated and RT-PCR gel blot analyses of the full-length second-step splicing factor

transcripts were performed (Figure 3.1). Early evidence indicated that many of the second-step splicing factor transcripts retained some or all of their introns in metal treated plants as indicated by the upper band in the panels labeled *At*PRP17-1, *At*PRP17-2, *At*PRP18A, *At*PRP22-1, *At*SLU7-1A and *At*SLU7-1B. In most of the *Arabidopsis* second-step splicing factor transcripts examined, the numbers and sizes of the introns varied substantially examination (Tables 3.3, 3.4). Closer inspections of the various splice junctions (Table 3.3) and intron lengths (Table 3.4) in these transcripts indicated that, while a few intron-exon junctions were unusual in the nucleotide preceding the 5' splice site, all contained canonical /gu…ag/ junctions. In addition, most of these transcripts contain at least one intron substantially longer than the average intron length of 101 bases in *Arabidopsis* (Arabidopsis Genome Initiative, 2000); these include *At*PRP16, *At*PRP17-1, *At*PRP17-2, *At*PRP18A, *At*PRP22-3 and *At*SLU7-1A but not *At*PRP18B, *At*PRP22-1, *At*PRP22-2 and *At*SLU7-1B that have only short introns or *At*SLU7-2 that has no introns.

Primers were designed to encompass smaller subsets of introns, and to overlap in a manner that every intron is accounted for (Figure 3.2). *Arabidopsis* were grown on ½ MS plates supplemented with increasing concentrations of $HgCl_2$ (10-30 μM), $Hg(OAc)_2$ (10-30 μM), $CdSO_4$ (2-15 μM), $CuSO_4$ (10-40 μM) and $ZnSO_4$ (20-60 μM) for three weeks. RT-PCR and Southern blot analysis revealed that pre-mRNA levels were increased in many cases as a result of being grown under metal-supplemented conditions, especially in plants treated with $Hg(OA)_2$ and $CdSO_4$ (Figure 3.3). Because pre-mRNA was detected across all subsets of introns in the extremely long *At*PRP16, *At*PRP22-2 and *At*PRP22-3 transcripts and because amplification of genomic DNA using full-length primers demonstrated a lack of product, it was determined that lack of pre-mRNA in Figure 3.1 is due to technical limitations caused by the length of these three

primary transcripts. The absence of DNA contamination in these samples was confirmed by control RT-PCR reactions of the splicing factor transcripts performed without reverse transcriptase and RT-PCR reactions of the internal control EF1α transcripts performed with reverse transcriptase, which failed to show any high molecular weight bands. Taken together, these data suggest that the higher molecular weight species in each of these blots are pre-mRNAs that have not been spliced.

The ratios of retained introns versus mature RNA vary from metal to metal and even within the same gene (Table 3.5A). This is likely a result of partial degradation of some transcripts, and selection for other regions, although this mechanism does not appear to be related to the intron-exon junctions of the second-step splicing factors (Tables 3.3, 3.4). Rather, certain metals (particularly 30 μM $HgCl_2$, 10 μM $Hg(OAc)_2$ and 10 μM $CdSO_{4)}$ were responsible for increases in intron retention.

Accumulation of some transcripts as a result of heavy metal treatment

In some cases, the overall amount of second-step splicing factor transcript being made in response to treatment with heavy metals is altered (Table 3.5B). This is further complicated by the fact that the cases in which the mature transcript of the second-step splicing factors is also altered, this may not be directly proportional to the total amount of transcript (Table 3.5C). Determination of the relatedness between total transcript accumulation and mature-mRNA retention might lend some clues as to the nature of the intron retention observed when the plants are grown under mercuric or cadmium stress.

Expression of these genes was examined in two different ways. First, total expression was analyzed in comparison to untreated samples by assessing the total amount of transcript regardless of it being fully spliced or not (Table 3.5B). The expression of *At*PRP16 drops in 30

μM $HgCl_2$ and 30 μM $Hg(OAc)_2$ while 10 μM $Hg(OAc)_2$ results in an increase of overall expression. *At*PRP16 is also expressed in higher levels in the presence of 10 μM $CdSO_4$ and its expression increases as the concentration of $ZnSO_4$ increases. *At*PRP17-1 expression is higher with both concentrations of $HgCl_2$ and $Hg(OAc)_2$. *At*PRP17-2 is decreased with 10 μM $HgCl_2$, 30 μM $Hg(OAc)_2$ and 20 μM $CuSO_4$ and increased in 10 μM $CdSO_4$ and as $ZnSO_4$ concentrations increase. Both *At*PRP18A and *At*PRP18B peak in expression with 10 μM $HgCl_2$ and 10 μM $Hg(OAc)_2$. *At*PRP18A also increases when treated with all levels of $CdSO_4$, 20 μM $CuSO_4$ and increasing concentrations of $ZnSO_4$. In addition to the mercury samples mentioned above, *At*PRP18B, on the other hand, peaks only at 10 μM $CdSO_4$. *At*PRP22-1 expression peaks at 10 μM $Hg(OAc)_2$ and drops at 10-20 μM $CuSO_4$. *At*PRP22-2 expression decreases at both concentrations of $HgCl_2$ and at 30 μM $Hg(OAc)_2$ and increases at 15 μM $CdSO_4$, 40 μM $CuSO_4$ and 60 μM $ZnSO_4$. *At*PRP22-3 expression is only increased in the 30 μM $Hg(OAc)_2$ sample. *At*SLU7-1A and *At*SLU7-1B expression do not considerably change. *At*SLU7-2 expression is decreased in 10 μM $HgCl_2$ and 30 μM $Hg(OAc)_2$ but increased in 10 μM $Hg(OAc)_2$. Its expression is also increased at 10-15 μM $CdSO_4$ and decreased in all treatments of $CuSO_4$. At the lower concentrations of $ZnSO_4$, *At*SLU7-2 expression is decreased, while at 60 μM, its expression is increased.

Second, the accumulation of mature transcript was examined (Table 3.5C). No significant change in the presence of mature mRNA was detected in *At*PRP16, *At*PRP17-1, *At*PRP17-2, *At*PRP18A, *At*PRP22-3, *At*SLU7-1A or *At*SLU7-1B. *At*PRP18B mature transcripts were decreased in samples treated with 10-15 μM $CuSO_4$. *At*PRP22-1 mature transcript levels were increased in samples treated with 10-30 μM $Hg(OAc)_2$ and decreased in samples treated with 10-20 μM $CuSO_4$. *At*PRP22-2 mature transcripts were decreased in 10-30 μM $HgCl_2$ and increased

in response to treatment with 15 μM $CdSO_4$, 10 and 40 μM $CdSO_4$ and all concentrations of $ZnSO_4$. Because there is no intron present in *At*SLU7-2, the processing of that gene is not called into question and there is no discerning between the accumulation of mature transcript and the increased expression of overall transcript in metal-treated plants.

Effects of heavy metals on other splicing factors

In order to determine if the retention of pre-mRNA as a result of heavy metal treatment is specific to the second-step splicing factors, SR protein transcripts are examined under heavy metal treatment conditions. Each of the proteins is involved in various steps of splicing and is alternatively spliced. RSp31 is a plant-specific SR protein (Lopato et al., 1996) with nucleolar localization (Tillemans et al., 2006), SR33 is a plant-specific SR protein that interacts with U1-70K (Golovkin and Reddy, 1999), SRp34/SR1 homologs are important for 5' splice site recognition and interact with U1-70K and U11-35K (Lorković et al., 2004) and U1-70K is U1 snRNA-specific and binds to various other SR proteins (Golovkin and Reddy, 1998). Each of these alternatively spliced isoforms is depicted in Figure 3.4 with product A being the smallest mature mRNA and the most abundant splice form. Briefly, RSp31 has alternative 3' splice site recognition (B), partial intron retention (C) and second alternative 3' splice site recognition (D). SR30 has an alternative 3' splice site (B). SR33 has alternative 3' splice site recognition (B), alternative intron retention (C), alternative 5' splice site recognition (D) and intron retention (E). SRp34/SR1 has alternative 3' splice site (B) and a different alternative 3' splice site combined with partial intron retention (C). U1-70K has intron retention (B).

Under metal stress conditions, each of the major alternative splice products is observed (Figure 3.4), but no pre-mRNA is detected in any of the genes based on comparison to amplified genomic DNA (data not shown). The presence of heavy metals causes alterations in the

alternative splice product ratios (Table 3.6A) and expression (Table 3.6B). All of these alternatively spliced species are predicted products and typically the increases in alternative splice products are a result of retention of the full-length of the alternative intron. In contrast to the second-step splicing factors, the total expression levels (total product of all alternative splice products) of the SR proteins observed are not significantly altered (Table 3.6B).

Effects of heavy metals on non-splicing-related alternatively spliced proteins

In order to further understand the targeting of alternative splicing by the heavy metal stress on *Arabidopsis*, the alternative splicing effects of genes predicted to have multiple alternatively spliced species was examined (Wang and Brendel, 2006). These genes include the transcription factor, MYB59; the plant defensin gene family protein, PDF2.2 and the whirly family protein WHY1 (Figure 3.5). There is no published evidence of alternative splicing in PDF2.2 or WHY1 and our own data show no alternative splice products (Figure 3.5).

In contrast, MYB59 alternative splicing results in four products (Li et al., 2006a) with three products detected in this study and the fourth product below the level of detection. The two alternative products that can be detected are the alternate 5' splice site (B) and retention of the first intron (C) (Figure 3.5). Changes in the levels of alternative splice products are minimal, but there is a reproducible slight increase in the intron-retention product (C) when treated with $Hg(OAc)_2$ and $CdSO_4$. Total transcript levels of MYB59 are also not significantly altered under heavy metal treatment (Table 3.7B).

PDF2.2 transcripts retaining the only intron in its gene (B) are present in control and heavy metal-treated plants (Figure 3.5). The ratio of the intron retained product to the fully-spliced product drops drastically when plants are treated with $HgCl_2$ and significantly increase when plants are treated with $Hg(OAc)_2$ and $CdSO_4$ (Table 3.7A). WHY2, which contains six

introns between the two primers used in this study, shows a nearly undetectable accumulation of pre-mRNA in untreated, $Hg(OAc)_2$ and $CdSO_4$ and there are no significant changes in accumulation of the pre-mRNA or total transcript levels of WHY2 (Tables 3.7A, B).

3.3 Discussion

Initial efforts at analysis of the levels of the full length second-step splicing factor transcripts revealed retention of higher molecular weight species in many of the samples. In order to determine what these amplification products are, gel blot analysis was carried out on smaller subsets of the introns within the gene. This method ensures that each intron is examined for possible alternative splicing or intron retention as a result of pre-mRNA accumulation. Because the overall intron check was directly compared to amplified genomic DNA and no-RT controls were performed on each set of cDNAs used in this experiment (data not shown), we determined that additional products detected were a result of intron retention in the primary transcripts derived from the second-step splicing factor genes. No bands of additional sizes were detected, eliminating the possibility of metal stress-related alternative splicing of the second-step splicing factors.

Initially, to analyze the response of the second-step splicing factor genes to metal stresses on the plants, total expression levels of the genes was analyzed. This involved assessing the total amount of product, as compared to untreated plants, including both pre-mRNA and mature transcript and assessing the total, mature transcript levels that accumulate in the plant. Commonly, total levels of gene expression were not mirrored in the levels of mature transcript. This indicates that there is a blockage in the processing of the RNA so that some of the pre-mRNA is never processed into mature (and presumably functional) mRNA.

Intron retention in second-step splicing factor transcripts tends to favor the 5' end of the gene in *At*PRP17-2, *At*PRP22-2 and *At*PRP22-3. *At*PRP22-2 and *At*PRP22-3 contain 12 and 15 introns, respectively, and *At*PRP17-2 only contains three introns of which the third is unusually long (720 base pairs). Genes containing multiple introns that do not exhibit this polar intron retention effect include *At*PRP16, *At*SLU7-1A and *At*SLU7-1B. None of these described genes exhibit unusual intron-exon junctions. While pre-mRNA accumulation of the internal control, EF1α (which does contain an intron), was not detected, it was still necessary to determine the extent of this effect. SR proteins that are alternatively spliced were examined under the same metal conditions.

The SR did not, in any case, exhibit an accumulation of pre-mRNA. As before, genomic DNA was amplified and compared to the products of the SR protein amplifications, and the size did not compare to any of the products observed until metal treatments (data not shown). Accumulation of different alternative splicing products, however, was observed. RSp31 product D (alternative 3' splice site) makes up less than 1% of the total product in untreated plants, but $HgCl_2$, $CdSO_4$ and $ZnSO_4$ treatments result in an accumulation of this product to making a total of 10% of the transcript. Product C (partial intron retention) accumulates in $ZnSO_4$ and product B (alternative 3' splice site) accumulates in $CdSO_4$ and $ZnSO_4$. This same effect is detected in SR33 product E (intron retention) when the plant is treated with $CdSO_4$. U1-70K, with a singular alternative splice product of a retained intron, has a 20% jump in that alternative splice product in the presence of $CdSO_4$.

I propose that the alteration of some alternative splice products of SR proteins is a result of metal treatment and that we do not observe retention of pre-mRNA because there are alternate options for splicing of SR proteins transcript. In other words, when normal splicing does not

work as a result of stress, there are bail-out splicing options in the alternatively spliced SR transcripts that the splicing machinery can rely on. In the case of the second-step splicing factor transcripts where there are no options for alternative splicing the splicing machinery has no option but to decrease splicing efficiency. .

Quantification of other non-splicing transcripts that are or are not alternatively spliced indicated that the splicing of PDF2.2 and WHY1, which do not contain alternative splice sites, is also diminished in the presence of some metals resulting in the accumulation of their pre-mRNAs. In contrast, MYB59, which does contain alternative splice sites, does not accumulate pre-mRNA but does have altered ratios of alternative splice products. Specifically, treatment with mercury or cadmium results in modulation between product C (intron retention) and product B (alternative 5' splice site).

Rather than the observed effects noted above being on the second-step splicing factor transcripts, it is more likely that there are two things acting in concert. First, splicing is perturbed. In genes that cannot be alternatively spliced, accumulation of pre-mRNA is observed. The SR proteins and other genes of alternatively spliced proteins undergo differential alternative splicing. Separately, RNA surveillance that would normally be responsible for degrading the accumulated, incorrectly processed mRNA is not functioning. We would expect that under conditions in which surveillance is not perturbed, we would not be able to detect increasing levels of pre-mRNA.

A clue that indicates that the surveillance system that not only relies on the presence of premature termination codon (PTC)-containing introns, but also regulates the level of normal transcripts as a method of translation regulation, is affected by the treatment of the plants with heavy metals is the set of observations on the *At*SLU7-2 transcript that does not contain any

introns. When the expression levels of this gene are analyzed under metal stress, transcript accumulation increases in cadmium and mercury stress. We expect that this would be the case with any gene that is susceptible to RNA surveillance regardless of the presence of an in-frame stop codon. Identification of the type of mRNA surveillance that is affected will be a pertinent step to understanding the mechanism of metal stress tolerance in the cell.

3.4 Materials and Methods

Plant growth conditions and material

Wild type *Arabidopsis thaliana* (Columbia ecotype) seeds were sterilized by submerging in 70% alcohol for 30 seconds, 20% bleach and a drop of Tween20 for 15 minutes, and washed three times with sterile water. Seeds were then plated on ½ MS agar plates (pH 5.7) that were supplemented with either no metal; 10 or 30 μM $HgCl_2$; 10 or 30 μM $Hg(OAc)_2$; 2, 10 or 15 μM $CdSO_4$; 10, 20 or 40 μM $CuSO_4$; and 20, 40 or 60 μM $ZnSO_4$ after autoclaving. Seedlings were incubated in the dark at 4°C for 24 hours then grown under long day conditions (22°C with 16-h-light and 8-h-dark cycles) for three weeks. Three-week-old plants were harvested directly from the plates, snap-frozen in liquid nitrogen, and stored at -80°C.

RNA isolation and cDNA synthesis

RNA was extracted by the beadbeater method described in Thimmapuram et al. (2005) including DNase treatment. The RNA was stored in 100 μl of water, snap frozen and stored at -80°C. cDNA was synthesized by Superscript III (Invitrogen) according to the user manual. Specifically, 1 μg of total RNA was used in a 20 μl reaction containing 200 ng of oligo$(dT)_{19}$, 1 μl 10 mM dNTP mix and 1 μl 0.1 M DTT. To obtain a final working concentration of cDNA, the sample was diluted to 40 μl in water.

Analysis of transcript levels

Semi-quantitative reverse transcriptase PCR (RT-PCR) primers are listed in Table 3.8 for the second-step splicing factors and in Table 3.9 for all other genes examined. When full-length gene products were amplified, the first forward primer for a gene and the last reverse primer for the same gene (Table 3.8) were used. cDNA was amplified by 25 cycles of PCR, with the exception of EF1a, which was amplified for 18 cycles. Each cycle consists of 94°C for 45 seconds, 58°C for 45 seconds and 72°C for 2 minutes. The first cycle is preceded by a 3 minute 95°C denaturation period and followed by a 10 minute 72°C elongation period. PCR products were run on 1.3% agarose gels, transferred to Hybond-N (Amersham-Pharmacia Biotech), and probed with ^{32}P-labelled DNA probe made with random primers (Invitrogen) and gene-specific cDNA template. Blots were hybridized and washed according to Lou et al (1993). Band strength was determined by ImageJ (NIH) (Rasband, W. S. ImageJ, U. S. National Institutes of Health, Bethesda, MD) after background signal was reduced with a rolling ball radius of 50-70 pixels. Transcript levels were normalized to the corresponding EF1α measurements and compared to untreated samples. These performed in biological duplicate and fold-change is plotted on a logarithmic scale. The p-value is less than 0.05.

Table 3.1 Arabidopsis SR proteins examined in this study.

Common Name	Locus	Accession number
RSp31	At2g46610	AAL38713
SRp30	At1g09140	AEE28400
SR33	At1g55310	AAW28547
SRp34/SR1	At1g02840	AAB71385
U1-70K	At3g50670	AAD12775

Table 3.2 Non-splicing related genes examined in this study.

Common Name	Locus	Accession number
MYB59	At5g59780	AAY97894
PDF2.2	At2g02100	AAK44170
WHY1	At1g14410	AAK43971

Table 3.3 Intron features of Arabidopsis second-step splicing factors. Unusual nucleotides are highlighted in pink.

Gene	Intron	5' splice site	3' splice site
***At*PRP16**	1	GGgu	agGU
	2	AGgu	agGU
	3	AGgu	agAA
	4	AUgu	agAU
	5	AGgu	agUG
	6	AGgu	agGG
	7	UGgu	agGU
	8	UGgu	agGU
	9	UGgu	agAU
	10	UGgu	agAU
	11	AGgu	agAU
	12	AGgu	agGU
	13	AGgu	agGU
	14	AGgu	agUA
	15	CGgu	agUG
	16	AGgu	agUG
***At*PRP17-1**	1	UGgu	agGG
***At*PRP17-2**	1	UGgu	agUU
	2	UGgu	agUU
	3	AGgu	agGU
***At*PRP18A**	1	AGgu	agGG
***At*PRP18B**	1	UAgu	agGU
	2	AGgu	agUG
***At*PRP22-1**	1	AAgu	agGU
***At*PRP22-2**	1	AGgu	agCA
	2	AGgu	agUG
	3	UCgu	agAA
	4	AGgu	agAG
	5	AGgu	agAU
	6	AGgu	agGU
	7	AGgu	agUG
	8	AGgu	agGG
	9	AGgu	agGC
	10	AGgu	agGC
	11	AGgu	agGG
	12	aggu - UTR	agau - UTR

Gene	Intron	5' splice site	3' splice site
***At*PRP22-3**	1	AGgu	agGC
	2	AAgu	agGU
	3	ACgu	agGG
	4	UGgu	agAU
	5	AGgu	agGA
	6	AGgu	agGU
	7	GUgu	agGU
	8	AAgu	agAG
	9	AUgu	agCU
	10	UGgu	agAU
	11	AGgu	agUA
	12	AGgu	agGU
	13	AGgu	agAU
	14	AGgc	agGU
	15	AGgu	agCG
***At*SLU7-1A**	1	AGgu	agUU
	2	CGgu	agAG
	3	AAgu	agCU
	4	AGgu	agGA
	5	UGgu	agGG
	6	AGgu	agGA
	7	AGgu	agAG
	8	AGgu	agGA
	9	AUgu	agGU
***At*SLU7-1B**	1	AGgu	agUU
	2	CGgu	agAG
	3	AGgu	agGA
	4	UGgu	agGG
	5	AGgu	agGA
	6	AGgu	agAG
	7	AGgu	agGA

Table 3.4 Intron lengths in *Arabidopsis* second-step splicing factors. In each subset of primers used for intron subgroup checks, the lengths of the included introns are listed (from left to right).

Primer set	Unspliced (bp)	fully spliced (bp)	intron lengths (5' - 3') (bp)				
***At*PRP16 I1-4**	1084	544	109	81	158	189	
***At*PRP16 I5-8**	1270	593	195	150	182	150	
***At*PRP16 I9-12**	1365	750	297	106	82	130	
***At*PRP16 I13-16**	1045	604	112	156	87	86	
***At*PRP17-1 I1**	850	466	384				
***At*PRP17-2 I1-2**	1200	985	121	94			
***At*PRP17-2 I3**	1090	370	720				
***At*PRP18A I1**	660	270	390				
***At*PRP18B I1**	660	500	100				
***At*PRP22-1 I1**	485	394	91				
***At*PRP22-2 I1-4**	1227	838	102	86	98	103	
***At*PRP22-2 I5-8**	1686	1175	116	103	119	173	
***At*PRP22-2 I9-12**	852	567	90	86	109		
***At*PRP22-3 I1-3**	730	455	105	92	78		
***At*PRP22-3 I4-8**	1806	610	121	92	120	553	310
***At*PRP22-3 I9-12**	1282	562	175	83	113	349	
***At*PRP22-3 I12-15**	1480	693	349 (same intron as above)	198	132	108	
***At*SLU7-1A I1-3**	1535	859	426	93	157		
***At*SLU7-1A I4-5**	970	768	108	94			
***At*SLU7-1A I6-9**	1150	698	125	143	87	97	
***At*SLU7-1B I1-4**	1671	1396	89	92	94		
***At*SLU7-1B I5-7**	960	654	100	106	100		

Table 3.5 Calculated levels of second-step splicing factors under metal treatment. (**A**) **Band ratios.** The ratio of pre-mRNA to mature mRNA found in each transcript. Colors range based on the amount of product detected for pre-mRNAs in the following order: white, light yellow, yellow, light green, green, light blue, dark blue, lavender, purple. (**B**) **Transcript levels of second-step splicing factors under metal treatments** This includes analysis of all possible transcripts. The measurements for both pre-mRNA and mature transcripts are combined and compared to untreated samples **(C) Mature transcript levels of second-step splicing factors.** Transcript levels of mature mRNA only are compared to untreated samples.

A

	no treatment	10 μM $HgCl_2$	30 μM $HgCl_2$	10 μM $Hg(OAc)_2$	30 μM $Hg(OAc)_2$	2 μM $CdSO_4$	10 μM $CdSO_4$	15 μM $CdSO_4$	10 μM $CuSO_4$	20 μM $CuSO_4$	40 μM $CuSO_4$	20 μM $ZnSO_4$	40 μM $ZnSO_4$	60 μM $ZnSO_4$
PRP16 I1-4	0.12/0.88	0.013/0.99	0.024/0.98	0.19/0.81	0.12/0.88	0.019/0.98	0.25/0.75	0.19/0.81	0.024/0.98	0.018/0.98	0.047/0.95	0.022/0.98	0.0024/0.99	0.060/0.94
PRP16 I5-8	0.13/0.87	0.056/0.94	0.39/0.61	0.096/0.90	0.15/0.85	0.060/0.94	0.42/0.58	0.30/0.70	0.023/0.98	0.0017/0.99	0.054/0.95	0.080/0.92	0.037/0.96	0.11/0.89
PRP16 I9-12	0.032/0.97	0.0016/0.99	0.017/0.98	0.35/0.65	0.043/0.96	0.014/0.99	0.13/0.87	0.073/0.93	0.0026/0.99	0.0022/0.99	0.0081/0.99	0.018/0.92	0.0009/0.99	0.0088/0.99
PRP16 I13-16	0.037/0.96	0.0076/0.99	0.0087/0.99	0.15/0.85	0.017/0.98	0.12/0.88	0.086/0.91	0.061/0.94	0.0091/0.99	0.026/0.97	0.027/0.97	0.037/0.96	0.057/0.94	0.045/0.95
PRP17-1 I1	0.047/0.95	0.0002/0.99	0.016/0.84	0.22/0.78	0.012/0.99	0.15/0.85	0.24/0.91	0.16/0.84	0.0073/0.99	0.0001/0.99	0.031	0.051/0.95	0.0086/0.99	0.14/0.86
PRP17-2 I1-2	0.45/0.55	0.29/0.71	0.28/0.72	0.72/0.28	0.52/0.48	0.38/0.62	0.71/0.29	0.65/0.35	0.22/0.78	0.25/0.75	0.29/0.71	0.37/0.63	0.29/0.71	0.55/0.45
PRP17-2 I3	0.24/0.76	0.0085/0.99	0.071/0.93	0.46/0.54	0.16/0.84	0.14/0.86	0.49/0.51	0.37/0.63	0.033/0.97	0.012/0.99	0.081/0.92	0.17/0.83	0.077/0.92	0.23/0.77
PRP18A I1	0.070/0.93	0.047/0.95	0.047/0.95	0.16/0.84	0.031/0.97	0.041/0.96	0.23/0.77	0.15/0.85	0.027/0.97	0.014/0.99	0.052/0.95	0.045/0.96	0.036/0.94	0.097/0.90
PRP18B I1	0.14/0.86	0.040/0.96	0.054/0.95	0.26/0.74	0.029/0.97	0.074/0.93	0.28/0.72	0.17/0.83	0.020/0.98	0.0088/0.99	0.038/0.96	0.053/0.95	0.032/0.97	0.13/0.87
PRP22-1 I1	0.15/0.85	0.020/0.98	0.041/0.96	0.27/0.73	0.044/0.96	0.036/0.96	0.26/0.74	0.13/0.87	0.062/0.94	0.17/0.83	0.046/0.96	0.065/0.94	0.040/0.96	0.21/0.79
PRP22-2 I1-4	0.40/0.60	0.31/0.69	0.26/0.74	0.67/0.33	0.51/0.49	0.45/0.55	0.80/0.20	0.48/0.52	0.28/0.72	0.37/0.63	0.21/0.79	0.16/0.84	0.34/0.66	0.40/0.60
PRP22-2 I5-8	0.31/0.69	0.34/0.66	0.27/0.73	0.35/0.65	0.27/0.73	0.29/0.71	0.44/0.56	0.39/0.61	0.15/0.85	0.26/0.74	0.17/0.83	0.16/0.84	0.23/0.77	0.24/0.76
PRP22-2 I9-12	0.21/0.71	0.033/0.97	0.050/0.95	0.35/0.65	0.044/0.96	0.041/0.96	0.43/0.57	0.23/0.77	0.032/0.97	0.022/0.98	0.058/0.94	0.056/0.94	0.041/0.96	0.24/0.76
PRP22-3 I1-3	0.37/0.63	0.087/0.92	0.19/0.81	0.57/0.43	0.32/0.68	0.29/0.71	0.42/0.58	0.41/0.69	0.26/0.74	0.053/0.95	0.13/ 0.87	0.23/0.77	0.20/0.80	0.38/0.62
PRP22-3 I4-8	0.045/0.96	0.0082/0.99	0.014/0.99	0.23/0.77	0.031/0.97	0.024/ 0.98	0.20/0.80	0.064/0.94	0.0088/0.99	0.0070/0.99	0.020/0.98	0.0070/0.99	0.017/0.98	0.037/0.96
PRP22-3 I9-12	0.033/0.67	0.0077/0.99	0.019/0.98	0.13/0.87	0.026/0.97	0.021/0.98	0.12/0.88	0.052/0.95	0.011/0.99	0.0049/0.99	0.014/0.99	0.0068/0.99	0.013/0.99	0.024/0.98
PRP22-3 I12-15	0.083/0.92	0.035/0.96	0.024/0.98	0.33/0.67	0.026/0.97	0.10/0.90	0.20/0.80	0.21/0.79	0.0072/0.99	0.017/0.98	0.031/0.97	0.024/0.98	0.021/0.98	0.12/0.88
SLU7-1A I1-3	0.092/0.91	0.14/0.86	0.067/0.93	0.11/0.89	0.17/0.83	0.14/0.86	0.21/0.79	0.076/0.92	0.045/0.96	0.064/0.94	0.024/0.98	0.049/0.95	0.061/0.94	0.12/0.88
SLU7-1A I4-5	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1
SLU7-1A I6-9	0/1	0.11/0.89	0/1	0/1	0/1	0/1	0.25/0.75	0/1	0/1	0/1	0/1	0/1	0/1	0/1
SLU7-1B I1-4	0.026/0.97	0.012/0.99	0.0051/0.99	0.066/0.93	0.013/0.99	0.012/0.99	0.050/0.95	0.035/0.97	0.0040/0.99	0.0076/0.99	0.0051/0.99	0.0094/0.99	0.0066/0.99	0.035/0.97
SLU7-1B I5-7	0.018/0.98	0.0037/0.97	0.0081/0.99	0.092/0.91	0.0068/0.99	0.0068/0.99	0.094/0.91	0.071/0.93	0.0079/0.99	0.0083/ 0.99	0.018/0.98	0.017/0.98	0.0087/0.99	0.020/0.98

Table 3.5 (cont.)

Table 3.5 (cont.)

B

	no treatment	10 μM $HgCl_2$	30 μM $HgCl_2$	10 μM $Hg(OAc)_2$	30 μM $Hg(OAc)_2$	2 μM $CdSO_4$	10 μM $CdSO_4$	15 μM $CdSO_4$	10 μM $CuSO_4$	20 μM $CuSO_4$	40 μM $CuSO_4$	20 μM $ZnSO_4$	40 μM $ZnSO_4$	60 μM $ZnSO_4$
PRP16 I1-4	1.00	0.45	0.62	1.42	0.35	0.48	1.28	0.98	1.83	0.67	1.03	1.64	1.14	2.59
PRP16 I5-8	1.00	1.16	0.78	1.84	0.63	0.67	1.36	1.08	0.67	0.86	0.79	0.98	1.07	1.14
PRP16 I9-12	1.00	162.98	0.84	1.30	0.63	0.91	1.32	1.30	0.97	1.11	1.11	0.84	0.84	1.37
PRP16 I13-16	1.00	1.40	0.68	1.80	0.99	1.68	1.48	1.36	0.96	1.87	0.92	0.73	1.42	2.30
PRP17-1 I1	1.00	1.72	1.76	2.37	1.60	1.47	1.62	1.39	1.02	1.47	1.29	1.04	1.10	1.90
PRP17-2 I1-2	1.00	0.48	1.34	5.29	0.86	0.69	2.54	2.08	0.91	0.50	1.11	1.37	1.10	3.31
PRP17-2 I3	1.00	0.67	0.53	1.37	0.71	0.80	2.90	1.81	0.80	0.67	1.16	1.01	0.84	1.33
PRP18A I1	1.00	1.33	0.59	1.53	0.89	1.78	1.52	1.49	1.07	2.12	1.16	0.82	1.44	2.28
PRP18B I1	1.00	1.19	0.82	2.52	0.97	0.68	1.26	0.82	0.53	0.46	0.82	0.68	0.93	1.34
PRP22-1 I1	1.00	1.15	1.06	3.16	1.52	1.25	1.24	1.18	0.31	0.34	1.67	1.85	0.92	1.73
PRP22-2 I1-4	1.00	0.63	0.50	1.53	0.37	1.98	2.49	4.47	1.42	1.90	2.92	4.00	0.82	2.62
PRP22-2 I5-8	1.00	0.39	0.89	0.94	0.77	1.38	0.80	2.01	1.57	0.53	2.32	3.33	1.29	1.08
PRP22-2 I9-12	1.00	0.38	0.53	0.99	0.36	0.36	1.23	1.75	1.36	0.41	1.83	1.58	1.52	2.61
PRP22-3 I1-3	1.00	0.95	0.87	1.92	0.79	0.61	1.09	0.49	0.06	0.02	0.09	0.75	0.59	1.57
PRP22-3 I4-8	1.00	2.18	2.39	3.40	3.27	0.97	0.36	0.75	0.66	2.19	1.35	0.64	0.89	0.54
PRP22-3 I9-12	1.00	0.94	0.81	1.13	1.23	1.26	1.02	1.16	0.53	0.41	0.60	0.49	0.63	0.85
PRP22-3 I12-15	1.00	0.79	1.39	1.89	1.29	0.80	1.60	1.28	1.05	0.91	1.55	1.35	0.79	0.96
SLU7-1A I1-3	1.00	0.80	0.95	1.40	0.64	0.60	0.68	1.21	1.03	1.24	1.70	1.03	1.06	1.20
SLU7-1A I4-5	1.00	0.74	0.93	1.36	0.62	0.97	1.08	1.30	0.87	0.88	1.27	1.29	0.90	0.84
SLU7-1A I6-9	1.00	0.32	0.92	0.98	1.09	0.66	2.49	1.15	1.62	0.68	1.61	1.90	1.14	0.70
SLU7-1B I1-4	1.00	0.93	1.48	1.67	0.81	0.93	1.35	1.74	0.95	1.24	1.79	1.38	1.02	1.27
SLU7-1B I5-7	1.00	1.16	1.08	1.44	1.10	1.04	1.49	1.50	0.87	1.15	1.49	0.90	0.97	1.29
SLU7-2	1.00	0.07	0.34	2.86	0.19	0.44	4.14	3.31	0.25	0.26	0.77	0.50	0.20	1.79

C

	no treatment	10 µM $HgCl_2$	30 µM $HgCl_2$	10 µM $Hg(OAc)_2$	30 µM $Hg(OAc)_2$	2 µM $CdSO_4$	10 µM $CdSO_4$	15 µM $CdSO_4$	10 µM $CuSO_4$	20 µM $CuSO_4$	40 µM $CuSO_4$	20 µM $ZnSO_4$	40 µM $ZnSO_4$	60 µM $ZnSO_4$
PRP16 I1-4	1.00	0.57	0.78	0.97	0.36	0.50	0.63	0.74	1.91	0.70	1.07	1.74	1.21	2.39
PRP16 I5-8	1.00	1.30	0.84	1.24	0.64	0.72	0.89	0.87	0.76	1.00	0.89	1.06	1.21	1.17
PRP16 I9-12	1.00	1.11	0.86	1.07	0.63	0.91	1.11	1.26	1.00	1.12	1.16	0.86	0.85	1.37
PRP16 I13-16	1.00	1.44	0.70	1.50	1.01	1.41	1.36	1.35	0.98	1.85	0.93	0.74	1.36	2.18
PRP17-1 I1	1.00	1.75	1.75	1.78	1.60	1.20	1.24	1.29	1.09	1.58	1.34	1.08	1.17	1.67
PRP17-2 I1-2	1.00	0.61	1.80	2.28	0.74	0.83	1.23	1.39	1.37	0.77	1.95	1.83	1.54	3.12
PRP17-2 I3	1.00	0.93	0.71	1.02	0.88	0.82	1.55	1.37	0.94	0.81	1.28	1.02	0.93	1.23
PRP18A I1	1.00	1.06	0.82	1.57	1.84	1.18	1.30	1.19	1.16	1.36	1.20	0.88	1.13	1.34
PRP18B I1	1.00	1.32	0.89	1.76	1.09	0.73	0.92	0.80	0.60	0.53	0.92	0.76	1.05	1.35
PRP22-1 I1	1.00	1.37	1.22	2.22	1.79	1.38	1.01	1.19	0.36	0.38	1.80	1.98	1.00	1.37
PRP22-2 I1-4	1.00	0.66	0.58	0.83	0.36	3.91	1.02	6.38	2.56	3.39	5.75	8.48	1.43	4.79
PRP22-2 I5-8	1.00	0.38	0.94	0.97	0.87	1.72	0.73	1.94	2.14	0.60	3.33	4.42	1.71	1.33
PRP22-2 I9-12	1.00	0.52	0.66	0.73	0.45	0.41	0.67	1.50	1.57	0.45	2.00	1.77	1.71	1.90
PRP22-3 I1-3	1.00	1.06	1.20	1.33	0.94	0.93	1.46	1.48	1.31	1.77	0.47	0.88	1.24	0.44
PRP22-3 I4-8	1.00	1.34	1.15	1.48	1.31	1.23	0.93	1.42	1.18	1.79	1.07	0.95	1.43	1.17
PRP22-3 I9-12	1.00	1.01	0.86	0.79	1.31	1.22	0.78	1.03	0.58	0.45	0.66	0.54	0.69	0.81
PRP22-3 I12-15	1.00	0.66	1.61	1.58	1.50	0.92	1.12	1.16	1.26	0.74	1.47	1.76	0.92	0.48
SLU7-1A I1-3	1.00	0.77	0.99	1.36	0.59	0.56	0.56	1.17	1.03	1.24	1.76	1.02	1.06	1.12
SLU7-1A I4-5	1.00	0.74	0.93	1.36	0.62	0.97	1.08	1.30	0.87	0.88	1.27	1.29	0.90	0.84
SLU7-1A I6-9	1.00	0.85	0.73	0.81	0.90	0.58	1.99	1.75	1.43	1.41	1.75	1.23	1.56	2.01
SLU7-1B I1-4	1.00	0.94	1.52	1.60	0.82	0.94	1.30	1.71	0.96	1.24	1.81	1.39	1.03	1.24
SLU7-1B I5-7	1.00	1.17	1.09	1.30	1.12	1.06	1.36	1.43	0.88	1.16	1.49	0.90	0.98	1.29
SLU7-2	1.00	0.07	0.34	2.86	0.19	0.44	4.14	3.31	0.25	0.26	0.77	0.50	0.20	1.79

Table 3.5 (cont.)

Table 3.6 (A) SR protein band ratios. Ratios of alternative splice products **(B) SR protein total transcript levels.** Total transcript levels as compared to untreated (½ MS) samples.

A

		no treatment	$HgCl_2$	$Hg(OAc)_2$	$CdSO_4$	$CuSO_4$	$ZnSO_4$
RSp31	**D**	0.086	0.10	0.098	0.18	0.072	0.13
	C	0.0089	0.011	0.022	0.025	0.0095	0.16
	B	0.085	0.099	0.096	0.19	0.070	0.13
	A	0.91	0.90	0.90	0.81	0.93	0.87
SR30	**B**	0.34	0.31	0.41	0.48	0.38	0.42
	A	0.66	0.69	0.59	0.52	0.62	0.58
SR33	**E**	0.063	0.035	0.031	0.13	0.038	0.051
	D	0.0043	0.0064	0.0091	0.012	0.0058	0.0074
	C	0.12	0.12	0.12	0.20	0.11	0.11
	B	0.027	0.019	0.029	0.043	0.015	0.016
	A	0.78	0.82	0.81	0.62	0.83	0.81
SRp34/ SR1	**C**	0.79	0.78	0.68	0.46	0.67	0.63
	B	0.0099	0.0156	0.0086	0.023	0.018	0.054
	A	0.20	0.21	0.31	0.52	0.31	0.31
U1-70K	**B**	0.17	0.14	0.12	0.37	0.084	0.22
	A	0.83	0.86	0.88	0.63	0.92	0.78

B

	no treatment	$HgCl_2$	$Hg(OAc)_2$	$CdSO_4$	$CuSO_4$	$ZnSO_4$
RSp31	1	0.63	0.75	1.06	0.97	0.74
SR30	1	0.66	0.71	0.70	0.79	0.89
SR33	1	0.69	0.75	0.92	0.80	0.68
SRp34/SR1	1	0.57	0.69	0.74	0.66	0.83
U1-70K	1	0.47	0.58	1.52	0.69	1.49

Table 3.7 (A) Band ratios of non-splicing related proteins. Ratios of alternative splice products. **(B) Total transcript levels of non-splicing related proteins.** Total transcript levels as compared to untreated (½ MS) samples.

A

		no treatment	$HgCl_2$	$Hg(OAc)_2$	$CdSO_4$	$CuSO_4$	$ZnSO_4$
MYB59	**C**	0.12	0.10	0.23	0.23	0.09	0.22
	B	0.43	0.49	0.38	0.39	0.45	0.40
	A	0.45	0.41	0.39	0.38	0.46	0.38
PDF2.2	**B**	0.61	0.34	0.75	0.59	0.43	0.73
	A	0.39	0.65	0.25	0.41	0.57	0.27
WHY1	**B**	0.04	0.02	0.13	0.15	0.03	0.04
	A	0.96	0.98	0.87	0.85	0.97	0.96

B

	no treatment	$HgCl_2$	$Hg(OAc)_2$	$CdSO_4$	$CuSO_4$	$ZnSO_4$
MYB59	1	0.93	1.47	1.65	1.78	1.83
PDF2.1	1	0.70	2.21	2.20	0.31	1.66
WHY2	1	0.94	1.33	1.27	1.27	1.22

Table 3.8 Intron check primers.

Gene	Introns	Forward Primer	Reverse Primer
PRP16	1-4	gctcaaatcgtcgactgag	gtatctcctacgcttttctcc
	5-8	cgagaagaacataggcgtg	ctctttgccagttcagtttcc
	9-12	cgaagggaactcactgttt	gtaagttgagtggtctttcc
	13-16	gagaaaaccaggtgatagtg	ctgcctcaatctcatcttgc
PRP17-1	1	gcttgctgcacagagtttgg	cgagattgatacaacactaaag
PRP17-2	1-2	ggatctcatctgcaatcg	ctctaacatattcatgcacg
	3	ggtcaaagcaagggttatcg	ggtgacgtttgatgcaagg
PRP18A	1	ggaaacaggagcttgatgc	ccatagtcactccaatagg
PRP18B	1	ggtttcgttattcatgatgtgag	ggtctcatgaatgataggaag
PRP22-1	1	cgacgtgctcaagatgtgagg	ggtcatacaacggctcaatcc
PRP22-2	1-4	gaatttcttgccccgttcg	cctgcacatgaacagctttgg
	5-8	ggagtgttcatactgatgttc	ccacactgagcacggcaac
	9-12	ggagagttgcattcgcttgg	cgtttggaaatggtgaggag
PRP22-3	1-4	ggagaaaccggttctggg	ggtacattcagcacaggg
	5-8	ggtttctgaattcttttcgg	ggcaagagaagttcgttgg
	9-12	gttatcgtttatacccattgg	ctaacatctttgacaaacacc
	12-15	catgatgaagattctaatcttcc	gagacatcaaatttatttctgtcc
SLU7-1A	1-3	gaagtcgaagcttcacgatcg	cagcttgaaagatctttgcacc
	4-5	cgaagtacacaaacatgaac	ccttttataatcctccctgc
	6-9	gttcaaacagctaaacatcc	ggaaatctttcattggatcc
SLU7-1B	1-4	cctctactcgaagcttaacg	ctcaatttgtctttcactttgg
	5-7	ccatcccaagctgaattgc	ggaaatttctcattgggtcc

Table 3.9 Non splicing-related primers.

Gene	Name	Forward Primer	Reverse Primer
EF1alpha	AT5G60390	accaccaagtactactgcac	gacctctcaatcatgttgtc
GRP7	At2g21660	gctcactactctcactgtaatcc	ccaagtagaagcataacaagc
MAGO	At1g02140	ggacacggagaggtaatctcg	gtcacaaagtcctcatgttcc
MYB59	At5g59780	atgaaacttgtgcaagaagaa	ctaaaggcgaccactaccatg
PDF2.2	At2g02100	gcaatgaagctctctatgcg	ggtagatcagcaatgtctgg
RSp31	At3g61860	cgagtatgaaactcgccagtcgg	ccagtcccaagaggaagaccttg
SRp30	At1g09140	gtagccgatggaatcgtacg	gtacagccagtggccagttttc
SR33	At1g55310	gaggggaaggagctacactc	gctcttacctctgactggag
SRp34/SR1	At1g02840	gcagtcgttcgagtagaacc	ggactcctagtgtggatagg
SLU7-2	At3g45950	cgtgttcgaacaactgatgg	cgagtttgattccgttggc
U1-70K	At3g50670	gagactccggcgatcctttc	cgaacatactctcgcgattc
UPF1	At5g47010	cctaatgctgacagaagagg	ccaagttttgtctttctcacc
UPF2	At2g39260	ccattgcaaacgtactgaagc	ccaactattaccgcttttgac
UPF3	At1g33980	caaggaaaatttagggttttcg	ggacaaaatccagtggaatgg
Why1	At1g14410	cgtatgtcgcaactcttatcg	cctccatagttaggcgaagc
ZAP-like	At5g62760	cctcatcgatctactcgtcc	tccatgtccatatccacctcc

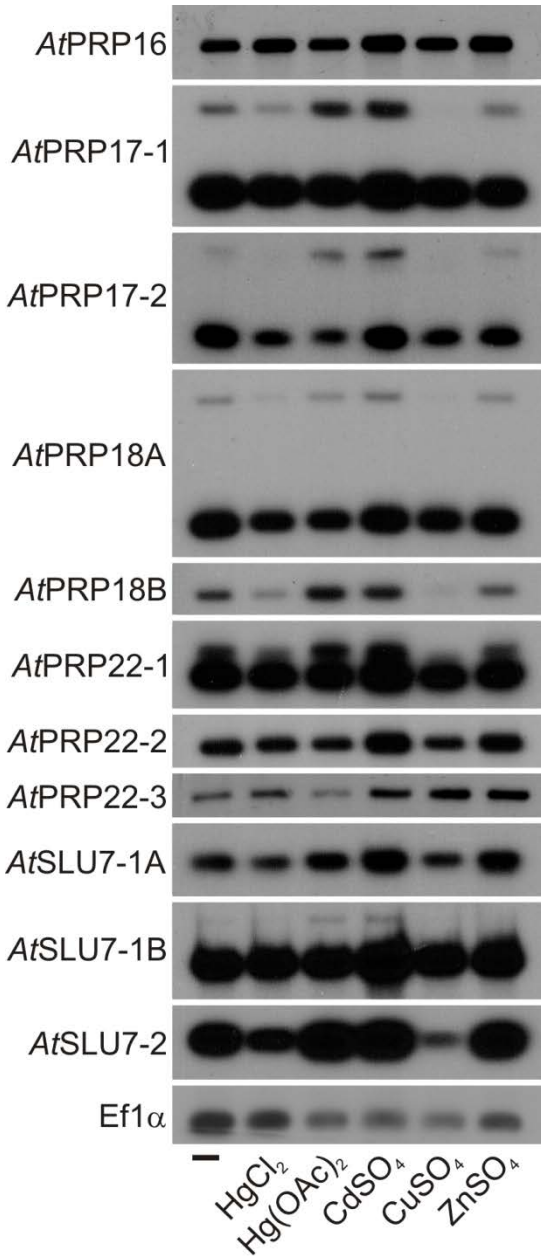


Figure 3.1 Amplification of full-length second-step splicing factor genes under metal stress conditions. Southern blots with ^{32}P-labeled cDNA probes. In panels that include two bands, upper band is pre-mRNA and lower band is mature mRNA. In the cases of *At*PRP16, *At*PRP22-2, *At*PRP22-3, gene-specific primers are unable to detect genomic, and in turn, pre-mRNA.

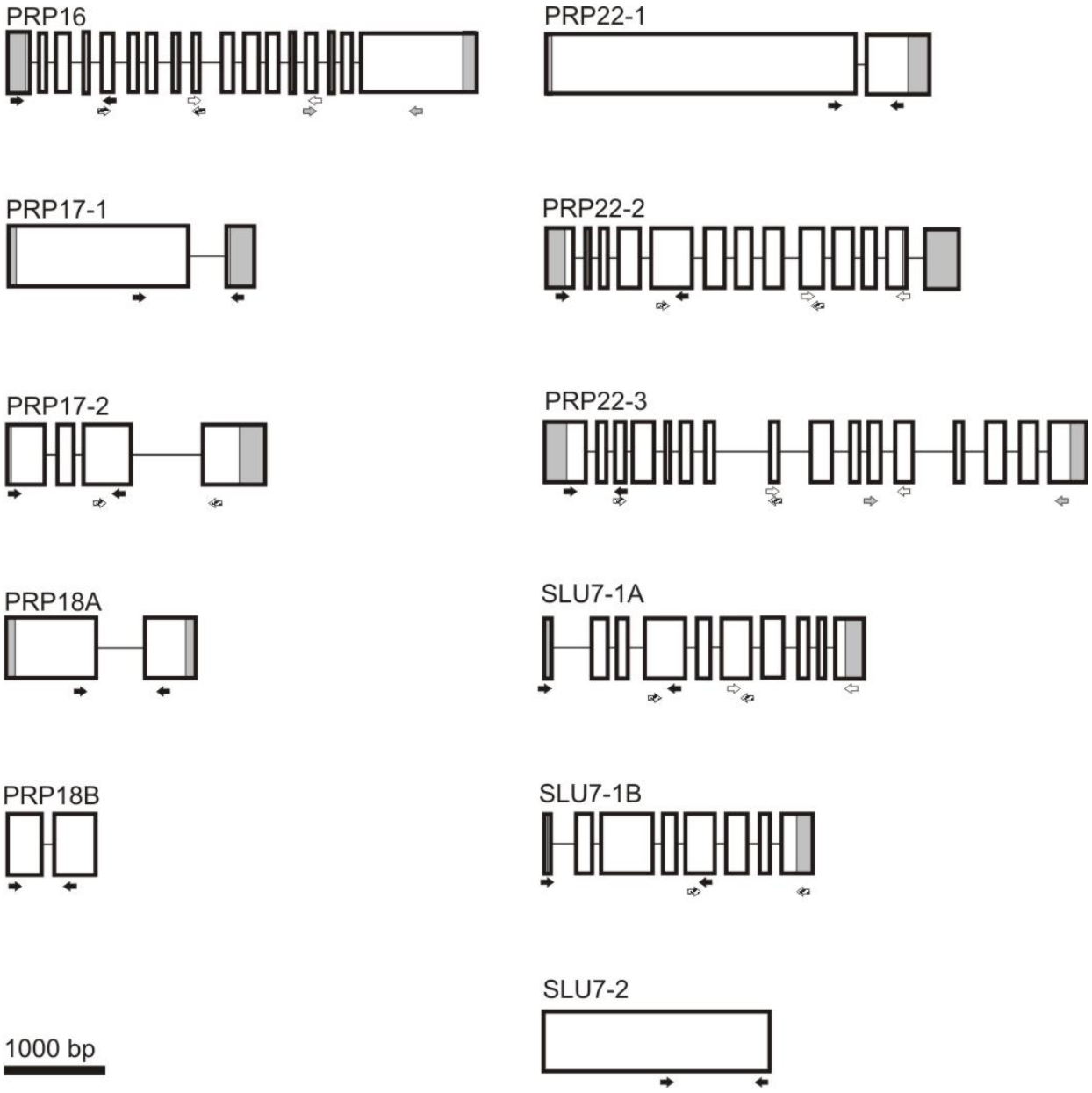


Figure 3.2 Gene structures and primer locations of second-step splicing factors. Gray areas represent the UTRs, open boxes are exons and black bars are introns. Arrows below the genes represent the primers used to check each set of introns for retention. Corresponding patterns within the arrows are paired together.

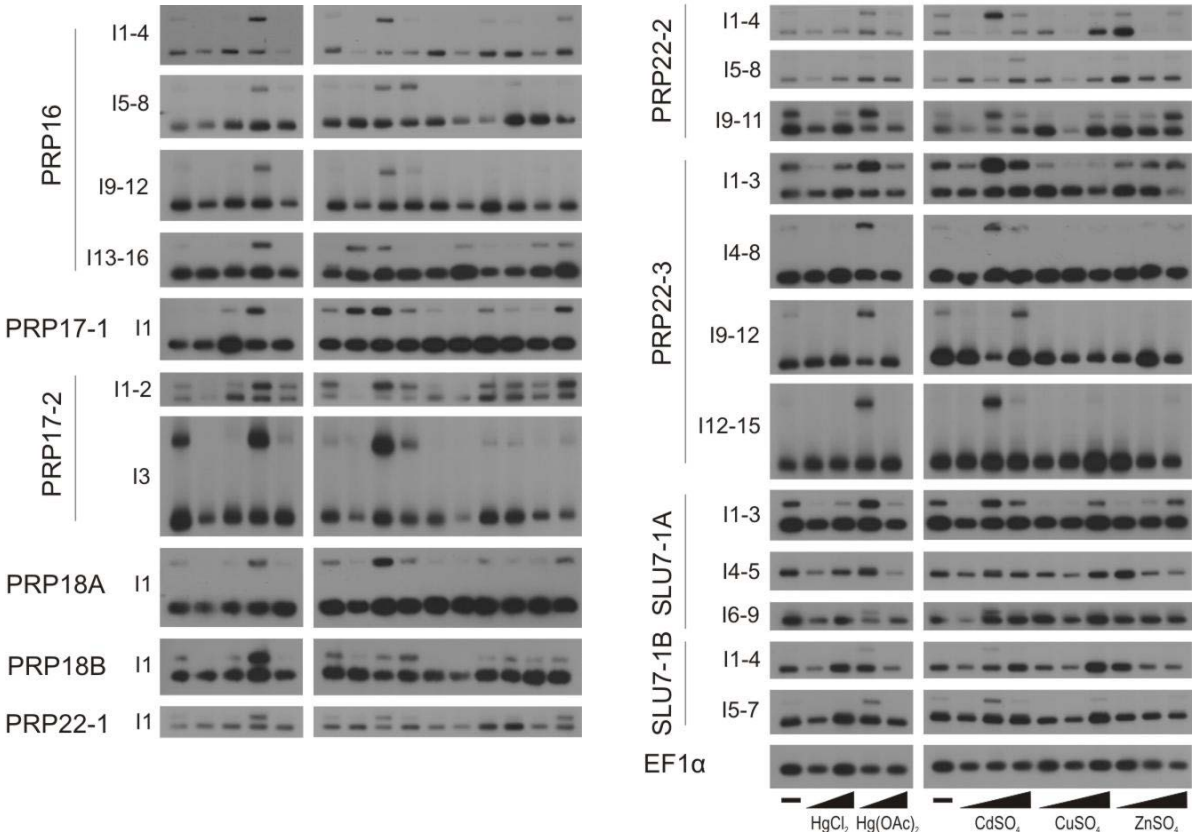


Figure 3.3 Effect of metal treatment on the expression of Arabidopsis second-step splicing factors. Southern blots with ^{32}P-labeled cDNA probes. With the top band representing the pre-mRNA and the bottom band representing the fully spliced mature mRNA. The dashed line samples are from plants grown on ½ MS plates and the metals are present in increasing concentrations.

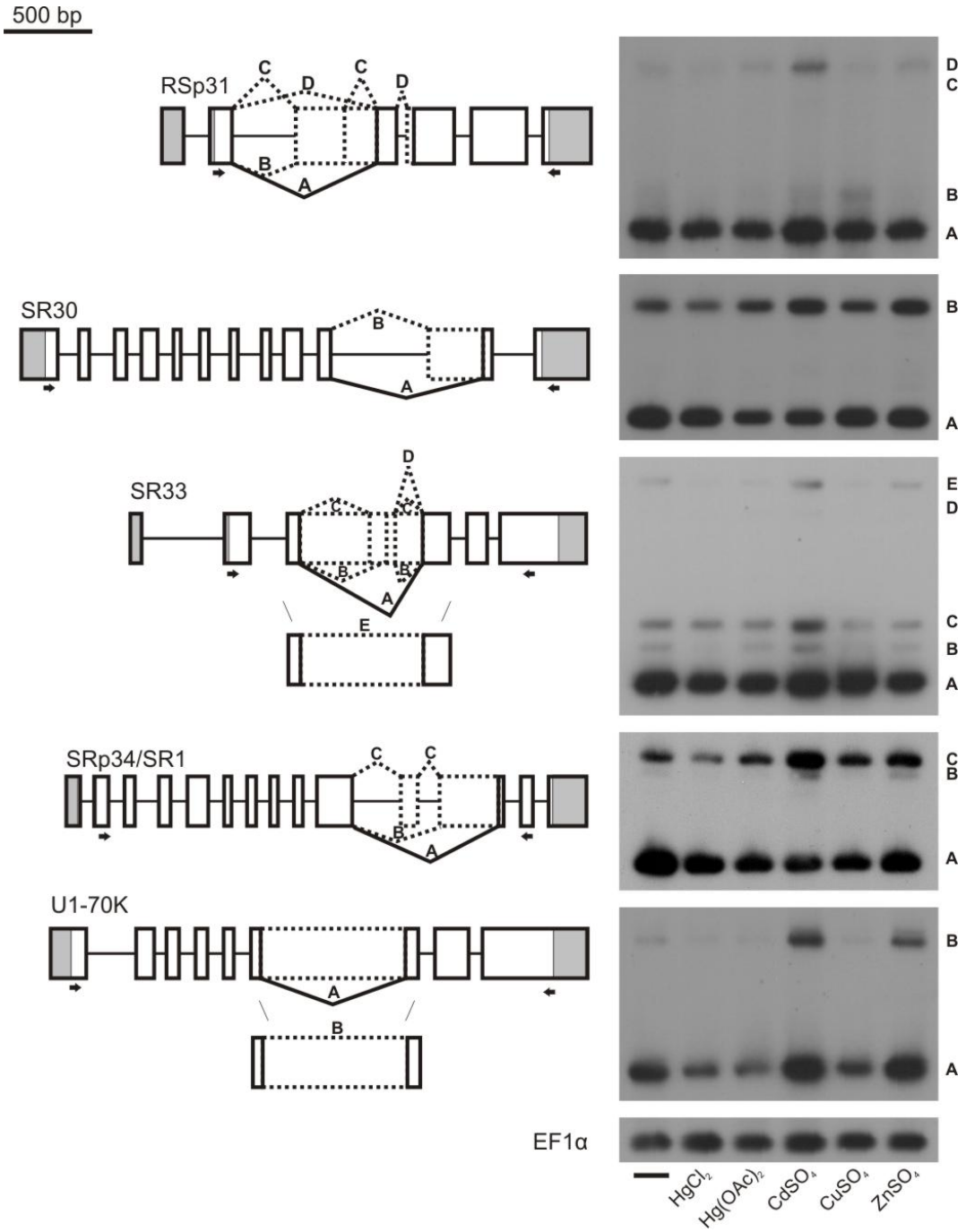


Figure 3.4 Gene structures and metal effects of SR proteins. Southern blots with ^{32}P-labeled cDNA probes with each band on the blot representing a different alternative splice product as labeled to the left with a letter. The corresponding product is pictured to the left in each gene-structure illustration. The dashed line samples are from plants grown on ½ MS plates and the metals are present in increasing concentrations.

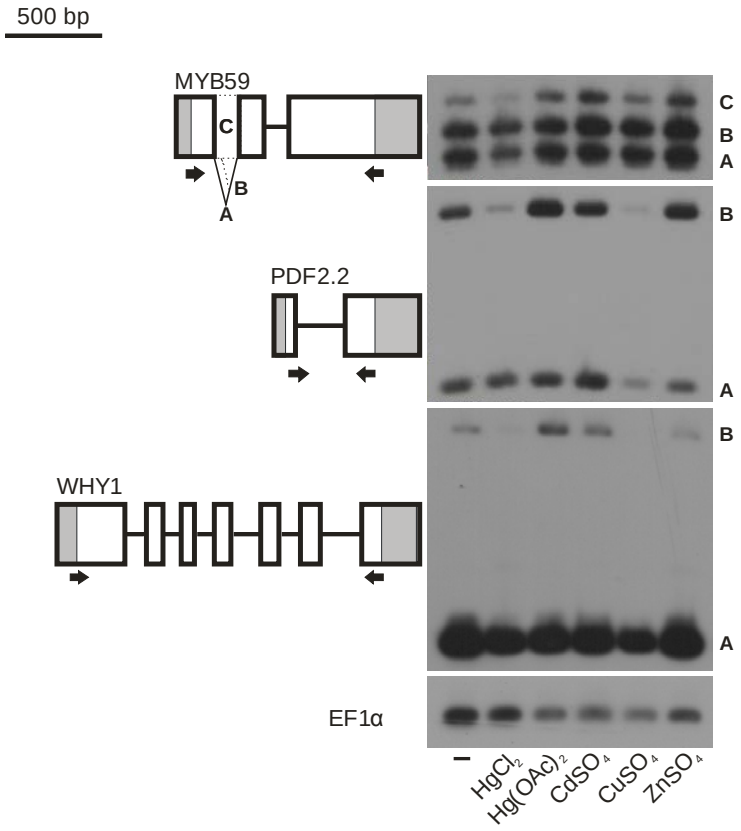


Figure 3.5 Gene structures and metal effects of non-splicing-related genes. Southern blots with ^{32}P-labeled cDNA probes with each band on the MYB59 blot representing a different alternative splice product as labeled to the left with a letter. Higher molecular weight bands on the PDF2.2 and WHY1 blots are representative of pre-mRNA. The corresponding product is pictured to the left in each gene-structure illustration. The dashed line samples are from plants grown on ½ MS plates and the metals are present in increasing concentrations.

3.5 References

Abdel-Ghany, S. E., Burkhead, J. L., Gogolin, K. A., Andrés-Colás, N., Bodecker, J. R., Puig, S., Peñarrubia, L. and Pilon, M. (2005). AtCCS is a functional homolog of the yeast copper chaperone Ccs1/Lys7. *FEBS Letters* **579**, 2307-2312.

Amaravadi, R., Glerum, D. M. and Tzagoloff, A. (1997). Isolation of a cDNA encoding the human homolog of COX17, a yeast gene essential for mitochondrial copper recruitment. *Human Genetics* **99**, 329-333.

Andrés-Colás, N., Sancenón, V., Rodríguez-Navarro, S., Mayo, S., Thiele, D. J., Ecker, J. R., Puig, S. and Peñarrubia, L. (2006). The *Arabidopsis* heavy metal P-type ATPase HMA5 interacts with metallochaperones and functions in copper detoxification of roots. *Plant Journal* **45**, 225-236.

Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796-815.

Balandin, T. and Castresana, C. (2002). AtCOX17, an *Arabidopsis* homolog of the yeast copper chaperone COX17. *Plant Physiology* **129**, 1852-1857.

Becher, M., Talke, I. N., Krall, L. and Krämer, U. (2004). Cross-species microarray transcript profiling reveals high constitutive expression of metal homeostasis genes in shoots of the zinc hyperaccumulator *Arabidopsis halleri*. *Plant Journal* **37**, 251-268.

Blaudez, D., Kohler, A., Martin, F., Sanders, D. and Chalot, M. (2003). Poplar metal tolerance protein 1 confers zinc tolerance and is an oligomeric vacuolar zinc transporter with an essential leucine zipper motif. *Plant Cell* **15**, 2911-2928.

Broadley, M. R., White, P. J., Hammond, J. P., Zelko, I. and Lux, A. (2007). Zinc in plants. *New Phytologist* **173**, 677-702.

Chiang, H.-C., Lo, J.-C. and Yeh, K.-C. (2006). Genes Associated with Heavy Metal Tolerance and Accumulation in Zn/Cd Hyperaccumulator *Arabidopsis halleri*: A Genomic Survey with cDNA Microarray. *Environmental Science & Technology* **40**, 6792-6798.

Cho, U.-H. and Park, J.-O. (2000). Mercury-induced oxidative stress in tomato seedlings. *Plant Science* **156**, 1-9.

Chu, C. C., Lee, W. C., Guo, W. Y., Pan, S. M., Chen, L. J., Li, H. M. and Jinn, T. L. (2005). A copper chaperone for superoxide dismutase that confers three types of copper/zinc superoxide dismutase activity in *Arabidopsis*. *Plant Physiology* **139**, 425-436.

Clemens, S. (2001). Molecular mechanisms of plant metal tolerance and homeostasis. *Planta* **212**, 475-486.

Cuypers, A., Karen, S., Jos, R., Kelly, O., Els, K., Tony, R., Nele, H., Nathalie, V., Suzy, V. S., Frank, V. B. et al. (2011). The cellular redox state as a modulator in cadmium and copper responses in *Arabidopsis thaliana* seedlings. *Journal Plant Physiology* **168**, 309-316.

DalCorso, G., Farinati, S., Maistri, S. and Furini, A. (2008). How plants cope with cadmium: staking all on metabolism and gene expression. *Journal Integrative Plant Biology* **50**, 1268-1280.

del Pozo, T., Cambiazo, V. and González, M. (2010). Gene expression profiling analysis of copper homeostasis in *Arabidopsis thaliana*. *Biochemical and Biophysical Research Communications* **393**, 248-252.

Desbrosses-Fonrouge, A.-G., Voigt, K., Schröder, A., Arrivault, S., Thomine, S. and Krämer, U. (2005). *Arabidopsis thaliana* MTP1 is a Zn transporter in the vacuolar membrane which mediates Zn detoxification and drives leaf Zn accumulation. *FEBS Letters* **579**, 4165-4174.

Drążkiewicz, M., Skórzyńska-Polit, E. and Krupa, Z. (2004). Copper-induced oxidative stress and antioxidant defence in *Arabidopsis thaliana*. *BioMetals* **17**, 379-387.

Glerum, D. M., Shtanko, A. and Tzagoloff, A. (1996). Characterization of COX17, a yeast gene involved in copper metabolism and assembly of cytochrome oxidase. *Journal Biological Chemistry* **271**, 14504-14509.

Golovkin, M. and 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**, 1637-1647.

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

Grill, E. (1987). Phytochelatins, the heavy metal binding peptides of plants: characterization and sequence determination. *Experientia Supplementum* **52**, 317-322.

Grotz, N., Fox, T., Connolly, E., Park, W., Guerinot, M. L. and Eide, D. (1998). Identification of a family of zinc transporter genes from *Arabidopsis* that respond to zinc deficiency. *Proceedings National Academy of Sciences USA* **95**, 7220-7224.

Guerinot, M. L. (2000). The ZIP family of metal transporters. *Biochimica Biophysica Acta (BBA) - Biomembranes* **1465**, 190-198.

Ha, S.-B., Smith, A. P., Howden, R., Dietrich, W. M., Bugg, S., O'Connell, M. J., Goldsbrough, P. B. and Cobbett, C. S. (1999). Phytochelatin synthase genes from *Arabidopsis* and the yeast *Schizosaccharomyces pombe*. *Plant Cell* **11**, 1153-1164.

Halliwell, B. and Gutteridge, J. M. (1984). Oxygen toxicity, oxygen radicals, transition metals and disease. *Biochemical Journal* **219**, 1-14.

Herbette, S., Taconnat, L., Hugouvieux, V., Piette, L., Magniette, M. L. M., Cuine, S., Auroy, P., Richaud, P., Forestier, C., Bourguignon, J. et al. (2006). Genome-wide transcriptome profiling of the early cadmium response of *Arabidopsis* roots and shoots. *Biochimie* **88**, 1751-1765.

Himelblau, E., Mira, H., Lin, S.-J., Cizewski Culotta, V., Peñarrubia, L. and Amasino, R. M. (1998). Identification of a functional homolog of the yeast copper homeostasis gene ATX1 from *Arabidopsis*. *Plant Physiology* **117**, 1227-1234.

Howden, R., Goldsbrough, P. B., Andersen, C. R. and Cobbett, C. S. (1995). Cadmium-sensitive, cad1 mutants of *Arabidopsis thaliana* are phytochelatin deficient. *Plant Physiology* **107**, 1059-1066.

Javot, H. and Maurel, C. (2002). The role of aquaporins in root water uptake. *Annals Botany* **90**, 301-313.

Kampfenkel, K., Kushnir, S., Babiychuk, E., Inzé, D. and Van Montagu, M. (1995). Molecular characterization of a putative *Arabidopsis thaliana* copper transporter and its yeast homologue. *Journal Biological Chemistry* **270**, 28479-28486.

Kobae, Y., Uemura, T., Sato, M. H., Ohnishi, M., Mimura, T., Nakagawa, T. and Maeshima, M. (2004). Zinc transporter of *Arabidopsis thaliana* AtMTP1 is localized to vacuolar membranes and implicated in zinc homeostasis. *Plant Cell Physiology* **45**, 1749-1758.

Korshunova, Y. O., Eide, D., Clark, W. G., Guerinot, M. L. and Pakrasi, H. B. (1999). The IRT1 protein from *Arabidopsis thaliana* is a metal transporter with a broad substrate range. *Plant Molecular Biology* **40**, 37-44.

Krämer, U. (2005). MTP1 mops up excess zinc in *Arabidopsis* cells. *Trends Plant Science* **10**, 313-315.

Lagriffoul, A., Mocquot, B., Mench, M. and Vangronsveld, J. (1998). Cadmium toxicity effects on growth, mineral and chlorophyll contents, and activities of stress related enzymes in young maize plants (*Zea mays* L.). *Plant and Soil* **200**, 241-250.

Li, J., Li, X., Guo, L., Lu, F., Feng, X., He, K., Wei, L., Chen, Z., Qu, L.-J. and Gu, H. (2006a). A subgroup of MYB transcription factor genes undergoes highly conserved alternative splicing in *Arabidopsis* and rice. *Journal Experimental Botany* **57**, 1263-1273.

Li, Y., Dankher, O. P., Carreira, L., Smith, A. P. and Meagher, R. B. (2006b). The shoot-specific expression of γ-glutamylcysteine synthetase directs the long-distance transport of thiol-peptides to roots conferring tolerance to mercury and arsenic. *Plant Physiology* **141**, 288-298.

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

Lorković, Z. J., Lopato, S., Pexa, M., Lehner, R. and Barta, A. (2004). Interactions of *Arabidopsis* RS domain containing cyclophilins with SR proteins and U1 and U11 small nuclear ribonucleoprotein-specific proteins suggest their involvement in pre-mRNA splicing. *Journal Biological Chemistry* **279**, 33890-33898.

Maitani, T., Kubota, H., Sato, K. and Yamada, T. (1996). The composition of metals bound to class III metallothionein (phytochelatin and its desglycyl peptide) induced by various metals in root cultures of *Rubia tinctorum*. *Plant Physiology* **110**, 1145-1150.

Montillet, J.-L., Cacas, J.-L., Garnier, L., Montané, M.-H., Douki, T., Bessoule, J.-J., Polkowska-Kowalczyk, L., Maciejewska, U., Agnel, J.-P., Vial, A. et al. (2004). The upstream oxylipin profile of *Arabidopsis thaliana*: a tool to scan for oxidative stresses. *Plant Journal* **40**, 439-451.

Ortega-Villasante, C., Hernández, L. E., Rellán-Álvarez, R., Del Campo, F. F. and Carpena-Ruiz, R. O. (2007). Rapid alteration of cellular redox homeostasis upon exposure to cadmium and mercury in alfalfa seedlings. *New Phytologist* **176**, 96-107.

Pence, N. S., Larsen, P. B., Ebbs, S. D., Letham, D. L., Lasat, M. M., Garvin, D. F., Eide, D. and Kochian, L. V. (2000). The molecular physiology of heavy metal transport in the Zn/Cd hyperaccumulator *Thlaspi caerulescens*. *Proceedings Naional Academy of Sciences* **97**, 4956-4960.

Puig, S., Mira, H., Dorcey, E., Sancenon, V., Andres-Colas, N., Garcia-Molina, A., Burkhead, J. L., Gogolin, K. A., Abdel-Ghany, S. E., Thiele, D. J. et al. (2007). Higher plants possess two different types of ATX1-like copper chaperones. *Biochemical Biophysical Research Communications* **354**, 385-390.

Reddy, A. S. N. (2004). Plant serine/arginine-rich proteins and their role in pre-mRNA splicing. *Trends Plant Science* **9**, 541-547.

Rellán-Álvarez, R., Ortega-Villasante, C., Álvarez-Fernández, A., Campo, F. and Hernández, L. (2006). Stress responses of *Zea mays* to cadmium and mercury. *Plant and Soil* **279**, 41-50.

Roth, U., von Roepenack-Lahaye, E. and Clemens, S. (2006). Proteome changes in *Arabidopsis thaliana* roots upon exposure to Cd^{2+}. *Journal Experimental Botany* **57**, 4003-4013.

Sancenón, V., Puig, S., Mira, H., Thiele, D. J. and Peñarrubia, L. (2003). Identification of a copper transporter family in *Arabidopsis thaliana*. *Plant Molecular Biology* **51**, 577-587.

Schützendübel, A. and Polle, A. (2002). Plant responses to abiotic stresses: heavy metal-induced oxidative stress and protection by mycorrhization. *Journal Experimental Botany* **53**, 1351-1365.

Semane, B., Dupae, J., Cuypers, A., Noben, J.-P., Tuomainen, M., Tervahauta, A., Kärenlampi, S., Van Belleghem, F., Smeets, K. and Vangronsveld, J. (2010). Leaf proteome responses of *Arabidopsis thaliana* exposed to mild cadmium stress. *Journal of Plant Physiology* **167**, 247-254.

Shanmugam, V., Lo, J.-C., Wu, C.-L., Wang, S.-L., Lai, C.-C., Connolly, E. L., Huang, J.-L. and Yeh, K.-C. (2011). Differential expression and regulation of iron-regulated metal transporters in *Arabidopsis halleri* and *Arabidopsis thaliana* – the role in zinc tolerance. *New Phytologist* **190**, 125-137.

Srinivasan, C., Posewitz, M. C., George, G. N. and Winge, D. R. (1998). Characterization of the copper chaperone Cox17 of *Saccharomyces cerevisiae*. *Biochemistry* **37**, 7572-7577.

Talke, I. N., Hanikenne, M. and Krämer, U. (2006). Zinc-dependent global transcriptional control, transcriptional deregulation, and higher gene copy number for genes in metal homeostasis of the hyperaccumulator *Arabidopsis halleri*. *Plant Physiology* **142**, 148-167.

Tennstedt, P., Peisker, D., Böttcher, C., Trampczynska, A. and Clemens, S. (2009). Phytochelatin synthesis is essential for the detoxification of excess zinc and contributes significantly to the accumulation of zinc. *Plant Physiology* **149**, 938-948.

Thomine, S., Wang, R., Ward, J. M., Crawford, N. M. and Schroeder, J. I. (2000). Cadmium and iron transport by members of a plant metal transporter family in *Arabidopsis* with homology to Nramp genes. *Proceedings National Academy of Sciences USA* **97**, 4991-4996.

Tillemans, V., Leponce, I., Rausin, G., Dispa, L. and Motte, P. (2006). Insights into nuclear organization in plants as revealed by the dynamic distribution of *Arabidopsis* SR splicing factors. *Plant Cell* **18**, 3218-3234.

Van Assche, F. and Clijsters, H. (1990). Effects of metals on enzyme activity in plants. *Plant, Cell and Environment* **13**, 195-206.

van der Zaal, B. J., Neuteboom, L. W., Pinas, J. E., Chardonnens, A. N., Schat, H., Verkleij, J. A. C. and Hooykaas, P. J. J. (1999). Overexpression of a novel *Arabidopsis* gene related to putative zinc-transporter genes from animals can lead to enhanced zinc resistance and accumulation. *Plant Physiology* **119**, 1047-1056.

Wang, B.-B. and Brendel, V. (2006). Genomewide comparative analysis of alternative splicing in plants. *Proceedings National Academy of Sciences USA* **103**, 7175-7180.

Weber, M., Harada, E., Vess, C., Roepenack-Lahaye, E. v. and Clemens, S. (2004). Comparative microarray analysis of *Arabidopsis thaliana* and *Arabidopsis halleri* roots identifies nicotianamine synthase, a ZIP transporter and other genes as potential metal hyperaccumulation factors. *Plant Journal* **37**, 269-281.

Wintz, H., Fox, T., Wu, Y.-Y., Feng, V., Chen, W., Chang, H.-S., Zhu, T. and Vulpe, C. (2003). Expression profiles of *Arabidopsis thaliana* in mineral deficiencies reveal novel transporters involved in metal homeostasis. *Journal Biological Chemistry* **278**, 47644-47653.

Wintz, H. and Vulpe, C. (2002). Plant copper chaperones. *Biochemical Society Transactions* **30**, 732-735.

CHAPTER 4

The effect of heavy metal stresses on the surveillance of second-step splicing factor transcripts

4.1 Introduction

Germination and growth of *Arabidopsis* for three weeks on heavy metal-treated media results in alteration of splicing of gene products (Chapter 3). Genes that are known to go through alternative splicing undergo an alteration to the ratios of their alternative splice products. Conversely, genes that are not alternatively spliced undergo accumulation of pre-mRNA. These effects were most prominent in SR proteins and second-step splicing factors that were treated with 10 μM $Hg(OAc)_2$ or 10 μM $CdSO_4$.

It would be expected that accumulation of pre-mRNA in cases in which there is no alternative splicing should not happen due to the multiple mRNA surveillance mechanisms that are in place in the cell. Because this accumulation still persists, we propose that mRNA surveillance is damaged in the presence of heavy metal stress. There are three known types of mRNA surveillance in eukaryotes: non-stop decay (NSD), no-go decay (NGD) and nonsense-mediate decay (NMD). NSD degradation occurs in transcripts that lack a stop codon as a result of transcription abortion, point mutations that eliminate the stop codon or when premature polyadenylation occurs resulting the ribosome stalling on the mRNA during the elongation phase of translation (Klauer and van Hoof, 2012; van Hoof et al., 2002; Wu et al., 2011). NSD has been best characterized in yeast and it is hypothesized that degradation of the mRNA (and likely the nascent peptide) occur as the ribosome reaches the end of the mRNA. NGD quality control serves to degrade mRNA that has stalled on the ribosome during the elongation phase of translation (Doma and Parker, 2006). These mRNAs are subjected to endolytic cleavage and

subsequent degradation by the exosome. Stalling of the ribosome during elongation is not uncommon, but if this occurs for a prolonged period of time, an empty A site on the ribosome may be bound by the Dom34/Hbs1 complex (Harigaya and Parker, 2010). This can lead to mRNA cleavage, release of the ribosome from the mRNA and/or degradation of the nascent peptide. Prolonged stalling that leads to NGD can result from a structural barrier in the mRNA resulting from strong secondary structure (like hairpin loops, pseudoknots). NMD is activated due to the presence of premature termination codons (PTCs). These are stop codons located within the open reading frame of a gene and are in-frame. When a gene harbors a PTC (PTC+), it is targeted for NMD. A PTC can be found in the intron(s) of unspliced or misspliced genes, but it can also be the result of a nonsense mutation. In *Arabidopsis*, NMD is carried out via three proteins (along with other, yet to be identified, factors): UPF1, UPF2 and UPF3. UPF1 is an ATP-dependent RNA helicase that, along with UPF3, associates with the EJC (Huang and Wilkinson, 2012) and UPF2 is a phosphoprotein that binds both UPF1 and UPF3 and promotes phosphorylation of UPF1 (Huang and Wilkinson, 2012). Knockouts of UPF1 or UPF3 result in the knockout of NMD and accumulation of certain transcripts that would typically be susceptible to degradation (Arciga-Reyes et al., 2006; Hori and Watanabe, 2005).

In addition to the described mechanisms for degradation of nonsense mRNA, NMD can be used for the regulation of normal, non-mutated genes in *Arabidopsis* (Huang and Wilkinson, 2012). Genes that would be likely targets of this type of mRNA are identified by certain characteristics. These include upstream open reading frames (uORFs) in the 5' untranslated region (UTR) (Nyiko et al., 2009), a lengthy 3' UTR (Hori and Watanabe, 2005; Schwartz et al., 2006), the presence of an intron in the 3' UTR (leading to an exon-junction complex (EJC)

(Kertesz et al., 2006) downstream from the stop codon) and a long distance between the final EJC and stop codon (Le Hir et al., 2001)

Heavy metal stress on three-week-old *Arabidopsis* plants leads to an accumulation of pre-mRNA transcript in transcripts that are not alternatively spliced. The sequence of each of the second-step splicing factors reveals that PTCs are present in each retained intron and in every frame (Table 4.1). Additionally, the cDNA libraries that were generated to test this data were created from oligo(dT) primers, primers designed to amplify all transcripts containing a polyadenylated (polyA) tail. Taken together, we predict that pre-mRNA retention in the heavy metal-treated plants is due to disruption of NMD.

It has been observed that certain transcripts, such as MAGO, GRP7 and ZAP-like are affected when NMD is inactivated in UPF1 or UPF3 knockouts or by chemical inactivation using the translation inhibitor cycloheximide (CHX). These transcripts were analyzed to determine if the same effects were observed when the plants are treated with metal alone. Additionally, each of the UPF transcripts and second-step splicing factor transcripts were analyzed when NMD was inactivated using CHX with and without heavy metal stress. The combination of both heavy metal and CHX treatments is intended to offer insight into the many layers of intricacy in regulation. These data indicate that some of the second-step splicing factor transcript levels are regulated by NMD while others are not. Because most of these genes contain *cis* features that indicate that they could be susceptible to this type of regulation, we propose that these rules are not strict and serve only as a guide for prediction of NMD regulation.

4.2 Results

Mature transcript levels of MAGO, the EJC protein is affected by heavy metal stress

When NMD is inactivated in knockouts of either UPF1 or UPF3, EJC protein expression is decreased (Kerenyi et al., 2008; Kim et al., 2009). To determine the extent to which the effects of external metal stress is mediated by NMD, the expression of MAGO transcripts, encoding one member of the EJC complex, were determined under heavy metal stress conditions (Figure 4.1A). MAGO mature mRNA accumulation was lowest in the presence of copper and zinc, two metals that did not induce significant pre-mRNA accumulation of the second-step splicing factor transcripts, (previous data), was not affected in the presence of mercury and was increased in the presence of cadmium, a metal that did result in an accumulation of pre-mRNA (Figure 4.1B). MAGO pre-mRNA accumulation was increased with $HgCl_2$ and $Hg(OAc)_2$ and, to some extent, with $CdSO_4$. Analysis of the ratio of pre-mRNA and mature mRNA indicates that MAGO processing is affected under several treatment conditions (Figure 4.1C) with substantially more pre-mRNA accumulating with $HgCl_2$, $Hg(OAc)_2$ and ZnSO4 and less pre-mRNA accumulating with $CdSO_4$.

Alternative splicing of NMD-relevant genes

GRP7 is an alternatively spliced transcript that exhibits increased stability of higher molecular weight alternatively spliced products in the absence of UPF1 and UPF3 (Heintzen et al., 1997; Schöning et al., 2008). Its expression levels were analyzed with gel blot (Figure 4.2A). In the presence of both $HgCl_2$ and $Hg(OAc)_2$, this alternative splice product is stabilized (Figure 4.2B), but it is less abundant in $CuSO_4$-treated plants. Other metal treatment conditions resulted in little to no change in alternative splice products in the GRP7 transcript.

ZAP-like is a gene of unknown function that has sequence homology to the ZAP family of proteins and its transcript levels were analyzed in the presence of the metal treatments (Figure 4.3A). It is predicted to be alternatively spliced with its alternatively spliced isoform containing a PTC (PTC+) (Hori and Watanabe, 2005). When NMD is inactivated in UPF3 knockout plants, the PTC+ isoform levels increase greatly (Hori and Watanabe, 2005). The PTC+ form (upper band) of the ZAP-like mRNA did accumulate in the $CdSO_4$-treated plant, but not in any other treated plants (Figure 4.3B).

UPF1, UPF2 and UPF3 are variably expressed in Hg- and Cd-treated plants

When NMD is knocked out, it is expected that both aberrant transcripts containing PTCs and normal transcripts that are regulated by NMD will accumulated in the plant. In order to gain insight into the possibility that metal stress is causing the accumulation of pre-mRNA in plants, these effects are compared between plants treated with 10 μM $Hg(OAc)_2$ or 10 μM $CdSO_4$ and plants that are NMD-deficient. Treatment of *Arabidopsis* with CHX induces a halt in translation without releasing the ribosome from the nascent mRNA transcript or decapping the transcript (Schneider-Poetsch et al., 2010; Xu et al., 2006). Because mRNA surveillance is dependent upon translation, this also serves to halt NMD. Plants that are grown on agar plates cannot be completely exposed to CHX treatment because the entire root tissue is not exposed to the chemical. Because of this, *Arabidopsis* seedlings were germinated and grown for two weeks in rich, liquid ½ MS media (no treatment) and then treated with 10 μM of $Hg(OAc)_2$, 10 μM $CdSO_4$ or water for one week before treatment with CHX or the control of water. Plants were not germinated in the metal via this method and as a result of this technical difference, pre-mRNA accumulation was minimal. A general effect resulting from the treatment of CHX was, however, observed.

UPF1 and UPF3 are necessary for NMD in *Arabidopsis* (Arciga-Reyes et al., 2006; Hori and Watanabe, 2005; Yoine et al., 2006), and while the role of UPF2 in *Arabidopsis* has not been determined experimentally, it is necessary for NMD in yeast (Cui et al., 1995; He and Jacobson, 1995). The transcript accumulation of all three genes are not significantly altered in the presence of each $Hg(OAc)_2$ type (Figure 4.4A). The expression level of each increases approximately 1.5-fold in the presence of $CdSO_4$ (Figure 4.4B). When treated with CHX alone, each of the UPF transcripts accumulates. When combining CHX and the metals, the expression levels alter when comparing the accumulation of transcript to that of no metal and no CHX. But when the expression levels are compared within the metal-treated and metal-treated with CHX subset (Figure 4.4C), $Hg(OAc)_2$ levels are not significantly altered. When comparing the expression levels of the UPF genes in the presence of $CdSO_4$ with or without CHX (Figure 4.4D), UPF1 accumulation increases by 2-fold, UPF2 accumulation increases by 1.5-fold, and UPF3 accumulation decreases.

Treatment of plants with CHX causes accumulation of some second-step splicing factors

These above described conditions were used for the analysis of second-step splicing factors. Some *Arabidopsis* second-step splicing factors exhibit features that are implicated in NMD-mediated regulation of expression levels. These include an upstream ORF or a start codon in the 5' UTR (Nyiko et al., 2009), a lengthy 3' UTR (Hori and Watanabe, 2007), and intron in the 3' UTR (Kurihara et al., 2009) and a short distance between the final EJC and the stop codon of the gene (Scofield et al., 2007). The presence of these features found in the second-step splicing factor transcripts are detailed in Table 4.2. Because it is currently predicted that 13% of the *Arabidopsis* genome is regulated by NMD (Kalyna et al., 2012), and many of the second-step

splicing factors exhibit features that are common in NMD-regulated genes, we tested to see if any of these genes are regulated via NMD regardless of the presence of aberrant processing or mutation resulting in a PTC or nonsense codon (Figure 4.5A). The gene transcripts that accumulate in the presence of CHX alone are PRP16, PRP22-1, PRP22-2, SLU7-1A, SLU7-1B and SLU7-2 (Figure 5.5B). When treated with metal in combination with CHX, the accumulation of each transcript varies with no discernible pattern. Two genes are not affected by the presence of CHX alone: PRP17-1 and PRP18B. PRP22-3 is downregulated in the presence of CHX, and effect that could be related to downstream effects of either loss of NMD or translation. Three special cases exist (Figure 5.5C). PRP22-3 levels decline with CHX treatment and both PRP17-2 and PRP18B have differentially accumulating pre-mRNA transcripts (Figures 5.5D, E) which could offer some insight into the observed effect of pre-mRNA accumulation in plants that are germinated in the presence of metals. PRP17-2 accumulation of pre-mRNA is not significantly altered with respect to mature mRNA, but the overall mature transcript levels are increased with CHX treatment. PRP18B precursor mRNA levels do accumulate with CHX, regardless of the presence of either metal. Additionally, the mature transcript of PRP18B decreases as compared to untreated plants to nearly undetectable levels.

Combination of CHX treatment and heavy metal treatment can cause some accumulation of pre-mRNA

Under these experimental conditions, treatment of the plants with 10 μM $Hg(OAc)_2$ does not result in significant changes of mature transcript accumulation. When CHX is added to $Hg(OAc)_2$-treated plants (Figure 5.5F), the level of accumulation does not significantly increase or decrease for many of the transcripts. Notable exceptions include PRP17-2, PRP22-2 and PRP22-3. Both PRP17-2 and PRP22-2 accumulate with treatment of $Hg(OAc)_2$ and CHX at the

same levels as treatment with CHX (no metal). PRP22-3 transcript levels decrease with $Hg(OAc)_2$ and CHX but not to the extent that they do with CHX alone.

Plants that are treated with $CdSO_4$ alone also exhibit varying levels of transcript accumulation (Figure 5.5F). PRP17-1 and PRP22-2 accumulate with metal treatment alone while PRP22-3 and SLU7-2 transcript levels decline with metal treatment. At the addition of CHX to the $CdSO_4$-treated plants, no change is observed in the accumulation of PRP16, PRP17-1, PRP17-2, PRP22-1 and SLU7-1B relative to the transcript levels in $CdSO_4$ treatment alone. The only transcript that accumulates is SLU7-2, while PRP18A, PRP22-2, PRP22-3 and SLU7-1A transcript levels decrease relative to no CHX treatment. These data do not follow an obvious pattern, but they are detailed in Table 3.

4.3 Discussion

It is evident that the mode of heavy metal treatment on the plants is important with regards to the resulting effects on the splicing and accumulation of transcript. As previously demonstrated, plants that are germinated and grown for three weeks in the presence of varying concentrations of heavy metals have an alteration of splicing, resulting in accumulation of pre-mRNA in genes that do not undergo alternative splicing. This accumulation can also be evidence of another issue with the metabolism of RNA; the accumulation of this pre-mRNA is taking place because the PTC+ pre-mRNA transcripts are not being degraded.

When plants are treated with 10 μM $Hg(OAc)_2$ or 10 μM $CdSO_4$ after growing in liquid media for two weeks, they do not exhibit as much precursor accumulation as plants that are germinated under these conditions. This allows for examination of the accumulation of mature transcript under metal treatment conditions as well as CHX treatment conditions with and without metal. Additionally, we observed that metal treatment caused alterations of the splice

products of the SR proteins (previous work). Many SR proteins are alternatively spliced, and the accumulation of alternative splice products can be regulated, in part, by NMD (NAS-nonsense-associated alternative splicing) (Ali and Reddy, 2008)

MAGO and GRP7 each are affected by loss of NMD through UPF knockouts. Each of these effects is mirrored in Hg-treatment, but not in Cd-treatment. The effects of ZAP-like under UPF knockout conditions are present in Cd-treated plants only. This indicates that each of these metals do cause a loss of NMD, but that this effect is partial, and it is might be specific to different types of genes.

When the transcript accumulation levels of each of the UPF genes are examined with CHX treatment alone, they accumulate to varying degrees. When treated with $Hg(OAc)_2$ alone, there is no significant change in transcript accumulation. However, when treated with $CdSO_4$ alone, each of these transcripts accumulates by approximately 50%. When treated with $Hg(OAc)_2$ or $CdSO_4$ and CHX, the levels of UPF1 and UPF2 transcripts do not significantly alter and UPF3 transcript accumulation decreases; it is as if normal NMD is rescued by the presence of the heavy metals.

There is evidence of rescue of nonsense mutations via exon skipping being developed as a means as a type of therapy in human diseases. Among these are engineering of specific exon-skipping via the use of modified small nucleolar RNA (snRNA)-containing vectors (Chaouch et al., 2009; Goyenvalle et al., 2009) One limitation to this method of therapy is the fact that PTC+ transcripts are targeted for degradation before the possibility of read-through. The rescue of nonsense mutations by stabilization of PTC+ transcripts via the use of the drugs Amlexanox (Gonzalez-Hilarion et al., 2012), PTC124 (Welch et al., 2007), and gentamicin (Bidou et al., 2004; Linde et al., 2007) have been effective in this manner. Additionally, alternative

polyadenylation can function as a pathway for rescue from NMD (Gilat and Shweiki, 2007). It would be exciting to examine the possibility of metal treatment as a method for effecting NMD-related defects in gene expression. This could be an inexpensive and easily executed method for accumulating pre-mRNA, especially at specific time-points in the plant life cycle.

4.4 Materials and Methods

Plant growth conditions and material

Wild type *Arabidopsis thaliana* (Columbia ecotype) seeds were sterilized by submerging in 70% alcohol for 30 seconds, 20% bleach and a drop of Tween20 for 15 minutes, and washed three times with sterile water. Seeds were then grown in ½ MS liquid media (pH 5.7) under long day conditions (22°C with 16-h-light and 8-h-dark cycles) and shaken at 30 rpm for 3 weeks. After the first 2 weeks, they were supplemented with 10 μM $Hg(OAc)_2$ or 10 μM $CdSO_4$ and grown for 1 additional week. After three weeks, CHX in a concentration of 50 μM was added to the liquid cultures for 3 hours. Plants were rinsed with sterile water and blotted dry. They were then snap-frozen in liquid nitrogen and RNA was extracted immediately.

RNA isolation and cDNA synthesis

RNA was extracted by the beadbeater method described in Thimmapuram et al. (2005) including DNase treatment. The RNA was stored in 100 μl of water, snap frozen and stored at -80°C. cDNA was synthesized by Superscript III (Invitrogen) according to the user manual. Specifically, 1 μg of total RNA was used in a 20 μl reaction containing 200 ng of $oligo(dT)_{19}$, 1 μl 10 mM dNTP mix and 1 μl 0.1 M DTT. To obtain a final working concentration of cDNA, the sample was diluted to 40 μl in water.

Analysis of transcript levels

Semi-quantitative reverse transcriptase PCR (RT-PCR) primers are listed in Table 4.4. cDNA was amplified by 25 cycles of PCR, with the exception of EF1α, which was amplified for 18 cycles. Each cycle consists of 94°C for 45 seconds, 58°C for 45 seconds and 72°C for 2 minutes. The first cycle is preceded by a 3 minute 95°C denaturation period and followed by a 10 minute 72°C elongation period. PCR products were run on 1.3% agarose gels, transferred to Hybond-N (Amersham-Pharmacia Biotech), and probed with ^{32}P-labeled DNA probe made with random primers (Invitrogen) and gene-specific cDNA template. Blots were hybridized and washed according to Lou et al (1993). Band strength was determined by ImageJ (NIH) (Rasband, W. S. ImageJ, US National Institutes of Health, Bethesda, MD, USA) after background signal was reduced with a rolling ball radius of 50-70 pixels. Transcript levels were normalized to the corresponding EF1α measurements and compared to untreated samples. These performed in biological duplicate and fold-change is plotted on a logarithmic scale. The p-value is less than 0.05.

Table 4.1 Summary of PTCs in the second-step splicing factors. The PTCs listed below are the first, in-frame PTCs encountered when each intron is retained alone. Positions are indicated within either an intron or an exon, with the first position being the first nucleotide of that region (*i.e.* the first nucleotide within an intron is the nucleotide immediately subsequent to the intron/exon junction). I = intron, nt = nucleotide

PRP16 – At5g13010

Intron retained	**PTC**	**Position of first nt of PTC**
Intron 1	TAA	43 of I1
Intron 2	TGA	17 of I2
Intron 3	TAG	36 of I2
Intron 4	TGA	73 of I4
Intron 5	TGA	51 of I5
Intron 6	TAA	66 of I6
Intron 7	TAG	31 of I7
Intron 8	TAA	2 of I8
Intron 9	TGA	63 of I9
Intron 10	TAG	6 of I10
Intron 11	TGA	58 of I11
Intron 12	TGA	28 of I12
Intron 13	TAG	49 of I13
Intron 14	TGA	33 of I14
Intron 15	TAA	62 of I15
Intron 16	TAG	2 of I16

PRP17-1 – At1g10580

Intron retained	**PTC**	**Position of first nt of PTC**
Intron 1	TAA	91 of I1
Intron 2	TGA	85 of I2

PRP17-2 – At5g54520

Intron retained	**PTC**	**Position of first nt of PTC**
Intron 1	TAG	39 of I1
Intron 2	TGA	26 of E3
Intron 3	TAG	70 of I3

PRP18A – At1g03140

Intron retained	**PTC**	**Position of first nt of PTC**
Intron 1	TAA	25 of I1

PRP18B – At1g54590

Intron retained	**PTC**	**Position of first nt of PTC**
Intron 1	TAA	100 of I1
Intron 2	TAA	70 of I2

PRP22-1 – At3g26560

Intron retained	PTC	Position of first nt of PTC
Intron 1	TAG	6 of I1

PRP22-2 – At2g26370

Intron retained	PTC	Position of first nt of PTC
Intron 1	TGA	83 of I1
Intron 2	TAA	15 of I2
Intron 3	TGA	49 of E4
Intron 4	TAA	2 of I4
Intron 5	TAA	19 of I5
Intron 6	TAG	13 of I6
Intron 7	TGA	129 of E8
Intron 8	TGA	41 of I8
Intron 9	TAA	58 of I9
Intron 10	TAG	20 of I10
Intron 11	TAA	31 of I11
Intron 12	This intron falls in the 3’ UTR	

PRP22-3 – At1g27900

Intron retained	PTC	Position of first nt of PTC
Intron 1	TGA	62 of I1
Intron 2	TAA	17 of I2
Intron 3	TAA	2 of I3
Intron 4	TGA	14 of E5
Intron 5	TGA	31 of I5
Intron 6	TGA	25 of I6
Intron 7	TAA	2 of I7
Intron 8	TAA	2 of I8
Intron 9	TAA	27 of I9
Intron 10	TAG	10 of E11
Intron 11	TGA	59 of I11
Intron 12	TGA	169 of I12
Intron 13	TAG	2 of I13
Intron 14	TAG	43 of I14
Intron 15	TAA	2 of I15

Table 4.1 (cont.)

SLU7-1A – At1g65660

Intron retained	**PTC**	**location**
Intron 1	TAA	Position 27 of I1
Intron 2	TAA	Position 7 of I2
Intron 3	TAA	Position 131 of I3
Intron 4	TGA	Position 74 of I4
Intron 5	TGA	Position 7 of I5
Intron 6	TGA	Position 55 of I6
Intron 7	TAA	Position 41 of I7
Intron 8	TAA	Position 58 of I8
Intron 9	TAA	Position 13 of I9

SLU7-1B – At4g37120

Intron retained	**PTC**	**location**
Intron 1	TAA	Position 42 of I1
Intron 2	TGA	Position 49 of I2
Intron 3	TAA	Position 65 of I3
Intron 4	TGA	Position 10 of I4
Intron 5	TAG	Position 4 of I5
Intron 6	TGA	Position 12 of I6
Intron 7	TAG	Position 12 of E8

Table 4.1 (cont.)

Table 4.2 Properties of second-step splicing factors found in transcripts that are NMD-regulated.

	ATGs in 5' UTR[1]	3' UTR[2] (>200-300 triggers NMD)	intron in 3' UTR[2]	Previous evidence?	distance from final intron to stop codon (55 bp rule)[3]	half life[4]
PRP16	no	201	no	no	1600	3-6
PRP17-1	no	279	no	no	4	1-3
PRP17-2	no	267	no	accum. UPF3 KO[5]	370	1-3
PRP18A	no	99	no	no	415	unknown
PRP18B	no	undetermined	no	no	420	unknown
PRP22-1	no	218	no	no	390	1-3
PRP22-2	no	538	yes, 173 bp	no	185	1-3
PRP22-3	yes, 2	162	no	no	145	3-6
SLU7-1A	no	188	no	no	75	3-6
SLU7-1B	no	171	no	no	150	6-12
SLU7-2	no	undetermined	no	no	N/A	unknown

[1] **Nyiko, T., Sonkoly, B., Merai, Z., Benkovics, A. H. and Silhavy, D.** (2009). Plant upstream ORFs can trigger nonsense-mediated mRNA decay in a size-dependent manner. *Plant Molecular Biology* **71**, 367-378.

[2] **Hori, K. and Watanabe, Y.** (2007). Context analysis of termination codons in mRNA that are recognized by plant NMD. *Plant Cell Physiology* **48**, 1072-1078.

[3] **Scofield, D. G., Hong, X. and Lynch, M.** (2007). Position of the final intron in full-length transcripts: determined by NMD? *Molecular Biology Evolution* **24**, 896-899.

[4] **Narsai, R., Howell, K. A., Millar, A. H., O'Toole, N., Small, I. and Whelan, J.** (2007) Genome-wide analysis of mRNA decay rates and their determinants in *Arabidopsis thaliana*. *Plant Cell* **19**, 3418-3436.

[5] **Kurihara, Y., Matsui, A., Hanada, K., Kawashima, M., Ishida, J., Morosawa, T., Tanaka, M., Kaminuma, E., Mochizuki, Y., Matsushima, A. et al.** (2009). Genome-wide suppression of aberrant mRNA-like noncoding RNAs by NMD in *Arabidopsis*. *Proceedings National Academy of Sciences USA* **106**, 2453-2458.

Table 4.3 Summary of expression levels of second-step splicing factors with and without CHX or metal treatment

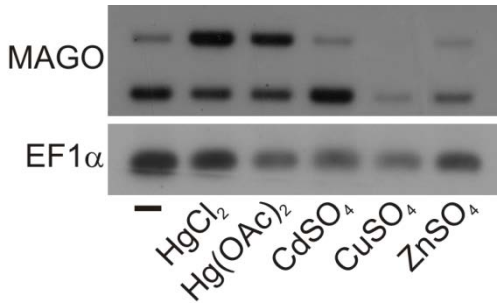
	no metal	no metal +CHX	$Hg(OAc)_2$	$Hg(OAc)_2$ +CHX	$CdSO_4$	$CdSO_4$ +CHX	
PRP16	-	↑	-	-	-	-	PRP16 accumulates when NMD is knocked out, but this effect is negated by the presence of metal
PRP17-1	-	-	-	↑	↑	↑	PRP17-1 does not accumulate when NMD is knocked out, does accumulate in the presence of Hg and CHX and does accumulate in the presence of Cd with or without CHX
PRP17-2	-	↑	-	-	-	-	PRP17-2 accumulates when NMD is knocked out, but this effect is negated by the presence of metal
PRP18A	-	-	-	-	-	-	PRP18 levels are unaffected
PRP18B	-	-	-	-	-	-	Mature PRP18B levels are undetectable, PRP18B pre-mRNA accumulates in the presence of CHX regardless of metal presence
PRP22-1	-	↑	-	-	-	-	PRP22-1 accumulates when NMD is knocked out, but this effect is negated by the presence of metal
PRP22-2	-	↑	-	↑	-	-	PRP22-2 accumulates when NMD is knocked out and does accumulate in the presence of Hg and CHX
PRP22-3	-	↓	-	↓	↓	↓	PRP22-3 levels are decreased in the presence of CHX alone, in the presence of Hg +CHX, in the presence of Cd regardless of the presence of CHX
SLU7-1A	-	↑	-	-	-	-	SLU7-1A accumulates when NMD is knocked out, but this effect is negated by the presence of metal
SLU7-1B	-	↑	-	-	-	-	SLU7-1B accumulates when NMD is knocked out, but this effect is negated by the presence of metal
SLU7-2	-	↑↑	-	↑	↓	↑↑	SLU7-2 accumulates when NMD is knocked out with no metal and with both metals

All genes are present in stable amounts with $Hg(OAc)_2$ alone

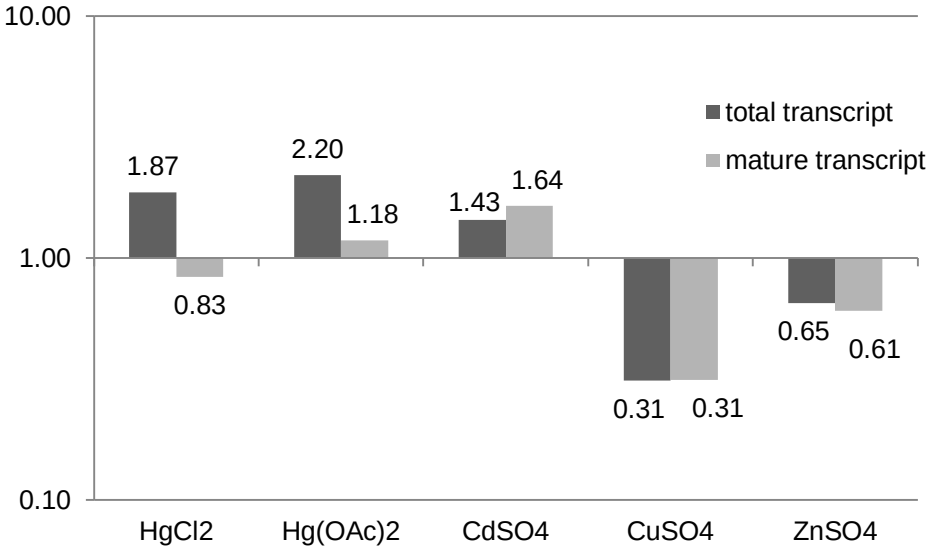
Table 4.4 Southern blot primers.

Gene	Forward Primer	Reverse Primer
PRP16	gagaaaaccaggtgatagtg	ctgcctcaatctcatcttgc
PRP17-1	gcttgctgcacagagtttgg	cgagattgatacaacactaaag
PRP17-2	ggtcaaagcaagggttatcg	ggtgacgtttgatgcaagg
PRP18A	ggaaacaggagcttgatgc	ccatagtcactccaatagg
PRP18B	ggtttcgttattcatgatgtgag	ggtctcatgaatgataggaag
PRP22-1	cgacgtgctcaagatgtgagg	ggtcatacaacggctcaatcc
PRP22-2	ggagagttgcattcgcttgg	cgtttggaaatggtgaggag
PRP22-3	catgatgaagattctaatcttcc	gagacatcaaatttatttctgtcc
SLU7-1A	gttcaaacagctaaacatcc	ggaaatctttcattggatcc
SLU7-1B	ccatcccaagctgaattgc	ggaaatttctcattgggtcc
SLU7-2	cgtgttcgaacaactgatgg	cgagtttgattccgttggc
UPF1	cctaatgctgacagaagagg	ccaagttttgtctttctcacc
UPF2	ccattgcaaacgtactgaagc	ccaactattaccgcttttgac
UPF3	caaggaaaatttagggttttcg	ggacaaaatccagtggaatgg
MAGO	ggacacggagaggtaatctcg	gtcacaaagtcctcatgttcc
GRP7	gctcactactctcactgtaatcc	ccaagtagaagcataacaagc
ZAP-like	cctcatcgatctactcgtcc	tccatgtccatatccacctcc
EF1α	accaccaagtactactgcac	gacctctcaatcatgttgtc

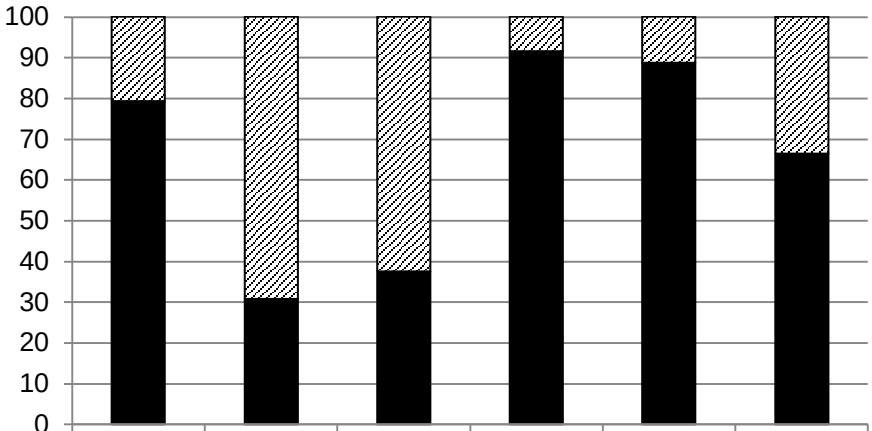
A



B



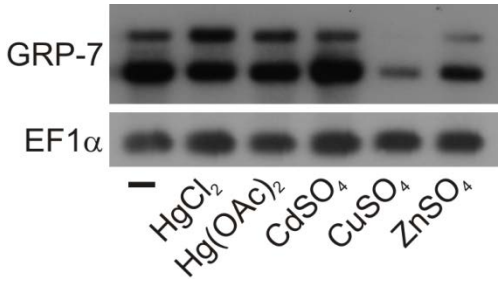
C



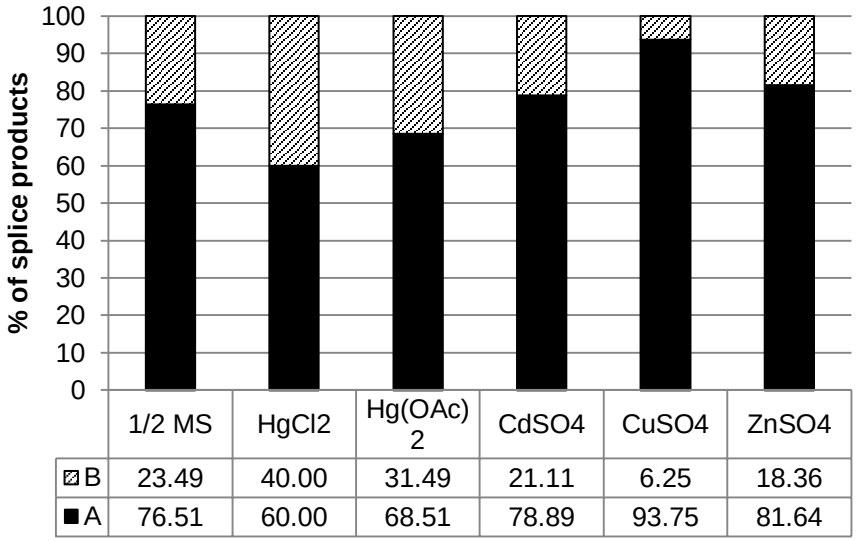
	1/2 MS	HgCl2	Hg(OAc)2	CdSO4	CuSO4	ZnSO4
▨B	20.69	69.26	62.46	8.36	11.29	33.58
■A	79.31	30.74	37.54	91.64	88.71	66.42

Figure 4.1 MAGO expression with metal treatments. (A) Southern blots with ^{32}P-labeled cDNA probes. Upper band is pre-mRNA, lower band is mature mRNA. **(B)** Levels of MAGO transcript. Expression is normalized to EF1α and relative to untreated plants. **(C)** Percent of mature mRNA levels (A) and pre- mRNA levels (B).

A



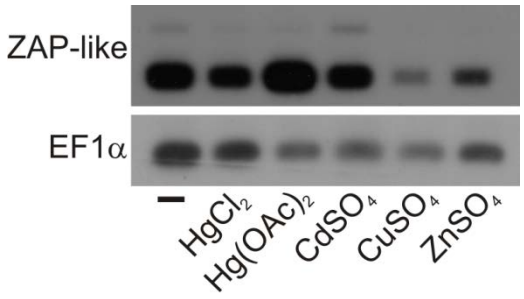
B



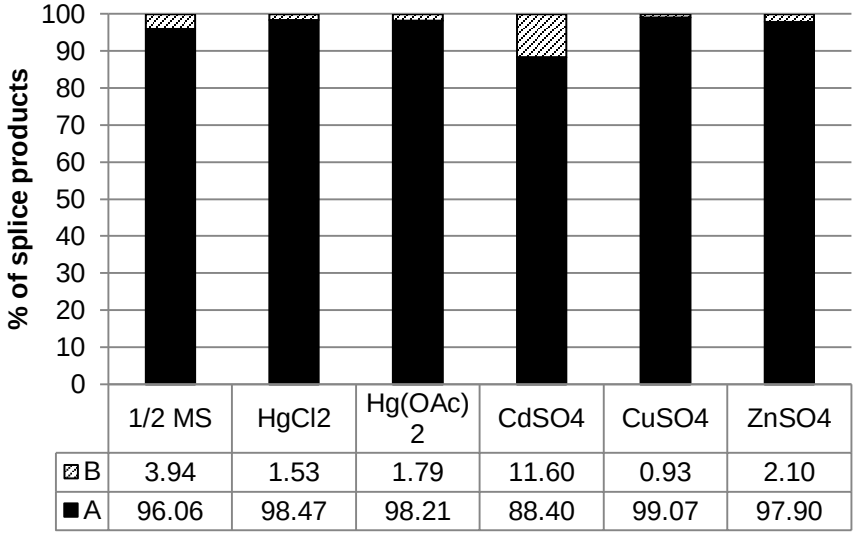
	1/2 MS	HgCl2	Hg(OAc)2	CdSO4	CuSO4	ZnSO4
B	23.49	40.00	31.49	21.11	6.25	18.36
A	76.51	60.00	68.51	78.89	93.75	81.64

Figure 4.2 GRP7 expression with metal treatments. (**A**) Southern blots with ^{32}P-labeled cDNA probes. Upper band is pre-mRNA, lower band is mature mRNA. (**B**) Percent of mature mRNA levels (A) and pre- mRNA levels (B).

A



B



	1/2 MS	HgCl2	Hg(OAc)2	CdSO4	CuSO4	ZnSO4
▨B	3.94	1.53	1.79	11.60	0.93	2.10
■A	96.06	98.47	98.21	88.40	99.07	97.90

Figure 4.3 ZAP-like expression with metal treatment. (**A**) Southern blots with ^{32}P-labeled cDNA probes. Upper band is pre-mRNA, lower band is mature mRNA. (**B**) Percent of mature mRNA levels (A) and pre- mRNA levels (B).

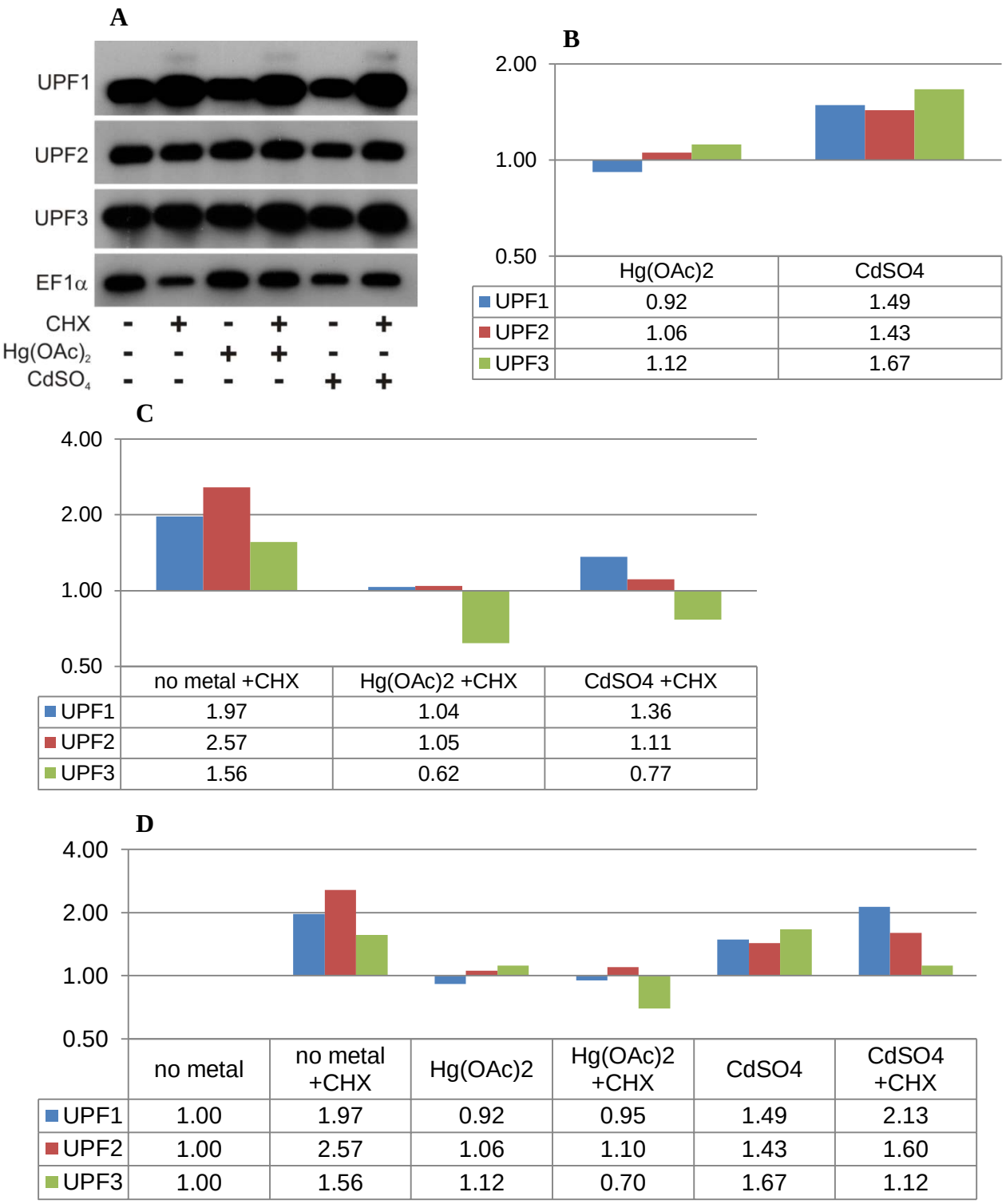


Figure 4.4 UPF1, UPF2 and UPF3 expression with metal +/- CHX treatment. (**A**) Southern blots with ^{32}P-labeled cDNA probes. (**B**) Transcript levels of UPF genes under metal treatments relative to no treatment. (**C**) Transcript levels of UPF genes under with CHX treatment relative to no treatment. (**D**) Transcript levels of UPF genes relative to no treatment.

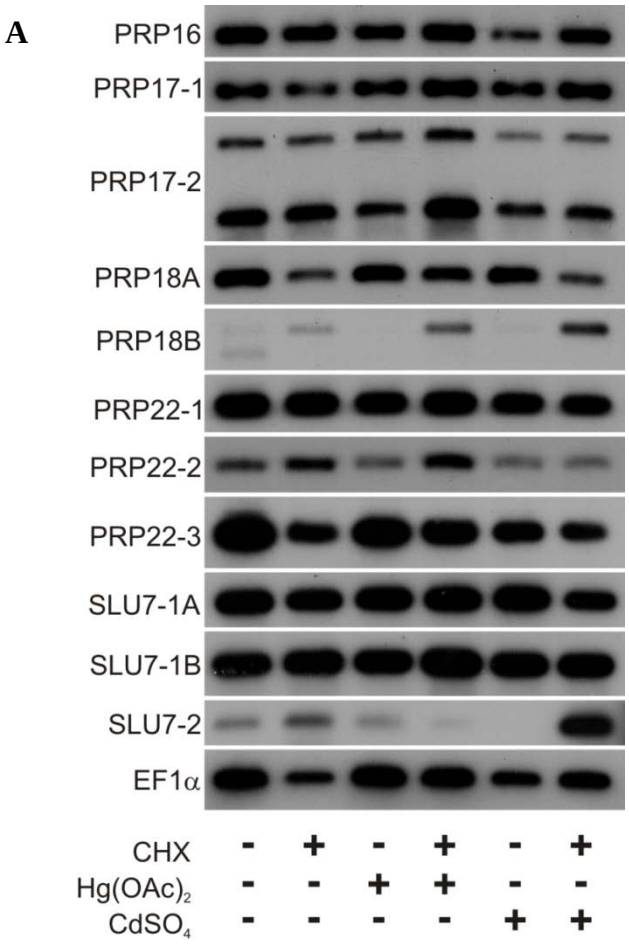
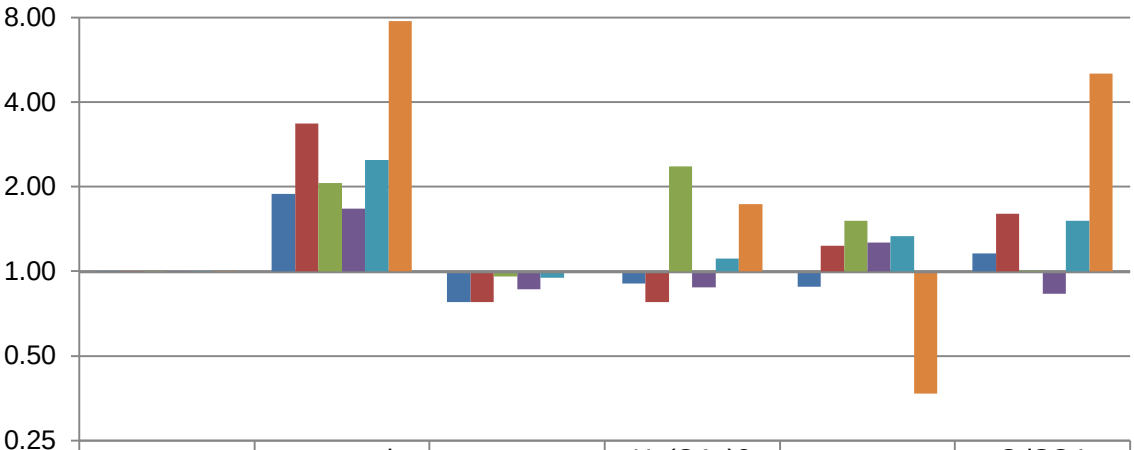


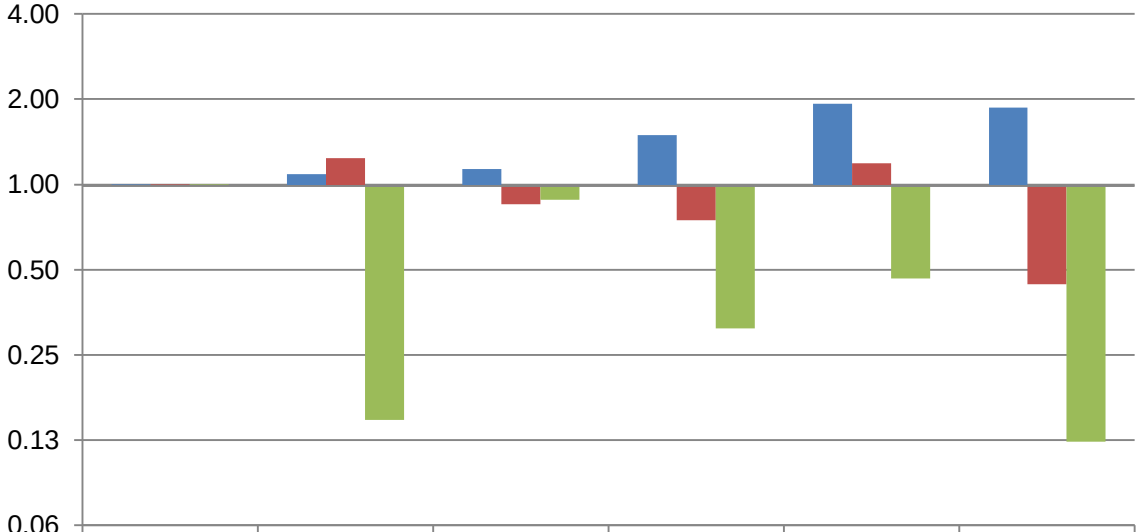
Figure 4.5 Expression levels of second-step splicing factors with metal +/- CHX treatment. (**A**) Southern blots with ^{32}P-labeled cDNA probes. (**B**) Transcript levels of second-step splicing factors that accumulate with knockout of NMD relative to no treatment. (**C**) Transcript levels of second-step splicing factors that are not affected by the knockout of NMD relative to no treatment. (**D**) Percent pre-mRNA accumulation of PRP17-2. (**E**) Percent pre-mRNA accumulation of PRP18B. (**F**) Transcript accumulation of second-step splicing factors relative to no treatment.

B



	no metal	no metal +CHX	Hg(OAc)2	Hg(OAc)2 +CHX	CdSO4	CdSO4 +CHX
PRP16	1.00	1.88	0.78	0.91	0.88	1.16
PRP22-1	1.00	3.35	0.78	0.78	1.23	1.60
PRP22-2	1.00	2.06	0.96	2.36	1.51	1.00
SLU7-1A	1.00	1.67	0.86	0.88	1.27	0.83
SLU7-1B	1.00	2.50	0.95	1.11	1.34	1.51
SLU7-2	1.00	7.78	1.00	1.73	0.37	5.05

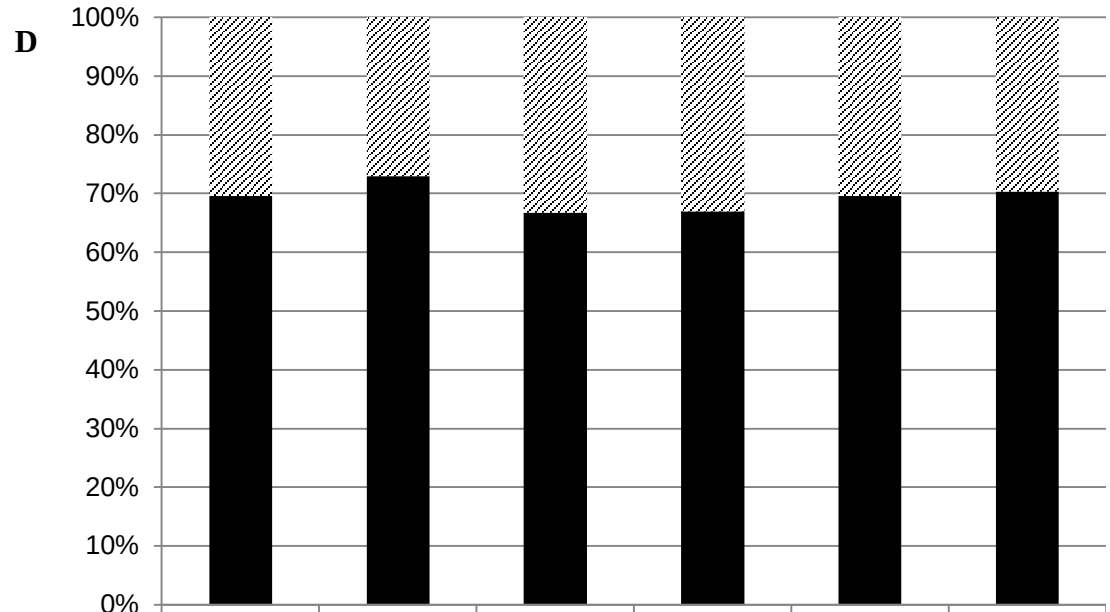
C



	no metal	no metal +CHX	Hg(OAc)2	Hg(OAc)2 +CHX	CdSO4	CdSO4 +CHX
PRP17-1	1.00	1.09	1.13	1.50	1.93	1.87
PRP18A	1.00	1.24	0.85	0.75	1.19	0.45
PRP22-3	1.00	0.15	0.88	0.31	0.47	0.12

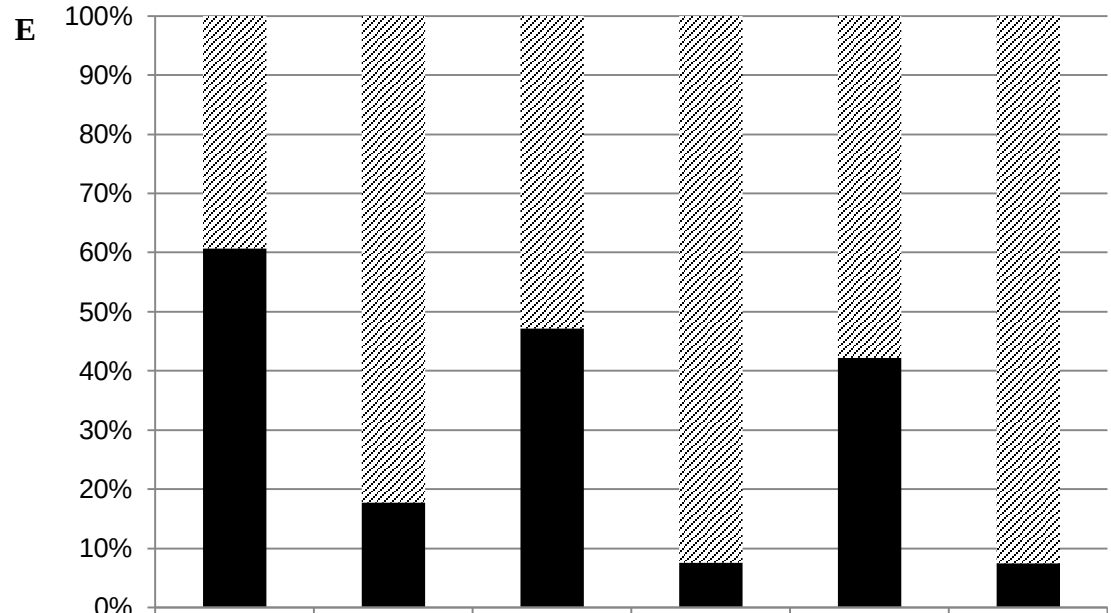
Figure 4.5 (cont.)

D



	no metal	no metal +CHX	Hg(OAc)2	Hg(OAc)2 +CHX	CdSO4	CdSO4 +CHX
PRP17-2 precursor	30.45	27.07	33.27	33.04	30.50	29.73
PRP17-2 mature	69.55	72.93	66.74	66.97	69.50	70.27

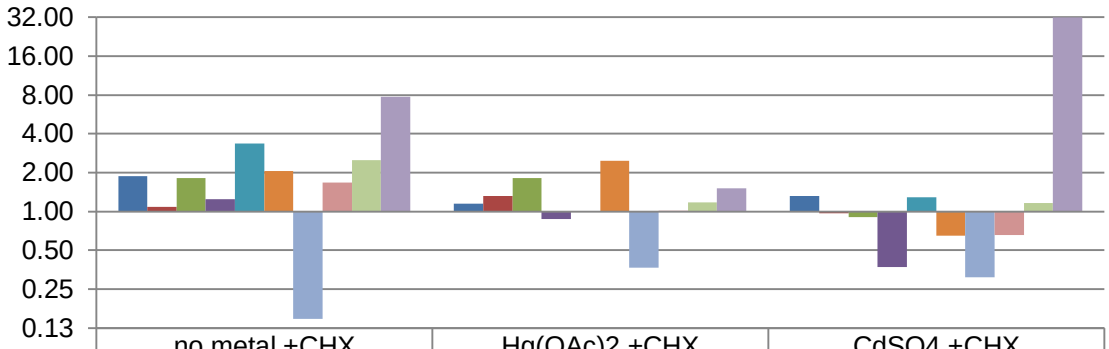
E



	no metal	no metal +CHX	Hg(OAc)2	Hg(OAc)2 +CHX	CdSO4	CdSO4 +CHX
PRP18B precursor	39.22	82.16	52.78	92.39	57.72	92.47
PRP18B mature	60.78	17.84	47.22	7.61	42.28	7.53

Figure 4.5 (cont.)

F



	no metal +CHX	Hg(OAc)2 +CHX	CdSO4 +CHX
PRP16	1.88	1.15	1.31
PRP17-1	1.09	1.32	0.97
PRP17-2	1.82	1.81	0.91
PRP18A	1.24	0.88	0.37
PRP22-1	3.35	1.00	1.29
PRP22-2	2.06	2.47	0.65
PRP22-3	0.15	0.36	0.31
SLU7-1A	1.67	1.00	0.66
SLU7-1B	2.50	1.17	1.16
SLU7-2	7.78	1.52	31.83

Figure 4.5 (cont.)

4.5 References

Ali, G. S. and Reddy, A. S. (2008). Regulation of alternative splicing of pre-mRNAs by stresses. *Current Topics Microbiology and Immunology* **326**, 257-275.

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

Bidou, L., Hatin, I., Perez, N., Allamand, V., Panthier, J. J. and Rousset, J. P. (2004). Premature stop codons involved in muscular dystrophies show a broad spectrum of readthrough efficiencies in response to gentamicin treatment. *Gene Therapy* **11**, 619-627.

Chaouch, S., Mouly, V., Goyenvalle, A., Vulin, A., Mamchaoui, K., Negroni, E., Di Santo, J., Butler-Browne, G., Torrente, Y., Garcia, L. et al. (2009). Immortalized skin fibroblasts expressing conditional MyoD as a renewable and reliable source of converted human muscle cells to assess therapeutic strategies for muscular dystrophies: validation of an exon-skipping approach to restore dystrophin in Duchenne muscular dystrophy cells. *Human Gene Therapy* **20**, 784-790.

Cui, Y., Hagan, K. W., Zhang, S. and Peltz, S. W. (1995). Identification and characterization of genes that are required for the accelerated degradation of mRNAs containing a premature translational termination codon. *Genes and Development* **9**, 423-436.

Doma, M. K. and Parker, R. (2006). Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature* **440**, 561-564.

Gilat, R. and Shweiki, D. (2007). A novel function for alternative polyadenylation as a rescue pathway from NMD surveillance. *Biochemical Biophysical Research Communications* **353**, 487-492.

Gonzalez-Hilarion, S., Beghyn, T., Jia, J., Debreuck, N., Berte, G., Mamchaoui, K., Mouly, V., Gruenert, D. C., Deprez, B. and Lejeune, F. (2012). Rescue of nonsense mutations by amlexanox in human cells. *Orphanet Journal Rare Diseases* **7**, 58-72.

Goyenvalle, A., Babbs, A., van Ommen, G. J., Garcia, L. and Davies, K. E. (2009). Enhanced exon-skipping induced by U7 snRNA carrying a splicing silencer sequence: Promising tool for DMD therapy. *Molecular Therapy* **17**, 1234-1240.

Harigaya, Y. and Parker, R. (2010). No-go decay: a quality control mechanism for RNA in translation. *Wiley Interdisciplinary Reviews: RNA* **1**, 132-141.

He, F. and Jacobson, A. (1995). Identification of a novel component of the nonsense-mediated mRNA decay pathway by use of an interacting protein screen. *Genes and Development* **9**, 437-454.

Heintzen, C., Nater, M., Apel, K. and Staiger, D. (1997). AtGRP7, a nuclear RNA-binding protein as a component of a circadian-regulated negative feedback loop in *Arabidopsis thaliana*. *Proceedings Naional Academy of Sciences USA* **94**, 8515-8520.

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

Hori, K. and Watanabe, Y. (2007). Context analysis of termination codons in mRNA that are recognized by plant NMD. *Plant Cell Physiology* **48**, 1072-1078.

Huang, L. and Wilkinson, M. F. (2012). Regulation of nonsense-mediated mRNA decay. *Wiley Interdisciplinary Reviews: RNA* **3**, 807-828.

Kalyna, M., Simpson, C. G., Syed, N. H., Lewandowska, D., Marquez, Y., Kusenda, B., Marshall, J., Fuller, J., Cardle, L., McNicol, J. et al. (2012). Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in *Arabidopsis*. *Nucleic Acids Research* **40**, 2454-2469.

Kerenyi, Z., Merai, Z., Hiripi, L., Benkovics, A., Gyula, P., Lacomme, C., Barta, E., Nagy, F. and Silhavy, D. (2008). Inter-kingdom conservation of mechanism of nonsense-mediated mRNA decay. *EMBO Journal* **27**, 1585-1595.

Kertesz, S., Kerenyi, Z., Merai, Z., Bartos, I., Palfy, T., Barta, E. and 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**, 6147-6157.

Kim, S. H., Koroleva, O. A., Lewandowska, D., Pendle, A. F., Clark, G. P., Simpson, C. G., Shaw, P. J. and Brown, J. W. (2009). Aberrant mRNA transcripts and the nonsense-mediated decay proteins UPF2 and UPF3 are enriched in the *Arabidopsis* nucleolus. *Plant Cell* **21**, 2045-2057.

Klauer, A. A. and van Hoof, A. (2012). Degradation of mRNAs that lack a stop codon: a decade of nonstop progress. *Wiley Interdisciplinary Reviews: RNA* **3**, 649-660.

Kurihara, Y., Matsui, A., Hanada, K., Kawashima, M., Ishida, J., Morosawa, T., Tanaka, M., Kaminuma, E., Mochizuki, Y., Matsushima, A. et al. (2009). Genome-wide suppression of aberrant mRNA-like noncoding RNAs by NMD in *Arabidopsis*. *Proceedings National Academy of Sciences USA* **106**, 2453-2458.

Le Hir, H., Gatfield, D., Izaurralde, E. and Moore, M. J. (2001). The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO Journal* **20**, 4987-4997.

Linde, L., Boelz, S., Nissim-Rafinia, M., Oren, Y. S., Wilschanski, M., Yaacov, Y., Virgilis, D., Neu-Yilik, G., Kulozik, A. E., Kerem, E. et al. (2007). Nonsense-mediated mRNA decay affects nonsense transcript levels and governs response of cystic fibrosis patients to gentamicin. *Journal Clinical Investigation* **117**, 683-692.

Narsai, R., Howell, K. A., Millar, A. H., O'Toole, N., Small, I. and Whelan, J. (2007) Genome-wide analysis of mRNA decay rates and their determinants in *Arabidopsis thaliana*. *Plant Cell* **19**, 3418-3436.

Nyiko, T., Sonkoly, B., Merai, Z., Benkovics, A. H. and Silhavy, D. (2009). Plant upstream ORFs can trigger nonsense-mediated mRNA decay in a size-dependent manner. *Plant Molecular Biology* **71**, 367-378.

Schneider-Poetsch, T., Ju, J., Eyler, D. E., Dang, Y., Bhat, S., Merrick, W. C., Green, R., Shen, B. and Liu, J. O. (2010). Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. *Nature Chemical Biology* **6**, 209-217.

Schöning, J. C., Streitner, C., Meyer, I. M., Gao, Y. and Staiger, D. (2008). Reciprocal regulation of glycine-rich RNA-binding proteins via an interlocked feedback loop coupling alternative splicing to nonsense-mediated decay in *Arabidopsis*. *Nucleic Acids Research* **36**, 6977-6987.

Schwartz, A. M., Komarova, T. V., Skulachev, M. V., Zvereva, A. S., Dorokhov Iu, L. and Atabekov, J. G. (2006). Stability of plant mRNAs depends on the length of the 3'-untranslated region. *Biochemistry (Moscow)* **71**, 1377-1384.

Scofield, D. G., Hong, X. and Lynch, M. (2007). Position of the final intron in full-length transcripts: determined by NMD? *Molecular Biology Evolution* **24**, 896-899.

van Hoof, A., Frischmeyer, P. A., Dietz, H. C. and Parker, R. (2002). Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science* **295**, 2262-2264.

Welch, E. M., Barton, E. R., Zhuo, J., Tomizawa, Y., Friesen, W. J., Trifillis, P., Paushkin, S., Patel, M., Trotta, C. R., Hwang, S. et al. (2007). PTC124 targets genetic disorders caused by nonsense mutations. *Nature* **447**, 87-91.

Wu, X., Liu, M., Downie, B., Liang, C., Ji, G., Li, Q. Q. and Hunt, A. G. (2011). Genome-wide landscape of polyadenylation in *Arabidopsis* provides evidence for extensive alternative polyadenylation. *Proceedings National Academy of Sciences USA* **108**, 12533-12538.

Xu, J., Yang, J. Y., Niu, Q. W. and Chua, N. H. (2006). *Arabidopsis* DCP2, DCP1, and VARICOSE form a decapping complex required for postembryonic development. *Plant Cell* **18**, 3386-3398.

Yoine, M., Ohto, M. A., Onai, K., Mita, S. and Nakamura, K. (2006). The *lba1* mutation of UPF1 RNA helicase involved in nonsense-mediated mRNA decay causes pleiotropic phenotypic changes and altered sugar signalling in *Arabidopsis*. *Plant Journal* **47**, 49-62.