THESIS

# ALTERNATIVE SPLICING AND ITS REGULATORY MECHANISMS IN PHOTOSYNTHETIC EUKARYOTES

Submitted by

Alicia Link

Department of Biology

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2011

Master's Committee:

Advisor: A. S. N. Reddy

Stephen Stack
Nora Lapitan

## ABSTRACT

### ALTERNATIVE SPLICING AND ITS REGULATORY MECHANISMS IN PHOTOSYNTHETIC EUKARYOTES

In recent years, alternative splicing (AS) of pre-mRNAs, which generates multiple transcripts from a single gene, has emerged as an important process in general proteome diversity and in regulatory gene expression in multicellular eukaryotes. In Arabidopsis over 40% of intron-containing genes are alternatively spliced. However, mechanisms by which AS is regulated in plants are not fully understood, primarily due to the lack of an *in vitro* splicing system derived from plants. Furthermore, the extent of AS in simple unicellular photosynthetic eukaryotes from which plants have evolved is also not known. My research addresses these two attributes of splicing in plants.

In Part 1 of my thesis, I have investigated an aspect of AS regulation in plants. We have previously shown that an SR-related splicing regulator called SR45 regulates AS of pre-mRNAs in Arabidopsis by altering splice site selection (Ali et al. 2007). In this work using bimolecular fluorescent complements, I have demonstrated that SR45 interacts with U2AF$^{35}$, an important spliceosomal protein involved in 3' splice site selection in plant cells. This interaction takes place in the nucleus, specifically in the subnuclear domains called speckles,

which are known to contain splicing regulators and other proteins involved in transcription. My work has shown that SR45 interacts with both paralogs of U2AF$^{35}$ and I mapped the domains in SR45 that are involved in its interaction with U2AF$^{35}$. In addition, my studies have revealed interaction of the paralogs as hetero- and homodimers. Interestingly, U2AF$^{35}$ was found to interact with U1-70K, a key protein involved in 5' splice site selection. Based on this work and previous work in our laboratory, a model is proposed that explains the role of SR45 in splice site selection.

In the second part of my work I studied the extent of alternative splicing (AS) in the unicellular green alga *Chlamydomonas*, that shares a common ancestor with land plants. In collaboration with Dr. Asa Ben Hur's lab, we have performed a comprehensive analysis of AS in *Chlamydomonas reinhardtii* using both computational and experimental methods. Our results show that AS is common in *Chlamydomonas*, but its extent is less than what is observed in land plants. However, the relative frequency of different splicing events in *Chlamydomonas* is very similar to higher plants. We have found that a large number of genes undergo alternative splicing, and together with the simplicity of the system and the use of available molecular and genetic tools. This organism is an experimental system to investigate the mechanisms involved in alternative splicing. To further validate predicted splice variants, we performed extensive analysis of AS for two genes, which not only confirmed predictions but also revealed novel splice variants, suggesting that the extent of AS is higher than we predicted.

AS can also play a role in the regulation of gene expression through processes such as regulated unproductive splicing and translation (RUST) that involves nonsense-mediated decay (NMD), a mechanism of mRNA surveillance that degrades transcripts containing premature termination codons (PTCs). The basic mechanism of NMD relies upon many factors, but there are three critical proteins, termed the UP-frameshift (UPF) proteins due to their ability to up-regulate suppression of nonsense transcripts. UPF1, UPF2, and UPF3 appear to be conserved across animals and plants. Our analysis of AS has found that in Chlamydomonas, many splice variants have a premature termination codon (PTC). However, to date, the mechanism of NMD has not been investigated in Chlamydomonas. Analysis of the Chlamydomonas genome sequence shows that UPF1, 2, and 3 proteins are present, and we have shown that they share some sequence similarity with both plants and humans, indicating that the process of NMD may be present in this organism. To address the role of UPFs in NMD in Chlamydomonas, we have utilized the artificial miRNA approach. I have generated stably transformed Chlamydomonas cell lines that are expressing amiRNA for UPF1 and UPF3 that will be useful in analyzing NMD of selected genes as well as all PTC-containing transcripts globally.

ACKNOWLEDGEMENTS

Completing this thesis has been one of the most difficult things in my life I've ever done, but also one that I am most proud of. First of all, I'd like to thank Dr. Reddy for his outstanding role as my advisor and seemingly perpetual happiness throughout my education in graduate school at CSU. He took me under his wing when continuing school seemed futile, and gave me the chance I needed to accomplish my dream. For that, I am forever indebted to him. Second of all I'd like to thank the Reddy lab for their help and guidance. In particular, I'd like to thank Dr. Irene Day and Julie Thomas, whose support and amazing fortitude carried me throughout my degree. I would also like to thank Dr. Salah Abdel-Ghany, who showed up at the end of my term, but was eager to help and truly meets the definition of an "upright man." I'd also like to thank Dr. Asa Ben Hur and his students for their collaboration on the *Chlamydomonas* AS project. I'd like to thank my committee not only for being there for me and giving me support, but for being the inspiring people they are. I chose them because I looked up to them. I'd also like to thank Dr. June Medford for her support during the tough times.

Most importantly, I'm sending thanks to Dr. Paul Kugrens, who passed on during my time at CSU, but left an everlasting impression on me and will always be in my heart. His passion for science was contagious, and if I hadn't met him, I never would have wondered what "pond scum" was. Science really hasn't been the same for me without him, but I believe he provided for me the tools for success in life. I'd

also like to thank his family for their continuing support and for being a great part of my life. I know I wouldn't have succeeded without them.

Next, I thank my parents. I couldn't have asked for a better family, and I wake up each day grateful to have them in my life. Without my mother, I never would have had the perseverance to finish, and without my father, I wouldn't have the confidence and "guts" to succeed. I'd also like to thank my brother for his support. My grandparents also have been extremely supportive of my education, and when I grow to be their age, I hope to have lived as rich of a life as they have.

I never would have made it without my fiancé, Paul. He has been my lighthouse in the middle of the storm, my constant, and my true partner. He continues to inspire me everyday with his innovative thinking, joyful presence, and determination. Meeting him has changed my life, and I know if he is with me, I can face anything and succeed. I'd also like to thank my best friend Stephanie Long for her friendship and early morning calls. She kept me sane through this experience and has through many others. She is a wonderful friend and I'm lucky to have her. Last but not least, I'd like to thank Tallulah the Cat and Charlie the Dog. Even though they can't talk I know they love me.

I read one of Dr. Kugrens' student's thesis once and was greatly moved by the quote in their acknowledgements section, so I shall end mine in the same way with this quote by Antony van Leeuwenhoek: "*This was for me, among all the marvels that I have discovered in nature, the most marvelous of all…no more pleasant sight has come before my eyes than these many thousands of living creatures, seen all alive in a drop of water.*"

TABLE OF CONTENTS

# CHAPTER 1: *IN VIVO* INTERACTION OF A SPLICING REGULATOR (SR45) WITH U2 AUXILLARY FACTORS ($U2AF^{35}$)

(The results presented here are from a manuscript titled "Interactions of SR45 with $U2AF^{35}$ and U1-70K: Insight into the spliceosome assembly on 5' and 3' splice sites by Day, Golovkin, Link, Ali, and Reddy" which has been submitted for publication)

## Introduction

Precursor mRNA (pre-mRNA) in eukaryotes contains coding sequences known as exons, and intervening noncoding sequences known as introns. Pre-mRNA splicing is essential for the expression of most genes in eukaryotic organisms. This process involves the removal of introns and the covalent joining of exons in pre-mRNA. Four conserved core elements, which include 5' and 3' splice sites, a branch point, and a polypyrimidine tract upstream of the 3' splice site that are necessary for splicing. Multicellular eukaryotes are known to have less conserved sequences around the splice sites. As a result, additional regulatory sequences adjacent to the splice sites, which are called enhancers or repressors, are needed for correct and efficient definition of splice sites (Ali et al. 2007; Ast 2004; Reddy 2007).

Pre-mRNA splicing takes place in a large complex known as the spliceosome, which is made up of 5 small nuclear ribonucleoparticles (U1,2,4,5,6 SnRNPs) and many non-snRNP factors(Reddy 2007; Reddy 2001; Sharp 1994; Zhou et al. 2002). A fully assembled spliceosome contains around 300 different proteins and is one of the most complex cellular components investigated so far (Rappsilber et al. 2002; Zhou et al. 2002). The first step in spliceosomal assembly is the formation of the ATP-independent

early (E) complex, and this is where the 5' splice site is recognized and subsequently bound by U1 snRNP. The branch site and 3' splice site are then recognized by U2 snRNP and U2 auxiliary factors (U2AFs). Following this step, the A complex (the association of U2 with the branch site/3' splice site region) is altered in an ATP-dependent fashion in order to stabilize binding to this region (Figure 1.1). Then, the U4/U6/U5 tri-snRNP joins the assembly of the spliceosome to form the B-complex. This complex then undergoes a series of conformational rearrangements, one of them involving a tri-snRNP associated protein kinase (SRPK2), which phosphorylates the RS domain of Prp28, an RNA helicase. This step is required for the stable association of Prp28 with the tri-snRNP as well as for tri-snRNP association with assembling spliceosomes during B complex formation (Mathew et al. 2008). After the dissociation of U1 snRNP, the U4/U6 base-paired interaction is unwound by Brr2, which is a U5-assocated helicase (Wahl et al. 2009; Will 2006). Then, U1 and U4 leave the spliceosome, and U6 replaces U1 at the 5' splice site (Figure 1.1). This then marks the progress of the B complex moving towards being the catalytically active B* complex, which performs the first step of splicing. However before the B* complex is formed however, the Bact complex is the transitional state between the B and B* complex. The Bact complex is void of U1, U4/U6 snRNPs, but contains the NineTeen Complex proteins (NTC) (Fabrizio et al. 2009). The Bact complex is then transitioned to B* following ATP hydrolysis by RNA helicase Prp2, which dissociates and gives the final B* complex. A protein called Prp16 transitions the B* complex to the C complex, which contains U5, U6 and U2 and is capable of completing the second step of splicing (Valadkhan and Jaladat 2010) (Figure 1.1).

In contrast to constitutive splicing, a process called regulated alternative splicing (AS) produces multiple transcripts that encode different proteins from the same gene,

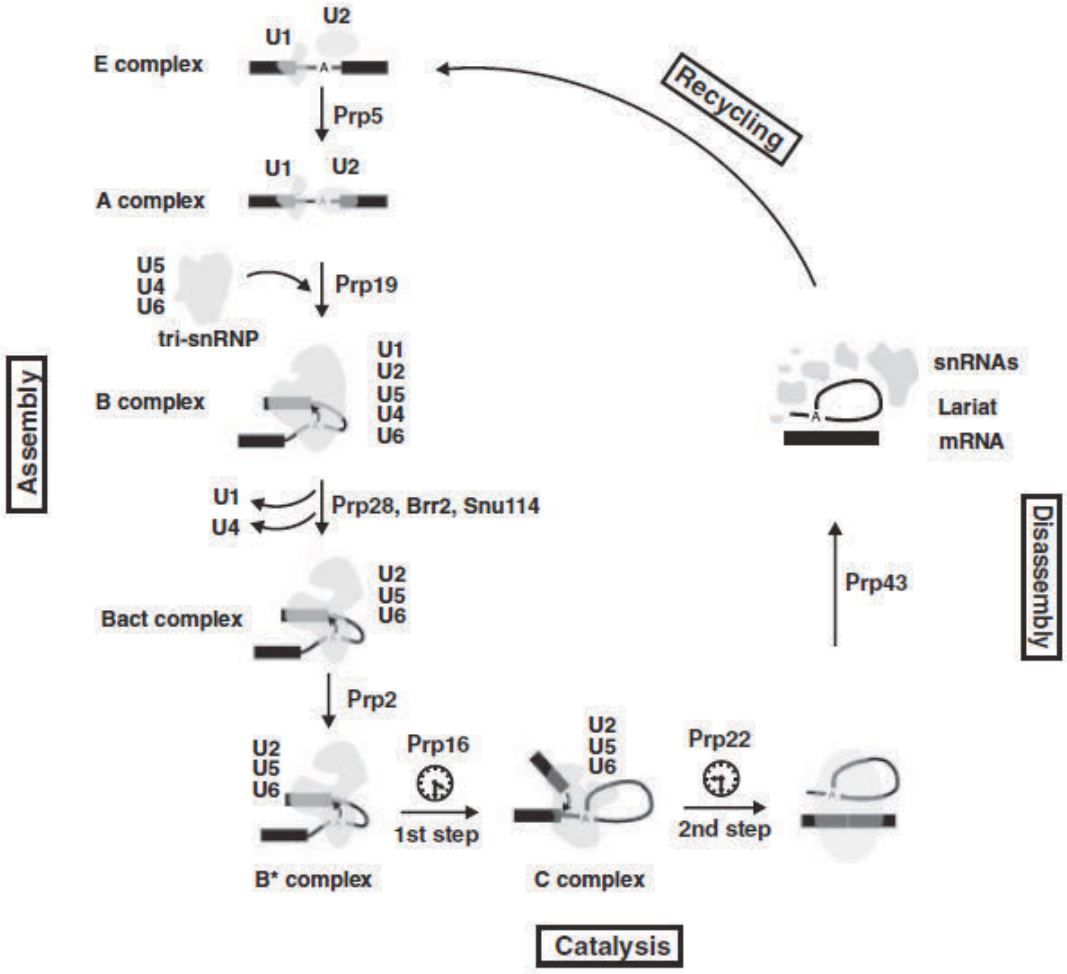



**Figure 1.1. Spliceosomal Assembly** (Valadkhan and Jaladat 2010).

which then aids in post-transcriptional gene regulation and generates proteomic diversity. Both of these processes are essential for the function of many genes in eukaryotes (Ast 2004; Graveley 2001; Maniatis and Tasic 2002; Reddy 2007; Reddy 2001). Recent human genome-wide studies have suggested that 95% of pre-mRNAs from intron-containing genes are alternatively spliced (Pan et al. 2008; Wang et al. 2008). In addition to this, genome-wide studies in Arabidopsis have indicated that over 40% of intron-containing genes undergo AS (Filichkin et al. 2010 ).

The process of AS lengthens or shortens exons by altering the position of one of their splice sites, which results in both alternative 3' and 5' splice sites. AS also involves exon skipping or intron retention, where important regulatory events can be controlled by the failure to recognize an exon or excise an intron. It has been shown that abiotic/biotic stresses effect the pattern of splicing, showing that AS is regulated and the pattern of splicing is adjusted according to what kind and level of stress a plant experiences (Iida et al. 2004; Palusa et al. 2007). It has also been shown that AS has a critical role in flowering time, where temperature changes regulate this action, and it has been shown that varying AS transcripts are part of making this process proceed smoothly (Balasubramanian et al. 2006). In addition, wound-response is affected by patterns of AS, and a separate study linked AS to disease resistance in plants (Bove et al. 2008; Dinesh-Kumar and Baker 2000). Thus, it seems that AS in plants is a posttranscriptional regulatory mechanism, which in turn eventually effects gene expression and subsequently plant form and function (Reddy 2007).

In metazoans, the key players for the recruitment of U1 to the 5' splice site and U2 to the 3' splice site are members of what is known as the serine/arginine-rich (SR) protein family (Nilsen 2003; Reddy 2001). SR proteins are currently characterized in the following ways: they must contain one or two N-Terminal RNA recognition motifs

(RRMs) followed by a downstream serine/arginine rich region known as the RS domain of at least 50 amino acids and a minimum of 20% RS or SR dipeptides (Barta et al. 2010). These proteins aid in identifying the 5' splice site by interacting with U1 and pre-mRNA concurrently (Graveley 2000; Reddy 2004). In addition to this, SR proteins are thought to regulate the selection of weak alternative splice sites. SR proteins bind to these less conserved sequences and therefore promote the recruitment of U1 to the correct 5' splice site (Eperon et al. 1993; Graveley 2000; Jamison et al. 1995; Kohtz et al. 1994; Reddy 2004; Zahler and Roth 1995). All of these functions combined lend the SR protein family the reputation of being a major player for increasing transcriptome complexity as well as proteomic diversity (Graveley 2000). Cell-free extracts have been utilized to study splicing in animals, and it has been found that there are RNA-RNA, RNA-protein, and protein-protein interactions involved in the splicing process. Serine/arginine (SR) proteins are responsible for many different roles in mRNA splicing (Barta et al. 2008; Graveley 2000; Lorkovic et al. 2008). About 18 SR proteins have been identified, while only 12 are present in humans (Barta et al. 2010). SR proteins have been found to be localized in interchromatin granule clusters in the nucleus as well as in the nucleoplasm (Ali et al. 2008a; Ali and Reddy 2008a; Fang et al. 2004; Lorkovic et al. 2008).

Plant introns differ from animal introns in their size, nucleotide composition, branch point sequence and polypyrimidine tract (Reddy 2001). Because of this, plant intron-containing transcripts are usually either not processed correctly or not processed at all in mammalian splicing extracts (McCullough et al. 1991). There is no plant-derived cell-free splicing extract available for plant splicing factors analysis, and because of this, *in vivo* methods have to be used to investigate the interaction network of proteins involved in splicing.

SR45 is an SR-like protein with two RS domains, one being N-terminal and the other C-terminal. A second SR-like protein, known as SR45a, has been identified and shares the same domains as SR45, but only shows a 26% identity to SR45 with most of this being in the RS domains (Ali et al. 2007; Tanabe et al. 2007). SR45 has been shown to be an essential splicing factor in complementation assays, and has orthologs in other flowering plants, but none in algae. SR45 interacts with U1-70K, one of the U1 snRNP specific proteins, and their association in nuclear speckles was shown using bimolecular fluorescence complementation (BiFC) (Ali et al. 2007; Ali et al. 2008a). BiFC is a technique where two halves of a fluorescent protein are separately fused to two putative interaction proteins. When the two proteins interact, the two halves of the protein come together resulting in reconstitution of fluorescent protein. The fluorescent signal can then be detected through confocal microscopy (Hu et al. 2002). SR45 is alternatively spliced and produces two splice forms. In the SR45 mutant, *sr45-1*, the splicing patterns of many other SR genes are affected, suggesting its role in AS (Ali et al. 2007). The mutant phenotype indicates that SR45 has a role in multiple plant-specific developmental processes, including plant size, flowering time, and organ morphology. In addition to this, *sr45-1* plants show delayed flowering, altered leaf morphology, and flowers with abnormal petal and stamen numbers, (Ali et al. 2007). One splice form of SR45 rescues the flower phenotype while the other splice form rescues the root phenotype, as shown in a gene complementation study (Mount and Zhang 2009).

To further understand the role of SR45 in splicing, we used it in a yeast two hybrid system, which resulted in isolation of $U2AF^{35}$, a spliceosomal protein. It is part of the U2AF complex, which is involved in spliceosomal assembly. The large subunit of the U2AF complex is $U2AF^{65}$, which binds to the polypyimidine (Py) tract. $U2AF^{35}$ is the small subunit of the U2AF complex, which is involved in the recognition of the 3' splice

site. This recognition occurs through the contact of the AG dinucleotide at the 3' splice site. For introns with weak Py tracts, the U2AF$^{35}$ interaction with the 3' splice site is critical for U2AF (Merendino et al. 1999; Wu et al. 1999; Zorio and Blumenthal 1999). In Arabidopsis, two paralogs of U2AF$^{35}$ have been characterized, known as U2AF$^{35a}$ and U2AF$^{35b}$. I investigated *in vivo* interactions of U2AF$^{35}$ with SR45 using BiFC. Here, I have shown that SR45 interacts with both paralogs and mapped the domains in SR45 that are involved in its interaction with U2AF$^{35}$. In addition, my studies have shown interaction of the paralogs as hetero- and homodimers. A protein alignment of the paralogs of U2AF$^{35}$ present in Arabidopsis, humans, and rice was performed which showed that there is a C-terminal domain that is unique to plants (Figure 1.2). We have tested a truncated version of this domain to investigate its role in U2AF$^{35}$ interaction with SR45. My studies have also shown that U2AF$^{35}$ and U1-70K interact. Since SR45 interacts with both U1-70K and U2AF$^{35}$, it is likely that it has a role in 3' to 5' splice site selection and could possibly bridge the 5' and 3' components of the spliceosome.

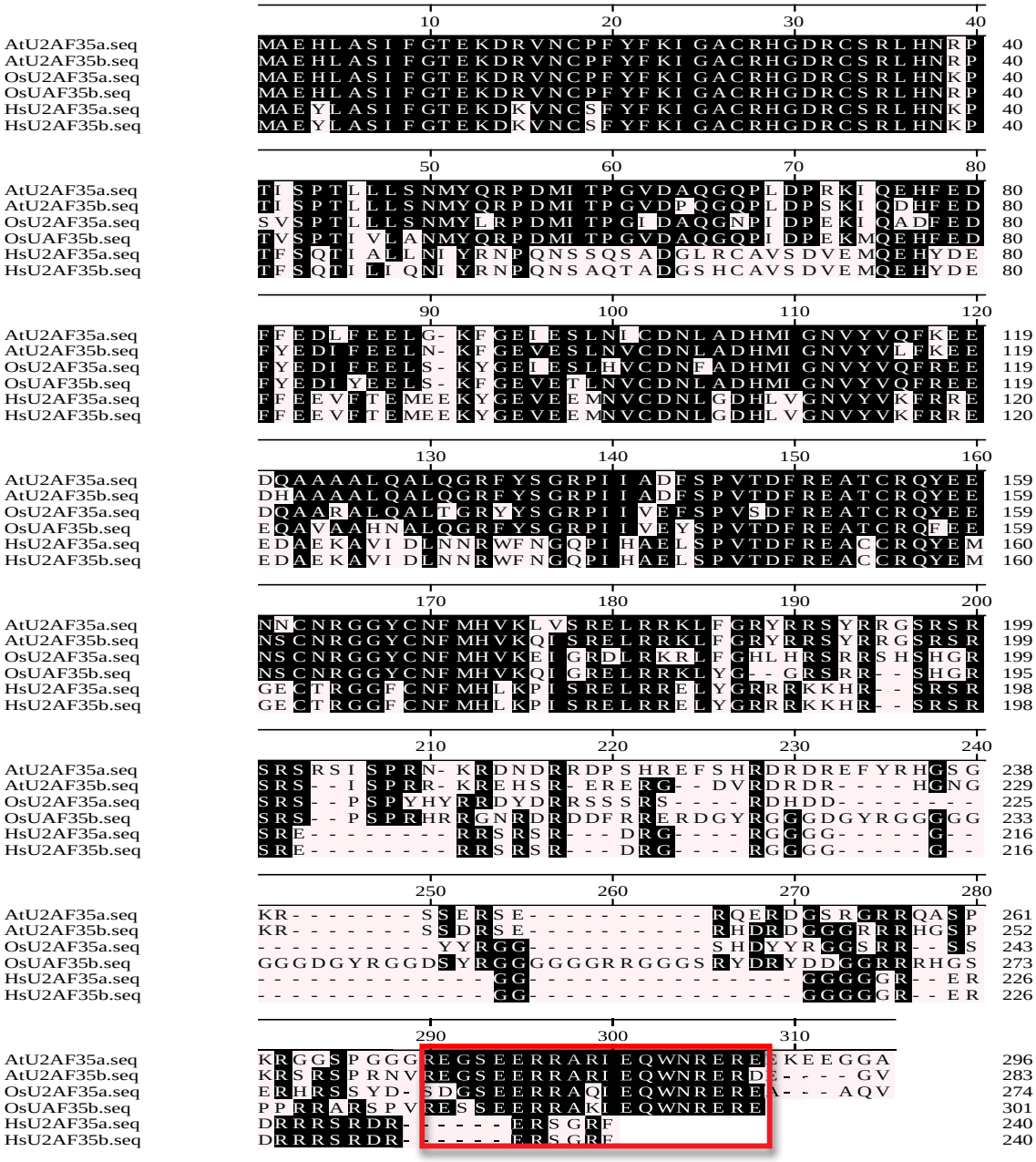



**Figure 1.2. U2AF[35] proteins from Arabidopsis, Rice, and Human are aligned using MegaAlign. Identical amino acids are shown in reverse contrast lettering. Red boxes indicate that the C-terminal region is highly conserved in plants, but does not share the same identity with human U2AFs.**

# MATERIALS AND METHODS

## Constructs

Full-length and truncated mutants of SR45 were amplified with forward and reverse primers containing *Sal*I and *Xma*I sites, respectively and cloned into pSPYNE-35S/pUC-SPYNE and pSPYCE-35S/pUCSPYCE (Ali and Reddy 2008b). Primers listed in Day et al 2011 (in preparation) were used to amplify (U2AF$^{35}$a, c and Ctrb using DNA as a template. PCR was done using Takara Ex Taq (Fisher Scientific) according to manufacturer's specifications (two-step). Fragments were digested with *Sal*I/*Kpn*I and ligated into pSPYNE-35S/pUC-SPYNE and pSPYCE-35S/pUCSPYCE vectors digested with the same enzymes. Full length U1-70K/YFP$^{N}$ was constructed previously (Ali et al. 2008b). Plasmid DNA was prepared using a Qiagen Midi-Prep kit or a Maxi-Prep kit. DNA concentration was quantified using a spectrophotometer.

## Transient Expression of BiFC constructs in protoplasts

### A) Protoplast isolation

Protoplasts were prepared from healthy leaves from 4-week-old WT *Arabidopsis thaliana* ecotype Columbia. Leaves were sliced into thin strips and were immersed in enzyme solution (0.4 M mannitol, 20 mM KCl, 20 mM MES pH 5.7) containing 2.0% cellulase and 0.2% macroenzyme. After vacuum infiltration the mixture was shaken in a 500 ml flask at low speed, RT for 3-4 hours and filtered through 75 µm mesh. The protoplasts were harvested by centrifugation (200xg for 2 minutes), resuspended in W5 medium (154 mM NaCl, 25 mM $CaCl_2$, 5 mM KCl and 2 mM MES pH 5.7) and placed on ice for 30 minutes. Protoplasts were pelleted again and resuspended in MMG (0.4 M mannitol, 15 mM $MgCl_2$, 4 mM Mes pH 5.7) medium at a concentration of $2x10^6$/ml.

**B) Transfection**

Plasmid DNA was added at a concentration of 1 mg/ml and an equal amount of 40% PEG was added, the tubes were inverted several times to mix, followed by incubation at RT for 30 minutes. The protoplasts were washed twice with two volumes of W5 and resuspended in 1 ml of W5 they were dispensed into a 6-well plate and kept in the dark in a 22°C incubator for 16 hours to allow for expression. A 50 µl aliquot was transferred to a glass bottom petri dish with a 30 to 70 mm cover slip and observed immediately under the microscope.

**C) Confocal Microscopy**

Transfected protoplasts were examined using a Zeiss LSM 510 Meta laser scanning confocal microscope. All samples were viewed using the YFP channel (Ali et al. 2008a). Protoplasts were first viewed at 40x to see transformation efficiency, and then single protoplasts were viewed and photographed using the 63x, N.A. 1.4 oil immersion apochromat objective. The YFP filter was set up with excitation at 514nm, 458/514 dichroic, and emission 560-615 BP filter.

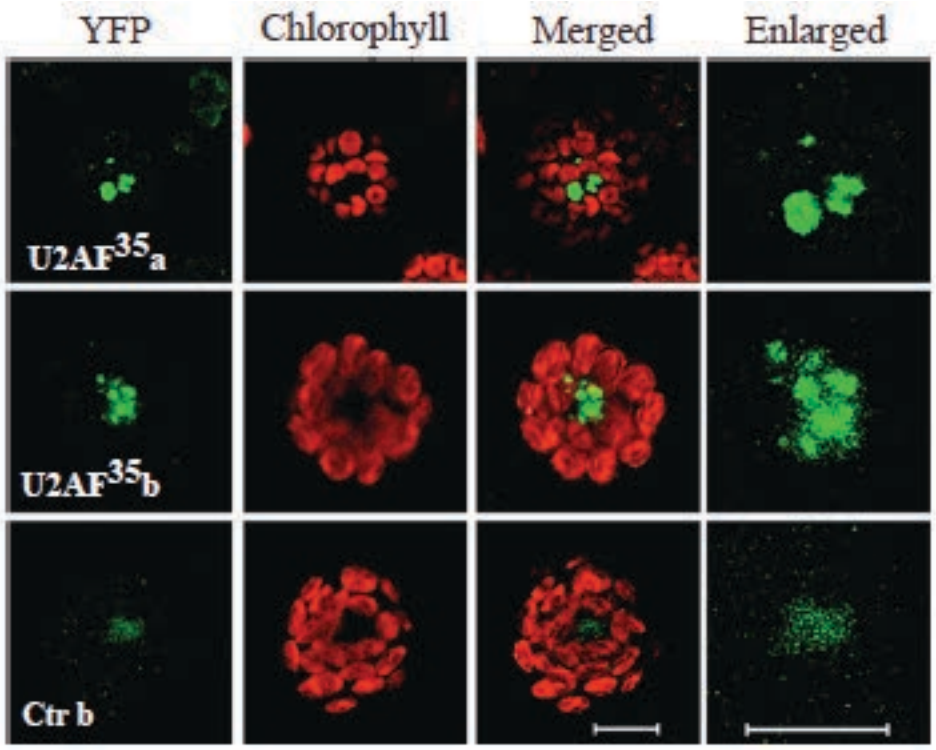



**Figure 1.3. BiFC assay using SR45 and U2AF$^{35}$. Confocal images of Arabidopsis protoplasts expressing full-length SR45-YFPc paired with U2AF$^{35a}$-YPFn, full length U2AF$^{35b}$-YFPn, or U2AF$^{35Ctrb}$-YFPN as indicated on the right. The images in the right-most panels show a zoomed in view of the nuclear region shown in the images under the YFP column. Bars = 10 µm.**

# RESULTS

## *In vivo* interaction of SR45 with both U2AF$^{35}$ paralogs

Previously, SR45 has been shown to co-localize with U1-70K in speckles in the nucleus (Ali et al. 2008a). In addition, both paralogs of U2AF$^{35}$ have also been detected in nuclear speckles (Wang and Brendel 2006b). Recently, we have isolated U2AF$^{35b}$ as an interacting partner of SR45 using yeast two-hybrid screens. The aim of my work is to determine where in the cell they interact and which domains of SR45 are responsible for this interaction. To determine the location of the interaction between SR45 and U2AF$^{35}$, we utilized BiFC as an in vivo proof of association. BiFC is based upon the reconstitution of two split halves of yellow fluorescent protein (YPF), which upon reconstitution, results in fluorescence. Each half of YFP is fused to two putative interacting proteins, and if those two proteins interact, the split halves come together and YFP can be visualized (Walter et al. 2004). Fusions to the N-terminal region of YFP (YFPn) were constructed by cloning U2AF$^{35a}$, U2AF$^{35b}$ (full length), and U2AF$^{35Ctrb}$, which lacks the C-terminal region) into a BiFC vector.

In previous studies, SR45 had been cloned in a similar way as a fusion to the C-terminal region of YFP (SR45/YFPc) into a BiFC vector. Arabidopsis protoplasts were than transformed with both constructs and examined for fluorescence using confocal microscopy. Figure 1.3 shows fluorescence in protoplasts transformed with SR45/YFPc and U2AF$^{35a}$/YPFn or U2AF$^{35Ctrb}$/YFPn. YFP fluorescence is largely seen in nuclear speckles with some fluorescence visualized in the nucleoplasm, which is similar to the localization of SR45 (Ali et al. 2008a). These results support the interaction of SR45 with both U2AF$^{35a}$ and U2AF$^{35b}$. Since plant U2AF$^{35}$ has a plant-specific C-terminal region (Figure 1.3), we tested if this region is necessary for U2AF interaction and

localization using a truncated version of U2AF$^{35b}$. Interestingly, this construct showed YFP reconstitution but it is mostly in the nucleoplasm as seen in Figure 1.3, supporting the importance of the C-terminal domain for the localization of U2AFs in speckles.

**Protein Alignment of U2AF$^{35}$ paralogs**

A protein alignment was performed for U2AF$^{35}$ paralogs using MegaAlign for Arabidopsis, Rice, and Human. The two paralogs of Arabidopsis share 84% similarity, and they are about 65% similar to rice U2AFs. However, AtU2AF$^{35}$s share only 26% similarity with human U2AFs and plant U2AFs have a short conserved C-terminal domain that is absent in humans.

**U2AF$^{35}$ proteins form hetero- and homodimers**

U2AF is a heterodimer consisting of U2AF$^{35}$, the smaller subunit, and U2AF$^{65}$, the larger subunit. Förster resonance energy transfer (FRET) has been recently shown that the smaller subunit, U2AF$^{35}$ in animals interacts with itself (Chusainow et al. 2005). BiFC was utilized to test if plant U2AF$^{35}$ subunits interact with each other, as well as with its paralogs. U2AF$^{35}$ proteins (a,b, and Ctrb) cloned into BiFC vectors as YFP N-terminal and YFP C-terminal fusions were used for BiFC studies. Protoplasts were transformed as before with U2AF$^{35a}$/YFPc, U2AF$^{35b}$/YFPc, or U2AF35$^{Ctrb}$/YPFc together with U2AF$^{35a}$/YFPn, U2AF$^{35b}$/YFPn, or U2AF35$^{Ctrb}$/YFPn. Figure 1.4 shows protoplasts transformed with each set of U2AF$^{35}$ proteins mentioned above. Each shows fluorescence indicating that U2AF$^{35}$ proteins can form both homo-and heterodimers, and that the C-terminal domain, which is not present in the truncated b form (Ctrb), is not essential for this interaction. It also appears that the localization of the dimer pairs is different amongst the protoplast transformations. U2AF$^{35b}$ dimers appear in speckles but with more fluorescence in the nucleoplasm than when dimerized with U2AF$^{35a}$.

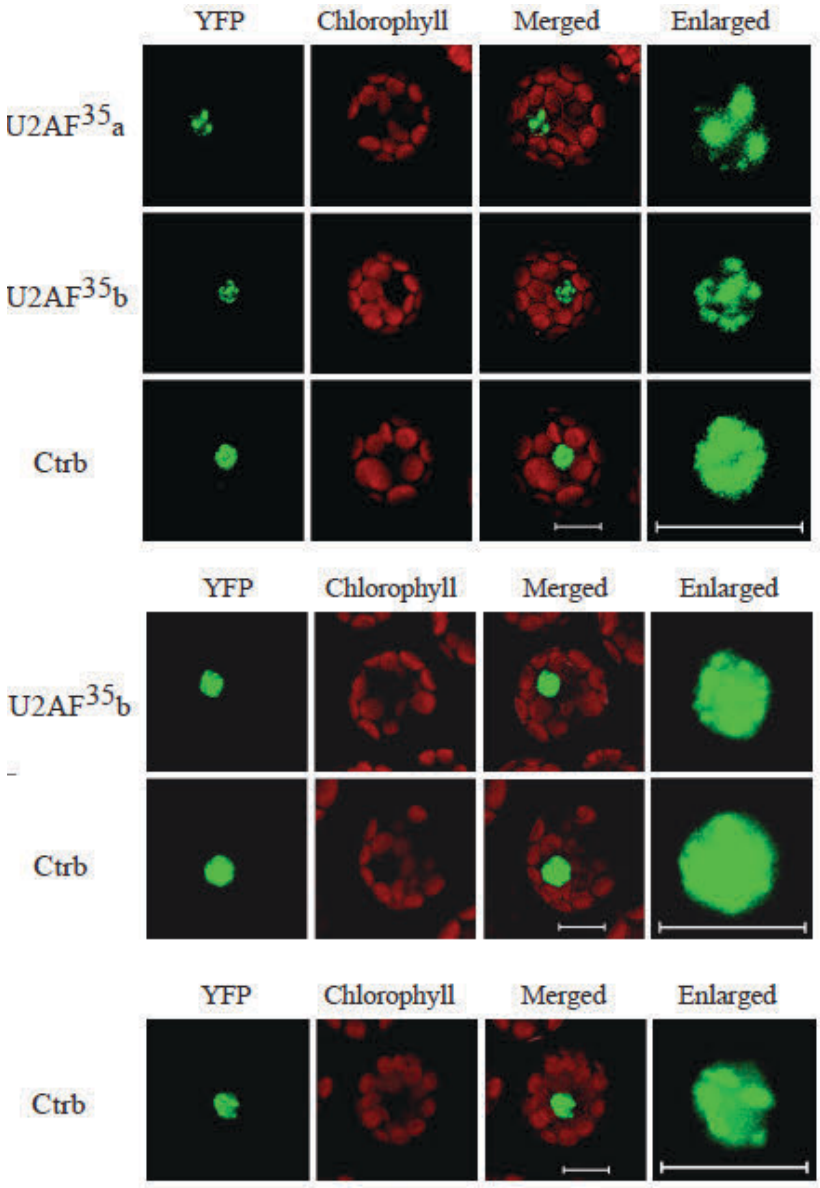


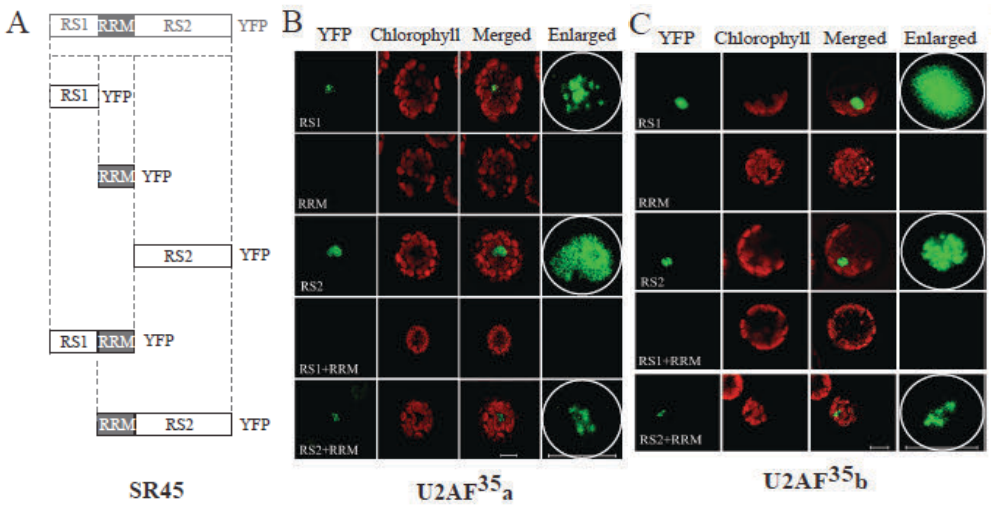

**Figure 1.4. Confocal images show Arabidopsis protoplasts expressing (Top Pictures) U2AF$^{35a}$-YFPn paired with full-length U2AF$^{35a}$-YFPc or U2AF$^{35b}$-YFPc or U2AF$^{35Ctrb}$-YFPc, (Middle pictures) full-length U2AF$^{35b}$-YFPn paired with U2AF$^{35b}$-YFPc or U2AF$^{35Ctrb}$-YFPc, or (Bottom pictures) U2AF$^{35Ctrb}$-YFPn paired with U2AF$^{35Ctrb}$-YFPc. Images in the right most panels show a zoomed in view of the nuclear region shown in the images under the YFP column. Bars = 10 µm.**

The U2AF$^{35b}$ and the U2AF$^{35Ctrb}$ dimer showed even more diffuse fluorescence in the nucleus.

**RS1 and RS2 of SR45 associate with U2AF$^{35}$s independently.**

SR45 is comprised of an N-terminal RS domain (RS1), a central RRM domain, and a C-terminal RS domain (RS2). BiFC was utilized to identify the domains of SR45 that interacts with U2AF$^{35}$ proteins. In order to do this, a series of SR45 deletion mutants were introduced in a BiFC vector as fusions to YFPc, and then used in BiFC assays with the U2AF$^{35}$ proteins (Figure 1.5). Constructs included were: RS1/YFPc, RRM/YFPc, RS2/YFPc, RS1/RRM/YFPc, and RRM/RS2/YFPc (Ali et al. 2008a). The SR45/YFPc deletion constructs were tested with U2AF$^{35a}$/YFPn, U2AF$^{35b}$/YFPn, and U2AF$^{35Ctrb}$/YFPn. Protoplasts transformed with each variant of U2AF$^{35}$/YFPn and those that contained either RS1 or RS2 showed fluorescence, which indicated an interaction of these two domains in SR45 with U2AFs (Figure 1.5B-D). Protoplasts transformed with the RRM/YFPc domain and each variant of U2AF$^{35}$YFPn exhibited no fluorescence, indicating RRM is not involved in SR45 interactions with U2AFs. While the RS2+RRM/YFPc domain showed some diminished fluorescence with each U2AF$^{35}$/YFPn protein (Figure 1.5B-D), it is interesting to note that the RS1+RRM/YFPn domain exhibited fluorescence only when paired with U2AF$^{35Ctrb}$ (Figure 1.5B,C,D).

In the case of U2AF$^{35a}$/RS2, fluorescence was much more diffuse throughout the nucleus when compared to the full-length SR45 (Figure 1.2B), although small speckles were present (Figure 1.5B). Protoplasts were also transformed with U2AF$^{35b}$+RS1, and the fluorescence was very diffuse throughout the nucleus, with very fine speckles. In contrast to this, U2AF$^{35b}$+RS2 shows fluorescence predominantly in the speckles (Figure 1.5C). U2AF$^{35Ctrb}$/RS1 was similar to U2AF$^{35b}$, showing a more diffuse pattern with RS1,



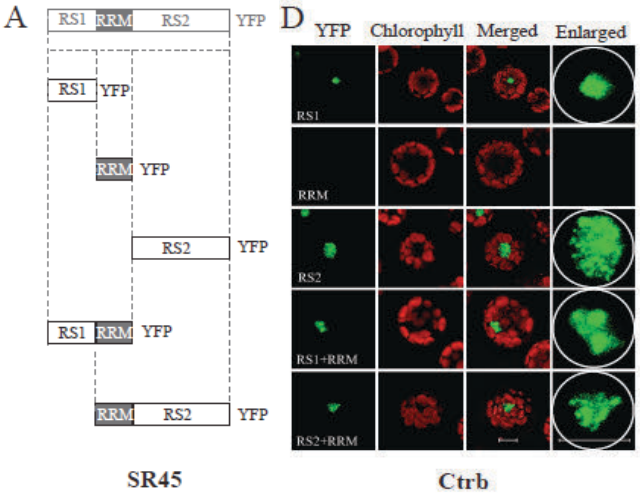




**Figure 1.5. Interaction of SR45 domains with U2AF$^{35}$ proteins. A. Schematic diagrams of the SR45 domains used in BiFC. B-D. Confocal images of Arabidopsis protoplasts expressing U2AF$^{35}$a-YFPn (B), full-length U2AF$^{35}$b-YFPn (C), or U2AF35Ctrb-YFPN (D) paired with the deletion mutant indicated on each panel. The images in the right most panels show a zoomed-in view of the nuclear region shown in the images under the YFP column. Bars = 10 µm.**

but in contrast, the RS2, RS1+RRM, and RS2+RRM constructs exhibited speckles which were much more visible with some still remaining in the nucleoplasm (Figure 1.5D). These results combined suggest that U2AF$^{35}$ can interact with the RS1 and RS2 domains independently, but other domains of the protein alter the strength of that interaction.

**U2AF$^{35}$ interacts with U1-70K**

U1-70K and SR45 both interact with U2AF$^{35}$ and have the same localization pattern in the nucleus. Because of this, we investigated whether U1-70K and U2AF$^{35}$ interact utilizing BiFC. The full-length and truncated versions of U2AF$^{35b}$ showed interactions with U1-70K, but U2AF$^{35a}$ did not. Each of the U2AF$^{35b}$ types that interacted with U1-70K showed fluorescence localized in the nuclear speckles, with some diffusion of fluorescence into the nucleoplasm (Figure 1.6).

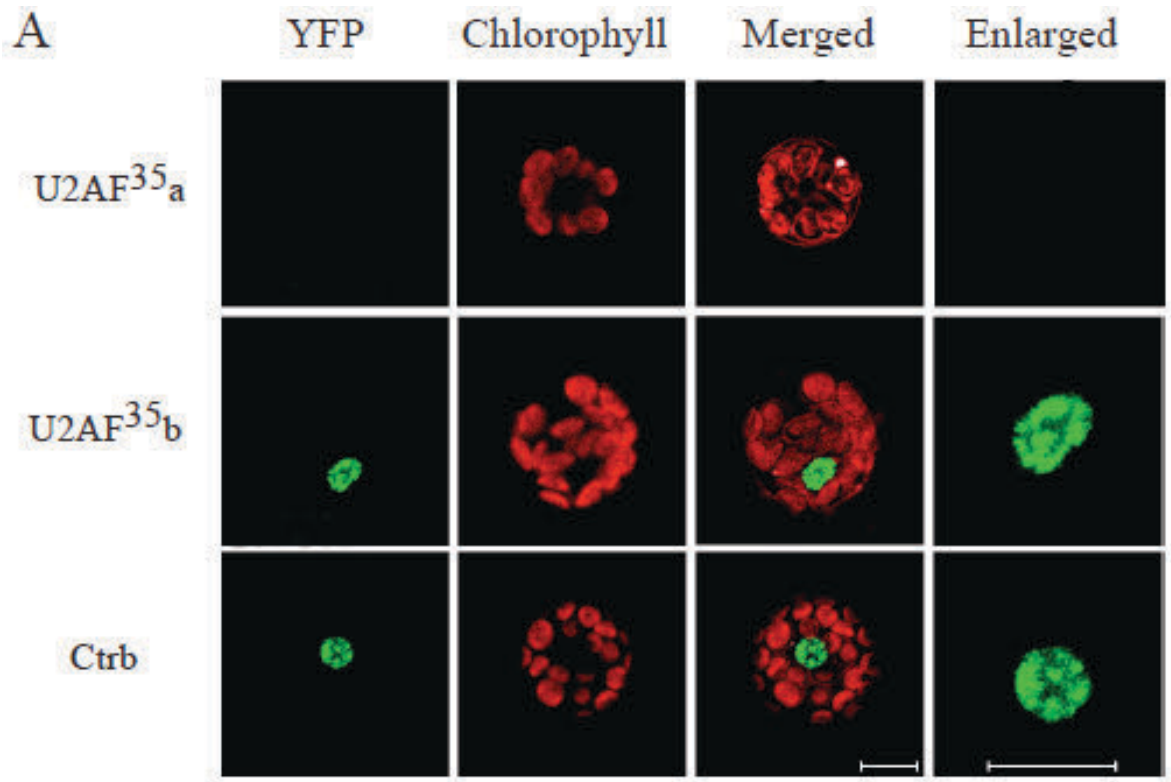



**Figure 1.6. Interaction of U1-70K and U2AF$^{35}$. A. Confocal images of Arabidopsis protoplasts expressing U1-70K-YFPn paired with U2AF$^{35}$a-YFPc, full-length U2AF$^{35}$b-YFPc, or U2AF$^{35}$Ctrb-YFPc as indicated on the left. Images in the right most panels show a zoomed-in view of the nuclear region shown in the images under the YFP column. Bars = 10 µm.**

## DISCUSSION

Eighteen SR proteins have been identified in Arabidopsis to date (Barta et al. 2010), whereas there are only 12 SRs in humans. In general, plants have many more SRs as compared to animals. Recent studies support evidence that there could be some functional importance to this variance as some aspects of pre-mRNA splicing in plants could vary from animals (Lazar and Goodman 2000; Lopato et al. 2002; Lorkovic et al. 2008; Reddy 2004; Tanabe et al. 2007). Previous studies have shown that SR45 binds to U1-70K, and hence, it could be involved in the 5' site recognition in splicing. In this study, we identified an interaction between SR45 and U2AF$^{35}$, which points to a connection between the 5' and 3' binding site proteins. This interaction is supported by *in vivo* studies using BiFC. We have seen through this study that SR45 interacts with U2AF$^{35}$a and U2AF$^{35}$b. More evidence to support this interaction comes from the interaction observed through BiFC utilizing mutant phenotypes of SR45 and U2AF$^{35}$. It is known that SR45 knockout plants have altered development and delayed flowering in long- and short-day photoperiods, along with an abnormal number of floral organs (Ali et al. 2007). Similarly, plants with a T-DNA insertion in U2AF$^{35}$a or U2AF$^{35}$b RNAi also are late flowering under similar conditions and have abnormal flowering morphology (Wang and Brendel 2006b). In addition to this, the expression of *FLC* (Flowering Locus C) has been found to be much higher in both *U2AF$^{35}$b* and *SR45* mutants (Ali et al. 2007; Wang and Brendel 2006b). Another study has shown that the SR protein SR45a interacts with U2AF$^{35}$b but not with U2AF$^{35}$a in yeast two-hybrid assays (Tanabe et al. 2007). Both SR45a and SR45 have two RS domains (N- and C-terminal to the RRM domain), but share only 26% similarity.

In animals U2AF, a heterodimer consisting of U2AF$^{35}$ and U2AF$^{65}$, is involved in 3' splice site recognition (Mollet et al. 2006; Wu et al. 1999). We report here that in

Arabidopsis both paralogs of U2AF$^{35}$ form homo- and heterodimers *in vivo*. Human U2AF$^{35}$ has also previously shown dimerization utilizing FRET analysis (Chusainow et al. 2005). These results suggest that two components of U2AF$^{35}$ may pair with U2AF$^{65}$ that make up the U2AF complex. BiFC analyses showed that this plant-specific domain is not essential for either interaction with SR45 or dimerization. Nevertheless, there were differences in the pattern of localization to speckles and nucleoplasm for proteins lacking this domain. The human U2AF$^{35}$ has three splice forms and there are two U2AF$^{35}$ proteins in humans but none of these have the C-terminal domain found in plant U2AF$^{35}$ (Mollet et al. 2006), suggesting a plant specific function associated with this C-terminal domain.

BiFC analysis of the SR45 domains and U2AF$^{35}$ interactions showed that both the RS1 and RS2 domains of SR45 interact independently with both U2AF$^{35}$ paralogs. There were some differences in the localization of the U2AF$^{35}$/SR45 interactions with different domains of SR45 and/or different paralogs of U2AF$^{35}$, where the distribution between the speckles and nucleoplasm was altered. Previous studies have shown a similar result when the same domains of SR45 were utilized to study U1-70K interactions with SR45 (Ali et al. 2008a). The fact that there was no interaction observed with any RRM construct with SR45/U1-70K, but there was observed fluorescence with U2AF$^{35}$a, b, and Ctrb along with RS2+RRM of SR45 and RS1+RRM with U2AF35Ctrb suggests that interactions between SR45 with either U1-70K or U2AF35 are regulated differently. However, in both cases it seems as though all three domains are necessary for the specificity found with the full-length SR45. In the case of SR45a, the RS1 domain by itself did not interact with U2AF$^{35}$b, which suggests that SR45a and SR45 interact with U2AF$^{35}$ differently (Tanabe et al. 2007).

BiFC also has suggested that both U2AF$^{35}$a and U2AF$^{35}$b interact with U1-70K. Previous studies of SR proteins in animal systems have shown SR proteins interact with both U2AF$^{35}$ and U1-70K, and may function as bridging factors between the 5' and 3' splice site factors (Wu and Maniatis 1993). However, no evidence of human U1-70K and U2AF35 interaction has been suggested recently and a FRET analysis was negative for interaction (Ellis et al. 2008). In contrast to this, our studies with BiFC indicate that both Arabidopsis U2AF$^{35}$b and U1-70K associate. Arabidopsis contains three genes that encode for the large U2AF subunit (U2AF$^{65}$) (Wang and Brendel 2006b), along with two paralogs of U2AF$^{35}$ that may interact in very specific ways to modulate both splicing and alternative splicing. As we have shown here, SR45 interacts with both U2AF$^{35}$a and U2AF$^{35}$b, but in contrast, SR45a only interacts with U2AF$^{35}$b. The possibility that the five similar human U2AF35 proteins (of these three are isoforms) and the single U2AF65 subunit may form heterodimers with different functional activities has previously been suggested (Kielkopf et al. 2004).

Based on our results we proposed a model (Figure 1.7) illustrating the roles of SR45, U1-70K, and U2AFs in splice site selection (Day et al in preparation). Both U2AF$^{35}$ and U2AF$^{65}$ have RRM-like motifs, which are a novel class of protein recognition motifs called UHMs (U2AF homology motif), that bind RNA weakly and need accessory proteins in order to assist proper binding (Kielkopf et al. 2004). Experiments with U2AF$^{35}$, SR proteins, and enhancer sequences have revealed that U2AF$^{35}$ mediates interactions between U2AF$^{65}$ and proteins that are bound to enhancers (Zhou et al. 2002). In addition to this, several human SR proteins have been found to interact with both U1-70K and U2AF$^{35}$ as a bridge between the 5' and 3' splice sites. This interaction has been confirmed for 2 of the SR proteins (SRSF1/ASF and SRSE2) utilizing FRET studies in vivo (Ellis et al. 2008; Wu and Maniatis 1993). We have shown using BiFC, as

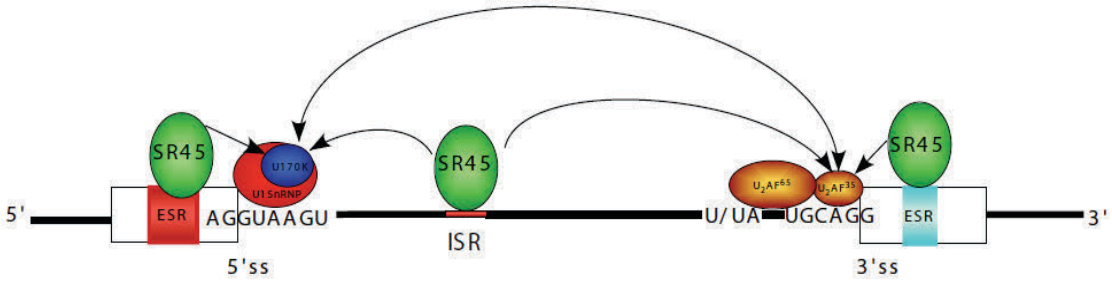



**Figure 1.7. Model of SR45 roles in splicing.** The RRM domain of SR45 binds to RNA and may do so at either exonic (ESR) or intronic (ISR) splicing regulators. The RS domains then interact with splicing factors U1-70K and U2AF$^{35}$ to recruit them to the 5' and 3' splice site, respectively. Interaction of both proteins with SR45 may bridge the 5' and 3' splice sites. White boxes indicate exons and horizontal line between and on either side of each exon indicate introns. Consensus sequences at 5' and 3' splice sites in plants are shown. Colored boxes in exons and intron represent ESRs and an ISR, respectively. SR45 may also interact with ESR/ISR through other SR proteins such as SR33, which is known to interact with SR45 (Golovkin and Reddy 1999). (From Day et al. submitted)

well as yeast two-hybrid assays and immunoprecipitation studies (Day et al in preparation), that SR45 interacts *in vivo* and *in vitro* with both U1-70K and U2AF35, and subsequently may bridge these two sites. It could be that SR45 binds to specific ESRs and/or ISRs while other SR and SR-like proteins may bind to others, providing specificity to the splicing of genes. Methods such as RNA-ChIP (Chromatin immunoprecipitation) and CLIP (Crosslinking and Immunoprecipitation) are powerful assays (Niranjanakumari et al. 2002; Ule et al. 2005) that may provide some insight for *in vivo* RNA targets of RNA-binding proteins towards the identification of RNA sequences recognized by SR45.

# CHAPTER 2: ALTERNATIVE SPLICING IN *CHLAMYDOMONAS REINHARDTII*

(The results presented here were published in “Genome-wide Analysis of Alternative Splicing in Chlamydomonas” BMC Genomics, (Labradorf et al, 2010))

## Introduction

The coding regions called exons in eukaryotic genes are disrupted by intervening non-coding sequences called introns. The process of pre-mRNA splicing which removes introns and covalently joins exons is both efficient and precise, and this is an important step for gene expression. Pre-mRNA splicing, whether constitutive or alternative, is carried out by macromolecular machinery known as the spliceosome, which consists of U1, U2, U4/U6, and U5 small ribonucleoprotein particles (snRNPs), and many other non-snRNP protein factors (Wahl et al. 2009). Years of research have established the accepted pathway for this stepwise process that allows for the spliceosome to become fully assembled. The assembly begins with the binding of U1 snRNP to the 5’ splice site, followed by the binding of U2 snRNP to the branchpoint at the 3’ splice site. Following this, the U4/6:U5 tri-snRNPs are joined to form the spliceosome (Matlin and Moore 2007; Smith et al. 2008). Aside from constitutive splicing, another process known as alternative splicing (AS) has been found to take place in many higher eukaryotic transcripts. This process has the ability to generate multiple transcripts from the same gene, thus potentially increasing the proteomic diversity and also introducing new ways in which gene expression may be regulated. In addition to this, the importance of AS is

further supported by recent evidence linking it to important biological pathways in development and disease (Cooper et al. 2009; Orengo and Cooper 2007). Effects AS has on biological pathways may be due to the fact that AS effects protein production and stability. Protein isoforms generated by a splice variant may either lose or gain a function, have altered subcellular localization, and/or posttranslational modifications (Black 2003; Reddy 2007). In addition to this, AS regulates gene expression through processes such as regulated unproductive splicing and translation (RUST) and mRNA recruitment (Brenner et al. 2007a; Brenner et al. 2007b). Alternative splicing also plays a role in the evolution of organisms (Blencowe et al. 2007).

Recently, complete genome sequences of many multicellular eukaryotic organisms have become available, along with large sets of full-length cDNAs and expressed sequence tags (ESTs), which have permitted a comprehensive analysis of AS. Additionally, new techniques such as splicing-sensitive microarrays and next generation sequencing tools have provided the opportunity for a global analysis of AS (Blencowe et al. 2007; Blencowe et al. 2008; Johnson et al. 2003). These analyses have shown that pre-mRNAs in humans undergo AS in ~95% of multi-exon genes (Brenner et al. 2007a), whereas genome-wide studies in Arabidopsis have recently indicated that over 40% of intron-containing genes undergo AS (Filichkin et al. 2010 ). These levels may be an underestimate due to the low level of ESTs available in plants compared to animals, and because some AS events are currently not represented or may be under-represented in EST collections since they occur only in specific cells, tissues, growth conditions, or developmental stages (Barbazuk et al. 2008; Hirose et al. 1993; Reddy 2007; Yoshimura et al. 2002).

Alternative splicing in gene families which encode for serine/arginine rich (SR) proteins has been shown to be quite extensive, giving a five-fold increase in

transcriptome complexity due to AS (Palusa et al. 2007). It has been found in mammalian systems that exon skipping is the most dominant type of AS, and in contrast 55% of AS events in flowering plants is due to intron retention (Campbell et al. 2006; Kim et al. 2007; Reddy 2007; Wang and Brendel 2006a). These differences in frequencies of splicing variations between plants and animals may be due to the differences in gene architecture and a regulatory mechanism that controls splicing (Reddy 2007; Wang and Brendel 2006a).

So far AS has not been studied extensively in unicellular autotrophs. The model green alga *Chlamydomonas* is of particular interest because it may be similar to the unicellular ancestor of land plants. Recently, the *Chlamydomonas* genome has been sequenced and a large number of available ESTs provide an opportunity to investigate post-transcriptional events including AS on a global level (Liang et al. 2008; Merchant et al. 2007; Vallon and Dutcher 2008). *Chlamydomonas* is a unicellular green alga that contains multiple mitochondria, two anterior flagella for mating as well as motility, and a single chloroplast that contains the photosynthetic apparatus along with many other critical metabolic pathways. *Chlamydomonas* diverged from land plants about one billion years ago, but still retains some animal as well as plant characteristics (Merchant et al. 2007). *Chlamydomonas*, like land plants, is an autotroph. It is similar to animals because it is also heterotrophic and is mobile (Harris 2001).

Comparative genomic analysis has traced *Chlamydomonas* genes back to the plant-animal common ancestor. Many *Chlamydomonas* genes are derived from the plant-animal common ancestor, and are have homologs in plants. Genes that are shared by *Chlamydomonas* and animals are derived from the last plant-animal common ancestor, although many genes that were once shared with animals have been lost in

angiosperms, namely those that encode for the eukaryotic flagellum and associated basal bodies (Li et al. 2004).

Another attribute which makes *Chlamydomonas* a useful model organism is that unlike plants, which are sessile, it is able to inhabit and survive in a wide variety of environments and conditions, due to regulatory genes that allow for extensive metabolic flexibility (Grossman et al. 2007).

For the last five decades, *Chlamydomonas* has been used as a model organism for photosynthesis, flagella function and structure, and many other biological processes. In addition to this, recent studies have utilized *Chlamydomonas* to investigate biofuels production, hydrogen production, and making human protein therapeutics. Although these investigations are just beginning, they look promising (Ghasemi et al. 2010; Mayfield et al. 2010; Rupprecht 2009).

*Chlamydomonas reinhartii* has a 120Mb genome, of which about 93% has been fully sequenced (Merchant et al. 2007; Vallon and Dutcher 2008). There are about 16,709 protein-coding genes predicted in the most recent version (v4) of the *Chlamydomonas reinhardtii* genome and of these about half have cDNA/EST support. These protein coding genes contain on average 8.3 exons per gene and are intron-rich when compared with both unicellular eukaryotes and land plants. Interestingly, the average *Chlamydomonas* intron is much longer than that of Arabidopsis (~373bp), and although *Chlamydomonas* is a protist, the average number and size of its introns are more similar to multicellular organisms. In addition to this, only 1.5% of introns are short (<100 bp) and it was found that the bimodal intron size distribution, which is typical of most eukaryotes, was not observed (Merchant et al. 2007). Intron length has been

positively correlated in the past to an increased number of splicing events, resulting in separate isoforms.

Utilizing *Chlamydomonas* to analyze AS will allow for a much-needed comparison of AS between unicellular photosynthetic eukaryotes and their related more complex flowering plants. This analysis will also allow us to gain insight into to how much AS has evolved during the evolution of land plants. In this regard, we have used both computational and experimental methods for a comprehensive analysis of AS in *Chlamydomonas reinhardtii*. Our results show that AS is common in *Chlamydomonas*, but its extent is less than in land plants. However, the relative frequency of different splicing events in *Chlamydomonas* is very similar to higher plants. Detailed results from the computational analysis are available on our "*Chlamydomonas AS"* site http.//combi.cs.colostate.edu/as/chlamy.

## MATERIALS AND METHODS

### Cultures and Strains

Wild-type strain cc1690 and wall-less strain cc503 were obtained from the Chlamydomonas Center culture collection at Duke University. These were then used to inoculate 75ml of autoclaved TAP media in Erylenmeyer flasks (Harris 2009) with cotton stoppers. The cells were maintained at 22℃ on a shaking platform in a growth chamber on a 12:12 light/dark cycle. Cells were subcultured during log phase at a starting density of $1x10^5$ cells/mL (Harris 2009). In order to obtain a cell pellet for RNA isolation, a 2ml aliquot was collected during log phase in 2ml tubes and centrifuged at 0.2 g for 2 minutes. Supernatant was removed and the procedure repeated until a combined pellet of 4-6mls was obtained. The pellet was then frozen immediately in liquid $N_2$ and stored at -20℃ until RNA isolation.

### RNA Isolation

Total RNA was isolated using an RNeasy Plant Mini Kit (Qiagen, http://www.qiagen.com/). Prior to RNA isolation, the cell pellet was thawed on ice and frozen in liquid N2. This procedure was repeated 2-3 times in order to lyse the cells and the total RNA was isolated according to the protocol provided by the kit manufacturer. RNA amount was quantified spectrophotometrically at 260 nm. The RNA sample was treated with DNase I according to the manufacturer's instructions (Invitrogen). The quality of RNA was verified by running an aliquot on a 1% agarose gel.

### cDNA Synthesis

DNase-treated RNA (1.5µg) was used to synthesize first-strand cDNA with an oligo (dT) primer in a 20ul reaction volume using SuperScriptII (Invitrogen). After DNase treatment, 1 µl of oligo (dT) primer was added and the sample was centrifuged at 10,000

rpm for 30 seconds. This was incubated first at 65℃ for 10 minutes and then on ice for 5 minutes. A cocktail solution was prepared for each sample which consisted of 4 µl 5x buffer, 2 µl 100mM DTT, 1 µl 10mM dNTP, 1 µl RNase Out enzyme, and 1 µl SuperScript II. This cocktail was added to the sample after the ice incubation, and then it was kept at 42℃ for 1 hour. As a final step, the sample was kept at 65℃ for 10 minutes before PCR. Samples of cDNA were stored at -20℃ prior to PCR.

**PCR of Ornithine Decarboxylase 1 (*ODC1*) and Asparagine Synthase (*ASyn*) transcripts**

One-twentieth of the first-strand cDNA was used for PCR amplification in a reaction volume of 20 ul. The primers were designed using the Primer3 Input (http://frodo.wi.mit.edu/) software. The control primers for *TUA1* were designed according to Bisova et al 2005. Touchdown PCR (TD-PCR) was performed using a temperature range of 50-60℃ based upon the primer Tm (Korbie et al 2008). An extended hot-start method was utilized in which the PCR sample was allowed to incubate at 95℃ for 1.5 hrs prior to PCR cycling. The following TD-PCR conditions were used: initial denaturation performed at 95℃ for 3 minutes, followed by 10 cycles where denaturation was at 95℃ for 30 seconds, and an annealing temperature is 60℃ for 45 seconds. The annealing temperature was set to decrease 0.5℃ every cycle until the 10 cycles are complete. Extension was done at 72℃ for 3.5 minutes. The next 20 cycles have a denaturing temperature of 95℃ for 30 seconds, an annealing temperature of 50℃ for 45 seconds, and an elongation temperature of 72℃ for 3.5 minutes. The final extension occurs at 72℃ for 5 minutes. Amplified PCR products were resolved by electrophoresis in 1% agarose gels. All PCR reactions were performed using Takara EX

Taq™ polymerase. Bands were extracted using a razor blade and stored at -20℃ until gel extraction.

**TOPO Cloning and Sequencing**

Gel extraction was performed prior to TOPO cloning using the GeneJET™ Gel Extraction kit (Fermentas). After DNA was extracted, the sample was dried using the Speed-Vac (Savant SC110) and then the pellet was resuspended in 4 µl of water. The DNA was cloned using the TOPO TA Cloning Kit (Invitrogen). To a sterile tube 2.0 µl of PCR product, 0.5 µl of a kit-provided salt solution, and 0.5 µl of TOPO vector were added. This was then incubated for 10 minutes at RT and put on ice. To 2 µl of this mixture, 12.5 µl of TOP10 cells was added and heat shocked for 30 seconds at 42℃. This was then put on ice and 250 µl of kit provided SOC media was added. The sample was then shaken horizontally for 37℃. After 1 hr, the cells were spread on pre-warmed LB plates supplemented with carbinomycin, IPTG, and X-Gal. The plates were incubated overnight at 37℃. The next day white colonies were picked from the plates to inoculate a 5 mL overnight culture was shaken at 37℃. Plasmid isolation was performed using the QIAprep Spin Miniprep kit (Qiagen). Digestion was then performed with 20 µl plasmid, 2.0 µl water, 0.5 µl EcoRI, and 2.5 ul buffer. The sample was digested at 37℃ for 8 hours. The digestion was resolved by electrophoresis in 1% agarose gels. Upon confirmation of an insert, the plasmids were then sequenced at the Colorado State Macromolecular Center.

# RESULTS

## Computational Analysis

Our collaborator Dr. Ben Hur and his graduate students in computer sciences developed a pipeline for the detection and visualization of AS in *Chlamydomonas* utilizing the BLAT system to produce EST-to-genome alignments (Kent 2002), paired with a modified version of the Sircah tool for AS detection software (Harrington and Bork 2008). EST data for this analysis utilized a recently constructed EST dataset containing 252,484 ESTs that were processed using cDNA termini to anchor transcripts to their correct positions in the genome (Liang et al. 2008). The alignment and AS detection pipeline utilized in this study generated 498 ESTs aligned to the genome, which showed 611 AS events. These AS events were then summarized into splice graphs (2009a; Heber et al. 2002).

## Experimental Verification

To verify the predicted splicing events, we chose two genes corresponding to ornithine decarboxylase 1 (*ODC1*, gene ID:OVA_SAN_estEXT_fgenesh2_kg.C_340012) and asparagine synthase (*ASyn*, gene ID: estExt_fgenesh2_kg.C_280076) with splice graphs seen in Figures 2.1 and 2.2. We then performed reverse transcription PCR (RT-PCR) to detect the splice variants that were predicted computationally. When DNAse-treated RNA with primers corresponding to these genes was used in PCR, there was no amplification, suggesting no DNA contamination within the RNA. The RT-PCR analysis performed with primers corresponding to the first and last exons of *ODC1* revealed six splice variants (Figure 2.3B). An RT-PCR analysis performed for *ASyn* produced two splice variants (Figure 2.3C). When compared with the computational analysis, alignments of previously available ESTs predicted two AS events in *ODC1* and three in


**Figure 2.1. ODC splice graph. Splice graph with the relevant EST evidence for the ODC1 gene that exhibits intron retention and alternate 3' splice site. This figure was generated by Sircah as part of our pipeline.**

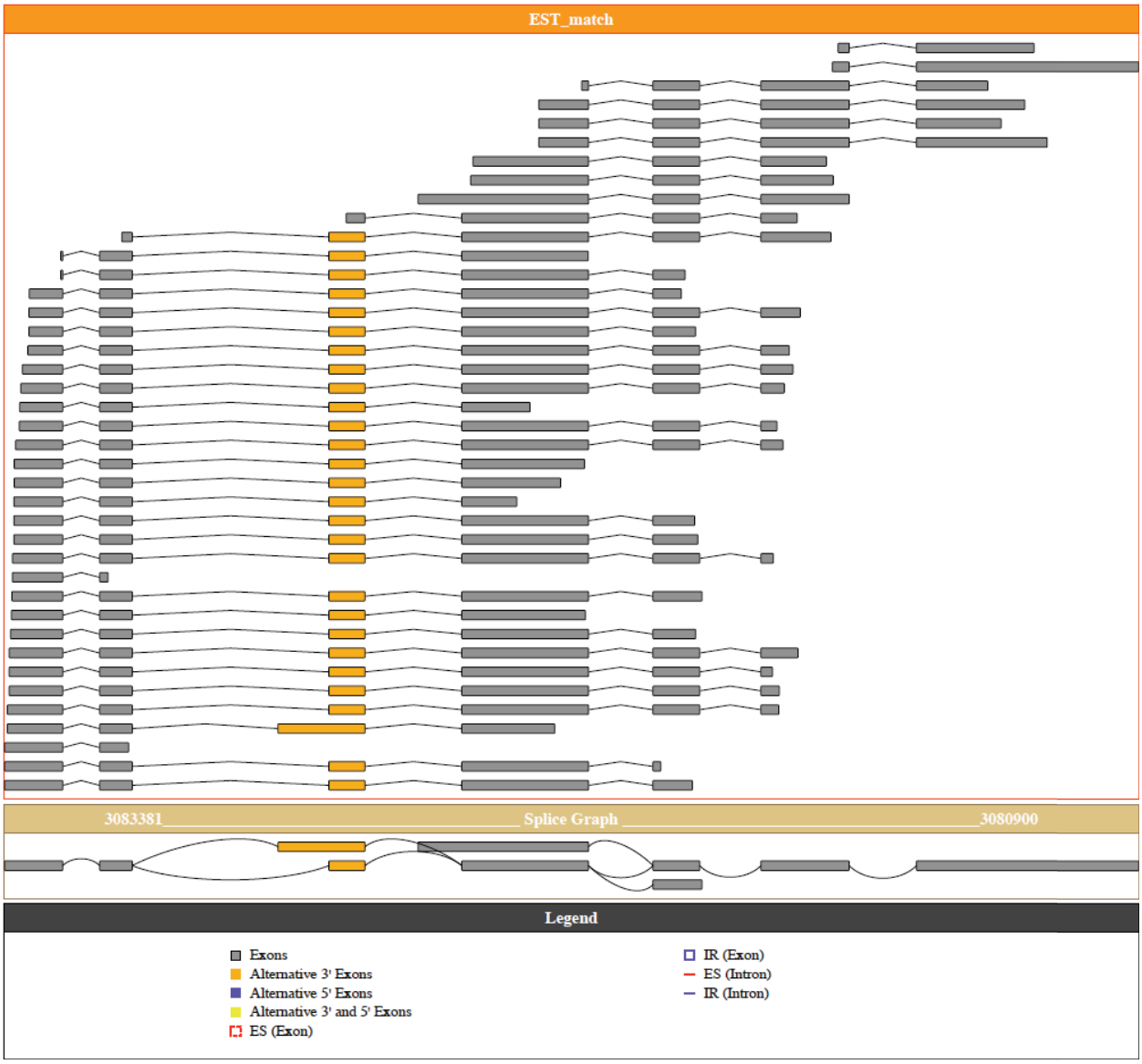



**Figure 2.2. Asyn splice graph. Shown is a splice graph with the relevant EST evidence for the Asyn gene, which exhibits alternative 3' splice site. This figure was generated by Sircah as part of our pipeline.**

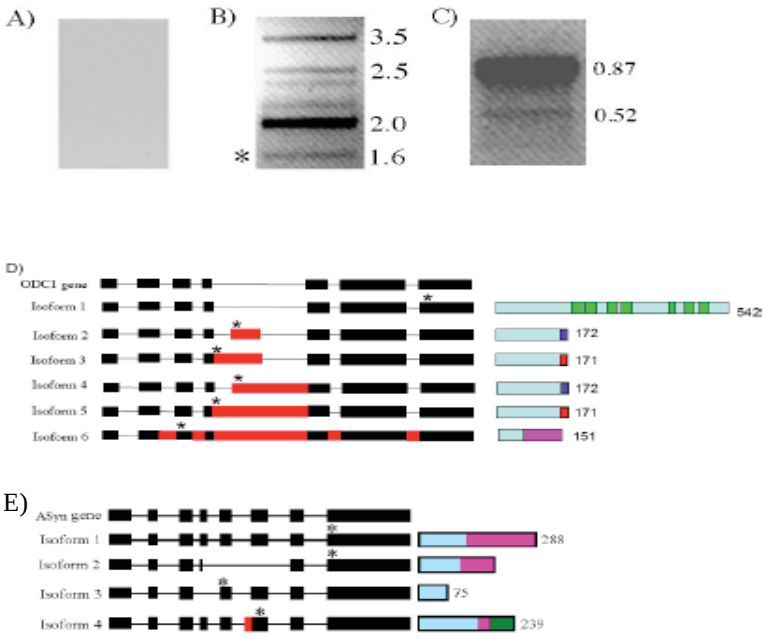



**Figure 2.3 Analysis of ornithine decarboxylase 1 (ODC1) and asparagine synthase (ASyn) splice variants in Chlamydomonas using RT-PCR**. A) DNAase-treated RNA was used in PCR with ODC1 primers. B) cDNA amplified with ODC1 primers. An asterisk indicates the spliced form for the full-length protein. Numbers on the right indicate amplified product size in Kb. C) cDNA amplified with ASyn primers. Numbers on the right indicate amplified product size in Kb. D) Diagram showing splicing events in six splice variants (left) and predicted proteins (right) for ODC1. Gene is indicated on top and all six splice variants are shown under the gene. Black boxes indicate constitutively spliced exons and red boxes indicate the included regions in different isoforms. Asterisk indicates the position of translation termination codon. Isoforms 1 to 6 correspond to the bottom to top bands in 3B. The number next to each predicted protein indicates the length of the protein. Conserved signature motifs in ODC1 are represented as green boxes in the full-length protein. Red, blue and magenta colors in truncated proteins represent amino acids unique to them. E) Diagram showing splicing events in four splice variants (left) and predicted proteins (right) for the gene ASyn. The representation of the gene and its splice variants is the same as in D. A conserved domain in ASyn is shown in pink. The green region in isoform 4 represents a unique sequence.

*ASyn*. To verify these results, we cloned each isoform that was amplified and sequenced these products. The types of AS events we discovered from this sequencing along with their effect on the predicted proteins is shown in Figures 2.3D and 2.3E. The RT-PCR results we obtained in this study show that *ODC1* produces more isoforms than predicted by EST alignments. This suggests that although there is a considerable number of ESTs available for analysis, there still are more to be discovered and currently all AS events in a gene cannot be predicted based on the current collection alone.

A sequence analysis was performed on all six isoforms found in *ODC1* and these studies revealed that five out of the six isoforms observed are due to AS of the 4$^{th}$ intron, which happens to be the largest in the ODC1 gene. In these isoforms, the AS events observed included intron retention, and Alt5' and Alt3' events. Of the six isoforms observed, only one produced the functional full-length protein product of 542 amino acids, which contains all of the seven conserved signature motifs of *ODC1*. Each of the remaining five isoforms were predicted to produce three truncated forms of this full-length protein containing 152-172 amino acids due to in-frame translation termination codons. All of these truncated forms do not contain conserved regions found in *ODC1* and thus they are unlikely to be functional proteins. However five of the six isoforms contain a premature termination transcript (PTC), a component that is involved in the Nonsense-Mediated Decay (NMD) mRNA transcript surveillance system. It is interesting to note that all five of these splice variants with PTCs are highly expressed, and in some cases, expressed at a much higher level than the functional transcript (Figure 2.3C, compare the lowest band to the rest of the bands). For the *ASyn* gene, two of the four splice variants also encode truncated proteins (Figure 2.3E). Of these truncated proteins, isoform 3 and 4 contain PTCs and are likely to be targets of the NMD system.

Although there were three predicted splice variants for *ASyn*, we verified only one of these predicted isoforms (Isoform 1). We were able to detect a novel isoform, which was not detected with ESTs (Isoform 2).

## DISCUSSION

### Properties of Introns

The vast majority of genes in *Chlamydomonas* have introns (~88%). Interestingly, although *Chlamydomonas* is considered to be a very simple organism, the percentage of intron-containing genes is much higher when compared with plants and humans. This is noteworthy, given that *Chlamydomonas* contains both animal and plant characteristics, which elicits provocative evolutionary questions about gene architecture and evolution. Previous studies, which compare gene architecture in flowering plants and animals, have shown that there are many significant differences between the two groups (Filipowicz et al. 2000; Reddy 2001). It has been shown that genes of land plants are not only shorter than animal genes, but that land plant genes also contain fewer exons with shorter introns (Reddy 2007). There are also differences in gene architecture between *Chlamydomonas* and land plants. For example, the average number of introns in *Chlamydomonas* is more similar to humans, whereas the median size of exons (132 bases) and introns (232 nucleotides) is more similar to flowering plants. Plant introns are rich in T or T/A, and this compositional bias is necessary for the recognition of splice sites and the efficient splicing of pre-mRNAs (Filipowicz et al. 2000; Reddy 2001).

The *Chlamydomonas* genome has a GC content of 64%, which is much higher than that of multicellular organisms. Four signals within the introns of protein coding genes of metazoans are necessary for precise splicing of mRNAs. These include two consensus sequences at the 5' and 3' splice sites with conserved GT and AG dinucleotides, a polypyrimidine tract at the 3' end of the intron, and a branch point found 17-40 nucleotides upstream of the 3' splice site (Black 2003). In land plants a branch point is not that obvious, and the 3' end of plant introns is very rich in T nucleotides

(Reddy 2001). In contrast to this, in *Chlamydomonas*, the 3' end of introns is enriched in C in place of a polypyrimidine tract.

**Extent and types of alternative splicing**

Our computational analysis has shown that 498 clusters resulted in 611 AS events. Each of these splicing events is summarized in a splice graph similar to Figures 2.1 and 2.2. A website with splice graphs of all alternatively spliced genes and additional information is available at http://combi.cs.colostate.edu/as/chlamy (Reddy et al. 2010). Out of the clusters that showed AS, 484 were associated with the genes predicted in the 4.0 version of the *Chlamydomonas* genome (2009b). Each of the observed AS events were classified into groups: Intron Retention (IR), Alternative 5' splice site (Alt5'), Alternative 3' splice site (Alt3'), events where both 5' and 3' ends of an intron are alternative spliced (AltB), and exon skipping (ES). Our studies revealed that the relative frequency of each of these various types of splicing events is very similar to those observed in other plant species, with IR making up ~50% of those events (Table 2.1) (2009b).

**Splice site strength**

Splice sites that participate in AS are usually weaker than constitutive splice sites, and our observations were consistent with this trend in other organisms (Zheng et al. 2005), and all differences were statistically significant (Table 2.2). Of these differences, the most significant was shown to be at the 3' splice site of Alt3' events. The most prevalent form in each AS event was identified as the one supported by the largest number of ESTs. In the case of Alt5' and Alt3' events, it was found that the splice sites for the non-prevalent forms were weaker than those observed in constitutive splicing (Table 2.3). Each of these differences also proved to be highly statistically significant.

**Table 2.1 The prevalence of different types of alternative splicing events.**

| | Chlamydomonas | Arabidopsis | Rice |
|---|---|---|---|
| IR | 305 (50.0%) | 4635 (56.1%) | 7774 (53.5%) |
| ES | 73 (11.9%) | 666 (8.1%) | 2004 (13.8%) |
| Alt5′ | 71 (11.6%) | 845 (10.2%) | 1642 (11.3%) |
| Alt3′ | 158 (25.8%) | 1810 (26.0%) | 2201 (15.5%) |
| AltB | 4 (0.7%) | 308 (3.7%) | 921 (6.3%) |
| Total | 611 | 8264 | 14542 |

This table shows the number and frequency of each type of alternative splicing. Percentage of the total number of events is shown in parenthesis. The statics for Arabidopsis and rice are from [11]

**Table 2.2. Splice site strength in Alternative and Constitutive Splicing.**

| Event | 5′ site | | 3′ site | |
|---|---|---|---|---|
| | motif score | *p*-value | motif score | *p*-value |
| Intron Retention | 7.790 | 7.492 (6.258e-44) | 7.165 | 6.925 (8.734e-11) |
| Exon Skipping | 6.701 | 6.156 (1.465e-09) | 7.735 | 6.921 (4.402e-12) |
| Alt 5′ | 7.304 | 6.373 (7.175e-20) | 7.448 | 7.097 (0.0517) |
| Alt 3′ | 8.594 | 8.434 (0.00176) | 5.460 | 3.478 (1.686e-80) |
| Constitutive | 8.822 | N/A | 7.574 | N/A |

Average splice site scores and *p*-values for alternative splicing events and constitutive splicing are shown here. All scores are computed with respect to the splice site motif of the constitutive splice form, following the protocol used in [29]. In all cases, the scores for the alternatively spliced form are lower than for constitutive splicing. The *p*-values are based on a comparison of the scores for each type of alternative splicing event with the scores for constitutive splicing, and are computed using the Wilcoxon signed-ranks test. Except for the case of exon skipping, the 5′ and 3′ sites refer to the splice sites of an excised intron. In exon skipping the 5′ and 3′ sites are the splice sites flanking the skipped exon.

**(Tables taken from Labradorf et al 2010)**

**Table 2.3. Splice site strength for prevalent and non-prevalent forms.**

| AS event | non-prevalent vs prevalent | | | prevalent vs constitutive | | |
|---|---|---|---|---|---|---|
| | non-prevalent avg. score | *p*-value | prevalent avg. score | prevalent avg. score | *p*-value | constitutive avg. score |
| Alt5′ | 5.88 | 1.98e-07 | 8.14 | 7.51 | 1.67e-07 | 8.85 |
| Alt3′ | 4.17 | 2.30e-12 | 6.73 | 6.31 | 6.95e-18 | 7.57 |

Splice site scores and *p*-values for the 5′ and 3′ sites of prevalent and non-prevalent splice forms. The table shows data that support two hypotheses: (i) non-prevalent splice sites are weaker than splice sites associated with the prevalent splice form; (ii) prevalent splice sites are weaker than splice sites associated with constitutive splicing. The "avg. score" columns provide the average score of splice site occurrences. For the comparison of non-prevalent with prevalent splice forms, the scores are computed with respect to a motif model of prevalent instances; for the comparison of prevalent and constitutive splicing the scores are computed with respect to a model of the constitutive splice sites.

**Table 2.4. The effect of splicing on predicted proteins.**

| | | AS in Coding Sequence | | | |
|---|---|---|---|---|---|
| | | ORF Shortened By | | | |
| Event | # events | bp | % | AS in UTR # events | Total |
| IR | 30 | 276.52 | 54.52% | 1 | 31 |
| ES | 4 | 270.00 | 59.74% | 0 | 4 |
| Alt5′ | 6 | 353.62 | 60.67% | 4 | 11 |
| Alt3′ | 22 | 476.32 | 51.37% | 9 | 31 |
| Total/Avg. | 62 | 359.52 | 54.84% | 14 | 77 |

We considered a subset of the clusters with a full-length cDNA, a single alternative splicing event, and a published start codon in the JGI version 4.0 genome annotation. For these clusters we show the number of events in a UTR and the number of events in the coding sequence, where UTRs were detected by the location of where AS occurred with respect to the published start codon and the first stop codon in the reading frame. For events in the coding sequence, we show the average reduction in the length of the predicted ORF that results when comparing the prevalent splice form with the non-prevalent splice form. In all cases but one, the non-prevalent splice form is shorter as a result of a premature termination codon. For IR, the prevalent form is always the one where the intron is spliced, and the non-prevalent form retains the intron. For ES, the prevalent form always contained the exon while the non-prevalent form skipped it.

**(Tables taken from Labradorf et al 2010)**

**Length and GC content of retained introns and skipped exons**

We compared the length of retained introns and skipped exons with those that didn't exhibit AS. Our computational analysis revealed that retained introns are much shorter than those that do not exhibit AS. We found that the median size of retained introns is 127 bp compared to a median size of 232 bp in excised introns (Reddy et al. 2010). It was found that this difference was much more pronounced in *Chlamydomonas*, where median sizes are 93bp, compared with *Arabidopsis*, where median sizes are 100-200bp (Ner-Gaon et al. 2007). In addition to this, skipped exons were found to be shorter than exons that are not alternatively spliced. *Chlamydomonas* alternatively spliced exons exhibited a median size of 84 bp compared to the median size of 132 bp in constitutively spliced exons. Interestingly, it is known that in land plants introns have high AT content, whereas exons are GC rich. It has been reported that a high percentage of A/T or T is an important factor for efficient splicing of introns in flowering plants. The presence of proteins that bind to U-rich sequences has also been found in plants (Lorkovic et al. 2000a; Lorkovic et al. 2000b). In our computational studies, we found that in *Chlamydomonas*, retained introns have a GC content of 57%, and that excised introns have a GC content of 62%. Moreover, short in-frame introns have a much lower GC content of 56%. We observed a similar pattern for skipped exons, which have a low GC content of 63%, when compared with constitutive exons, which have a GC content of 66%. These differences are highly statistically significant (Reddy et al. 2010). Interestingly, the opposite pattern is observed in Arabidopsis, where retained introns have a higher GC content (Ner-Gaon et al. 2007).

**Impact of AS on predicted proteins**

Alternative splicing usually results in the generation of a premature termination codon (PTC) (Black 2003; Brenner et al. 2007a). mRNA transcripts that contain a PTC are subject to degradation through NMD pathway (Chang et al. 2007; Maquat 2004). Many studies support that AS of pre-mRNAs is coupled to mRNA degradation though regulated unproductive splicing and translation (RUST) (Brenner et al. 2007a; Brenner et al. 2007b; Palusa and Reddy 2010). Our computational analysis has shown that out of the 498 clusters showing AS, 483 of these correspond to the annotated genes. 416 of these have published start codons and 77 have a single AS event including a stop codon within a full-length EST. When alternative splicing occurred in coding region, 76 out of 77 clusters' non-prevalent splice form led to a shorter protein because of the presence of a PTC (Table 4). Comparatively in Arabidopsis, 50% of AS events that occur in the coding region have a PTC (Wang and Brendel 2006a). It has been shown in plants that transcripts with a PTC undergo NMD, and some of the machinery involved in the NMD pathway has been reported in plants (Davies et al. 2006; Kurihara et al. 2009; Palusa and Reddy 2010; Schoning et al. 2008). The *Chlamydomonas* predicted proteome also contains components of NMD such as UPF1 and UPF3, along with exon-junction complex proteins, which point towards NMD playing a role in the gene expression of *Chlamydomonas*.

**Alternative splicing motifs**

It is thought that other sequences may be involved in regulated splicing due to the presence of four loosely conserved signals. Protein factors such as SR proteins and hnRNPs have been shown to regulate splicing by binding to splicing regulatory elements either in exons or introns and to enhance or to prevent the utilization of a splice site in

both metazoans and land plants (Burge et al. 2002; Caceres and Long 2009). Our computational analysis found constitutively spliced introns consistently produced motifs that contained tandem repeats of di-nucleotides or tri-nucleotides. The top scoring consensus sequence was TGCTGCTG (Reddy et al. 2010). This is relevant because it has been observed in previous studies that simple repetitive elements have been shown to bind splicing regulatory proteins such as SRs and hnRNPs, and thus contribute to regulated splicing (Majewski and Ott 2002). Interestingly, *Chlamydomonas* has several SR and hnRNP proteins that share significant sequence similarity with splicing regulators found in multicellular organisms.

**Experimental verification of alternative splicing**

In previous studies, it has been found that in most SR genes, which undergo extensive AS in Arabidopsis, the longest intron is involved in generating multiple transcripts (Palusa et al. 2007). In humans, splice isoforms that contain a PTC at more than 50 nucleotides upstream from the last 3' exon-exon junction are found to be degraded by NMD. In the case of *ODC1*, it is interesting to note that all five splice variants containing a PTC are abundant, and occasionally in higher levels than the functional transcript. This observation may suggest that the isoforms containing a PTC may perform some sort of regulatory role in gene expression. The *ASyn* gene also has isoforms that result in truncated proteins. Two of these *ASyn* isoforms also contain a PTC, indicating that they are likely to be targets of NMD. The *ASyn* gene's role is to transfer the amide group of glutamine to aspartate to form asparagine, which plays a role in nitrogen metabolism (Inokuchi et al. 2002). Land plants have also been found to have one or more genes similar to this, but it is not known whether or not they undergo AS, as in the case of *Chlamydomonas* (Nakano et al. 2000).

*ODC1* is gene that produces ornithine decarboxylase, a key rate-limiting enzyme in the biosynthesis of polyamines that are required for both cell growth and division in *Chlamydomonas* as well as in other organisms (Theiss et al. 2002). Ornithine decarboxylase's role in biosynthesis is to catalyze the formation of purtrescine from ornithine. This enzyme is present in both algae and animals, but not found in higher plants such as Arabidopsis, due to a different polyamine synthesis pathway (Hanfrey et al. 2001). In animals, ODC pre-mRNA also undergoes pre-mRNA splicing, but the 5' untranslated region is alternatively spliced and this event is responsible for control of ribosomal association with the ODC mRNA (Pyronnet et al. 2000). In *Chlamydomonas*, no physiological significance of *ODC1* isoforms has been studied. Due to the fact that some isoforms of *ODC1* contain a PTC, it can be postulated that these isoforms may be involved in the regulation of the level of the functional isoform through NMD. In addition to the genes we have reported here, other *Chlamydomonas* genes have had splice variants predicted and verified (Beligni et al. 2004; Croft et al. 2007; Falciatore et al. 2005; Grossman et al. 2007; Schroda et al. 2001; Willmund et al. 2007).

This study has shown that *Chlamydomonas* has a small percentage of alternatively spliced genes, some of which have been verified experimentally. As the *Chlamydomonas* alternative splicing field expands, it is likely that there are many more alternatively spliced genes that were not represented in this study. Deep sequencing of the *Chlamydomonas* transcriptome paired with various stress conditions should also reveal more information on the extent of alternative splicing in *Chlamydomonas*. We found that 3% of genes in *Chlamydomonas* are alternatively spliced. This is a small number compared to Arabidopsis, which has been shown to have 40 genes that undergo AS (Filichkin et al. 2010). There are few documented cases of alternative splicing in *Chlamydomonas* (Beligni et al. 2004; Croft et al. 2007; Falciatore et al. 2005; Gonzalez-

Ballester et al. 2008; Schroda et al. 2001; Willmund et al. 2007), and in several cases, the protein coded by splice isoforms was found to be different, suggesting that proteins generated by alternative splicing may have different functions. To support this, alternative splicing in some genes has been shown to have a physiological role, and the events reported here are likely to be important in the regulation of gene expression and protein function. In addition, AS may also contribute to the regulation of functional transcript levels. We have shown in this analysis that alternative splicing is prevalent in *Chlamydomonas reinhardtii*. We have found that a large number of genes undergo alternative splicing, and together with the simplicity of the system and largely available molecular and genetic tools, our studies suggest that this organism is a viable experimental system that can be utilized to investigate the mechanisms involved in alternative splicing.

# CHAPTER 3: ANALYSIS OF NONSENSE-MEDIATED DECAY IN *CHLAMYDOMONAS*

## Introduction

Nonsense-mediated decay (NMD) is a mechanism of mRNA surveillance that degrades transcripts containing premature termination codons (PTCs). PTCs are stop codons that occur prior to the normal termination codon of a transcript and result in production of a truncated protein. These may be harmful, so cells have a way to detect and degrade these transcripts. PTCs can arise at the DNA level from mutations in the gene sequence of the genome or due to insertions or deletions that lead to frameshift mutations in genes. At the RNA level, PTC transcripts can be a product of alternative splicing mutations or transcriptional errors (Nicholson and Muhlemann 2010). NMD prevents production of abnormal truncated proteins (Bhuvanagiri et al. 2010). NMD also functions to protect cells from errors that may occur during transcription and splicing that may lead to deleterious effects.

NMD has been found in animals as well as plants (Hwang and Maquat 2011; Palusa and Reddy 2010). In animals, NMD functions to protect many heterozygous carriers of defective genes that encode PTCs, thus preventing the expression of dominantly inherited disorders. A genome-wide tiling array study done with mutants of Arabidopsis UPF1 and UPF3, which are key NMD factors, has confirmed that NMD plays a key role in the genome-wide suppression of aberrant transcripts, further supporting NMD's role as a key regulator in plants (Kurihara et al. 2009). The basic mechanism of NMD relies upon many factors, but there are three critical proteins that

appear to be conserved across animals and plants. These proteins are termed the UP-frameshift (UPF) proteins (UPF1, UPF2, and UPF3), and are considered to be core components of the NMD surveillance system needed for proper NMD function (Figure 3.1). There are other proteins that are required for successful NMD surveillance and these are called the suppressor of morphological defects on genitalia proteins (SMGs) (Anderson et al. 1999a; Hodgkin et al. 1989; Pulak and Anderson 1993). The role of these proteins is to determine the phosphorylation status of UPF1. SMG1 is responsible for catalyzing phosphorylation of UPF1 at several serine residues (Ohno et al. 2006; Yamashita et al. 2001). SMG5, SMG6, and SMG7 provide the opposite effect and are responsible for the dephosphorylation of UPF1 (Anderson et al. 1999b).

NMD can be divided into several steps. In the nucleus, the Exon Junction Complex (EJC), that includes UPF3, is deposited on the transcript during pre-mRNA splicing. The mRNA is then exported to the cytoplasm, where UPF2 binds UPF3, and then binds to pre-mRNA. As the ribosomes are moving along with transcript, producing nascent peptides, it may encounter a PTC. If this happens, UPF1 is then recruited, forming the SURF complex, thus terminating translation. The SURF complex is made up of SMGI, UPF1, eRF1 (eukaryotic releasing factor 1), and eRF3 (eukaryotic releasing factor 3) (Figure 3.1). The SMG1 factor then phosphorylates UPF1 after it is associated with EJC-bound UPF2, resulting in the repression of further round of translation initiation on the NMD target by binding to the translation initiation factor eIF3 of the 43S pre-initiation complex located at the translation initiation codon, which recruits mRNA decay activities (Kashima et al. 2006). The mRNA is then degraded by SMG5 and SMG7, which degrade the NMD target from either or both ends, and in addition SMG6 provides endonucleolytic activity (Nicholson et al. 2010). This process appears to be conserved throughout eukaryotic organisms (Amrani et al. 2004; Muhlemann 2008; Silva and

Romao 2009). When termination occurs "normally," the poly (A) binding protein (PABP) binds to the eukaryotic releasing factor 3 (eRF3), that then completes termination.

In addition, other mechanisms of NMD also exist. For instance, if an NMD *cis* element is present in the 3' UTR, eRF3 will fail to bind PABP and instead will interact with UPF1 (Amrani et al. 2004). UPF1 then recruits both UPF2 and 3 and thus forms the NMD complex, which is recruited to the PTC of the mRNA. This signals or "tags" the mRNA transcript for rapid degradation, as described in detail above. It has been found that in yeast and invertebrates long 3' UTRs are the primary NMD *cis* elements, and in contrast, animals have 3' UTR introns as the primary *cis* elements (Amrani et al. 2004; Gatfield et al. 2003; Le Hir et al. 2000; Longman et al. 2007; Muhlemann 2008; Nagy and Maquat 1998; Silva and Romao 2009). Recent results indicate that 3' UTR introns induce NMD in *Drosophila* (Sauliere et al. 2010). Although a great deal of research has been conducted on the NMD process for over 20 years, NMD still is not a fully understood process.

NMD is important in animals because it serves as a surveillance system to monitor the somatic-cell rearrangement and hypermutation of immunoglobin or T-cell receptor genes that are responsible for maintenance of the immune system (Bhuvanagiri et al. 2010; Nicholson and Muhlemann 2010). It has been found in zebrafish, mouse, and human experiments that NMD is an essential process for embryonic development. (Anastasaki et al. 2011; Azzalin and Lingner 2008; Bhuvanagiri et al. 2010; Blencowe et al. 2010; Gecz et al. 2007; Medghalchi et al. 2001; Porse et al. 2008; Wittkopp et al. 2009). For example, in human males, it has been found that when UPF1 or UPF3 factors of NMD are mutated, a mild to severe X-linked mental retardation phenotype paired with physical anomalies is observed in (Gecz et al. 2007). Another example is that in mice, when NMD is completely inhibited, embryos resorb soon after uterine

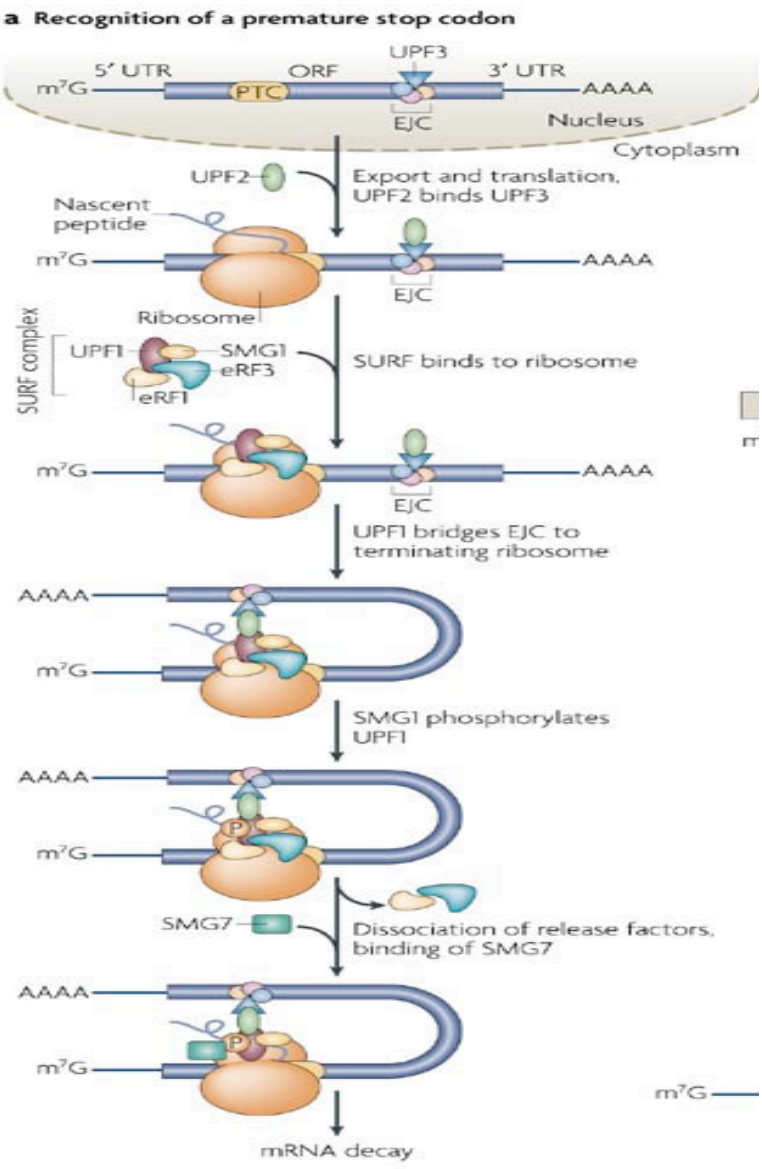



**Figure 3.1. Diagram of NMD Mechanism (Garneau et al. 2007).**

implantation and pre-plantation blastocysts undergo apoptosis in cell culture (Medghalchi et al. 2001). More recently, it has been found that in zebrafish, downregulating NMD results in a multitude of developmental abnormalities and delays that ultimately lead to increased mortality rates (Wittkopp et al. 2009). In plants, it appears that NMD also has a critical function, though the pathway has not been studied as extensively as it has been in animals. In UPF1 or UPF3 mutants, plants displayed multiple developmental abnormalities (Davies et al. 2006; Yoine et al. 2006a), severe seedling lethal phenotypes (Davies et al. 2006), and partial loss-of-function mutants for UPF1 displayed altered sugar responses in addition to the accumulation of mRNAs encoding for important transcription factors and metabolic enzymes (Yoine et al. 2006b). In this study, we have utilized *Chlamydomonas* as a model organism to investigate NMD. The reasons for this choice are outlined in the introduction of Chapter 2.

Alternative splicing (AS) is a process in which splicing of the coding regions can produce more than one transcript from the same gene. AS can also play a role in regulation of gene expression through processes such as regulated unproductive splicing and translation (RUST), and NMD. In our recently published paper, we analyzed 77 alternatively spliced genes in *Chlamydomonas* for PTCs, and found that 76 out of the 77 contained a PTC. In addition, AS analysis of the candidate genes described in Chapter 2 revealed the presence of several splice variants with a PTC. To date, the mechanism of NMD has not been investigated in *Chlamydomonas*. A search through the *Chlamydomonas* genome shows that there are UPF1, 2, and 3 proteins present, and in this study, we have shown that they share some homology with both plants and humans, indicating that the process of NMD may also be present in this organism. The presence of PTC-containing transcripts in *Chlamydomonas* together with the presence of UPF proteins points to a possible NMD mechanism which may also be

present in protists (Reddy et al. 2010). This led to us to investigate the mechanism of NMD in this organism using artificial miRNA silencing to see if the process is similar to that in other multicellular organisms, possibly providing insight into the evolution of the mRNA degradation process that is so critical for cell maintenance and viability.

# MATERIALS AND METHODS

## Cultures and Strains

The arginine deficient strain cc425 was obtained from the Chlamydomonas Center culture collection at Duke University. These were then used to inoculate 75ml Erlenmeyer flasks containing autoclaved TAP + Arginine medium (Harris 2009). The cells were maintained at 22℃ on a shaking platform in a growth chamber on a 12:12 light/dark cycle. Cells were subcultured during log phase at a starting density of $1x10^5$ cells/mL (Harris 2009). In order to obtain cells for RNA isolation, a 2ml aliquot was collected during log phase in 2ml tubes and centrifuged at 0.2 g for 2 minutes. Supernatant was removed and the procedure repeated until 4-6mls made up a pellet. The pellet was then frozen immediately in liquid $N_2$ and stored at -20℃ until RNA isolation was performed.

## Design of artificial miRNAs

Artificial microRNA targeted to UPF1 and UPF3 were designed utilizing the software provided by Molnar et al 2009. A BLAST screen was performed for genes of interest using Arabidopsis UPFs and the sequence obtained was then placed into the DESIGNER provided by Molnar et al. 2009 at http://wmd2.weigelworld.org/cgi-bin/mirnatools.pl?page=3. This then gave an output of potential artificial miRNAs targeting the gene of interest. We chose the best potential candidates for each gene and then used the Oligo Design provided by Molnar et al 2009 at http://wmd2.weigelworld.org/cgi-bin/mirnatools.pl?page=4 to design our oligos for each gene. The target sequences we choose were UPF1- 5' TTGGAAACGGAAGTCCGACAG 3' and for UPF3 – 5' TAACGAAATTGATGTAGGCAC 3'. We then used these sequences to obtain long ds oligos from IDT for each gene.

Oligos for UPF1 were FW-5' ctagtCTGTCGGACTTCCGTTACCAAtctcgctgatcggcaccatgggggtggtggtgatcagcgctaTTGGAAACGGAAGTCCGACAGg 3' and RW - 5' ctagcCTGTCGGACTTCCGTTTCCAAtagcgctgatcaccaccacccccatggtgccgatcagcgagaTTGGTAACGGAAGTCCGACAGa 3'. The upper case letters in these sequences show the target sequences, and the lower case letters represent the sequences necessary for producing miRNAs. Oligos for UPF3 were FW-amiFor_mm 5' ctagtGTGCCTACATCAATTTGGTTAtctcgctgatcggcaccatgggggtggtggtgatcagcgctaTAACGAAATTGATGTAGGCACg 3' and RW- amiRev_mm 5' ctagcGTGCCTACATCAATTTCGTTAtagcgctgatcaccaccacccccatggtgccgatcagcgagaTAACCAAATTGATGTAGGCACa 3'. Upon receipt of these oligos, their concentration was adjusted with sterile water to 100uM.

**Oligo Annealing**

We followed Molnar et al 2009 for oligo annealing. We mixed 20µl 2x annealing buffer (20 µM Tris pH 8.0, 100 µM NaCl, 2 µM EDTA) with 10 µl each FW and RW dsDNA. This mixture was then boiled in a hot water bath for 5 min, then allowed to set overnight at RT for cooling. The dsDNA was than purified using the Fermentas GeneJet PCR purification kit. The DNA was eluted with 40µl of elution buffer and this was then repeated with the flow-through from the purification column. The DNA was then quantified using the NanoDropper.

**Oligo Phosphorylation**

To phosphorylate the ends, 7µl of the dsDNA was dried using the Speed-Vac (Savant SC110) for 0.7 µg/ul concentration. To this was added 2µl of 5x Invitrogen T4 DNA ligase buffer, 1µl T4 PNK, and 7µl sterile water. This mixture was than incubated for 30 minutes at 37°C, and then 20 minutes at 65°C for PNK inactivation.

**Vector Cloning**

We obtained plasmid pChlamiRNAi2 vector from Chlamy.org and streaked out on LB + CARB plate overnight at 37°C. The next day colonies were picked from the plates to inoculate a 5 mL culture and grown overnight at 37℃. Plasmid isolation was performed using the QIAprep Spin Miniprep kit (Qiagen). Plasmid digestion was then done with 1 µl SpeI enzyme, 5 µl enzyme buffer, 33.5 µl sterile water, and 10 µl plasmid. The mixture was then held at 37℃ for 4 hours. The enzyme was heat inactivated at 65℃ for 20 minutes, and then the mixture was stored at -20℃. An aliquot of this mixture was then run on a 1% gel to check digestion. The digested plasmid was then purified using the Gene JET Gel Extraction kit (Fermentas).

**De-Phosphorylation of Digested Vector**

Purified digested plasmid vectors were de-phosphorylated using the Antarctic enzyme de-phosphatase. A 60 µl reaction was prepared using 45 µl plasmid, 2 µl enzyme, 6 µl enzyme buffer, and 7 µl sterile water. This mixture was incubated at 37°C for 3 hours, then heat inactivated for 5 minutes at 65°C. An aliquot of the mixture was then run on a 1% agarose gel @ 40V for 3 hours to separate bands. The bands were extracted using the Gene JET Gel Extraction kit (Fermentas). After gel purification the bands were even further purified using the Gene JET PCR purification kit.

**DsDNA/Vector Ligation**

Ligation of the dsDNA with the vector was performed using either a 10x dilution or a 100x dilution of the dsDNA. Each reaction was set up with 2 µl 5x buffer, 1 µl T4 ligase, 1 µl plasmid vector, 1 µl of 100x or 10x dsDNA, and 5 µl sterile water. The ligations were incubated at RT for 1 hour and then incubated for 18 hours at 14°C. A small aliquot of the ligation was taken after 1 hour and ran on a 2% agarose gel at 100V to ensure that ligation was taking place.

### *E.Coli* Transformation by Electroporation

Recombinant plasmid constructs were introduced into electrocompetent DH5α cells. The reaction was set up with 25 µl cells, 2 µl DNA from each ligation, and then electroporated. These were then electroporated at 300V and the mixture was incubated with 800 µl SOC media at 37°C for 1 hour. The samples were then centrifuged and the pellet was re-suspended and the cells were plated on LB+CARB(50) plates. These were then left at 37°C overnight. Colonies were screened by colony PCR to check for the insert. The PCR reaction was set up with 5µl GO-TAQ, 0.1 µl of 100 µM FW primer 5'-GGTGTTGGGTCGGTGTTTTTG-3', 0.1 µl of 100 µM RW primer 5'-TAGCGCTGATCACCACCACCC-3', 2.8 µl sterile water, and 2 ul of bacteria mixture from the colonies. This was than put for PCR following the parameters given from (Molnar et al. 2009). The correct colonies were identified and then DNA was extracted from a 5ml overnight culture using the Qiagen Miniprep kit and verified orientation. Sequencing of the clones was done by the CSU Macromolecular Facility using the specifications described by (Molnar et al. 2009).

### Chlamydomonas Transformation

Transformation of cc425 *Chlamydomonas* strain with the amiRNA constucts was accomplished using the glass bead method. Cells were grown to 1-2 x $10^6$/ml and then centrifuged at 5 x 100g for 5 minutes in 50 ml conical tubes after counting. Supernatant was discarded and cells were re-suspended in 1/100th volume TAP(-ARG) leaving enough for 300 µl per transformation reaction. 1.5 ml tubes with glass beads were assembled and autoclaved as described in (Kindle 1990). The tubes were vortexed for 20 seconds and then the cells were pipetted onto TAP plates. The plates were then sealed with Parafilm and placed in a growth chamber with a 12:12 light:dark cycle at 22°C. Colonies were than verified by colony PCR in which a small amount of a colony

picked with a sterile toothpick and placed in 20µl sterile water. This was freeze/thawed with liquid nitrogen 2x and then 2µl of this was aliquoted for PCR. The PCR reaction was set up with 5µl GO-TAQ, 2.8µl sterile water, 0.1µl 100mM FW primer 5'-GGTGTTGGGTCGGTGTTTTTG-3' , 0.1 µl 100mM RW primer 5'-TAGCGCTGATCACCACCACCC-3', and 2µl of colony mixture. The PCR conditions were 96°C for 45 seconds, 96°C for 10 seconds, and 60°C for 4 minutes for a total of 25 cycles. The products were ran on a 2% gel at 100V and were positively identified at 182 bp.

**Culturing Chlamydomonas Transformants**

Putative transformants were used to inoculate 75ml Erlenmeyer flasks of autoclaved TAP media (Harris 2009) with cotton stoppers. The cells were maintained at 25°C on a shaking platform in a growth chamber on a 12:12 light/dark cycle. Cells were subcultured during log phase at a starting density of $1x10^5$ cells/mL (Harris 2009). In order to obtain a pellet for RNA isolation, a 2ml aliquot was collected during log phase in 2ml tubes and centrifuged at 0.2 g for 2 minutes. Supernatant was removed and the procedure repeated until 4-6mls made up a pellet. The pellet was than stored at -20°C until RNA isolation.

**RNA Isolation**

RNA was isolated from transformants using the TRIZOL reagent. Two to three ml of cells were pelleted in conical 1.5 ml tubes. 1ml of TRIZOL was added per 100 ul cells and the solution was resuspended and pipetted onto liquid nitrogen-cooled mortar for tissue grinding. The cells were ground with liquid N2 to a fine powder and then placed back into a 1.5ml tube. To this was added 800 µl TRIZOL and then the tube was incubated at RT for 5 minutes after shaking to homogenize. 200 µl chloroform was than added to the reaction and it was vortexed at low speed for 15 seconds. The reaction was then incubated at RT for 2-3 minutes. The tubes were then centrifuged at

13,000rpm for 15 minutes at 4°C. The aqueous phase was then extracted and placed into a new tube, and the chloroform extraction was repeated twice. Then 500 µl of isopropyl alcohol was added to the remaining aqueous phase and incubated at RT for 10 minutes. The reaction was centrifuged at maximum speed for 10 minutes at 4°C. The supernatant was removed and the pellet was washed with 500µl 95% EtOH and centrifuged at full speed for 10 minutes at 4°C. The RNA pellet was dried using Speed-Vac (Savant SC110) and 30 µl of RNAse-free water was added to resuspend. RNA amount was quantified spectrophotometrically at 260nm. The RNA sample was treated with DNase I according to the manufacturer's instructions (Invitrogen).

**cDNA Synthesis**

See methods section in Chapter 2 for constructing cDNA.

**PCR of miRNA expression**

Each transformed line was checked for miRNA expression by PCR . The reaction was set up using 1/20 of the first strand cDNA in a reaction volume of 25 µl. The forward primer was the same primer used previously for colony PCR and the reverse primer was designed based on the dsDNA sequence using the Primer3 Input (http://frodo.wi.mit.edu/) software. The PCR reaction was set up with 12.5 µl of Takara EmeraldAmp HS enzyme, 0.5ul of 10uM FW- 5'-GGTGTTGGGTCGGTGTTTTTG-3', 0.5 µl of 10µM RW – (need to add), 10.5 µl sterile water, and 1 µl cDNA. PCR conditions consisted of an extended hot start at 94°C followed by TD-PCR in the 50-60°C range for 30 cycles. The products were then separated on a 1% gel at 100V to give the product at 339bp.

**PCR of UPF1 and UPF3 transcripts**

One-twentieth of the first-strand cDNA was used for PCR amplification in a reaction volume of 20 µl. The primers were designed using the Primer3 Input (http://frodo.wi.mit.edu/) software. The control primers for TUA1 were designed according to Bisova et al 2005.

## RESULTS AND DISCUSSION

NMD is a complex but important process required proper expression of genes and for the survival and development of most eukaryotic organisms (Amrani et al. 2004; Muhlemann 2008; Silva and Romao 2009). Currently, it is an ongoing area of research in plants and animals, but has yet to be investigated in green algae, that share a common ancestor with the plant lineage. Although some NMD work has been done with plants, it cannot compare to the large breadth of studies conducted on NMD in animals, and *Chlamydomonas* provides a bridge between these two groups, possibly offering some insight into the evolution of this process. So far no studies have been performed with simple unicellular photosynthetic organisms. AS analysis indicates that many splice variants in *Chlamydomonas* have PTCs (Labradorf et al., 2010). In our recent study, we found through computational analysis that most genes that undergo at least one AS event also harbor a PTC. We verified this with experimental work, thereby further suporting a role for NMD in regulating gene expression in *Chlamydomonas* (Reddy et al. 2010). The presence of UPFs found in the sequenced *Chlamydomonas* genome further suggests that NMD may be an operational cellular process in this organism.

The key regulating factors in NMD are the UPFs. Previous studies in both plants and animals have shown that NMD is halted when UPF1, or UPF3 are mutated (Avery et al. 2011; Yoine et al. 2006a; Yoine et al. 2006b). In our investigation of the NMD pathway, we studied UPF1 and UPF3 due to their role as key-limiting factors for NMD in both plants and animals. Their role in gene regulation allows us to gain understanding of NMD on a global level via their downregulation in *Chlamydomonas*. Studying these genes in *Chlamydomonas* allows us to what, if any role, the UPF genes play in the process of NMD and in green algae.

UPF1 and UPF3 were identified by a BLAST search of the *Chlamydomonas* genome using the sequences of Arabidopsis UPF1 and UPF3. We found that *Chlamydomonas* has UPF1, UPF2, and UPF3 proteins. An alignment was conducted with each of these and their orthologs in humans and Arabidopsis, and it was found that UPF1 and UPF3 showed some sequence similarity while UPF2 showed much less. Because of this, we decided to concentrate our investigation on UPF1 and UPF3 in *Chlamydomonas*.

To identify sequence similarities of *Chlamydomonas* NMD factors UPF1, UPF2, and UPF3 with other organisms, an alignment was conducted utilizing the program CLUSTAL 2.1 comparing each of these proteins in *Chlamydomonas* to humans and Arabidopsis, to address the dual nature of *Chlamydomonas* genome makeup. Results can be seen in Figure 2 for UPF1 and UPF3. When UPF2 was aligned, there was little sequence similarity between the three organisms, so UPF2 was not pursued further. This could be due to an annotation problem, which we have encountered many times throughout our study, or it could be that UPF2 has diverged so much in these two lineages that UPF2 no longer has no function or has acquired a different role in the cellular processes of *Chlamydomonas.* In contrast, alignment for Chalmydomonas UPF1 shows 65% identity to Arabidopsis, and a 59% similarity to humans, which could indicate a shared function of UPF1 (Figure 3.2A). Percent identity was calculated for every pair of sequences aligned. The scores are the number of identities between the two sequences, divided by the length of the alignment, represented by a percentage. In regard to UPF3 alignment, not as much similarity was observed, with *Chlamydomonas* showing only a 21% identity to Arabidopsis, and only 15% identity to humans. However, there were some areas that were common to all three organisms so it seemed worthwhile to investigate UPF3 further (Figure 3.2B). The sequences of the genomic

## A) UPF1

```
CLUSTAL 2.1 multiple sequence alignment


Chlamydomonas      ------------------------------------------------------------
Arabidopis         ASQPDTVADEYTFLEFNTQGDSEFDYQDFGSPTAWPTPSDSISIADVADRGEGGAAADHH 60
Human              -MSVEAYGPSSQTLTFLDTEEAELLGADTQGSEFEFTDFTLPSQTQTPPGGPGGPGGGGA 59


Chlamydomonas      ------------------------------------------------------------
Arabidopis         SEASSPSSLSAGAGNGAKVGRGGVGGSGGVSSSSQVDALAAGVGNLNFEETGDDDGFDYG 120
Human              GSPGGAGAGAAAGQLDAQVGPEGILQNGAVDDS--VAKTSQLLAELNFEEDEED---TYY 114


Chlamydomonas      ---------SYCGIHNPSCVVKC--LSTNKWFCNGRVHGTGSCIILHLVKSKNKEVQLHR 49
Arabidopis         KNDFTEHACKYCGISNPACVVRCNVASCRKWFCNSRGNTSGSHIVNHLVRAKHKEVCLHR 180
Human              TKDLPIHACSYCGIHDPACVVYCN--TSKKWFCNGRGNTSGSHIVNHLVRAKCKEVTLHK 172
                             .**** :*:*** *   : .*****.* : :** *: ***::* *** **:

Chlamydomonas      DSPLGDTVLECYASGTRNLFVLGFVPVKSENTVVLLARDTPPNHPTIRDLNLDLSQWQPI 109
Arabidopis         DSPLGETILECYNCGCRNVFLLGFISAKTDSVVVLLCRDPCLNVNALKDMNWDLSQWCPL 240
Human              DGPLGETVLECYNCGCRNVFLLGFIPAKADSVVVLLCRQPCASQSSLKDINWDSSQWQPL 232
                   *.***:*:**** .* **:*:***:..*::..****.*:.  .  :::*:* * *** *:

Chlamydomonas      VEDRGLVPWLVKQPTEPELLRARHLKLDQINKLEEMWKTKPAAGVDDLDEKATVEGDGVT 169
Arabidopis         IDDRCFLPWLVKVPSEQEQLRARQISAQQINKIEELWKTNPDATLEDLEKPG---VDDEP 297
Human              IQDRCFLSWLVKIPSEQEQLRARQITAQQINKLEELWKENPSATLEDLEKPG---VDEEP 289
                   ::** ::.**** *:* * ****::. :****:**:** :* * ::**:: .    *  .

Chlamydomonas      QPVTLKYEDSAQYQAVFEPLVKLEADYDRSIKESQSRDDITLRWDWGLNAKRVAYFYFPR 229
Arabidopis         QPVQPKYEDAYQYQNVFAPLIKLEADYDKMMKESQSKENLTVRWDIGLNKKRVAYFVFPK 357
Human              QHVLLRYEDAYQYQNIFGPLVKLEADYDKKLKESQTQDNITVRWDLGLNKKRIAYFTLPK 349
                   * *  :***: *** :* **:*******: :****:::::*:*** *** **:*** :*:

Chlamydomonas      DDNELKLMQGDELKLRHKNASNRGAWEALGHVLTYQQ--SEEVCLELFTNVSDVPEDCTV 287
Arabidopis         EENELRLVPGDELRLRYSGDAVHPSWQSVGHVIKLTA--QEEVALELRANQG-VPIDVNH 414
Human              TDSDMRLMQGDEICLRYKGD-LAPLWKGIGHVIKVPDNYGDEIAIELRSSVG-APVEVTH 407
                    :.:::*: ***: **:..      *:.:***:.      :*:.:** :. . .* : .

Chlamydomonas      GFSVDFVWRGTSFDRMRNALNTFRKYSASISGYLYHLILGHPVESVTLKIPLPKAGLGVP 347
Arabidopis         GFSVDFVWKSTSFDRMQGAMKNFAVDETSVSGYIYHQLLGHEVEAQMVRNTLPRR-FGVP 473
Human              NFQVDFVWKSTSFDRMQSALKTFAVDETSVSGYIYHKLLGHEVEDVITKCQLPKR-FTAQ 466
                   .*.*****:.******:.*::.*   .:*:***:** :*** **    :  **:  : .

Chlamydomonas      SLPELNHSQLHAVKSVLQQPLSLIQGPPGTGKTVTSAAIVYHLAHSGTGQVLVAAPSNVA 407
Arabidopis         GLPELNASQVNAVKSVLQKPISLIQGPPGTGKTVTSAAIVYHMAKQGQGQVLVCAPSNVA 533
Human              GLPDLNHSQVYAVKTVLQRPLSLIQGPPGTGKTVTSATIVYHLARQGNGPVLVCAPSNIA 526
                   .**:** **: ***:***:*:****************:****:*:.* * ***.****:*

Chlamydomonas      VDQLAHKMDQTGLKVVRLCAKTREAVASPVEHLTLHYQVHGCVLVGR--LRKLLALRGAQ 465
Arabidopis         VDQLAEKISATGLKVVRLCAKSREAVSSPVEYLTLHYQVRHLDTSEKSELHKLQQLKDEQ 593
Human              VDQLTEKIHQTGLKVVRLCAKSREAIDSPVSFLALHNQIRNMDSMPE--LQKLQQLKDET 584
                   ****:.*:  ***********:***: ***..*:** *::      .  *:**  *:.

Chlamydomonas      GGLNASDEKELKSLRRRLEMEVLENADVVCTTCVGAGDPRLSHFRFQHVLIDESTQAAEP 525
Arabidopis         GELSSSDEKKYKNLKRATEREITQSADVICCTCVGAADLRLSNFRFRQVLIDESTQATEP 653
Human              GELSSADEKRYRALKRTAERELLMNADVICCTCVGAGDPRLAKMQFRSILIDESTQATEP 644
                   * *.::***. : *:*  * *:  .***:* *****.* **::::*: :*********:**

Chlamydomonas      ECLIPMVLGAKQVILVGDHCQLGPVIMCKKAAEAGLCQSLFERLRLLGVKPIRLQVQYRM 585
Arabidopis         ECLIPLVLGVKQVVLVGDHCQLGPVIMCKKAARAGLAQSLFERLVTLGIKPIRLQVQYRM 713
Human              ECMVPVVLGAKQLILVGDHCQLGPVVMCKKAAKAGLSQSLFERLVVLGIRPIRLQVQYRM 704
```

```
                    **::*:***.**::***********:******.***.*******  **::**********

Chlamydomonas       HPCLSEFPSNTFYEGTLQNGTGMGERRLAGVDFPWPNPDKPMMFWVQLGAEEISASSTSY 645
Arabidopis          HPALSEFPSNSFYEGTLQNGVTIIERQTTGIDFPWPVPNRPMFFYVQLGQEEISASGTSY 773
Human               HPALSAFPSNIFYEGSLQNGVTAADRVKKGFDFQWPQPDKPMFFYVTQGQEEIASSGTSY 764
                    **.** **** ****:****.   :*   *.** ** *::**:*:*  * ***::*.***

Chlamydomonas       LNRTEAAAVEKVVTRFLQNGMSPAQIGVITPYEGQRAHVVSVMVRNGTARQDLYKEIEVS 705
Arabidopis          LNRTEAANVEKLVTAFLKSGVVPSQIGVITPYEGQRAYIVNYMARNGSLRQQLYKEIEVA 833
Human               LNRTEAANVEKITTKLLKAGAKPDQIGIITPYEGQRSYLVQYMQFSGSLHTKLYQEVEIA 824
                    ******* ***:.* :*: *  * ***:********:::*. *  .*: : .**:*:*::

Chlamydomonas       SVDAFQGREKDIIVLSCVRSNEHSSIGFLSDPRRLNVALTRARFGLVVLGNPRVLSRQPL 765
Arabidopis          SVDSFQGREKDYIILSCVRSNEHQGIGFLNDPRRLNVALTRARYGIVILGNPKVLSKQPL 893
Human               SVDAFQGREKDFIILSCVRANEHQGIGFLNDPRRLNVALTRARYGVIIVGNPKALSKQPL 884
                    ***:******* *:*****:***..****.*************:*::::***:.**:***

Chlamydomonas       WNSLLQYFKEHGCLAEGPLTNLKASMVQLHKPKRV------------------------- 800
Arabidopis          WNGLLTHYKEHECLVEGPLNNLKQSMVQFQKPRKIYNDRRLFYGGGAGMIGNDNFGSGNP 953
Human               WNHLLNYYKEQKVLVEGPLNNLRESLMQFSKPRKLVNTINPGARFMTTAMYDAREAIIPG 944
                    ** ** ::**:  *.****.**: *::*: **:::

Chlamydomonas       ------------------------------------------------------------
Arabidopis          NADRRGSRGRAGGSYLPSGPPNGARPG--LHPAGYPIPRVPLSPFPGGPPSQPYAIPTRG 1011
Human               SVYDRSSQGRPSSMYFQTHDQIGMISAGPSHVAAMNIPIPFNLVMPPMPPPGYFGQAN-- 1002

Chlamydomonas       ------------------------------------------------------------
Arabidopis          PVGAVPHAPQPGNHGFGAGRGTSVGGHLPHQQATQHNVGTIGPSLNFPLDSPNSQPSPGG 1071
Human               --------------GPAAGRGTPKG-KTGRGGRQKNRFGLPGPSQTNLPNSQASQDVASQ 1047

Chlamydomonas       ------------------------------------------------------------
Arabidopis          PLSQPGYGSQAFRDGFSMGGISQDFLADDIKSQGSHDPYNMADFATQASPGGFAVDYATQ 1131
Human               P--------------FSQGALTQGYISMSQPSQMSQPGLSQPELSQDSYLG--------- 1084

Chlamydomonas       ------------------------------------------------------------
Arabidopis          GAHGAFPGNFMNQNSQGGYSRFSGINDFMSQEYMAHGGQGLFTQAGFIDSSQDDGQQNPY 1191
Human               -------------------DEFKSQIDVALSQDSTYQGERAYQHGGVTGLSQY------- 1118

Chlamydomonas       --------------------------------------------------
Arabidopis          GVNNPNLQSQGLPNSLYSQPFAHYNTQPLNLSGPQQSQPNQSSQNPKHPYNG 1243
Human               --------------------------------------------------
```

**B) UPF3**

```
CLUSTAL 2.1 multiple sequence alignment

Human             MKEEKEHR---------PKEKRVTLLTPAGATGSGGG-----TSGDSSKGEDKQDRNKE-- 45
Arabidopsis       MKEPLQKKKVVVRHLPPSLSQSDLLSQIDPRFADRYNWVSFRPGKSSYKNQKYSRAYVSF 60
Chlamydomonas     MKPPRQKTKVLVRKLPPAMSEDTFKSVLDSVAAGRYNWLSYYAGKVSLKRVASSRAYINF 60
                  **   ::         *  ..  : :  ..  :.        .*. *  .   .*

Human             --------------------KKEALSKVVIRRLP----PTLTKEQLQEHLQPMPEHDYFE 81
Arabidopsis       KAPEDVYEFAAFFNGHVFVNEKGAQFKAIVEYAPSQRVPKPSDKKDPREGSISKDPDYLE 120
Chlamydomonas     VSEEDVYNFKQRFDGHVFISRQGNQYRCAVEYAPLQKMPTLEAKPHPLEGTIDQGGGGKG 120
                                     .:    :  :.  *    *.   :    .       .

Human             FFSNDTSLYPHMYARAYINFKNQED--------------IILFRDR------FDGYVFLD 121
Arabidopsis       FLKVIAQPVENLPSAEIQLERREAEQSGASKAAPIVTPLMEFIRQKRATVMGPQGLSDIR 180
Chlamydomonas     RREQSGRQGKGQQAPAGAEAEAATAGRGK---------ASGKNGKAAAAAAAAAAAVSLE 171
                    .          :      .                          .         .   :

Human             NKGQEYPAIVEFAPFQKAAKKKTKK-------------RDTKVGTIDDDPEYRK------ 162
Arabidopsis       RGGRRTRVVSANKPSPRPSKRNSEKKKYVEKESSKNVPRKTTADVSSSKPDYRQSNSSGK 240
Chlamydomonas     EAGHEAHEPRERGAGRHGKGREREP-------------RGAPAAAAADADATAAG----- 213
                  . *:.        .  :   :: :              * : . .  .

Human             ------------------------------FLESYATDN-EKMTSTPETLLEEIEAKNRE 191
Arabidopsis       ELPGNETAAIIDSSPPGIALTMDSGKKKILLLRSKDRDNPDNPPPQPEQHIDTNLSRNST 300
Chlamydomonas     ------------------------------PSSSSRGERGAKGERSGKGARNAKAAAVAS 243
                                                 *   :.  :     :   :   :

Human             ------------------LIAKKT--TPLLS--FLKNKQRMREEKREDF----------- 218
Arabidopsis       DSRQNQKSDVGGRLIKGILLRNDS--RPSQSSTFVQSEQRVEPSEAENYKRPSRPANTRA 358
Chlamydomonas     ------------------LLGDDEGPAPSQQQPLVVRPQPTAAPKLLLMKGGARTTQVAA 285
                                   *: ..    *  .  ::   *     :

Human             ------------------------------------------------------------
Arabidopsis       GKDYHTSGTISEKQERRTRNKDRPDRVMWAPRRDGSEDQPLSSAGNNGEVKDRMFSQRSG 418
Chlamydomonas     QPADAEAGPAVTAAEKPGRKEPR------GHRQDSGAAQEPMYGTHKTVPACRLTRLHLY 339


Human             ------------------------------------------------------------
Arabidopsis       EVVNSSGGHTLENGSARHSSRRVGGRNRKEEVVIGEGKTSRRGSGGGPSSHEKQMWIQKP 478
Chlamydomonas     RALSSPYLFAFARGSCSTAYKTSCRRFSAHKP------------------------------ 371


Human             ----
Arabidopsis       SSGT 482
Chlamydomonas     ----
```

**Figure 3.2. Protein Alignment. Stars (*) indicates identical amino acids, Colons (:) indicate strongly similar amino acids, periods (.) indicate conservation between groups of weakly similar properties.**

UPF1 genomic clone

AGGCGGAGGCAGGGTAGACGGCGAGAGGATTACGGCGCGGTACGGAGCATCAGG
CTAATGTTGGGGGAGCAGACCCTTATGCCACAGCTACTTTCAAAGCCTCCCCGCCG
CTACCCCGGCTGACTGCACTCTCATCTGTGTGTCCGTGTGTGGGCGCGTGTGTATT
GCTTGTCACTATCGCTCCACATTTCTCCTTGCACAGTTACTGCGGCATCCACAACC
CCAGCTGCGTGGTCAAGTGCCTGAGCACCAACAAGTGGTTCTGCAACGGCCGCGT
GCACGGCACCGGCTCGTGCATCATCCTGCACCTGGTGTGTGGGCTTGCACGGGTA
ACCGGCTGGTAGCGTGGGTAGGAATGGAAGCCGCCGGTCAATCTTACCCATCCGT
ATTGTCGCAGGTTTGTTGCTAACACCTGACATTGACCTGTTATCAACGCACCGCAAA
CTTGCCGAGGACGCTGCATCCCTTCCCCTGGTCGCGGCGAACTCGAAACAGGTCA
AGAGCAAGAACAAGGAGGTGCAGCTGCACCGGGACTCGCCGCTGGGCGACACCG
TGCTTGAGTGCTATGCCAGCGGCACGCGCAACCTGTTCGTGCTGGGGTTCGTGCC
GGTCAAGAGCGAGAACACGGTGGTGCTGCTGGCGCGAGACACGCCGCCCAACCA
CCCCACCATCCGCGACCTCAATCTCGACCTGTCGCAGGTGCGGTGACAATCGGGT
CTGGGATCGGGGTAGGGACCTCAGCCTGGATACGATGGGGCGGAGAGGTTGACA
CAAACCCCAGGGTAGCTGGGATGCAGGCACAAGCGCATCTGAAGCGGAGGGATG
CGGGCACAAGTGCATCTGAAGCGGAGGGATGCGGGCACAAGTGCATCTGAAGCGT
TGCTGGGGCTCACCACGTCACTACTGCAAATACCCCCGTGTGCTAAAGTGTGCTTG
AGGCCGCTCAACCCCGCCGCCACCGCAGTGGCAGCCGATTGTGGAGGACCGCGG
CCTGGTGCCCTGGCTGGTGAAGCAGCCTACAGAGCCGGAGCTGCTGCGGGCGAG
GCACCTCAAGCTGGACCAGGTGGGCAGGGCAGCATTGGAAGGGGCGGGGCCGAT
CGAAGCTATTGTTTGGTAGCGGGGCGGCACACCGGTGCCGATACGAGGAAGTCTG
GGTTGCGCTGCGTTGCCTGGCCACAAACGATTCTAGTACTGTTGCTGGTGCTGCTG
CCGCGGTGCTGCTGCTTCTGCTGCTGGGCGGTGCAGCCGAGCTTGATCCTCCTGC
CGCGTGTGTTTTGCTGTTGCCCAGATCAACAAGCTGGAGGAGATGTGGAAGACCAA
GCCGGCGGCGGGTGTGGACGATCTGGATGAGAAGGCGACGGTGGAGGGGGACG
GCGTGACGCAGCCAGTGACGCTCAAGTACGAGGACTCGGCGCAGTACCAGGTAG
GTGCCTGCGGCAGCTGTAATGTGTGGTGTGTGTGTGTGTGTCTGGAAAAGACTC
ATCGCGTGTGATTACGGGCAACTGAACTAAGCGATTTGTGGTTGGTGCTTAATGAC
TGCCTCTGTGAAGAATTAGTTCGAAGATGTGTGGCTGCCAGTTCTCTTGTTTCGGCT
GCCGGGCGTCGTGAGCCGTTAGCATGATCCTTTCCTACACATCATACACAGGCCGT
GTTTGAGCCGCTGGTGAAGCTGGAGGCCGACTATGACCGCTCCATCAAGGAGAGC
CAGAGCCGCGATGACATCACGCTGCGGTGGGACTGGGGCCTCAACGCCAAGCGC
GTGGCGTACTTCTACTTCCCGCGCGACGACAACGAGCTCAAACTCATGCAGGTGC
GGGCTGCGGCTGCGGGGGAGGGCAAGTTGCTACAGGGGTCTGATACAGAATGCG
GCATTGCGATAGCATACAAGCGGACTGACATTCACATAAGTTGGTTTTGGATCAACA
GACGGCGGTGCCACGTGTTGCAACGTTTTTCCGTACCTGCCGAACCACTTCTTAAC
AGGGTGATGAGCTGAAGCTGCGTCACAAGAACGCGAGCAACCGCGGGGCGTGGG
AGGCGCTGGGCCATGTGCTCACATACCAGCAGAGCGAGGAGGTGTGCCTGGAGTT
GTTCACAAACGTGAGTGTGGCCCGCAAGTGCAGTGGCGCTATTCAATTGGCAGCG
CTTCATGGGTCTCCAGCCACCGCTCGTCGCCTCAGCAGCACACCGGTATTCGGCA
GTGCGAGCTTGAGCGTGCGTTCTTCGTTGATTGTTGAATATGGTGCAACACCCGGT
CAACCTCATGGACTTGTCCAAGAGCCAAGGGCCCCTGCCAGCCGCTGACCAATAT

GCGCCCGCCGCTCACGAACTAAACGTGTAGGATGTGCCCGAGGACTGCACTGTCG
GCTTCAGTGTGGACTTTGTGTGGCGCGGCACCAGCTTTGACCGCATGCGCAACGC
GCTCAACACCTTCCGCAAGTACTCCGCATCCATTTCGGGTGCGTGCCCCGGCATTG
AACCTGCACGGGCGGCTACGTCACTTGCGGACCGCTTCCTGACCCCTCATGTTCC
GGATCCCTCCACCCGTAACCAATGTGCCGGATGCGTTAACCTGCCTCCCCGAGTG
CCCCCAACGCGCCGCGTCCTTCCCCGCTCGTCATTCCTTCTCCACCGCCATCCAAC
TACGGCGTTCATGCCTGGAACCCCCGCCCCACCAGGCTACCTGTACCACCTCATTC
TGGGCCACCCGGTGGAGTCCGTGACGCTCAAGATCCCGCTGCCCAAGGCCGGTC
TGGGCGTGCCCTCGCTGCCCGAGCTCAACCACAGCCAGCTGCACGCCGTCAAGTC
CGTGCTGCAGCAGCCGCTGTCGCTCATCCAGGTGCGGTGGGGCCCGGGGGACTG
GAAGCTTCGCCGCGTGGGTTTGGAAAATGCCGCGGGTAAGCCGTATCCTGTCGCG
TGACTTACAGTAGTTCTGAGTTGCTGGGCTTGCTAGCATAATGCCTTGGCATGCATT
ATTAACCCCCGATTACACCTACACTTGCCGCCTATCCCGCCGTCAACATGCGTCAC
GTGCTCCACAGGGCCCTCCCGGCACGGGCAAGACGGTGACCAGTGCGGCCATCG
TGTACCACCTGGCGCACAGCGGCACGGGCCAGGTGCTGGTGGCGGCGCCCTCCA
ACGTGGCGGTGGACCAGCTGGCGCACAAGATGGACCAGACCGGGCTCAAGGTGG
TGCGCCTGTGCGCCAAGACGCGCGAGGCGGTGGCGTCGCCGGTGGAGCACCTCA
CGCTGCACTATCAGGTGCATGGCTGCGTGCTCGTGGGTGTCGTGAATGCGCATGC
CGTATGTAACACACGGCATAGGGGCGGAAGCGCGCTTACATCTTGCTGGCCCTGC
AACCTCCCGTTCTTCTGCCTGCTGTACCGTACCGTCATTCGGACTGCCCATCGCTC
TTTCACTGCGTGCGCCAGGTGACGCACATGGCTGTGCCTGAGGGCGAGCGGCTG
CGCAAGCTGCTGGCGCTGCGGGGCGCGCAGGGCGGCCTGAACGCCAGCGACGA
GAAGGAGCTCAAGAGCCTGCGGCGGCGGCTGGAGATGGAGGTGCTGGAGAACGC
CGACGTGGTGTGCACCACATGCGTGGGCGCAGGCGACCCGCGCCTGTCGCACTT
CCGTTTCCAGCACGTGAGTAGAGGGGTTGGGGCACGGGGCGGGAGACGCGGCTG
AGTGGCTGACGCTGGCCCCTCTGCAATTCTGCATCCTGGGTATGGTTGGGGCTCT
GGCGGCTGAAAGTATGATGTGCGCTTCAAATGTGCGGTGCAGGTGCTAATTGACG
AGAGCACCCAGGCCGCGGAGCCGGAGTGCCTCATTCCCATGGTGCTGGGAGCCA
AGCAGGTGCGTGTAGTGCTGAGCTTAGCACCGCATGGTCCTCCGTCTTGCGAGCA
GCGCATCTCATGACGCCTGCTCCCGCTTGCTTCTGTTGCCTCACTCCCGCCTGCGC
TCCCACTGCTCCCCACCCCCGCCCTCGGCCCTGCCCCTCCGCCGCCCCGCCCGC
AGGTGATCCTGGTGGGCGACCACTGCCAGCTGGGTCCGGTCATCATGTGCAAAAA
GGCGGCGGAGGCGGGGCTGTGCCAGTCGCTGTTCGAGCGCCTGCGCCTGCTGG
GCGTCAAGCCGATCCGCCTGCAGGTGGGAAGACTGGAAAGCCGGGCCGTGCGTG
GTGCCGAGCTGCAGACCTCCACAGCCACGATTTGCCAGGCCGCCGGGTTGGGCT
GGGCTGGGCTGGTAGCATGTCGACAACACTGCCTTTGCAATGTTGCCTTGGCGTA
CCACTGAGATCGCCCGCATCGCTGGCCTGCTCCGCCACCCTTCCCACAACCCGCA
GGTCCAGTACCGGATGCACCCCTGCCTGTCGGAGTTCCCCTCCAACACCTTCTAC
GAGGGCACGCTGCAGAACGGCACGGGCATGGGCGAGCGGCGGCTGGCGGGCGT
GGACTTCCCCTGGCCCAACCCCGACAAGCC**CATGATGTTCTGGGTGCAG**CTGGGC
GCGGAGGAGATCAGCGCCAGCAGTGAGAAGGGGCTGGGGGTAGCCGTCCGTGGA
CAACGCGAGCAAGGAACGGGGAGGCAGGGGTCCCAAACCATTCTTGAGTCTAGAG
GGAAGCCACTGCTGTTGGTGGTGCTGCCCTTATGGCTGGACTAGGCCGAAAGGGC
ACGGGAGCGGAAGCGCGGCACGCTGCCGATGCATCCGGTGTATGTAGTATGCTAC

CCGGCCAACAAGTGTACATTCTGCAGCGTTTTGACGTGGGTGGTATCGCCAATGCA
GGCACCTCGTACCTGAACCGCACGGAGGCGGCGGCAGTGGAGAAGGTGGTGACG
CGCTTCCTGCAGAACGGCATGAGCCCGGCCCAGATTGGCGTCATCACGCCGTACG
AGGGCCAGCGCGCGCACGTGGTGTCGGTCATGGTGCGCAACGGCACGGCCCGGC
AGGACCTGTACAAGGAGATTGAGGTGAGAGCAATTTGGTCGTGTGCTGCTTGTGC
GTGCGGGGCATGAGATTTGATAATGCTCTGTCCCCCAGACGCTCGTATTTCAAAGG
GCGCCCACAACATTGCTGCTGGCCTGTCTCGTGGCCAACTTCCTACCGCATCCTGA
TGGGAACACGTGCGCGCGTGTGCGTAGGTGTCGAGTGTGGACGCGTTCCAAGGC
CGCGAGAAGGACATCATCGTGCTGTCGTGCGTGCGCTCCAATGAGCACTCCTCCA
TCGGCTTTCTGTCCGACCCGCGCCGCCTCAACGTGGCGCTCACCCGTGCGCGCTT
CGGGTGAGTTGCTCTCTACCGTACTGATTCTTTCGACAGGCCAACTCCTTATGCCT
CATGCCTGCAATCCATGCCCTTGTCTAGAATTACGCCGCATGATATCGGTTTCGCTT
ACCGTAAATAACCGCCTAACCTCCCTCGGGGTTTTCCTCCCTCCACAGCCTGGTGG
TGCTGGGCAACCCGCGCGTGCTGTCGCGCCAGCCGCTGTGGAACAGCTTGCTGC
AGTACTTCAAGGAGCACGGCTGCCTGGCGGAGGGGCCGC**TCACAAACCTCAAAG**
**CCTCC**ATGGTGCAGCTGCACAAGCCCAAGCGGGTGAGGGCGGACGCCGCACCCA
CTCGGTCCGGCCCACATGTTTTGCCACCTGCTTCAAGGGGTTGGACGCTCTTACCG
TTGTACAACTAGGGATACTGAATATACCATGGCTACACCTTCCTTGCTGTGCAGGTG
TTTGACCGCATGTCGTTCGGGGTGGGTGCACTCACCACCAACCGCTTCCAGCCGC
CCGAGAAGGTCGGC

Green = Forward and Reverse primers
Blue = ForwardA and ReverseA Primers
Yellow = miRNA target sequence

>Chlamydomonas Upf1 Protein Sequence
SYCGIHNPSCVVKCLSTNKWFCNGRVHGTGSCIILHLVKSKNKEVQLHRDSPLGDTVLE
CYASGTRNLFVLGFVPVKSENTVVLLARDTPPNHPTIRDLNLDLSQWQPIVEDRGLVPW
LVKQPTEPELLRARHLKLDQINKLEEMWKTKPAAGVDDLDEKATVEGDGVTQPVTLKY
EDSAQYQAVFEPLVKLEADYDRSIKESQSRDDITLRWDWGLNAKRVAYFYFPRDDNEL
KLMQGDELKLRHKNASNRGAWEALGHVLTYQQSEEVCLELFTNVSDVPEDCTVGFSV
DFVWRGTSFDRMRNALNTFRKYSASISGYLYHLILGHPVESVTLKIPLPKAGLGVPSLPE
LNHSQLHAVKSVLQQPLSLIQGPPGTGKTVTSAAIVYHLAHSGTGQVLVAAPSNVAVDQ
LAHKMDQTGLKVVRLCAKTREAVASPVEHLTLHYQVHGCVLVGRLRKLLALRGAQGGL
NASDEKELKSLRRRLEMEVLENADVVCTTCVGAGDPRLSHFRFQHVLIDESTQAAEPE
CLIPMVLGAKQVILVGDHCQLGPVIMCKKAAEAGLCQSLFERLRLLGVKPIRLQVQYRM
HPCLSEFPSNTFYEGTLQNGTGMGERRLAGVDFPWPNPDKPMMFWVQLGAEEISASS
TSYLNRTEAAAVEKVVTRFLQNGMSPAQIGVITPYEGQRAHVVSVMVRNGTARQDLYK
EIEVSSVDAFQGREKDIIVLSCVRSNEHSSIGFLSDPRRLNVALTRARFGLVVLGNPRVL
SRQPLWNSLLQYFKEHGCLAEGPLTNLKASMVQLHKPKRV

Upf3 genomic clone sequence

CGGCTCAAACAACTCGCAACAGCCAGTTGCGAACTCTGTCATAGGATTAGCTGTTT

GGCTATCAGGCGCCTCCGGACTCAGAAGAGTTAGTGGGGCCGCATGGACTCCGAC
GCCATGTAAATGCTCATGCTCACAATTTTCTTCCAAGCGTACAGACGCCATACTCGT
TCATAGCACTTCGGTGACATTGCATTGCAAAGATGAAGCCCCCACGGCAGAAAAC
CAAAGTGCTGGTGCGCAAGCTGCCTCCTGCAATGTCAGAGGACACGTTCAAGAGC
GTCCTTGACAGCGTGGCAGCAGGACGCTATAATTGGCTGTCATATTATGCTGGTAA
AGTCAGGTGAGCAGCAGCACGGAGTTCAATGCAAACATCGCTCGGAAGCGGGAAC
GCCGCGCTCTGCGAGGCGTTTTGAATGCGTCAGTTTTTATCCTGCTGAATATACATT
GCTTTGACAAATGACATGCAGTCTGAAGCGCGTGGCGTCCTCTCGCGCCTACATCA
ATTTCGTTTCCGAGGAGGACGTCTACAACTTTAAGCAGCGCTTTGATGGACACGTTT
TTATCAGCCGGTAAGAGAGGTTTTGATGCATCAGGGCTCGGGGTTAGCGCCTATG
GACTTGCCCCAAGCCCGCAGCGCTGCAGACGCTCCCGTCCACCGTCATACGCTAC
GCCAACGCAGGCAGGGCAACCAGTACCGGTGCGCTGTGGAGTACGCGCCGCTGC
AGAAGATGCCCACGCTGGAGGCCAAGCCGCACCCGCTGGAGGGCACCATAGACC
AGGGTACGTACGTGGCATACACCGCCTGCACACCGGCCACCCTCTAGACGCTGTA
CTCACGCACACCTTCACTACCGGCATACGCGCAGATCCCGACTTCCTCGCATTTGC
GGAGGCCCTGGAGTCCGGCGCGCCCATGCCCGGCGGCATCACGCCCGCGGCCG
CCGTCGCGGCGCGCGCAACCGACCTGCCCACCGCCGCGGCCGCAGCCGCGGCC
GCGGCCGCGAACGGCGAGTCGGGCGGCCGCGTATCCCGCCTGATGGCCTACCTG
GCCTCAAAGTATGCGGAGGGCGGCAGCCGGCTGGGCGCCACGCGCGGCAGCAA
GGCCAAGCAGGGCCGGCAAGACAAGTCGGAGCTGTACCAGTCGGCACAGCAGGA
GAGCGGGGCCGAGGGCAGGGGCGGGCGGGCGGGGCGCGGCTCCAAGGCAAGC
GCGCAGGCCGCGGCGGCGGCAGCCAGCGCCGCCATTTCGGCAGTGGCGGGCAT
CCTGGGTGGGGAGGCAGGCGGCGGTGGGAAGGGCCGCCGCGAGCAGAGCGGGC
GACAGGGCAAGGGGCAGCAGGCGCCCGCCGGCGCGGAGGCGGAGGCGGCGAC
CGCCGGGAGGGGCAAAGCGTCCGGGAAGAATGGCAAGGCTGCGGCAGCTGCAG
CAGCTGCAGCGGCGGCCGTTTCGCTCGAGGAGGCGGGGCACGAGGCGCACGAG
CCGCGGGAGCGGGGTGCGGGGAGGCACGGCAAGGGGCGCGAGCGGGAGCCGC
GGGGTGCGCCGGCTGCAGCAGCAGCTGACGCCGACGCAACAGCTGCCGGGCCAT
CTTCGTCGTCCCGGGGGGAGCGTGGTGCTAAGGGGGAGCGCAGCGGGAAGGGC
GCCCGCAACGCCAAGGCGGCGGCGGTGGCTAGCCTGCTTGGTGACGACGAGGGG
CCCGCGCCGTCACAGCAGCAGCCGCTGGTGGTGCGGCCGCAGCCAACGGCGGCA
CCAAAACTGCTCCTGATGAAGGGCGGCGCGCGTACGACACAGGTGCGGGGAAGG
GGGCGGGCTAGAAGCTAAGGAACCGTGTGCATGACACATGTGCTGAATCTTCTTGT
AGGGAGAGGTATCGTCCCCACACTTGCCTGGGGCTCTTTCTCATGCCCGCAATGC
AGGTGGCCGCACAGCCGGCCGACGCAGAGGCCGGACCCGCCGTGACAGCCGCG
GAGAAGCCTGGGCGGAAGGAGCCGCGTGGTCACAGGCAGGACTCGGGCGCAGC
ACAGGAGCCGGTGGCGTCGGCGGCGTACCCGCAGCCGGAGAAAAAGCAGGTGAA
GCAGCAGCGGCCGCCAAAACAGGCGGCAGGAGGGGCGGCAGGAGGGGCAGCAG
CAGGGGTACCGGCGGGGCTGGCGGTGGCCGGGCCTGCCGACACTGCGGCCGCG
AGGCCGGCAGCACCTGAGCCGGCGAAGGGGAGGCCGGCGCGTGGCGCCTCCGC
AAAGGCGGCGGCCGCGCCTGCGTCTGCTGCGACTGCAGACCAAGCGGCACCGAC
GGCGTCAGGGAGGGGCGCGGGTAGCAGTGCAGCGGTGGCGGGCGGCGGCGGC
GGCGGCGGCACGGATGGCGACCTGGCGGCGGCGGCCGCTGCTGCCGCGGCGGC
GGCGATCGCCGCGTCGAGCGGCAAGAAGGTGCGGGCCGGCTTTCAGATGTATGT

GCCCGGCTCGCTGCGGAGGGGCGGAGGAGGTGGTGGCAGCGCGAACGGATGAA
GGGCGTCGGGAGCAAGTGGCTCTATGGAATACACACAGGCAGCACCACGGCACAT
CAGTGTGTCAGGGAACCCGTACAAAGACGGGTCGTACTGCCGGCCGTGGGCCTCA
GCGTACCGCACGCCGTGATCGCGCCAGCCTTGCGTGGAGGCATCGTAGCGGTTG
GCGCCAATGCACAGTGGGGCAGATGGCCGCTCGCCCTGATGCAAGACAGACGGG
GCCATGCAGCAGGCCAGTGCACGCACGCCGGGCTGCAACACAGCGATGCAGGCA
ATTATGTAAGGCTGCAACGGGCAGTGCGGTATTGTACGGTCGTGATTGCGGAGCAT
TAGCTTAGGGCAGGCCTGGCAGCGCTTGCTGGGGGACTAGCGTGGTCCAACCGG
GGATGCCTGGGTCGTTTGCTGGCAAGGAGCAGTGACATCCGAACCGGCTTCAGGA
GCGTCTACTTGGGAGCGATGAGCAGTTGTGCTGAGTGTTTGTGGTGATGTGCGGA
ATGACACCAGATGGATGCACCGCACAGAGTTGTGGACACTGCGTTGGCGTCTGCT
TGCGCGGCTGTAAGAAGACAGCCGAGCGACCCCCTCTGTCTTGCCCACATCCTAC
CTTGGGCCTAAGGCCAACAGCTAATAATCTCGACTGTCCGGCGCTGACAGCCTGT
GAGCCTGACAGCCATCGACAGCATGTTGTTTCAACGTGATCAATGGTCTAGCCATC
AGCACGGACGCGGCAGCCGGACGCGAAGAGCATCGAACAAGCCCCGAACCCGAA
TCAGACGCCACATACCGTGCTGCACACGTTCGGGTCAGACATCGTGGTGTCAGAC
AGATTCATGGTTATCACGGGTCATTTAACTCCCAATAATGCACCAACCGCCCTTCCT
CCTGCTCAGATGTATGGCACGCACAAGACCGTGCCCGCGTGCCGCCTCACTCGCC
TGCACCTGTACCGCGCCCTCTCCAGCCCATACTTGTTTGCCTTCGCGCGAGGATCC
TGCAGCACGGCCTACAAAACATCCTGCAGACGCT**TCTCCGCCCATAAACCATAG**TT
ACTACTACTATAAAATACACATCCTGCATGCCTTAAAAGAACATCCTGTGCACACCC
CGCGACGACGCCCAACCTGGGTTTAAAACACACGGCTATGTTACAAAAGCGTCTCA
CACTGCTCATGCACAGTCCACCCAAATCTCGCCCACCTCGTACGCCGCCACGCTCT
CGAACTCCGTGGCGCACCCGAACGACACC

Green = Forward/Reverse primers
Blue = ForwardA/ReverseA primers
Yellow = miRNA target sequence

Protein Sequence

Chlamydomonas UPF3
>MKPPRQKTKVLVRKLPPAMSEDTFKSVLDSVAAGRYNWLSYYAGKVSLKRVASSRAY
INFVSEEDVYNFKQRFDGHVFISRQGNQYRCAVEYAPLQKMPTLEAKPHPLEGTIDQG
GGGKGRREQSGRQGKGQQAPAGAEAEAATAGRGKASGKNGKAAAAAAAAAAAVSLE
EAGHEAHEPRERGAGRHGKGREREPRGAPAAAAADADATAAGPSSSSRGERGAKGE
RSGKGARNAKAAAVASLLGDDEGPAPSQQQPLVVRPQPTAAPKLLLMKGGARTTQVA
AQPADAEAGPAVTAAEKPGRKEPRGHRQDSGAAQEPMYGTHKTVPACRLTRLHLYRA
LSSPYLFAFARGSCSTAYKTSCRRFSAHKP*

Primers for amiRNA expression:

FWUPF1: 5’-ctagtCTGTCGGACTTCCGTT-3’

FWUPF3: 5’-ctagtGTGCCTACATCAATTT-3’

RW(both): 5’-TAGCGCTGATCACCACCACCC-3’

**Figure 3.3**. Sequences of each genomic clone with locations of primers and targets.

clones along with predicted protein sequences for UPF1 and UPF3 in *Chlamydomonas* are presented in Figure 3.2. To study NMD globally in *Chlamydomonas*, we utilized artificial miRNA (amiRNA) to knockdown expression of UPF1 and UPF3, the key regulators of NMD. There have been many other studies that have shown successful knockdown of these proteins, and subsequent inhibition of NMD. In animals, a recent study showed that knockdown of UPF1 and UPF2 in *Drosophila* resulted in inhibition of cell growth, and subsequent death in a UPF3-independent manner, due to the fly's inability to degrade aberrant transcripts (Avery et al. 2011). In plants, studies have shown that loss-of-function mutants in Arabidiopsis for UPF1 and UPF3 show accumulation of NMD transcripts that might otherwise be degraded, as well as multiple developmental abnormalities (Davies et al. 2006; Palusa and Reddy 2010).

There are many ways to suppress gene expression in *Chlamydomonas*. Utilizing artificial microRNA to suppress gene expression is a relatively new tool for investigating cellular processes in algae (Molnar et al. 2009). To date, only a handful of groups have utilized this technique in *Chlamydomonas* (Dymek et al. 2011; Godman et al. 2010; Manavella and Rubio-Somoza 2011; Schmollinger et al. 2010; Zhao et al. 2008). Traditional reverse-genetics methods have been highly successful in flowering plants (Alonso and Ecker 2006), but have yet to be developed fully for use in *Chlamydomonas*. It is difficult to perform techniques such as transposon tagging, insertional mutagenesis, and tilling because the entire genome must be saturated, and this requires large mutant populations, that are limited by the selectivity of mutational targeting. RNA silencing methods used for high-throughput analysis of gene function utilize a conserved cellular mechanism, that may have evolved as a response to viruses and transposons, and it has been adopted for endogenous gene regulation in a variety of eukaryotes (Baulcombe 2006). Small interfering RNAs (siRNAs) and microRNAs (miRNAs) are the

two main groups, which are part of the RNA silencing mechanism. Small interfering RNAs (siRNAs) are produced from double-stranded (ds) RNA by Dicer or Dicer-like enzymes, which are released as several ds intermediates ~21 nucleotides long, containing a two-nucleotide 3' overhang (Tuschl et al. 2001). Similarly miRNA intermediates are also released from Dicer, but they are released as 21-24 nt duplexes from a partly double-stranded region of an imperfectly matched foldback RNA (Ambros 2001). A 21-24-nt RNA duplex usually arises from the miRNA precursor, and in contrast, multiple forms of this molecule arise from siRNA precursors. In both the miRNA or siRNA pathways, strands with lower thermodynamic stability at their 5' ends are retained by an Argonaute (AGO) protein (Khvorova et al. 2003; Zamore et al. 2003) by way of a mechanism which is influenced by the 5' nucleotide (Mi et al. 2008). This complex is then guided to the target nucleic acid sequence for gene silencing through Watson-Crick base pairing with the bound small RNA, and the small RNA strand that is not incorporated into the AGO complex is rapidly degraded. Regulation of the target sequence occurs on both the transcriptional and post-transcriptional levels. The post-transcriptional level is much better understood than the transcriptional level, and its mechanisms have been used to develop many different methods for gene silencing. These processes can involve transcriptional arrest or targeted RNA degradation. This can happen as a result of mRNA destabilization or miRNA-guided cleavage (Bartel 2004). Destabilization occurs when small RNAs are only partial complementary to the target, and thus translational inhibition occurs. For cleavage, the small RNAs will have a complete or nearly complete match and thereby be more likely to direct mRNA cleavage. Animal miRNAs usually display complementarity to the target in a short region, which allows each miRNA to target many mRNAs, and thus not be very target specific (Bartel et al. 2005; Cohen et al. 2005; Lewis et al. 2005; Lim et al. 2005). On the other hand, plants are much more specific and have fewer mismatches to their targets, which

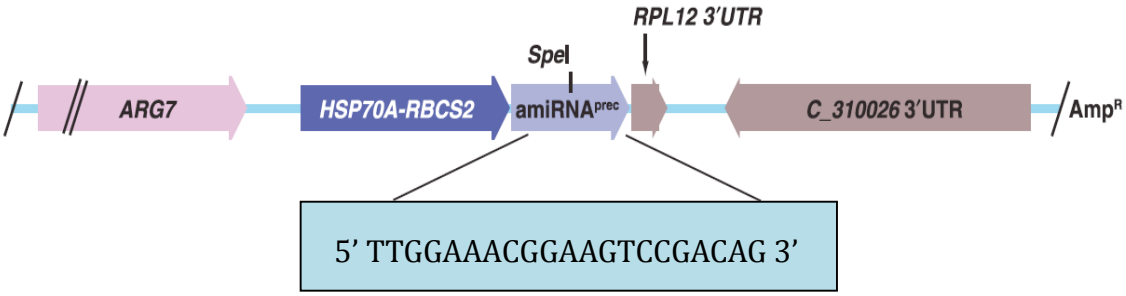
normally triggers transcript cleavage and eventual degradation of the mRNA targeted (Llave et al. 2002; Schwab et al. 2006).

Another approach for down-regulation of genes in *Chlamydomonas* is known as RNA interference or RNAi that is based on production of siRNAs targeted to a gene of interest. This mechanism is induced by expressing long dsRNA (Schroda 2006). This approach can be very effective, but sometimes the long transgenes are targets themselves for degradation by the transcriptional silencing mechanism, thus being unable to target in *trans* through the post-transcriptional mechanism (Rohr et al. 2004). In addition to this, there also can be off-target effects due to the fact that long ds pre-RNAs generate many siRNAs, and some of these may have the ability to effect genes that have no relation to the target of interest (Xu et al. 2006). Here we have chosen to utilize the approach of amiRNA to circumvent the problems associated with other methods of reverse genetics.

Although the amiRNA method is new to *Chlamydomonas*, there is no doubt to its degree of effectiveness as a more efficient way of silencing target genes. Recently, the amiRNA method has been utilized to gain insight into flagellar function, the physiological roles of hydrogenase-like genes, and how thermotolerance is regulated (Dymek et al. 2011; Godman et al. 2010; Schmollinger et al. 2010). Each of these approaches utilized the amiRNA approach and was successful in answering the question each study posed. Thus, we made the choice to utilize amiRNA to investigate the function of NMD due to ease of use and effectiveness.

**Construction of vectors and generation of transformants**A schematic diagram of amiRNA constructs that are generated for UPF1 and UPF3 is shown in Figure 3.4. Vectors for the amiRNA suppression were chosen from two kinds of amiRNA vectors

A) UPF1 amiRNA construct schematic diagram




B) UPF3 amiRNA construct schematic diagram

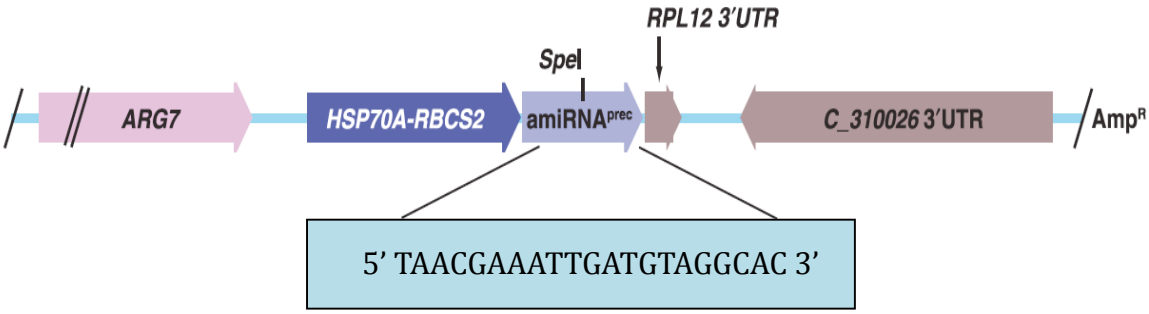



**Figure 3.4**. Schematic diagram of amiRNA vector with inserts for each gene shown.

currently available. We chose the pChlamyiRNA2 vector based upon simplicity of use and its strong promoter. The pChlamiRNA2 vector is under the control of the constituitive HSP70A-RBCS2 promoter (Figure 3.4) (Schroda et al. 2000). This vector also includes an inverted terminator sequence of transcript estEXt_fgenesh2_pg.C_310026 to terminate the transcription of any antisense strand transcript at the site of transgene integration, that may anneal with sense RNA to form dsRNA, thus silencing the amiRNA precursor itself (Molnar et al. 2009). The pChamiRNA2 vector also contains the ARG7 gene, which allows for selection of transformants on arginine deficient plates. Vectors for the transformation were made and target sequences for miRNA were chosen utilizing the methods and software provided (Molnar et al. 2009). The target sequences were ligated into the vectors and verified by colony PCR (Molnar et al. 2009). Once the amiRNA constructs were obtained and purified, they were transformed into the arginine-deficient strain cc425. Transformation was accomplished utilizing the glass bead method (Kindle 1990; Schroda 2006). Transformation proved to be successful as seen in Figure 3.5. An empty vector was utilized as a positive control and confirmed positive transformation of all mutants displaying the gained ability to grow on TAP plates lacking arginine. A negative control where no vector was added shows no growth on arginine deficient plates that were not transformed with the vector. Double transformations were performed utilizing linearized DNA and non-linearized DNA, and these are known as “cut” or “uncut” transformations. Linearized DNA appeared to produce about 30% more transformed colonies than circular DNA following transformations.

**Verificaction of integration and expression**

Transformations grown on selection plates (Figure 3.5) were tested for integration of the vector by colony PCR. In order to check both the correct orientation of amiRNA insert

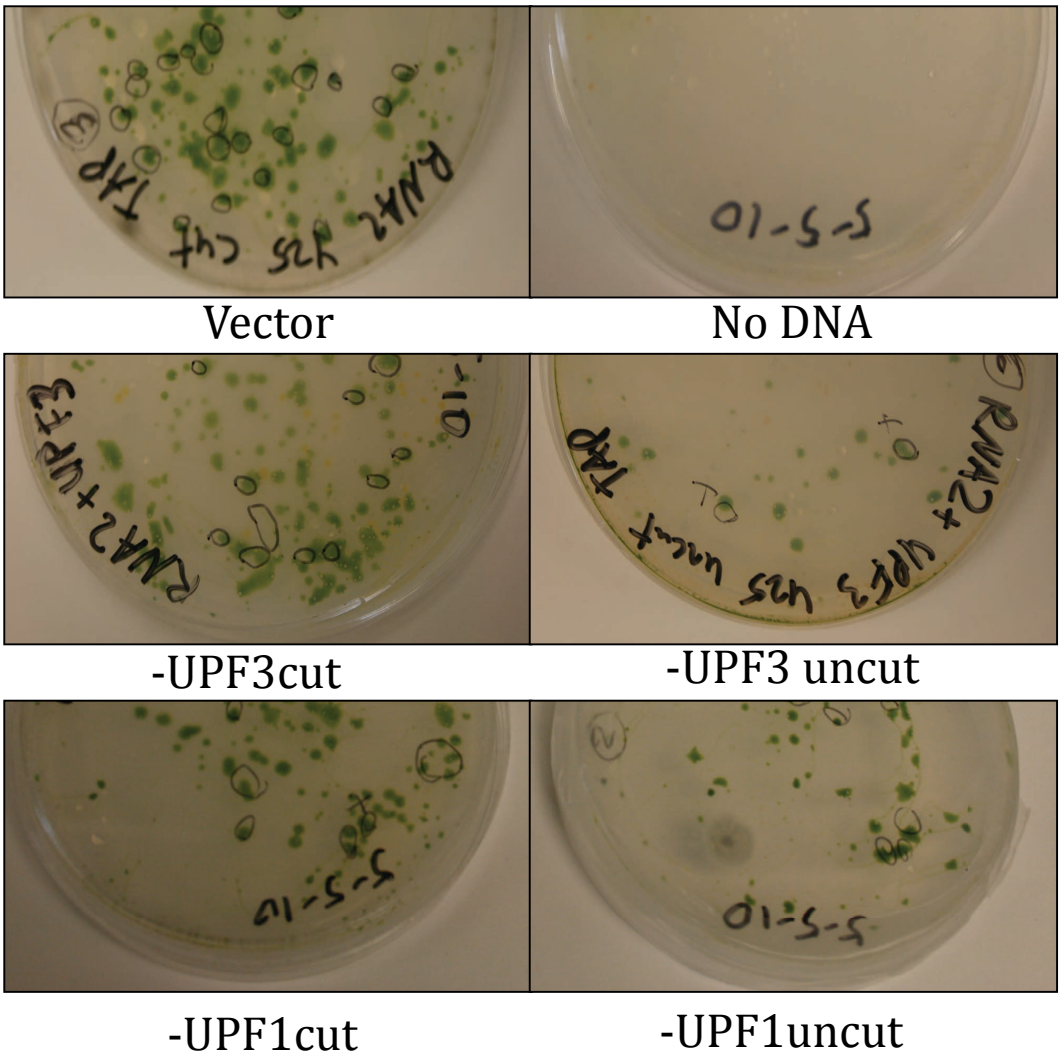



**Figure 3.5.** Plates showing transformed colonies growing on selection medium not containing arginine. Non-transformed cells in upper right show no growth on the same media.

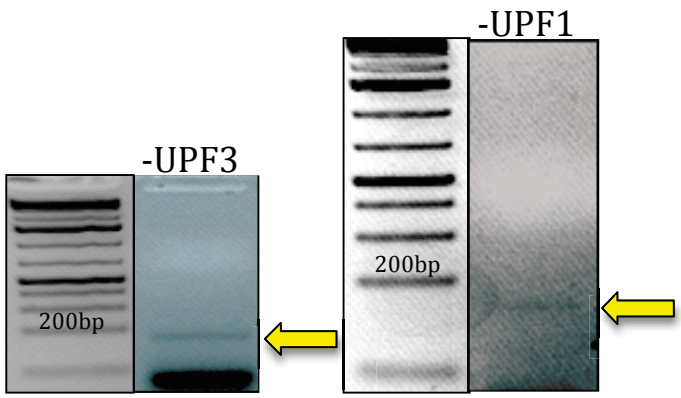



**Figure 3.6. Colony PCR to verify integration of insert. PCR product present at 182 bp contains both part of the insert as well as part of the vector.**

and the presence of the vector, a simple colony PCR was performed utilizing primers provided in a recent study (Molnar et al. 2009). A product of 182bp shown in Figure 3.6 contains both parts of the insert as well as the vector in the correct orientation, indicating successful transformation of the amiRNA vector. In order to make sure that the amiRNA was expressed, RT-PCR was performed and the results are shown in Figure 3.7. *TUA1*, which encodes tubulin-1 was utilized as a positive control for the quality and quantity of cDNA template. RT-PCR was performed using primers specific to the miRNA to be expressed (Figure 3.7), and it appears that each mutant line is expressing miRNA showing a band at 339bp, indicating a chance for suppression of UPF1 or UPF3. RT-PCR of the empty vector-transformed line showed no miRNA expression, and PCR of the vector including the insertion served as a positive control. DNAse treated RNA showed no signs of contamination, establishing confidence in the RT-PCR results.

**UPF1 and UPF3 expression**

Following this, efforts were then made to utilize RT-PCR to show expression of UPF1 and UPF3 in the WT DNA and cDNA as well as expression levels in the mutants. Genomic PCR was also performed to see if the primers were working. Primers were designed for both genes and the location of these primers can be seen in Figure 3.7. Genomic PCR of the genes in WT using two primer sets can be seen in Figure 3.8, but not all combinations of primers produced a genomic product for either gene. Efforts were than made to amplify the cDNA product of these primers in the WT strain, utilizing the same conditions for RT-PCR. Unfortunately, after multiple efforts, amplification of either gene using RT-PCR yielded no products.

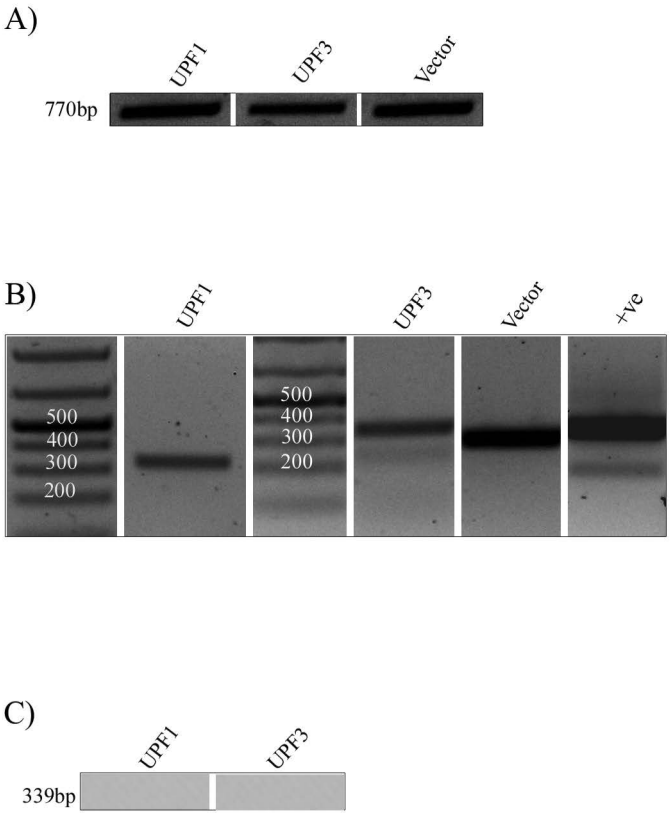



**Figure 3.7. Analysis of expression of miRNA using RT-PCR.** A) Beta-tubulin (TUA1) PCR to check quality of DNA. All lanes show high expression indicating high cDNA quality. B) PCR for amiRNA expression. *UPF1* and *UPF3* show expression of miRNA in transformed cells (339bp). C) DNAse treated RNA which shows no signs of DNA contamination.

A) UPF1

B)

UPF3

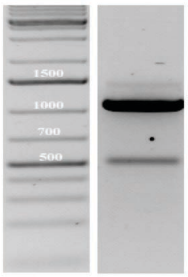
**Figure 3.8. Schematic Diagram of *UPF1* and *UPF3*.** Yellow arrows indicate forward primers and red arrows indicate reverse primers. The first yellow arrow from right to left is *UPF1*/*3*FW, the second yellow arrow from right to left is *UPF1*/*3*FWa. The first red arrow from right to left is *UPF1*/*3*RWa, the second red arrow from right to left is *UPF1*/*3*RW.

## Conclusions

In this study, I prepared amiRNA constructs to study NMD, generated and confirmed transformants, although we were unable to show UPF1 or UPF3 suppression due to time constraints. The lines which were transformed were checked for expression of Pre-miRNAs for amiRNA, which is the first step towards gene suppression, and it was verified in each transformed line that miRNA was being expressed. This evidence points to the possibility that the mutants are suppressing either UPF1 or UPF3. Unfortunately, the inability to show UPF1 or UPF3 expression in either the WT or the transformed lines has proven to be the Achilles' heel for this study. Although genomic PCR shows that these genes are represented in the WT, it is not clear why they are not represented in the mRNA. One possibility for is that certain conditions that not have been met are required in order for expression of UPF1 or UPF3 to occur. However, this seems doubtful due to the seemingly vital role of NMD, which involves these proteins in other organisms (Davies et al. 2006; Medghalchi et al. 2001; Wittkopp et al. 2009; Yoine et al. 2006b). The other possibility could be that the primers designed in the exonic region are not really exons due to mis-annotation of the genes. Instead, other approaches may be needed to verify both the presence of UPF1 and 3 in the WT and to show suppression of these in the transformant lines. The study that piloted amiRNA in *Chlamydomonas* circumvented this problem by utilizing a Northern Blot to detect the presence of amiRNA, and to show suppression of the targeted genes (Molnar et al. 2009). Since amiRNA presence was already determined by RT-PCR, Northern Blot could show both the presence of UPF1 or 3 in the WT as well as suppression, if any occurs, in the transformed lines. Quantitative PCR could be of use in this study as well to show suppression of the targeted genes, but our experience with non-specific products using RT-PCR to show UPF1/3 expression indicates a problem using this method.

This study has shown that amiRNA to UPF1 and UPF3 can be expressed in *Chlamydomonas*. These cell lines can be used to study NMD of target genes and for global analysis of NMD in *Chlamydomonas*. Although many questions remain to be answered, it is a start for beginning to investigate the importance of NMD in *Chlamydomona*s, its possible evolution, and the mechanism that still remains to be investigated in protists.

**A)**




**B)**

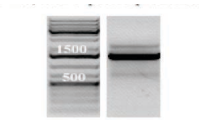



**Figure 3.9. Genomic PCR of UPF1 and UPF3.** A) Genomic PCR of UPF1 using primers FWa and RW (Figure 3 and 8), with expected product at 1076 bp. B) Genomic PCR of UPF3 using primers FWa and RW (Figure 3 and 8) , with expected product at 1551bp.

**LIST**

Alt3’ = Alternative 3’ splice site

Alt5’ = Alternative 5’ splice site

AltB = both 5’ and 3’ ends of an intron are alternative spliced

amiRNA = artificial microRNA

AS = Alternative splicing

Asyn = asparagine synthase gene

BiFC = bimolecular fluorescence complementation

ChIP = Chromatin immunoprecipitation

CLIP = Crosslinking and Immunoprecipitation

EJC = Exon Junction Complex

eRF1 = eukaryotic releasing factor 1

eRF3 = eukaryotic releasing factor 3

ES = exon skipping

ESR = exonic splicing regulator

EST = expressed sequence tag

FRET = Förster resonance energy transfer

IME = intron-mediated regulation of gene expression

IR = Intron Retention

ISR = intronic splicing regulator

miRNA = microRNA

NMD = nonsense mediated decay

NTC = NineTeen Complex proteins

ODC1 = ornithine decarboxylase 1 gene

PABP = poly (A) binding protein

PCR = polymerase chain reaction

Pre-mRNA = Precursor mRNA

PTC = premature termination codon

RRM = RNA recognition motif

RS1 = Arginine Serine domain 1

RS2 = Arginine Serine domain 2

RT-PCR = reverse transcription PCR

RUST = regulated unproductive splicing and translation

siRNA = small interfering RNA

SMG = suppressor of morphological defects on genitalia proteins

snRNP = small nuclear ribonucleoparticle

SR = Serine Arginine related protein

SURF = complex is made up of SMGI, UPF1, eRF1, and eRF3

U2AF = U2 auxiliary factor

UPF = UP-frameshift protein

YFP = yellow fluorescent protein

**Bibliography**

(2009a) Chlamydomonas AS Website.
(2009b) Chlamydomonas reinhardtii v4.0.
Ali GS, Palusa SG, Golovkin M, Prasad J, Manley JL, Reddy AS (2007) Regulation of plant developmental processes by a novel splicing factor. PLoS One 2: e471
Ali GS, Prasad KV, Hanumappa M, Reddy AS (2008a) Analyses of in vivo interaction and mobility of two spliceosomal proteins using FRAP and BiFC. PLoS One 3: e1953
Ali GS, Prasad KV, Hanumappa M, Reddy AS (2008b) Analyses of in vivo interaction and mobility of two spliceosomal proteins using FRAP and BiFC. PLoS One 3: e1953
Ali GS, Reddy AS (2008a) Spatiotemporal organization of pre-mRNA splicing proteins in plants. Curr Top Microbiol Immunol 326: 103-118
Ali GS, Reddy AS (2008b) Spatiotemporal organization of pre-mRNA splicing proteins in plants. Curr. Top. Microbiol. Immunol. 326: 103-118
Alonso JM, Ecker JR (2006) Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in Arabidopsis. Nature reviews. Genetics 7: 524-536
Ambros V (2001) microRNAs: Tiny regulators with great potential. Cell 107: 823-826
Amrani N, Ganesan R, Kervestin S, Mangus DA, Ghosh S, Jacobson A (2004) A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. Nature 432: 112-118
Anastasaki C, Longman D, Capper A, Patton EE, Caceres JF (2011) Dhx34 and Nbas function in the NMD pathway and are required for embryonic development in zebrafish. Nucleic Acids Res 39: 3686-3694
Anderson P, Cali BM, Kuchma SL, Latham J (1999a) smg-7 is required for mRNA surveillance in Caenorhabditis elegans. Genetics 151: 605-616
Anderson P, Page MF, Carr B, Anders KR, Grimson A (1999b) SMG-2 is a phosphorylated protein required for mRNA surveillance in Caenorhabditis elegans and related to Upf1p of yeast. Mol Cell Biol 19: 5943-5951
Ast G (2004) How did alternative splicing evolve? Nat Rev Genet: 773-782
Avery P, Vicente-Crespo M, Francis D, Nashchekina O, Alonso CR, Palacios IM (2011) Drosophila Upf1 and Upf2 loss of function inhibits cell growth and causes animal death in a Upf3-independent manner. Rna 17: 624-638
Azzalin CM, Lingner J (2008) Telomeres - The silence is broken. Cell Cycle 7: 1161-1165
Balasubramanian S, Sureshkumar S, Lempe J, Weigel D (2006) Potent induction of Arabidopsis thaliana flowering by elevated growth temperature. PLoS Genet 2: e106
Barbazuk WB, Fu Y, McGinnis KM (2008) Genome-wide analyses of alternative splicing in plants: opportunities and challenges. Genome Res 18: 1381-1392
Barta A, Kalyna M, Lorkovic ZJ (2008) Plant SR proteins and their functions. Curr Top Microbiol Immunol 326: 83-102
Barta A, Kalyna M, Reddy AS (2010) Implementing a rational and consistent nomenclature for serine/arginine-rich protein splicing factors (SR proteins) in plants. Plant Cell 22: 2926-2929
Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116: 281-297
Bartel DP, Farh KKH, Grimson A, Jan C, Lewis BP, Johnston WK, Lim LP, Burge CB (2005) The widespread impact of mammalian microRNAs on mRNA repression and evolution. Science 310: 1817-1821

Baulcombe DC (2006) Short silencing RNA: the dark matter of genetics? Cold Spring Harb Symp Quant Biol 71: 13-20
Beligni MV, Yamaguchi K, Mayfield SP (2004) Chloroplast elongation factor ts pro-protein is an evolutionarily conserved fusion with the s1 domain-containing plastid-specific ribosomal protein-7. Plant Cell 16: 3357-3369
Bhuvanagiri M, Schlitter AM, Hentze MW, Kulozik AE (2010) NMD: RNA biology meets human genetic medicine. Biochem J 430: 365-377
Black DL (2003) Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem 72: 291-336
Blencowe BJ, Calarco JA, Xing Y, Caceres M, Calarco JP, Xiao X, Pan Q, Lee C, Preuss TM (2007) Global analysis of alternative splicing differences between humans and chimpanzees. Genes Dev 21: 2963-2975
Blencowe BJ, McIlwain DR, Pan Q, Reilly PT, Elia AJ, McCracken S, Wakeham AC, Itie-Youten A, Mak TW (2010) Smg1 is required for embryogenesis and regulates diverse genes via alternative splicing coupled to nonsense-mediated mRNA decay. P Natl Acad Sci USA 107: 12186-12191
Blencowe BJ, Pan Q, Shai O, Lee LJ, Frey J (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet 40: 1413-1415
Bove J, Kim CY, Gibson CA, Assmann SM (2008) Characterization of wound-responsive RNA-binding proteins and their splice variants in Arabidopsis. Plant Mol Biol 67: 71-88
Brenner SE, Lareau LF, Brooks AN, Soergel DAW, Meng Q (2007a) The coupling of alternative splicing and nonsense-mediated mRNA decay. Alternative Splicing in the Postgenomic Era 623: 190-211
Brenner SE, Lareau LF, Inada M, Green RE, Wengrod JC (2007b) Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature 446: 926-929
Burge CB, Fairbrother WG, Yeh RF, Sharp PA (2002) Predictive identification of exonic splicing enhancers in human genes. Science 297: 1007-1013
Caceres JF, Long JC (2009) The SR protein family of splicing factors: master regulators of gene expression. Biochemical Journal 417: 15-27
Campbell MA, Haas BJ, Hamilton JP, Mount SM, Buell CR (2006) Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis. BMC Genomics 7: 327
Chang YF, Imam JS, Wilkinson MF (2007) The nonsense-mediated decay RNA surveillance pathway. Annu Rev Biochem 76: 51-74
Chusainow J, Ajuh PM, Trinkle-Mulcahy L, Sleeman JE, Ellenberg J, Lamond AI (2005) FRET analyses of the U2AF complex localize the U2AF35/U2AF65 interaction in vivo and reveal a novel self-interaction of U2AF35. Rna 11: 1201-1214
Cohen SM, Brennecke J, Stark A, Russell RB (2005) Principles of MicroRNA-target recognition. PLoS Biol 3: 404-418
Cooper TA, Wan L, Dreyfuss G (2009) RNA and disease. Cell 136: 777-793
Croft MT, Moulin M, Webb ME, Smith AG (2007) Thiamine biosynthesis in algae is regulated by riboswitches. P Natl Acad Sci USA 104: 20770-20775
Davies B, Arciga-Reyes L, Wootton L, Kieffer M (2006) UPF1 is required for nonsense-mediated mRNA decay (NMD) and RNAi in Arabidopsis. Plant Journal 47: 480-489
Dinesh-Kumar SP, Baker BJ (2000) Alternatively spliced N resistance gene transcripts: their possible role in tobacco mosaic virus resistance. P Natl Acad Sci USA 97: 1908-1913
Dymek EE, Heuser T, Nicastro D, Smith EF (2011) The CSC is required for complete radial spoke assembly and wild-type ciliary motility. Mol Biol Cell 22: 2520-2531

Ellis JD, Lleres D, Denegri M, Lamond AI, Caceres JF (2008) Spatial mapping of splicing factor complexes involved in exon and intron definition. J Cell Biol 181: 921-934
Eperon IC, Ireland DC, Smith RA, Mayeda A, Krainer AR (1993) Pathways for selection of 5' splice sites by U1 snRNPs and SF2/ASF. Embo J 12: 3607-3617
Fabrizio P, Dannenberg J, Dube P, Kastner B, Stark H, Urlaub H, Luhrmann R (2009) The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. Mol Cell 36: 593-608
Falciatore A, Merendino L, Barneche F, Ceol M, Meskauskiene R, Apel K, Rochaix JD (2005) The FLP proteins act as regulators of chlorophyll synthesis in response to light and plastid signals in Chlamydomonas. Genes Dev 19: 176-187
Fang Y, Hearn S, Spector DL (2004) Tissue-specific expression and dynamic organization of SR splicing factors in Arabidopsis. Mol Biol Cell 15: 2664-2673
Filichkin SA, Priest HD, Givan SA, Shen R, Bryant DW, Fox SE, Wong WK, Mockler TC (2010) Genome-wide mapping of alternative splicing in Arabidopsis thaliana. Genome Res 20: 45-58
Filipowicz W, Lorkovic ZJ, Kirk DAW, Lambermon MHL (2000) Pre-mRNA splicing in higher plants. Trends Plant Sci 5: 160-167
Garneau NL, Wilusz J, Wilusz CJ (2007) The highways and byways of mRNA decay. Nat Rev Mol Cell Biol 8: 113-126
Gatfield D, Unterholzner L, Ciccarelli FD, Bork P, Izaurralde E (2003) Nonsense-mediated mRNA decay in Drosophila: at the intersection of the yeast and mammalian pathways. Embo J 22: 3960-3970
Gecz J, Tarpey PS, Raymond FL, Nguyen LS, Rodriguez J, Hackett A, Vandeleur L, Smith R, Shoubridge C, Edkins S, Stevens C, O'Meara S, Tofts C, Barthorpe S, Buck G, Cole J, Halliday K, Hills K, Jones D, Mironenko T, Perry J, Varian J, West S, Widaa S, Teague J, Dicks E, Butler A, Menzies A, Richardson D, Jenkinson A, Shepherd R, Raine K, Moon J, Luo Y, Parnau J, Bhat SS, Gardner A, Corbett M, Brooks D, Thomas P, Parkinson-Lawrence E, Porteous ME, Warner JP, Sanderson T, Pearson P, Simensen RJ, Skinner C, Hoganson G, Superneau D, Wooster R, Bobrow M, Turner G, Stevenson RE, Schwartz CE, Futreal PA, Srivastava AK, Stratton MR (2007) Mutations in UPF3B, a member of the nonsense-mediated mRNA decay complex, cause syndromic and nonsyndromic mental retardation. Nat Genet 39: 1127-1133
Ghasemi Y, Morowvat MH, Rasoul-Amini S (2010) Chlamydomonas as a "new" organism for biodiesel production. Bioresource Technol 101: 2059-2062
Godman JE, Molnar A, Baulcombe DC, Balk J (2010) RNA silencing of hydrogenase(-like) genes and investigation of their physiological roles in the green alga Chlamydomonas reinhardtii. Biochem J 431: 345-351
Golovkin M, Reddy AS (1999) An SC35-like protein and a novel serine/arginine-rich protein interact with Arabidopsis U1-70K protein. J Biol Chem 274: 36428-36438
Gonzalez-Ballester D, Pollock SV, Pootakham W, Grossman AR (2008) The central role of a SNRK2 kinase in sulfur deprivation responses. Plant Physiol 147: 216-227
Graveley BR (2000) Sorting out the complexity of SR protein functions. Rna 6: 1197-1211
Graveley BR (2001) Alternative splicing: increasing diversity in the proteomic world. Trends Genet 17: 100-107
Grossman AR, Croft M, Gladyshev VN, Merchant SS, Posewitz MC, Prochnik S, Spalding MH (2007) Novel metabolism in Chlamydomonas through the lens of genomics. Curr Opin Plant Biol 10: 190-198
Hanfrey C, Sommer S, Mayer MJ, Burtin D, Michael AJ (2001) Arabidopsis polyamine biosynthesis: absence of ornithine decarboxylase and the mechanism of arginine decarboxylase activity. Plant Journal 27: 551-560

Harrington ED, Bork P (2008) Sircah: a tool for the detection and visualization of alternative transcripts. Bioinformatics 24: 1959-1960
Harris EH (2001) Chlamydomonas as a Model Organism. Annu Rev Plant Physiol Plant Mol Biol 52: 363-406
Harris EH (2009) The *Chlamydomonas* Sourcebook Second Edition. Elsevier
Heber S, Alekseyev M, Sze SH, Tang H, Pevzner PA (2002) Splicing graphs and EST assembly problem. Bioinformatics 18 Suppl 1: S181-188
Hirose T, Sugita M, Sugiura M (1993) cDNA structure, expression and nucleic acid-binding properties of three RNA-binding proteins in tobacco: occurrence of tissue-specific alternative splicing. Nucleic Acids Res 21: 3981-3987
Hodgkin J, Papp A, Pulak R, Ambros V, Anderson P (1989) A new kind of informational suppression in the nematode Caenorhabditis elegans. Genetics 123: 301-313
Hu CD, Chinenov Y, Kerppola TK (2002) Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. Mol Cell 9: 789-798
Hwang J, Maquat LE (2011) Nonsense-mediated mRNA decay (NMD) in animal embryogenesis: to die or not to die, that is the question. Curr Opin Genet Dev
Iida K, Seki M, Sakurai T, Satou M, Akiyama K, Toyoda T, Konagaya A, Shinozaki K (2004) Genome-wide analysis of alternative pre-mRNA splicing in Arabidopsis thaliana based on full-length cDNA sequences. Nucleic Acids Res 32: 5096-5103
Inokuchi R, Kuma KI, Miyata T, Okada M (2002) Nitrogen-assimilating enzymes in land plants and algae: phylogenic and physiological perspectives. Physiol Plant 116: 1-11
Jamison SF, Pasman Z, Wang J, Will C, Luhrmann R, Manley JL, Garcia-Blanco MA (1995) U1 snRNP-ASF/SF2 interaction and 5' splice site recognition: characterization of required elements. Nucleic Acids Res 23: 3260-3267
Johnson JM, Castle J, Garrett-Engele P, Kan Z, Loerch PM, Armour CD, Santos R, Schadt EE, Stoughton R, Shoemaker DD (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Science 302: 2141-2144
Kashima I, Yamashita A, Izumi N, Kataoka N, Morishita R, Hoshino S, Ohno M, Dreyfuss G, Ohno S (2006) Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. Genes Dev 20: 355-367
Kent WJ (2002) BLAT--the BLAST-like alignment tool. Genome Res 12: 656-664
Khvorova A, Reynolds A, Jayasena SD (2003) Functional siRNAs and miRNAs exhibit strand bias (vol 115, pg 209, 2003). Cell 115: 505-505
Kielkopf CL, Lucke S, Green MR (2004) U2AF homology motifs: protein recognition in the RRM world. Genes Dev 18: 1513-1526
Kim E, Magen A, Ast G (2007) Different levels of alternative splicing among eukaryotes. Nucleic Acids Res 35: 125-131
Kindle KL (1990) High-frequency nuclear transformation of Chlamydomonas reinhardtii. P Natl Acad Sci USA 87: 1228-1232
Kohtz JD, Jamison SF, Will CL, Zuo P, Luhrmann R, Garcia-Blanco MA, Manley JL (1994) Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. Nature 368: 119-124
Kurihara Y, Matsui A, Hanada K, Kawashima M, Ishida J, Morosawa T, Tanaka M, Kaminuma E, Mochizuki Y, Matsushima A, Toyoda T, Shinozaki K, Seki M (2009) Genome-wide suppression of aberrant mRNA-like noncoding RNAs by NMD in Arabidopsis. P Natl Acad Sci USA 106: 2453-2458
Lazar G, Goodman HM (2000) The Arabidopsis splicing factor SR1 is regulated by alternative splicing. Plant Mol Biol 42: 571-581

Le Hir H, Izaurralde E, Maquat LE, Moore MJ (2000) The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. Embo J 19: 6860-6869
Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120: 15-20
Li JB, Gerdes JM, Haycraft CJ, Fan Y, Teslovich TM, May-Simera H, Li H, Blacque OE, Li L, Leitch CC, Lewis RA, Green JS, Parfrey PS, Leroux MR, Davidson WS, Beales PL, Guay-Woodford LM, Yoder BK, Stormo GD, Katsanis N, Dutcher SK (2004) Comparative genomics identifies a flagellar and basal body proteome that includes the BBS5 human disease gene. Cell 117: 541-552
Liang C, Liu Y, Liu L, Davis AC, Shen Y, Li QQ (2008) Expressed sequence tags with cDNA termini: previously overlooked resources for gene annotation and transcriptome exploration in Chlamydomonas reinhardtii. Genetics 179: 83-93
Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature 433: 769-773
Llave C, Xie Z, Kasschau KD, Carrington JC (2002) Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. Science 297: 2053-2056
Longman D, Plasterk RH, Johnstone IL, Caceres JF (2007) Mechanistic insights and identification of two novel factors in the C. elegans NMD pathway. Genes Dev 21: 1075-1085
Lopato S, Forstner C, Kalyna M, Hilscher J, Langhammer U, Indrapichate K, Lorkovic ZJ, Barta A (2002) Network of interactions of a novel plant-specific Arg/Ser-rich protein, atRSZ33, with atSC35-like splicing factors. J Biol Chem 277: 39989-39998
Lorkovic ZJ, Hilscher J, Barta A (2008) Co-localisation studies of Arabidopsis SR splicing factors reveal different types of speckles in plant cell nuclei. Exp Cell Res 314: 3175-3186
Lorkovic ZJ, Wieczorek Kirk DA, Klahre U, Hemmings-Mieszczak M, Filipowicz W (2000a) RBP45 and RBP47, two oligouridylate-specific hnRNP-like proteins interacting with poly(A)+ RNA in nuclei of plant cells. Rna 6: 1610-1624
Lorkovic ZJ, Wieczorek Kirk DA, Lambermon MH, Filipowicz W (2000b) Pre-mRNA splicing in higher plants. Trends Plant Sci 5: 160-167
Majewski J, Ott J (2002) Distribution and characterization of regulatory elements in the human genome. Genome Res 12: 1827-1836
Manavella PA, Rubio-Somoza I (2011) Engineering elements for gene silencing: the artificial microRNAs technology. Methods Mol Biol 732: 121-130
Maniatis T, Tasic B (2002) Alternative pre-mRNA splicing and proteome expansion in metazoans. Nature 418: 236-243
Maquat LE (2004) Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. Nat Rev Mol Cell Biol 5: 89-99
Mathew R, Hartmuth K, Mohlmann S, Urlaub H, Ficner R, Luhrmann R (2008) Phosphorylation of human PRP28 by SRPK2 is required for integration of the U4/U6-U5 tri-snRNP into the spliceosome. Nat Struct Mol Biol 15: 435-443
Matlin AJ, Moore MJ (2007) Spliceosome assembly and composition. Adv Exp Med Biol 623: 14-35
Mayfield SP, Rasala BA, Muto M, Lee PA, Jager M, Cardoso RMF, Behnke CA, Kirk P, Hokanson CA, Crea R, Mendez M (2010) Production of therapeutic proteins in algae, analysis of expression of seven human proteins in the chloroplast of Chlamydomonas reinhardtii. Plant Biotechnol J 8: 719-733

McCullough AJ, Lou H, Schuler MA (1991) In vivo analysis of plant pre-mRNA splicing using an autonomously replicating vector. Nucleic Acids Res 19: 3001-3009
Medghalchi SM, Frischmeyer PA, Mendell JT, Kelly AG, Lawler AM, Dietz HC (2001) Rent1, a trans-effector of nonsense-mediated mRNA decay, is essential for mammalian embryonic viability. Hum Mol Genet 10: 99-105
Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, Terry A, Salamov A, Fritz-Laylin LK, Marechal-Drouard L, Marshall WF, Qu LH, Nelson DR, Sanderfoot AA, Spalding MH, Kapitonov VV, Ren Q, Ferris P, Lindquist E, Shapiro H, Lucas SM, Grimwood J, Schmutz J, Cardol P, Cerutti H, Chanfreau G, Chen CL, Cognat V, Croft MT, Dent R, Dutcher S, Fernandez E, Fukuzawa H, Gonzalez-Ballester D, Gonzalez-Halphen D, Hallmann A, Hanikenne M, Hippler M, Inwood W, Jabbari K, Kalanon M, Kuras R, Lefebvre PA, Lemaire SD, Lobanov AV, Lohr M, Manuell A, Meier I, Mets L, Mittag M, Mittelmeier T, Moroney JV, Moseley J, Napoli C, Nedelcu AM, Niyogi K, Novoselov SV, Paulsen IT, Pazour G, Purton S, Ral JP, Riano-Pachon DM, Riekhof W, Rymarquis L, Schroda M, Stern D, Umen J, Willows R, Wilson N, Zimmer SL, Allmer J, Balk J, Bisova K, Chen CJ, Elias M, Gendler K, Hauser C, Lamb MR, Ledford H, Long JC, Minagawa J, Page MD, Pan J, Pootakham W, Roje S, Rose A, Stahlberg E, Terauchi AM, Yang P, Ball S, Bowler C, Dieckmann CL, Gladyshev VN, Green P, Jorgensen R, Mayfield S, Mueller-Roeber B, Rajamani S, Sayre RT, Brokstein P, Dubchak I, Goodstein D, Hornick L, Huang YW, Jhaveri J, Luo Y, Martinez D, Ngau WC, Otillar B, Poliakov A, Porter A, Szajkowski L, Werner G, Zhou K, Grigoriev IV, Rokhsar DS, Grossman AR (2007) The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science 318: 245-250
Merendino L, Guth S, Bilbao D, Martinez C, Valcarcel J (1999) Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. Nature 402: 838-841
Mi S, Cai T, Hu Y, Chen Y, Hodges E, Ni F, Wu L, Li S, Zhou H, Long C, Chen S, Hannon GJ, Qi Y (2008) Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. Cell 133: 116-127
Mollet I, Barbosa-Morais NL, Andrade J, Carmo-Fonseca M (2006) Diversity of human U2AF splicing factors. Febs J 273: 4807-4816
Molnar A, Bassett A, Thuenemann E, Schwach F, Karkare S, Ossowski S, Weigel D, Baulcombe D (2009) Highly specific gene silencing by artificial microRNAs in the unicellular alga Chlamydomonas reinhardtii. Plant J
Mount SM, Zhang XN (2009) Two Alternatively Spliced Isoforms of the Arabidopsis SR45 Protein Have Distinct Roles during Normal Plant Development. Plant Physiol 150: 1450-1458
Muhlemann O (2008) Recognition of nonsense mRNA: towards a unified model. Biochem Soc Trans 36: 497-501
Nagy E, Maquat LE (1998) A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. Trends Biochem Sci 23: 198-199
Nakano K, Suzuki T, Hayakawa T, Yamaya T (2000) Organ and cellular localization of asparagine synthetase in rice plants. Plant and Cell Physiology 41: 874-880
Ner-Gaon H, Leviatan N, Rubin E, Fluhr R (2007) Comparative cross-species alternative splicing in plants. Plant Physiol 144: 1632-1641
Nicholson P, Muhlemann O (2010) Cutting the nonsense: the degradation of PTC-containing mRNAs. Biochem Soc Trans 38: 1615-1620
Nicholson P, Yepiskoposyan H, Metze S, Zamudio Orozco R, Kleinschmidt N, Muhlemann O (2010) Nonsense-mediated mRNA decay in human cells: mechanistic insights, functions beyond quality control and the double-life of NMD factors. Cell Mol Life Sci 67: 677-700

Nilsen TW (2003) The spliceosome: the most complex macromolecular machine in the cell? Bioessays 25: 1147-1149
Niranjanakumari S, Lasda E, Brazas R, Garcia-Blanco MA (2002) Reversible cross-linking combined with immunoprecipitation to study RNA-protein interactions in vivo. Methods 26: 182-190
Ohno S, Kashima I, Yamashita A, Izumi N, Kataoka N, Morishita R, Hoshino S, Ohno M, Dreyfuss G (2006) Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. Genes Dev 20: 355-367
Orengo JP, Cooper TA (2007) Alternative splicing in disease. Adv Exp Med Biol 623: 212-223
Palusa SG, Ali GS, Reddy AS (2007) Alternative splicing of pre-mRNAs of Arabidopsis serine/arginine-rich proteins: regulation by hormones and stresses. Plant J 49: 1091-1107
Palusa SG, Reddy AS (2010) Extensive coupling of alternative splicing of pre-mRNAs of serine/arginine (SR) genes with nonsense-mediated decay. New Phytol 185: 83-89
Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet 40: 1413-1415
Porse BT, Weischenfeldt J, Damgaard I, Bryder D, Theilgaard-Monch K, Thoren LA, Nielsen FC, Jacobsen SEW, Nerlov C (2008) NMD is essential for hematopoietic stem and progenitor cells and for eliminating by-products of programmed DNA rearrangements. Genes Dev 22: 1381-1396
Pulak R, Anderson P (1993) mRNA surveillance by the Caenorhabditis elegans smg genes. Genes Dev 7: 1885-1897
Pyronnet S, Pradayrol L, Sonenberg N (2000) A cell cycle-dependent internal ribosome entry site. Mol Cell 5: 607-616
Rappsilber J, Ryder U, Lamond AI, Mann M (2002) Large-scale proteomic analysis of the human spliceosome. Genome Res 12: 1231-1245
Reddy AS (2007) Alternative splicing of pre-messenger RNAs in plants in the genomic era. Annu Rev Plant Biol 58: 267-294
Reddy ASN (2001) Nuclear pre-mRNA splicing in plants. Crit Rev Plant Sci 20: 523-571
Reddy ASN (2004) Plant serine/arginine-rich proteins and their role in pre-mRNA splicing. Trends Plant Sci 9: 541-547
Reddy ASN, Labadorf A, Link A, Rogers MF, Thomas J, Ben-Hur A (2010) Genome-wide analysis of alternative splicing in Chlamydomonas reinhardtii. BMC Genomics 11
Rohr J, Sarkar N, Balenger S, Jeong BR, Cerutti H (2004) Tandem inverted repeat system for selection of effective transgenic RNAi strains in Chlamydomonas. Plant J 40: 611-621
Rose A (2008) Intron Mediated Regulation of Gene Expression. Springer, Heidelberg, Germany
Rupprecht J (2009) From systems biology to fuel-Chlamydomonas reinhardtii as a model for a systems biology approach to improve biohydrogen production. J Biotechnol 142: 10-20
Sauliere J, Haque N, Harms S, Barbosa I, Blanchette M, Le Hir H (2010) The exon junction complex differentially marks spliced junctions. Nat Struct Mol Biol 17: 1269-1271
Schmollinger S, Strenkert D, Schroda M (2010) An inducible artificial microRNA system for Chlamydomonas reinhardtii confirms a key role for heat shock factor 1 in regulating thermotolerance. Curr Genet 56: 383-389

Schoning JC, Streitner C, Meyer IM, Gao Y, 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 Res 36: 6977-6987
Schroda M (2006) RNA silencing in Chlamydomonas: mechanisms and tools. Curr Genet 49: 69-84
Schroda M, Blocker D, Beck CF (2000) The HSP70A promoter as a tool for the improved expression of transgenes in Chlamydomonas. Plant J 21: 121-131
Schroda M, Vallon O, Whitelegge JP, Beck CF, Wollman FA (2001) The chloroplastic GrpE homolog of Chlamydomonas: two isoforms generated by differential splicing. Plant Cell 13: 2823-2839
Schwab R, Ossowski S, Riester M, Warthmann N, Weigel D (2006) Highly specific gene silencing by artificial microRNAs in Arabidopsis. Plant Cell 18: 1121-1133
Sharp PA (1994) Split genes and RNA splicing. Cell 77: 805-815
Silva AL, Romao L (2009) The mammalian nonsense-mediated mRNA decay pathway: to decay or not to decay! Which players make the decision? FEBS Lett 583: 499-505
Smith DJ, Query CC, Konarska MM (2008) "Nought may endure but mutability": spliceosome dynamics and the regulation of splicing. Mol Cell 30: 657-666
Tanabe N, Yoshimura K, Kimura A, Yabuta Y, Shigeoka S (2007) Differential expression of alternatively spliced mRNAs of Arabidopsis SR protein homologs, atSR30 and atSR45a, in response to environmental stress. Plant Cell Physiol 48: 1036-1049
Theiss C, Bohley P, Voigt J (2002) Regulation by polyamines of ornithine decarboxylase activity and cell division in the unicellular green alga Chlamydomonas reinhardtii. Plant Physiol 128: 1470-1479
Tuschl T, Elbashir SM, Lendeckel W (2001) RNA interference is mediated by 21-and 22-nucleotide RNAs. Genes Dev 15: 188-200
Ule J, Jensen K, Mele A, Darnell RB (2005) CLIP: a method for identifying protein-RNA interaction sites in living cells. Methods 37: 376-386
Valadkhan S, Jaladat Y (2010) The spliceosomal proteome: at the heart of the largest cellular ribonucleoprotein machine. Proteomics 10: 4128-4141
Vallon O, Dutcher S (2008) Treasure hunting in the Chlamydomonas genome. Genetics 179: 3-6
Wahl MC, Will CL, Luhrmann R (2009) The spliceosome: design principles of a dynamic RNP machine. Cell 136: 701-718
Walter M, Chaban C, Schutze K, Batistic O, Weckermann K, Nake C, Blazevic D, Grefen C, Schumacher K, Oecking C, Harter K, Kudla J (2004) Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. Plant J 40: 428-438
Wang BB, Brendel V (2006a) Genomewide comparative analysis of alternative splicing in plants. P Natl Acad Sci USA 103: 7175-7180
Wang BB, Brendel V (2006b) Molecular characterization and phylogeny of U2AF35 homologs in plants. Plant Physiol 140: 624-636
Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB (2008) Alternative isoform regulation in human tissue transcriptomes. Nature 456: 470-476
Will CL, Luhrmann, R. (2006) Spliceosomal structure and function. The RNA World: 369-400
Willmund F, Muhlhaus T, Wojciechowska M, Schroda M (2007) The NH2-terminal domain of the chloroplast GrpE homolog CGE1 is required for dimerization and cochaperone function in vivo. J Biol Chem 282: 11317-11328

Wittkopp N, Huntzinger E, Weiler C, Sauliere J, Schmidt S, Sonawane M, Izaurralde E (2009) Nonsense-mediated mRNA decay effectors are essential for zebrafish embryonic development and survival. Mol Cell Biol 29: 3517-3528
Wu JY, Maniatis T (1993) Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell 75: 1061-1070
Wu S, Romfo CM, Nilsen TW, Green MR (1999) Functional recognition of the 3' splice site AG by the splicing factor U2AF35. Nature 402: 832-835
Xu P, Zhang Y, Kang L, Roossinck MJ, Mysore KS (2006) Computational estimation and experimental verification of off-target silencing during posttranscriptional gene silencing in plants. Plant Physiol 142: 429-440
Yamashita A, Ohnishi T, Kashima I, Taya Y, Ohno S (2001) Human SMG-1, a novel phosphatidylinositol 3-kinase-related protein kinase, associates with components of the mRNA surveillance complex and is involved in the regulation of nonsense-mediated mRNA decay. Genes Dev 15: 2215-2228
Yoine M, Nishii T, Nakamura K (2006a) Arabidopsis UPF1 RNA helicase for nonsense-mediated mRNA decay is involved in seed size control and is essential for growth. Plant and Cell Physiology 47: 572-580
Yoine M, Ohto MA, Onai K, Mita S, Nakamura K (2006b) The lba1 mutation of UPF1 RNA helicase involved in nonsense-mediated mRNA decay causes pleiotropic phenotypic changes and altered sugar signalling in Arabidopsis. Plant J 47: 49-62
Yoshimura K, Yabuta Y, Ishikawa T, Shigeoka S (2002) Identification of a cis element for tissue-specific alternative splicing of chloroplast ascorbate peroxidase pre-mRNA in higher plants. J Biol Chem 277: 40623-40632
Zahler AM, Roth MB (1995) Distinct Functions of Sr Proteins in Recruitment of U1 Small Nuclear Ribonucleoprotein to Alternative 5' Splice Sites. P Natl Acad Sci USA 92: 2642-2646
Zamore PD, Schwarz DS, Hutvagner G, Du T, Xu ZS, Aronin N (2003) Asymmetry in the assembly of the RNAi enzyme complex. Cell 115: 199-208
Zhao T, Wang W, Bai X, Qi Y (2008) Gene silencing by artificial microRNAs in Chlamydomonas. Plant J
Zheng CL, Fu XD, Gribskov M (2005) Characteristics and regulatory elements defining constitutive splicing and different modes of alternative splicing in human and mouse. Rna 11: 1777-1787
Zhou Z, Licklider LJ, Gygi SP, Reed R (2002) Comprehensive proteomic analysis of the human spliceosome. Nature 419: 182-185
Zorio DAR, Blumenthal T (1999) Both subunits of U2AF recognize the 3 ' splice site in Caenorhabditis elegans. Nature 402: 835-838