# Using Structural Biology to Characterize the *Schizosaccharomyces pombe* Spliceosome

By

Scott Edward Collier

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

In

Cell and Developmental Biology

August, 2015

Nashville, Tennessee

Approved:

Kathleen L. Gould, Ph.D.

Brandt F. Eichman, Ph.D.

Irina N. Kaverina, Ph.D.

Charles Sanders, Ph.D.

Melanie D. Ohi, Ph.D.

## ACKNOWLEDGEMENTS

Special thanks to the M. Ohi lab members both past and present for your friendship and daily conversations over the years. You all have made this time of my life truly special. A special thanks to two of the most wonderful lab managers, Melissa Chambers and Prashant Singh, for keeping the lab running smoothly. I also want to thank the P. Ohi and Gould labs for critically evaluating my work, our scientific collaborations and friendships over the years.

To my thesis committee: *Kathy Gould, Irina Kaverina, Chuck Sanders, and Brandt Eichman*. You have been a scientific inspiration and taught me many valuable life lessons, which I will forever cherish. I am blessed to have had you all as committee members.

To Mel, thank you for all that you have done for me over the years. You are truly the best boss I have ever had. You have afforded me so many opportunities through specialized training and collaborations, which have advanced my scientific training. Although the lab has changed over the years, you have always kept the lab environment a positive and friendly place to work and for that I will always be grateful.

To my previous mentors, Brent Jolicoeur, Dr. Ken O'Conner and Dr. Trevor Creamer, thank you for your guidance and fostering my love of science. Brent you taught high school chemistry and physics in such a fun and interactive way that it was impossible for me not to fall in love with science. Thanks you for participating in my long hypothetical discussions on things loosely associated with science. To Dr.

O’Conner, thank you for believing in me and seeing potential in me. To Dr. Creamer, thank you for your friendship and guidance.

Last and most importantly, I want to acknowledge my friends and family who have provided encouragement and faith over the years. To my wife, *Sarah Collier*, “you are a fire, you are dynamite” and I could not have made it through this journey without you.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| 2D | two-dimensional |
| 3D | three-dimensional |
| 5-FOA | 5-fluoroorotic acid |
| aa | amino acid |
| ACE | two-acetoxyethoxy |
| ATR | ataxia telangiectasia-mutated and Rad3-related |
| AU | analytical ultracentrifugation |
| bp | base-pair |
| BP | branchpoint |
| BSA | bovine serum albumin |
| CD | circular dichroism |
| CDC | cell division cycle |
| cDNA | complementary DNA |
| CS | cold sensitive |
| CS-Rosetta | chemical shift Rosetta |
| D3 | domain 3 |
| DDR | DNA damage response |
| DNA | deoxyribonucleic acid |
| dsRNA | double-stranded RNA |
| DSS | 4,4-dimethyl-4-silapentane-1-sulfonic acid |
| EM | electron microscopy |

| | |
|---|---|
| EMM | Edinburgh minimal media |
| HAT | half-a-TPR domain |
| hnRNP | heterogeneous nuclear ribonucleoprotein particle |
| *Hs* | *Homo sapiens* |
| HS | heat sensitive |
| HSQC | heteronuclear single-quantum correlation |
| HTH | helix-turn-helix |
| ILS | intron lariat spliceosome |
| ISL | inter-stem loop |
| LC-MS/MS | liquid chromatography-tandem mass spectrometry |
| LDS | lithium dodecyl sulfate |
| Loop | loop region found in the N-terminus of Cdc5 |
| LSm | like Sm |
| m3G | trimethyl guanosine |
| m7G | 7-methyl guanosine |
| MBP | maltose binding protein |
| MBP-MS2BP | maltose binding protein-MS2 binding protein |
| MID | middle region of Cdc5 |
| MLR | Myb-like repeat |
| mRNA | messenger RNA |
| *Ms* | *Mus musculus* |
| MS | mass spectrometry |

| | |
|---|---|
| Myb | myeloblastosis |
| NMR | nuclear magnetic resonance |
| NMT | no message in thiamine |
| NTC | NineTeen Complex |
| ORF | open reading frame |
| P-complex | post-catalytic complex |
| pI | isoelectric point |
| PPM | parts per million |
| pre-mRNA | precursor messenger RNA |
| PRP | precursor RNA processing |
| R1 | Myb repeat 1 |
| R2 | Myb repeat 2 |
| RMSD | root-mean-square deviation |
| RNA | ribonucleic acid |
| RRM | RNA recognition motif |
| RT | reverse transcriptase |
| *Sc* | *Saccharomyces cerevisiae* |
| S.E.M. | standard error of the mean |
| SR | serine-rich |
| snRNAs | small nuclear RNAs |
| snRNPs | small nuclear ribonucleoproteins |
| *Sp* | *Schizosaccharomyces pombe* |

| | |
|---|---|
| ssRNA | single-stranded RNA |
| SVAU | sedimentation velocity analytical ultracentrifugation |
| TAP | tandem affinity purification |
| TBE | tris-borate-EDTA |
| TCA | trichloroacetic acid |
| TPR | tetratricopeptide repeats |
| UV | ultraviolet |
| YE | yeast extract |
| ZnF | zinc-finger |

# CHAPTER I

## Introduction

A simplistic view of the central dogma of biology states that DNA is transcribed into RNA, which is then translated into protein (Fig. 1-1A). Some RNA molecules called precursor messenger RNAs (pre-mRNAs) need to be processed to a mature messenger RNA (mRNA) in order to be translated into protein. Pre-mRNAs are composed of both exons and introns. Exons are nucleotide sequences in a gene that remain after a pre-mRNA has been processed to mRNA. Introns are nucleotide sequences in a gene that are removed through a process called RNA splicing. Pre-mRNA processing via the splicing reaction is essential in eukaryotic organisms and although it has been extensively studied, there are still many unanswered questions regarding the molecular mechanisms of pre-mRNA splicing.

### Pre-mRNA Biology

Pre-mRNA transcripts are produced from a DNA template using an RNA polymerase and are modified before being translated into a protein. In order for pre-mRNAs to be processed they contain specific RNA sequences defining both the exons and introns within an RNA transcript (Fig. 1-1B). Generally speaking, the 5′ exon is defined by specific nucleotide(s) at the 5′ splice site (SS) while another set of nucleotide(s) defines the 3′ end of the intron at the 3′SS. Some types of pre-mRNAs also have an adenosine located in the intron known as the branchpoint (BP). Pre-mRNA molecules in eukaryotic organisms are not only spliced but are also modified co-

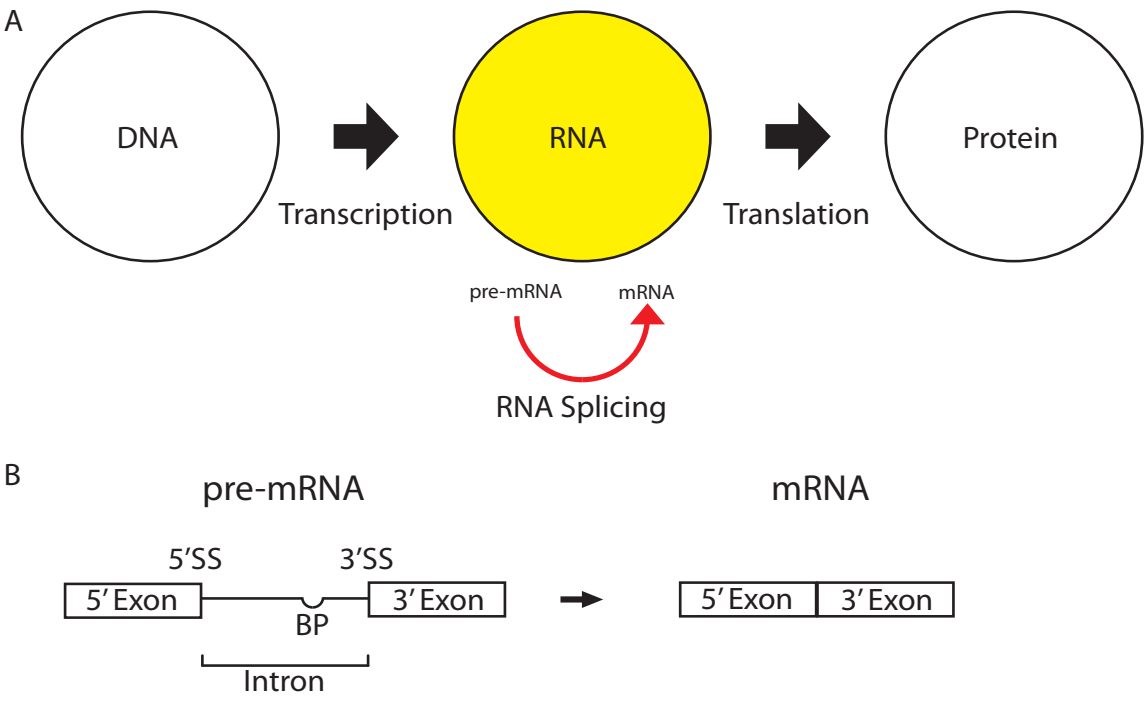



**Figure 1-1. Central Dogma and pre-mRNA Biology** (A) Simiplified view of the central dogma of biology. The yellow circle highlights the RNA step where pre-mRNA to mRNA processing (RNA splicing) occurs. The red arrow highlights the RNA splicing step. (B) General features of pre-mRNAs involed in RNA splicing. In the pre-mRNA the 5' and 3' exons (shown in boxes) are separated by an intron (solid line). The general location of the 5' and 3'SS (splice site) as well as the BP (branchpoint) are marked. After RNA processing a mRNA is produced consisting of exons.

transcriptionally at their 5′ and 3′ ends (Bentley 2002). The 5′ end of the pre-mRNA is capped with a 7-methylguanosine (m7G) nucleotide, protecting the RNA from degradation and initiating translation by the ribosome (Shatkin 1976). The 3′ end of a pre-mRNA is polyadenylated which stabilizes the transcript, and is important for pre-mRNA export and translation (Sarkar 1997).

In humans, pre-mRNA processing plays an essential role in cell biology. Humans have ~20,000 protein coding genes in their DNA which can to be transcribed to over 90,000 different proteins (Lander, Linton et al. 2001, Venter, Adams et al. 2001, Consortium 2004, Valdivia 2007). Higher eukaryotes utilize a form of splicing called alternative splicing which allows for a pre-mRNA to be processed differently in various cell types or under diverse cellular conditions (Roy, Haupt et al. 2013). Alternative splicing allows for one pre-mRNA transcript to generate a variety of protein molecules which can have distinct functions within an organism (Roy, Haupt et al. 2013). Mutations in pre-mRNAs can change SS selection resulting in proteins that have altered function, potentially giving rise to diseases such as inherited disorders and several types of cancers (Faustino and Cooper 2003, Ward and Cooper 2009, Bonomi, Gallo et al. 2013, Jhanwar 2014). Understanding how pre-mRNAs are processed is essential to understanding human health and disease.

Pre-mRNA processing can be catalyzed by large RNA ribozymes (such as Group I and Group II introns) or large dynamic protein-RNA complexes called spliceosomes. Although diverse in composition, these pre-mRNA splicing machines carry out similar

splicing chemistry. The goal of this work is to provide insight into the mechanism of pre-mRNA splicing catalyzed by the spliceosome.

**Self-Splicing RNAs**

Group I and Group II introns are ribozymes which catalyze the processing of pre-mRNA (Saldanha, Mohr et al. 1993, Lambowitz and Zimmerly 2011, Brown, Colas des Francs-Small et al. 2014, Hausner, Hafez et al. 2014). Group I and Group II introns are mobile genetic elements which insert themselves into genomic DNA and process pre-mRNA (Lambowitz and Zimmerly 2011). *In vitro* Group I and Group II introns are capable of self-splicing pre-mRNA by catalyzing self-removal from an RNA molecule in the absence of protein. The resulting RNA molecule can then be translated into protein. Self-splicing introns are found in prokaryotes and in fungal and plant organelles (Kück, Godehardt et al. 1990, Saldanha, Mohr et al. 1993, Brown, Colas des Francs-Small et al. 2014) and are thought to be a remnant from an RNA world, where biology was governed by RNA or a chemically similar molecule rather than DNA (Higgs and Lehman 2015). Group I and Group II introns share many similarities as both types of introns are large RNA molecules which process RNA using two transesterification reactions and they are composed of multiple structurally conserved domains with little to no sequence similarity (Harris-Kerr, Zhang et al. 1993, Hausner, Hafez et al. 2014) (Fig. 1-2). These introns are often found in genes that encode for transposase-like proteins which can copy the intron and insert it back into the genome (Braun, Mehlig et al. 2000).

There are clear differences between the mechanisms used by Group I and Group II introns to process pre-mRNA. Group I introns self-splice *in vitro* by a divalent cation-

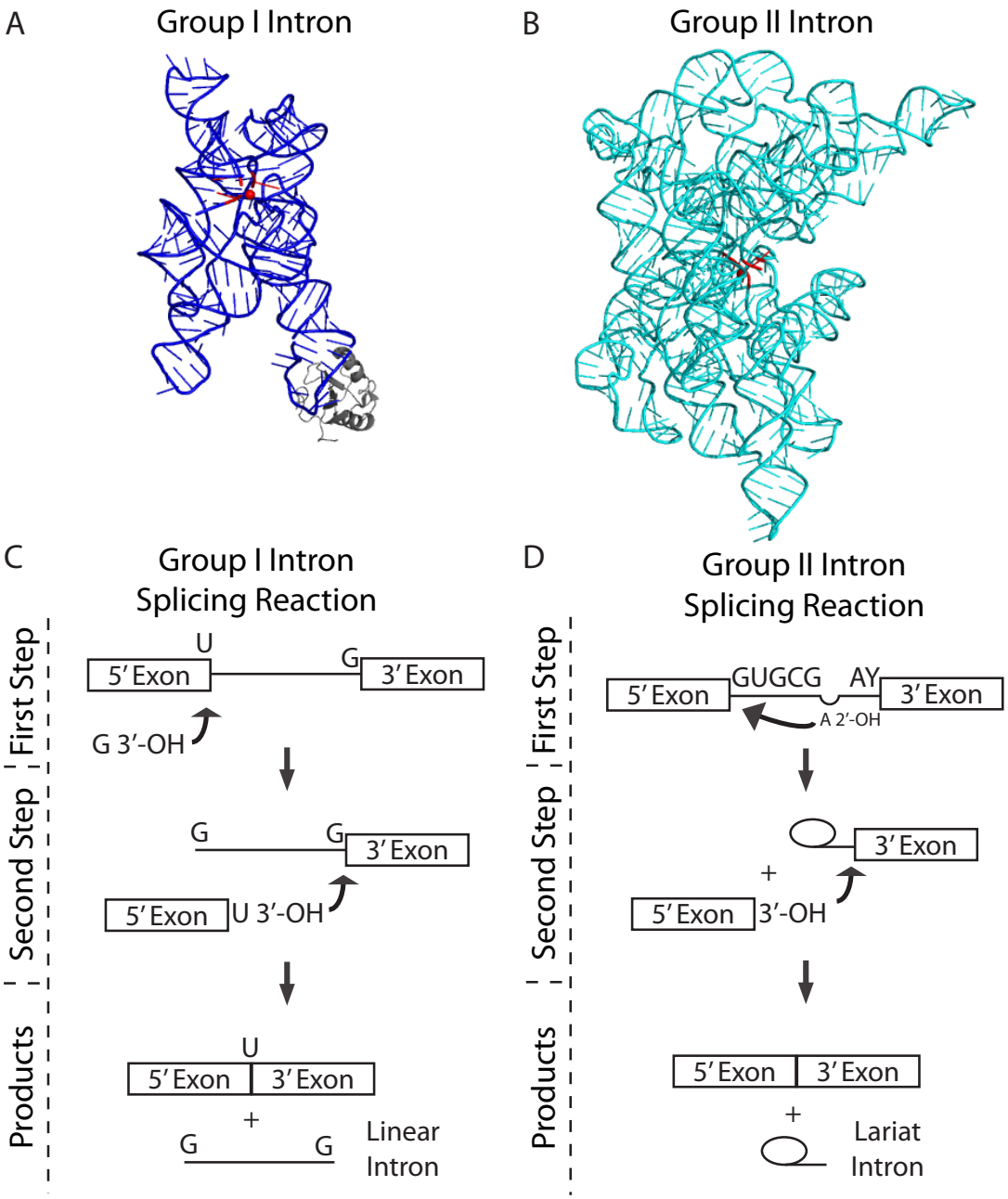



**Figure 1-2. Group I and Group II Self-splicing Introns** (A) Crystal structure of the *Azoarcus sp.* Group I intron at 3.1Å shown in blue. Active site residues are highlighted in red (Adams et al. 2004). (B) Crystal structure of a Group II intron from *Pylaiella littoralis* at 3.7Å shown in cyan. The active site is highlighted in red (Robart et al. 2014). (C) General splicing reaction carried out by the Group I intron. The two transesterification reactions are highlighted resulting in the ligation of two exons and linear intron. (D) Splicing reaction carried out by the Group II intron. The two transesterification reactions are highlighted resulting in the ligation of two exons and a lariat intron.

required folding event (Emerick and Woodson 1994, Zhang and Leibowitz 2001) to form the active site (Fig. 1-2A). Each Group I intron has a canonical recognition sequence including a 3′ uracil on the 5′ exon marking the 5′SS and a 3′ guanosine at the end of the intron indicating the 3′SS (Fig. 1-2C). Group I introns perform a 2-step transesterification reaction in which the 3′-OH of an exogenous GTP attacks the 5′-3′ phosphodiester bond at the 5′SS (Hausner, Hafez et al. 2014). This reaction is followed by an RNA structural rearrangement which allows for 3′SS cleavage at the 3′exon and the ligation of the 5′ and 3′ exons. The self-splicing reaction requires no external source of energy to complete splicing and only relies on the presence of $Mg^{2+}$ and GTP in the active site. The Group I self-splicing reaction results in exon ligation and a linear intron, which may undergo a third transesterification reaction to form a circular intron (Saldanha, Mohr et al. 1993, Zaug, McEvoy et al. 1993, Golden, Gooding et al. 1998).

Similar to Group I introns, Group II introns do not share sequence homology. However, Group II introns do have a conserved core structure composed of six domains (I-VI) (Saldanha, Mohr et al. 1993) (Fig. 1-2B). Group II introns have conserved recognition sequences similar to what is seen in pre-mRNAs spliced by the spliceosome in eukaryotes (Fig. 1-2D). The recognition sequences are found at the 5′ end of the intron (GUGYG, Y is any pyrimidine) and at the branchpoint (AY) (Saldanha, Mohr et al. 1993). Splicing is carried out in a two-step transesterification reaction wherein the first step the 2′-OH of an intronic adenosine cleaves at the 5′SS of the intron forming a lariat RNA with the 3′ exon. After a structural rearrangement removes the lariat intron from the active site, the 3′ exon is brought into position for the final transesterification reaction

(Harris-Kerr, Zhang et al. 1993). In the second step of splicing, the 3′-OH of the 5′ exon attacks the 3′ exon yielding a mRNA and a free lariat intron. The first transesterification reaction in Group II introns can also proceed through an efficient hydrolysis pathway *in vitro* (Podar, Chu et al. 1998). Both Group I and Group II self-splicing introns demonstrate protein-free pre-mRNA splicing. The chemistry of splicing found in Group I and Group II self-splicing introns, is conserved in eukaryotic organisms, wherein the spliceosome proteins have taken over some of the roles of the structural RNAs found in the self-splicing introns (Abelson 2008, Galej, Nguyen et al. 2014).

**Splicing by the Spliceosome**

The spliceosome is a dynamic macromolecular machine responsible for processing pre-mRNA to mRNA in eukaryotes. The spliceosome processes pre-mRNA using two transesterification reactions similar to Group II introns (Fig. 1-3A). In contrast to the self-splicing introns, the spliceosome is composed of several protein factors that facilitate pre-mRNA splicing. The spliceosome consists of both protein-RNA and protein-only subcomplexes. Five small nuclear RNAs (snRNAs; U1, U2, U4, U5, and U6) organize into snRNA-protein complexes called snRNPs (small nuclear ribonucleoproteins), where each snRNA interacts with a specific subset of proteins. The U4, U5, and U6 snRNPs can interact forming the tri-snRNP (Häcker, Sander et al. 2008). In addition to the snRNPs, the spliceosome contains a protein-only complex called the NineTeen Complex (NTC). The NTC is composed of a set of core components and a more loosely associated set of proteins called NTC-related components (Fig. 5). There

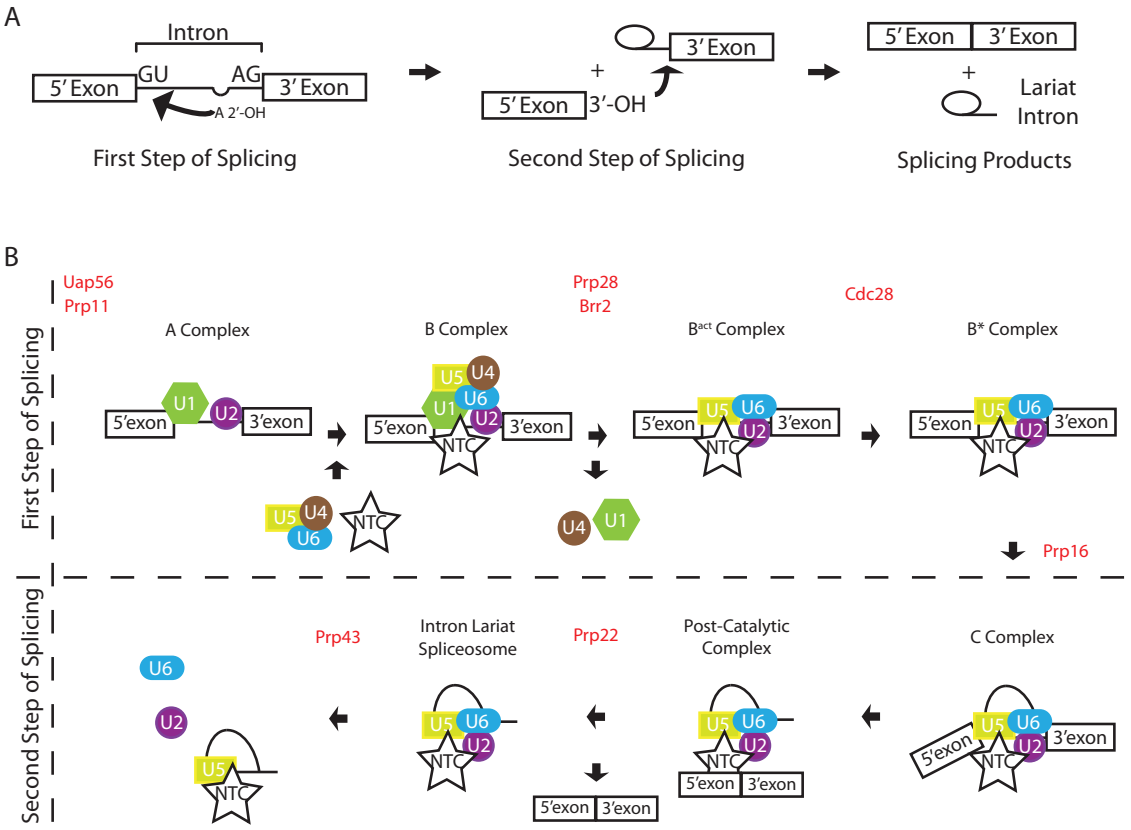



**Figure 1-3. Pre-mRNA Splicing by the Spliceosome** (A) Pre-mRNA splicing reaction carried out by the spliceosome in eukaryotes. Pre-mRNA splicing occurs using two transesterification reactions resulting in the ligation of two exons and the formation of a lariat intron. (B) *In vitro* model of the pre-mRNA splicing cycle catalyzed by the spliceosome. The splicing cycle is split into the two chemical reactions outlined in panel A. The first step of splicing is above the dotted line and the second step of splicing is below the dotted line. Each splicing complex is labeled with its corresponding name and the splicing RNA helicases are highlighted in red text. *S. pombe* nomenclature is used for the splicing RNA helicases.

are also a myriad of additional pre-mRNA splicing factors that associate with the spliceosome, although the functions of most of these proteins are not characterized.

The current *in vitro* model for pre-mRNA splicing in eukaryotes states that snRNPs interact with a pre-mRNA in a step-wise and dynamic manner that requires numerous protein and RNA conformational changes (Wahl, Will et al. 2009, Will and Luhrmann 2011, Hoskins and Moore 2012, Chen and Moore 2014) (Fig. 1-3B). In this model of spliceosome assembly, the 5′ and 3′ SSs are recognized by the U1 snRNP and Prp2, a protein component of the U2 snRNP, while other U2 snRNP components recognize the BP sequence to form the A complex. The arrival of the U4/U6.U5 tri-snRNP forms the B complex triggering the unwinding of the U4/U6 snRNA duplex and the disruption of the U1 snRNA base pairing at the 5′SS. The release of the U1 and U4 snRNPs and arrival of the NTC marks the formation of the $B^{act}$ complex. The $B^{act}$ complex transitions into the $B^{*}$ complex with the formation of catalytic structures between the pre-mRNA, U2 and U6 snRNAs, as well as the destabilization of several proteins from the U2 snRNP. The C complex forms after the first transesterification reaction resulting in 5′SS cleavage. The completion of the second transesterification reaction forms the mRNA and the P-complex (post-catalytic complex) which then becomes the ILS complex (intron lariat spliceosome) upon release of the mRNA.

The spliceosome is conserved from yeast to humans and has been primarily studied using *S. cerevisiae* (*Sc*), *S. pombe* (*Sp*), and *H. sapiens* (*Hs*) model systems. Although the ultimate goal of pre-mRNA splicing research is to understand how splicing functions in humans, the human model system is plagued with difficulties. With a

challenging diploid genome, this model system is expensive to manipulate and the spliceosome is far more complex than in yeast. The complexity of the human spliceosome is partially due to the reliance on alternative splicing which has allowed for the co-evolution of serine rich (SR) and heterogeneous nuclear ribonucleoproteins (hnRNPs) that facilitate SS selection (Busch and Hertel 2012). Both *S. cerevisiae* and *S. pombe* offer distinct experimental advantages for the study of pre-mRNA splicing. Yeast are relatively inexpensive, genetically tractable organisms which have both haploid and diploid states as well as simplified splicing machinery (Käufer and Potashkin 2000). Yeast model systems also have the advantage of a library of splicing mutants such as those isolated in "precursor RNA processing" (PRP) and "cell division cycle" (CDC) screens (Hartwell 1967, Nurse 1975, Nurse, Thuriaux et al. 1976, Hartwell 1978, Lundgren, Allan et al. 1996, Urushiyama, Tani et al. 1996, Gómez, Angeles et al. 2005). In yeast, early splicing work was primarily done in *S. cerevisiae* due to the availability of the *in vitro* splicing assay and the lack of defined molecular biology methods for *S. pombe*. Although *S. pombe* methods have been developed (Russell and Nurse 1986, Sabatinos and Forsburg 2010, Chen, Shulha et al. 2014), there is still no efficient *in vitro* splicing assay. *S. pombe* is more similar to humans than *S. cerevisiae* in that 43% of the *S. pombe* genome contains introns compared with *S. cerevisiae* at only 5% (Spingola, Grate et al. 1999, Wood, Gwilliam et al. 2002, Webb, Romfo et al. 2005). The spliceosome in *S. pombe* more closely resembles the protein components of the human spliceosome. Supporting this idea, there are several splicing components found in *S. pombe* which have either been lost or are highly divergent in *S. cerevisiae*. For example,

the is no homolog of the *Sp* Prp4 kinase in *S. cerevisiae* and a number of NTC components are either absent or are not essential in *S. cerevisiae* (Aravind, Watanabe et al. 2000, Kuhn and Käufer 2003) (Fig. 1-5A). For the reasons mentioned above, *S. pombe* has been used in this body of work as a model system to investigate the structure and function of the spliceosome. For clarity *S. pombe* nomenclature will be used throughout this work unless otherwise noted.

For over 15 years the spliceosome has been hypothesized to be a ribozyme (Collins and Guthrie 2000, Butcher 2009). This hypothesis is based on mechanistic as well as sequence and structural similarities between the RNA core of the spliceosome and Group II introns (Fig. 1-4A and C). For example, mechanistically, both splicing reactions result in the formation of a lariat intron and the ligation of two exons. There are also several key elements that are conserved between Group II introns and the spliceosome. The first conserved element is $Mg^{2+}$, which is required for the first and second catalytic steps of both splicing reactions (Gordon, Sontheimer et al. 2000, Fica, Tuttle et al. 2013, Fica, Mefford et al. 2014). The RNA structures of the Group II intron domain V and the U6 snRNA internal stem loop (ISL) of the spliceosome are similar (Huppler, Nikstad et al. 2002, Burke, Sashital et al. 2012, Russell, Jarmoskaite et al. 2013), with both stem loops coordinating a $Mg^{2+}$ ion required for catalysis (Yean, Wuenschell et al. 2000) (Fig. 1-4B and D). There are RNA sequence similarities between the two types of splicing including the AGC catalytic triad as well as the ACAGAGA sequence found in the U6 snRNA, which corresponds to the J2/J3 region of Group II introns (Toor, Keating et al. 2008) (Fig. 1-4A). In addition, a core splicing factor *Sp* Spp42 (*Sc* Prp8/*Hs* PRPF8) is

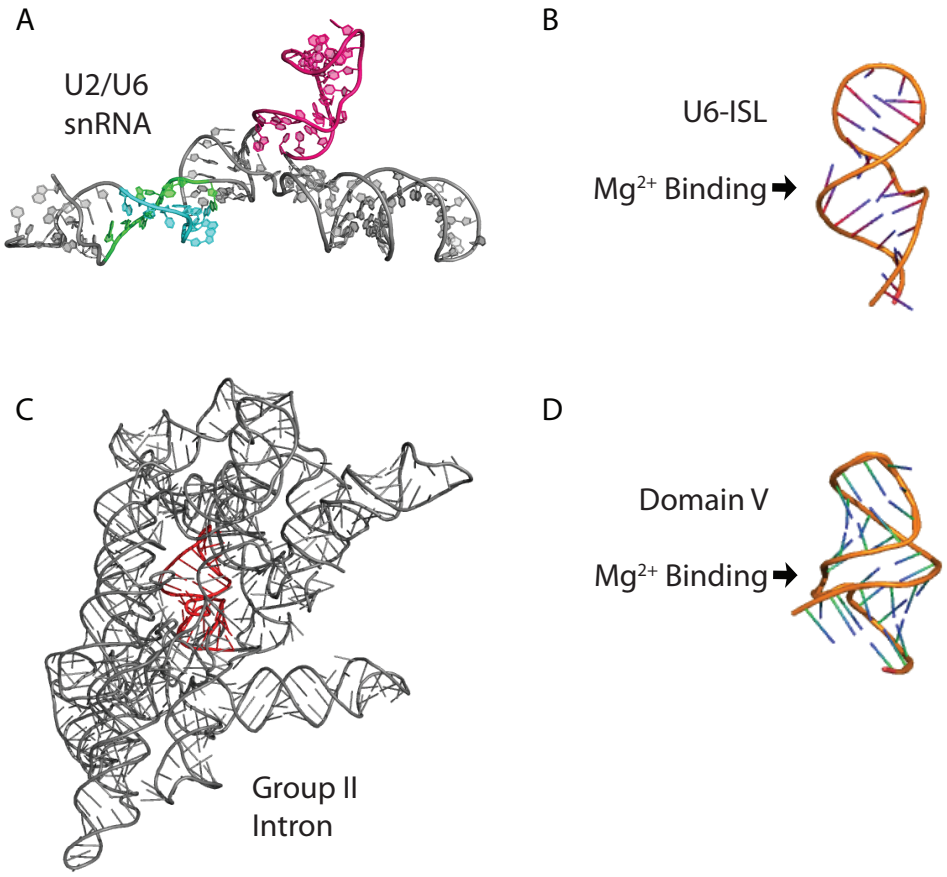



**Figure 1-4. Active site RNA Comparison between Group II Introns and the Spliceosome** (A) Model of the U2/U6 snRNA in the active site of the spliceosome (Burke et al. 2012; PDB 2LKR). Red highlights the U6-ISL helix, green is the ACAGAGA sequence and cyan is the U2 regions which interacts with the pre-mRNA. (B) Close up of the U6-ISL (panel A, red) with the $Mg^{2+}$ binding site marked in the helix. (C) Crystal structure of a Group II intron in grey where the red region corresponds to the to Domain V (Robart et al. 2014; PDB 4R0D). (D) Close up of Domain V (panel C, red) showing the $Mg^{2+}$ binding site.

hypothesized to be a remnant of Group II intron maturases, which have been evolutionally adapted as an RNA stabilization protein in the core of the spliceosome, replacing RNA domains found in Group II introns (Galej, Oubridge et al. 2013, Galej, Nguyen et al. 2014). Further supporting the "spliceosome as a ribozyme" hypothesis is evidence that a protein-free U2/U6 snRNA complex has residual catalytic activity *in vitro* (Valadkhan and Manley 2001, Valadkhan and Manley 2003, Valadkhan, Mohammadi et al. 2009). In the protein-free system the first cleavage step proceeds through hydrolysis, which has also been observed in Group II introns as well as in the spliceosome (Chu, Liu et al. 1998, Podar, Chu et al. 1998, Tseng and Cheng 2008). The ability of the U2/U6 snRNA to catalyze a splicing event in the absence of protein and the similarities between the RNA and protein within the active site of the spliceosome with Group II introns supports the hypothesis that the spliceosome is a ribozyme.

The spliceosome is a dynamic machine which is required for processing pre-mRNA into mRNA. Although the RNA found within the spliceosome may function as a ribozyme *in vitro*, this RNA-only reaction is inefficient and a number of protein components are necessary for proper and efficient splicing. The NTC and RNA helicases are two examples. The NTC complex is required for stabilizing RNA-RNA and protein-RNA interactions at the catalytic core of the spliceosome to promote proper pre-mRNA splicing (Tarn, Lee et al. 1993, Chan and Cheng 2005, Fabrizio, Dannenberg et al. 2009). RNA helicases are required to drive the splicing reaction and some ensure the fidelity of splicing (Cordin and Beggs 2013, Koodathingal and Staley 2013).

**The NineTeen Complex (NTC)**

The transition from an inactive to active spliceosome directly correlates with the stable binding of the NTC (Tarn, Lee et al. 1993, Chan and Cheng 2005, Fabrizio, Dannenberg et al. 2009). The NTC is conserved from yeast to humans and is a protein-only subcomplex of the spliceosome named after its founding member *Sc* Prp19 (Tarn, Hsu et al. 1994). The NTC becomes stably associated with the spliceosome prior to 5′SS cleavage before or during U4/U6 snRNA unwinding and remains associated with the spliceosome throughout the splicing cycle (Fig. 1-3B) (Fabrizio, Dannenberg et al. 2009). The NTC is required for the stable association of both the U5 and U6 snRNAs as well as the release of the like-Sm (LSm) ring from the U6 snRNA . The NTC also activates the spliceosome through promoting the stable interaction of the U5 and U6 snRNAs with the 5′SS (Chan and Cheng 2005). The NTC is also required for promoting the fidelity and efficiency of splicing (Villa and Guthrie 2005, Pleiss, Whitworth et al. 2007). Although the NTC is required for spliceosome stabilization, activation and efficiency, the exact mechanism for how the NTC accomplishes its function(s) has yet to be determined.

The NTC is composed of a number of different proteins which can be divided into NTC-core or NTC-associated proteins (Fig. 1-5A) (Hogg, McGrail et al. 2010). The NTC has been isolated from human, *Trypanosoma brucei* and *S. cerevisiae* (Tarn, Hsu et al. 1994, Ajuh, Kuster et al. 2000, Grote, Wolf et al. 2010, Ambrósio, Badjatia et al. 2015). In humans, the NTC is called the hPrp19/CDC5L complex and is composed of a similar, but not identical, set of proteins (Ajuh, Kuster et al. 2000, Grote, Wolf et al. 2010). The *Trypanosoma brucei* PRP19 complex was recently identified and is similar to the human hPrp19/CDC5L complex (Ambrósio, Badjatia et al. 2015). In *S. cerevisiae* a group of at

least ten proteins, of which nine are conserved in *S. pombe*, are often referred to as the NTC (Tarn, Hsu et al. 1994, Ohi and Gould 2002, Ren, McLean et al. 2011). The criteria for deciding which proteins are considered NTC-core versus NTC-associated are still debated; however, it is clear that these proteins play an important role in spliceosome activation (Chan and Cheng 2005), and to ensure both the fidelity and efficiency of the splicing reaction (Villa and Guthrie 2005, Pleiss, Whitworth et al. 2007).

Within the core of the NTC there are several protein-protein and protein-RNA interacting domains. Several of these domains have been characterized and have provided insight into how the NTC functions within the spliceosome (Fig. 1-5). For example, Prp19 is a core component of the NTC and acts as a scaffold for other core NTC components (Tarn, Hsu et al. 1994, Ohi and Gould 2002, Grote, Wolf et al. 2010, Ren, McLean et al. 2011) (Fig. 1-5B). Prp19 is composed of three domains, an N-terminal U-box, a central coiled-coil domain, and a C-terminal WD40 repeat. The N-terminal U-box domain is an E3 ubiquitin ligase domain (Ohi, Vander Kooi et al. 2003, Lu and Legerski 2007, Song, Werner et al. 2010) and forms a dimer similar to other U-box domains (Vander Kooi, Ohi et al. 2006). The coiled-coil region of Prp19 forms a tetramer using both *in vivo* and *in vitro* studies in human, *S. cerevisiae*, and *S. pombe* systems (Ohi, Vander Kooi et al. 2005, Grote, Wolf et al. 2010, Schmidt, Lenz et al. 2010, Chen, Shulha et al. 2014). The coiled-coil region also provides a scaffold for other NTC components *in vivo* (Ohi and Gould 2002, Grote, Wolf et al. 2010). The WD40 repeat is another protein-protein interaction domain, which has been shown to interact with splicing components such as *Sc* Cwc2, *Sc* Prp17 and *Sc* Urn1 (*Sp* Cwf2, *Sp* Prp17, and Dre4) (Ohi and Gould 2002,

Ren, McLean et al. 2011). Although Prp19 has not been shown to bind RNA directly, several components that directly interact with Prp19 have potential RNA binding domains, such as *Sp* Cwf2 and *Sp* Cdc5, which are capable of promoting the RNA-RNA and protein-RNA stabilization attributed to the NTC.

*Sp* Cwf2, another NTC component, interacts with the WD40 repeats of Prp19 (Fig. 1-5B). In humans and *S. cerevisiae,* homologs of *Sp* Cwf2 have been show to directly crosslink to RNA within the active site of the spliceosome. Cwf2 family members contain an N-terminal Zn-finger, an RNA recognition motif (RRM) domain (Lu, Lu et al. 2011, Schmitzová, Rasche et al. 2012) and a C-terminal domain which binds to the WD40 repeats of Prp19 (Vander Kooi, Ren et al. 2010). *Sc* Cwc2 (homolog of *Sp* Cwf2) interacts with both snRNA and pre-mRNA *in vitro,* and the RRM can bind RNA independent of the Zn-finger but to a reduced level (McGrail, Krause et al. 2009). *In vivo* both *Hs* RBM22 and *Sc* Cwc2 (*Sp* Cwf2 homologs) can be crosslinked to the U6-ISL and to regions of the U6 snRNA and the pre-mRNA close to the 5′SS (Rasche, Dybkov et al. 2012). Genetic interactions in *S. cerevisiae* suggest *Sc* Cwc2 (*Sp* Cwf2) also stabilizes the U2/U6 snRNA helix I prior to the second step of splicing where this interaction is disrupted by *Sc* Prp16 (Hogg, de Almeida et al. 2014). Cwf2 interacts with distinct regions of RNA in the spliceosome active site and is proposed to anchor the NTC to the active site of the spliceosome.

Within the NTC there are a number of proteins with distinct domains whose specific function within splicing are not well-characterized (Fig. 1-5A). For example, *Sp* Prp5 is an essential gene with a C-terminal WD40 repeat that has no known interacting

A

| | *H. sapiens* | *S. pombe* (*Sp*) | *S. cerevisiae* (*Sc*) | Essential | Domains |
|---|---|---|---|---|---|
| NTC core | PRP19 | Prp19 | Prp19 | *Sp* /*Sc* | WD40, U-box |
| | CDC5L | Cdc5 | Cef1 | *Sp* /*Sc* | Myb repeats |
| | BCAS2 | Cwf7 | Snt309 | *Sp* | |
| | PRLG1 | Prp5 | Prp46 | *Sp* /*Sc* | WD40 |
| | AD-002 | Cwf15 | Cwc15 | *Sp* | |
| | CTNNBL1 | SPAC1952.06c | UNK | | Armadillo-type fold |
| | RBM22 | Cwf2 | Cwc2 | *Sp* /*Sc* | Zn-finger, RRM |
| NTC-associated | XAB2 | Cwf3 | Syf1 | *Sp* /*Sc* | HAT repeat |
| | CRNKL1 | Cwf4 | Clf1 | *Sp* /*Sc* | HAT repeat |
| | RMB22 | Cwf5 | Ecm2 | *Sp* | RRM |
| | ISY1 | Cwf12 | Isy1 | | |
| | HSP73 | Sks2 | Ssa4 | | |
| | | | Ntc20 | | |

B

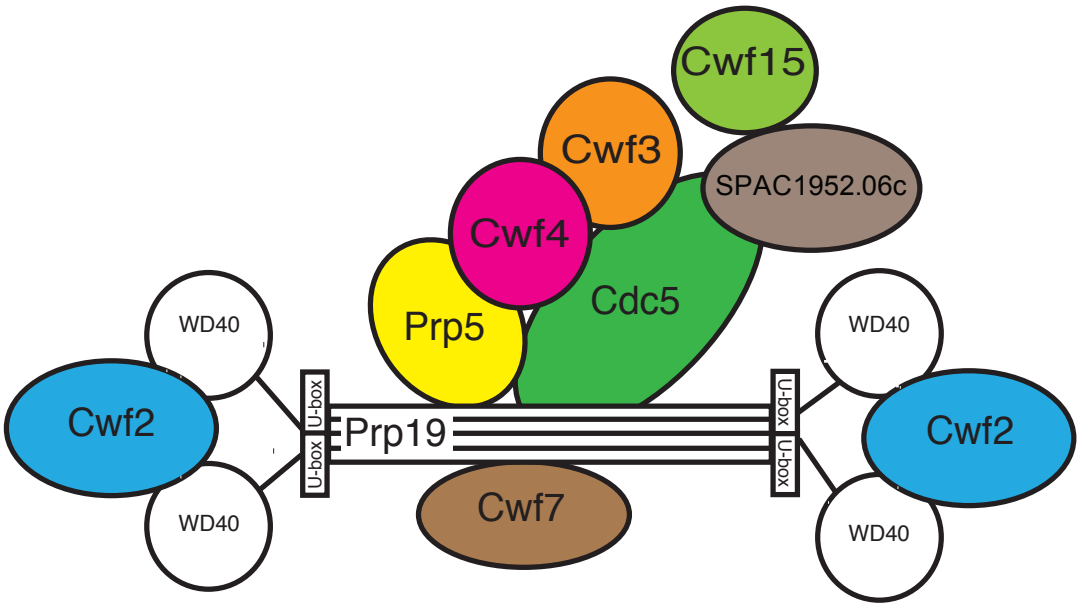



**Figure 1-5. NineTeen Complex (NTC)** (A) Components of the NTC from *H. sapiens*, *S. pombe*, and *S. cerevisiae*. Genes which are conserved in *S. pombe* are denoted with an "*Sp*" and genes which are conserved in *S. cerevisiae* are denoted with an "*Sc*". Domains for each protein are listed. Gene names and domains were acquired from www.pombase.org except for the human names of the hPRP19/CDC5L (NTC) (Grote et al. 2010). (B) A cartoon representation of a subset of NTC components. Each protein is represented by a different color. The tetramer, Prp19, as the core scaffold, is shown in more detail highlighting the N-terminal U-box domain, the central coil-coiled region, and the C-terminal WD40 repeat. *S. pombe* nomenclature is used in this cartoon.

partner(s), although the N-terminal region of the *S. cerevisiae* homolog of *Sp* Prp5 (*Sc* Prp46) interacts with NTC components (Ohi and Gould 2002, Albers, Diment et al. 2003). Additionally there are two components of the NTC, *Sp* Cwf3 and *Sp* Cwf4, with HAT (half-a-TPR) repeats that are thought to bind protein and/or RNA (Preker and Keller 1998, Liu, Rauhut et al. 2006, Hammani, Cook et al. 2012), but these HAT repeats have yet to be characterized. The NTC also has uncharacterized Myb repeats found in *Sp* Cdc5. Myb repeats are typically found in transcription factors where they interact with a specific sequence of DNA (Tanikawa, Yasukawa et al. 1993, Tahirov, Sato et al. 2002), but Myb repeats have also been shown to be inhibited by RNA polymers suggesting they bind RNA (Nordgård, Andersen et al. 2004). Understanding how these different domains function within the NTC will provide mechanistic insight into how the NTC functions to stabilize RNA-RNA and protein-RNA interactions and how the core of the spliceosome is organized.

The NTC may also function outside of the spliceosome where it is referred to as the PSO4 complex. The PSO4 complex is implicated in the DNA damage response (DDR), ataxia telangiectasia-mutated and Rad3-related **(**ATR) protein activation and in mitosis (de Andrade, Marques et al. 1989, Henriques, Vicente et al. 1989, Hofmann, Husedzinovic et al. 2010, Hofmann, Tegha-Dunghu et al. 2013, Wan and Huang 2014). While a role for PSO4 complex outside of the spliceosome has been suggested, recent work shows that mutations in the spliceosome can have specific functional defects, especially mitotic progression (Oka, Varmark et al. 2014, Sundaramoorthy, Vázquez-Novelle et al. 2014, van der Lelij, Stocsits et al. 2014, Watrin, Demidova et al. 2014).

These recent reports indicate that the roles for the PS04/NTC complex outside of the spliceosome are a result of a secondary splicing specific defect.

**Splicing RNA Helicases**

Pre-mRNA splicing by the spliceosome requires energy to drive the many RNA-RNA and protein-RNA rearrangements that occur throughout the splicing cycle (Staley and Guthrie 1998, Will and Lührmann 2001, van der Feltz, Anthony et al. 2012, Livesay, Collier et al. 2013). Each nucleotide triphosphate (NTP)-dependent step of splicing is accompanied by one or more of the eight conserved RNA helicases (*Sp* Uap56, *Sp* Prp11, *Sp* Prp28, *Sp* Brr2, *Sp* Cdc28, *Sp* Prp16, *Sp* Prp22 and *Sp* Prp43) (Fig. 6A) and one GTPase (*Sp* Cwf10). The RNA helicases are members of the superfamily2 (SF2) group of RNA helicases, which are typically monomers having two RecA-like domains flanked by variable N- and C-terminal extensions (Cordin and Beggs 2013) (Fig. 1-6B). Three splicing helicases belong to the DEAD-box family of RNA helicases (*Sp* Uap56, *Sp* Prp11 and *Sp* Prp28) and contain a Q-motif consisting of a nine amino acid sequence with an invariant glutamine denoting the helicase as an ATPase (Tanner, Cordin et al. 2003). Four splicing helicases are DEAH-box helicases (*Sp* Cdc28, *Sp* Prp16, *Sp* Prp22 and *Sp* Prp43), which share a winged helix, a ratchet domain and an OB-fold. Conversely the DEAH-box helicases do not have the Q-motif and can hydrolyze any NTP *in vitro*. A unique splicing helicase, *Sp* Brr2, of the Ski2-like family shares similarities with both DEAD and DEAH-box helicases containing a Q-motif found in DEAD-box helicases and the ratchet domain of the DEAH-box helicases. Brr2 is also unique in that it contains two helicase core domains. The Brr2 N-terminal helicase core is the functional helicase

A

| *H. sapiens* | *S. pombe* | *S. cerevisiae* | Group |
|---|---|---|---|
| DDX39B | Uap56 | Sub2 | DEAD |
| DDX46 | Prp11 | Prp5 | DEAD |
| DDX23 | Prp28 | Prp28 | DEAD |
| DHX16 | Cdc28 | Prp2 | DEAH |
| DHX38 | Prp16 | Prp16 | DEAH |
| DHX8 | Prp22 | Prp22 | DEAH |
| DHX15 | Prp43 | Prp43 | DEAH |
| SNRNP200 | Brr2 | Brr2 | Ski2-like |

B

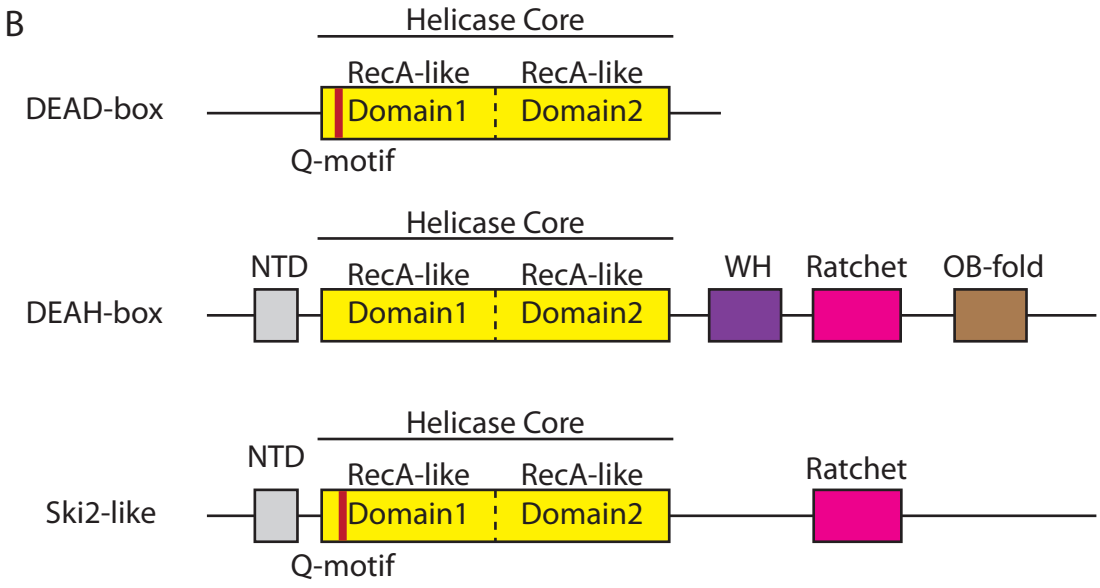



**Figure 1-6. Splicing RNA Helicases** (A) Table of splicing RNA helicases from *H. sapiens, S. pombe,* and *S. cerevisiae* and their corresponding class of RNA helicase. (B) Domain architecture of the DEAD-box, DEAH-box, and Ski2-like family of RNA helicases. Each RNA helicase has a helicase core composed of two RecA-like domains (shown in yellow) and variable N- and C-terminal extensions. DEAD-box helicases are distinguished by the presence of a Q-motif (shown in red) in the first RecA-like domain. DEAH-box helicases have three C-terminal domains; a WH (winged-helix; purple), Rachet (pink) and a OB fold (brown). The Ski2-like RNA helicase has a Q-motif similar to DEAD-box helicases and a Rachet (pink) domain similar to the DEAH-box helicases.

domain (Santos, Jovin et al. 2012) while the C-terminal helicase core is a protein-protein interaction domain (Cordin, Hahn et al. 2014). *Sp* Cwf10 is the only essential splicing GTPase and it shares sequence homology with ribosomal elongation factor 2 (EF2). *Sp* Cwf10 contains a conserved N-terminal extension not found in EF2 (Fabrizio, Laggerbauer et al. 1997, Brenner and Guthrie 2005, Livesay, Collier et al. 2013). The functions of the splicing helicases and GTPase are essential for pre-mRNA processing.

The splicing ATPases and GTPase are required to perform their function at distinct steps in pre-mRNA processing to drive the progression of the splicing reaction and to act with precise timing to ensure the fidelity of splicing (Cordin and Beggs 2013, Koodathingal and Staley 2013). Helicases are involved in every spliceosome transition from the A complex to the post-splicing complex (Fig. 1-3B). For example, early in RNA processing both *Sp* Uap56 and *Sp* Prp11 promote the formation of the A complex. In *S. cerevisiae* the *Sp* Uap56 homolog, *Sc* Sub2, acts on the pre-mRNA to remove *Sc* Msl5 (*Sp* Bpb1), a BP binding protein, which allows the BP RNA to be weakly bound by the U2 snRNP (Fleckner, Zhang et al. 1997, Kistler and Guthrie 2001, Zhang and Green 2001, Huang, Vilardell et al. 2002). In humans, *Hs* DDX39B (*Sp* Uap56) uses its helicase activity to remove *Hs* U2AF65 (*Sp* Prp2), allowing for the U2 snRNP to bind tightly resulting in the formation of the A complex (Fleckner, Zhang et al. 1997). Finally, *Hs* DDX39B and *Hs* U2AF65 (*Sp* Uap56 and *Sp* Prp2) stimulates the unwinding of the U4/U6 snRNA promoting the formation of the B complex (Shen, Zheng et al. 2008). In coordination with *Sp* Uap56, *Sp* Prp11 works directly on the U2 snRNP to promote the formation of the A complex (Xu, Newnham et al. 2004). The *Sp* Prp11 homolog, *Sc* Prp5,

binds to the U2 snRNP to activate the snRNP by stabilizing the U2 snRNA stem IIa conformation (Perriman, Barta et al. 2003, Perriman and Ares 2007, Perriman and Ares 2010). *Sc* Prp5 also proofreads the interaction of the U2 snRNP with the BP sequence selecting against suboptimal substrate sequences (Perriman and Ares 2007, Xu and Query 2007, Semlow and Staley 2012). Both Uap56 and Prp11 act to promote the stable association of the U2 snRNP with the pre-mRNA resulting in the A complex.

Catalytic activation of the spliceosome involves five ATPases and one GTPase (Fig. 1-3B). After the formation of the B complex, the *S. cerevisiae* homolog of *Sp* Prp28 (*Sc* Prp28) is required for displacing the U1 snRNP from the 5′SS and *Sc* Prp28 positions the 5′SS for pairing with the U6 snRNA (Staley and Guthrie 1999, Chen, Stands et al. 2001, Hage, Tung et al. 2009). Simultaneously, the human and *S. cerevisiae* homologs of *Sp* Brr2 (*Sc* Brr2 and *Hs* SNRNP200) works to unwind the U4/U6 snRNA to allow for U6 snRNA base pairing at the 5′SS (Laggerbauer, Achsel et al. 1998, Raghunathan and Guthrie 1998, Kim and Rossi 1999). The *Sc* Brr2 (*Sp* Brr2) is a component of the U5 snRNP where it associates with and is regulated by both *Sc* Prp8 and *Sc* Snu114 (*Sp* Spp42 and *Sp* Cwf10) (Häcker, Sander et al. 2008). Disruption of these interactions in *S. cerevisiae* results in a reduction of U4/U6 unwinding (Bartels, Klatt et al. 2002, Brenner and Guthrie 2006, Maeder, Kutach et al. 2009), which is required for $B^{act}$ complex formation. Mutants in the human homologs of *Sp* Brr2 and *Sp* Spp42 (*Hs* SNRNP200 and *Hs* PRPF8) may result in a degenerative eye disease called retinitis pigmentosa (Boon, Grainger et al. 2007, Zhao, Bellur et al. 2009). Although *Sc* Brr2 (*Sp* Brr2) can bind to the U4/U6 snRNA *in vivo* and unwind stem I (Maeder, Kutach et al. 2009, Hahn, Kudla

et al. 2012, Mozaffari-Jovin, Santos et al. 2012), *Sp* Brr2 or its homologs have not been shown to bind the U4/U6 snRNA stem II *in vivo* suggesting another helicase such as *Sp* Uap56 works in coordination with *Sp* Brr2 to fully unwind the U4/U6 snRNA *in vivo* (Shen, Zheng et al. 2008, Nielsen and Staley 2012). In other organisms, homologs of *Sp* Brr2 are present throughout the splicing cycle where they interact with other helicases modulating their activity (Silverman, Edwalds-Gilbert et al. 2003, Liu and Cheng 2012, Cordin, Hahn et al. 2014) and may act as a receptor for theses helicases at the core of the spliceosome (van Nues and Beggs 2001). Unlike in other organisms *Sp* Brr2 is not tightly associated with late splicing complexes (Chen, Shulha et al. 2014). Upon the concerted effort of *Sp* Prp28, *Sp* Brr2 and potentially *Sp* Uap56 the $B^{act}$ complex is formed where U1 and U4 snRNAs are removed from the spliceosome and the 5′SS and BP are primed to complete the first step of splicing.

Activation of the spliceosome begins with the recruitment of *Sp* Cdc28 to the $B^{act}$ complex (Fig. 1-3B). The *Sp* Cdc28 homolog (*Sc* Prp2), with the assistance of other protein factors, binds to nucleotides downstream of the BP and is believed to move 3′ to 5′ toward the BP where it displaces both SF3A and SF3B complexes (Kim and Lin 1993, Teigelkamp, McGarvey et al. 1994, Kim and Lin 1996, Warkocki, Odenwalder et al. 2009, Lardelli, Thompson et al. 2010, Liu and Cheng 2012). In *S. cerevisiae* the activity of *Sc* Prp2 (*Sp* Cdc28) has been proposed to expose the branched adenosine and binding sites for the stable association of *Sc* Yju2 and *Sc* Cwc25 (*Sp* Cwf16 and *Sp* Cwf25) (Warkocki, Odenwalder et al. 2009). After *Sp* Cdc28 activity the B* complex is formed and the spliceosome completes the first splicing reaction.

Upon completion of the first transesterification reaction with an optimal substrate, *Sp* Prp16 becomes tightly associated with the spliceosome, terminating the first step of splicing and activating the second step of splicing. *Sc* Prp16 (*Sp* Prp16) activity results in the release of protein factors (*Sc* Yju2/*Sp* Cwf16 and *Sc* Cwc25/*Sp* Cwf25) and RNA protection at the 3′SS, indicating that the 3′SS is bound within the active site of the spliceosome (Schwer and Guthrie 1991, Schwer and Guthrie 1992, Tseng, Liu et al. 2011). Although *Sp* Prp16 homologs do not normally act prior to the first step of splicing, *Sc* Prp16 was shown to have a kinetic proofreading ability with suboptimal substrates containing a 5′SS or BP mutation (Burgess, Couto et al. 1990, Burgess and Guthrie 1993, Koodathingal, Novak et al. 2010). The rate of splicing is slowed in the presence of suboptimal substrates which allows for *Sc* Prp16 to act before the first splicing reaction resulting in the rejection of suboptimal substrates. *Sp* Prp16 activity ensures the fidelity of the splicing reaction and with an optimal substrate promotes the formation of the C complex.

Once pre-mRNA splicing is complete the mRNA needs to be released and the spliceosome needs to be recycled for subsequent rounds of splicing. *Sp* Prp22 is the first helicase involved in spliceosome disassembly. In *S. cerevisiae Sc* Prp22 (*Sp* Prp22) binds to the 3′exon in a *Sc* Prp16 (*Sp* Prp16)-dependant manner and triggers the release of the mRNA and the U5 snRNP as it tracks along the RNA toward the 5′ exon (Company, Arenas et al. 1991, James, Turner et al. 2002, Schneider, Campodonico et al. 2004, Aronova, Bacíková et al. 2007, Schwer 2008). *Sc* Prp22, like *Sc* Prp16, plays a proofreading role. When an aberrant 3′SS competent for splicing is present *Sc* Prp22

hydrolyzes ATP and rejects the spliceosome (Mayas, Maita et al. 2006). These aberrant spliceosomes can either be degraded or a re-sampling can occur to correct misaligned transcripts. Once the mRNA is released *Sc* Prp43 (*Sp* Prp43) in complex with *Sc* Ntr1 (*Sp* SPAC1486.03c) and *Sc* Ntr2 (*Sp* Ntr2) hydrolyzes ATP and releases snRNAs (U2, U5, and U6), the NTC and the intron lariat (Tsai, Fu et al. 2005, Boon, Auchynnikava et al. 2006). *Sc* Prp43 not only plays a role in the disassembly of optimal substrates but it is also implicated in the dissociation of *Sc* Prp16 and *Sc* Prp22 rejected spliceosomes (Koodathingal, Novak et al. 2010).

The RNA helicases involved in pre-mRNA processing by the spliceosome are positioned at different stages to drive the progression of the splicing cycle. This makes RNA helicases a target for "trapping" and isolating splicing complexes (Lardelli, Thompson et al. 2010). Isolating these complexes allows for the interactions within each complex to be characterized and then compared to other complexes in the splicing cycle to observe on a molecular level how interactions between splicing components change. One way to characterize large splicing complexes is to use electron microscopy (EM). Using EM, several 2D projections and 3D structures of single snRNPs and large splicing complexes have been determined. To date several 2D projections of large spliceosome complexes from *S. cerevisiae* (B, $B^{act}$, B* and C) (Fabrizio, Dannenberg et al. 2009, Warkocki, Odenwalder et al. 2009, Wolf, Kastner et al. 2009), *S. pombe* (ILS or U5.U2/U6) (Ohi, Ren et al. 2007), *D. melanogaster* (B and C) (Herold, Will et al. 2009) and human (B and $B^{act}$) (Deckert, Hartmuth et al. 2006, Bessonov, Anokhina et al. 2010) have been obtained. Although only 2D projections, these complexes show the general

shape and size of the spliceosome and potentially how domains move relative to each other throughout the splicing cycle. However, without 3D structures these domain dynamics are just hypotheses. To obtain more information on spliceosome function a few 3D EM structures have been determined for a subset of purified complexes (A complex (Behzadnia, Golas et al. 2007), tri-snRNP complex (Sander, Golas et al. 2006), BΔ1 complex (Boehringer, Makarov et al. 2004), C complex (Jurica, Sousa et al. 2004, Golas, Sander et al. 2010), ILS (Ohi, Ren et al. 2007), 35S U5 snRNP (Golas, Sander et al. 2010), and the native spliceosome (Azubel, Wolf et al. 2004)) ( Fig.1-7A-G). Using helicase mutants, large splicing complexes can be isolated for characterization by EM. For example, in *S. cerevisiae* a mutant in the helicase *Sc* Prp2 (*Sp* Cdc28) was used to characterize the spliceosome by EM trapped both before and after catalytic activation (Warkocki, Odenwalder et al. 2009).

*S. pombe* is a great model organism to study the different stages of splicing using mutants of RNA helicases. There are *S. pombe* helicase mutants which have been characterized using biochemistry and genetics but have not been studied structurally. For example, the *Sp* Brr2 temperature sensitive mutant, *spp41-1*, was shown to be synthetic lethal with a mutant of *Sp* Prp1 (*Sp prp1-4*), a component of the tri-snRNP (Bottner, Schmidt et al. 2005). There are also *Sp* Prp11 (*Sp prp11-1*), *Sp* Cdc28 (*Sp cdc28-p8* and *Sp prp8-1*), and *Sp* Prp16 (*Sp prp14-1*/*Sp prp16-1*) mutants, which were identified in PRP and CDC screens and have been characterized as having a general splicing defect (Lundgren, Allan et al. 1996, Urushiyama, Tani et al. 1996). *Sp* Prp16 is particularly interesting because its homologs have been shown to act between the B* and C complex.

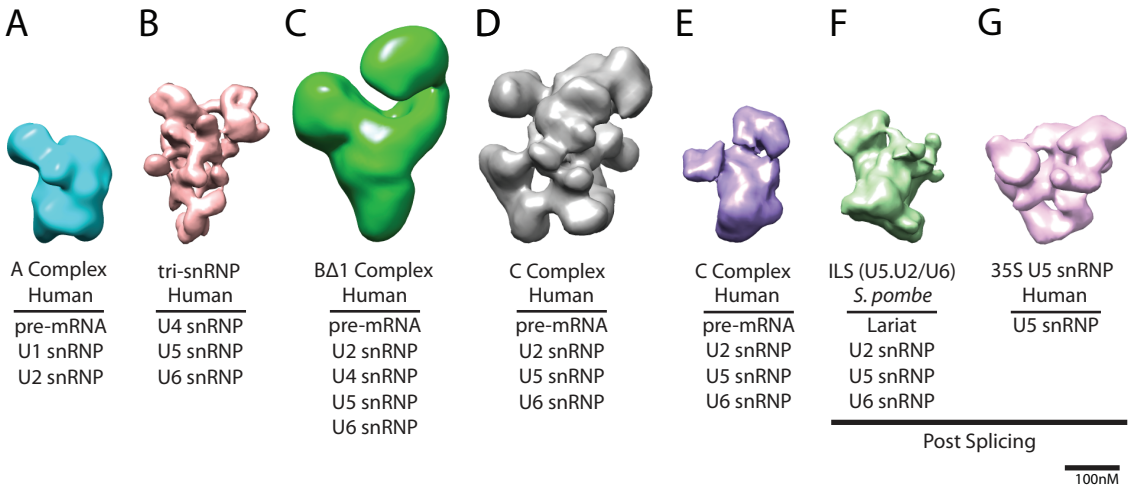



**Figure 1-7. Spliceosome Structures Determined by EM** Each EM structure A-F is listed with its corresponding name, pre-mRNA and snRNP composition. Each image is to scale and was generated using Chimera v1.6 (A) Human A complex (cyan; EMD 1325) (B) Human tri-snRNP complex (peach; EMD 1257) (C) Human BΔ1 complex (green; EMD 1066) (D) Human C complex (gray; EMD 1846; determined in the Luhrmann Lab) (E) Human C complex (purple; EMD 1062; determined in the Griegeroff lab) (F) *S. pombe* intron lariat spliceosome (ILS, also called the U5.U2/U6 complex)(light green; Ohi et al. PNAS 2007) (G) Human 35S U5 snRNP (pink;EMD 1847).

The *Sp* Prp16 mutant could potentially be used to isolate a complex which has completed the first step of splicing and is waiting for Prp16 activity to rearrange the spliceosome in order to complete the second step of splicing.

**Questions That Remain**

The goal of this work was to examine the fundamental protein-RNA interaction within the active site of the spliceosome using *S. pombe* as a model organism. The NTC contains several proteins whose domains have no known function in splicing. It is well-characterized that the NTC can interact with the active site of the spliceosome, with the homologs of the NTC component *Sp* Cwf2 binding the U6 snRNA directly (McGrail, Krause et al. 2009, Rasche, Dybkov et al. 2012). Here, I investigated the role of the N-terminus of *Sp* Cdc5 in binding RNA within the active site of the spliceosome *in vitro*. This study provides insight into how another component of the NTC can potentially bind RNA within the active site of the spliceosome, potentially directly promoting the RNA-RNA and protein-RNA stabilization needed for spliceosome activation.

In this work we also characterize a mutant of the *Sp* Prp16 helicase. The goal of this project is to isolate a large splicing complex other than the previously characterized *S. pombe* ILS. Using EM to determine a structure of this complex to ~12Å resolution will provide details of RNA within the spliceosome that are unavailable from any current 3D structure of the spliceosome. The characterization of the *Sp* Prp16 mutant is a first step toward understanding how the active site of the spliceosome is organized in a large splicing complex.

# CHAPTER II

## Structural and Functional Insights into the N-terminus of *Schizosaccharomyces pombe* Cdc5

### Introduction

The NTC (NineTeen Complex), named after its founding member *Saccharomyces cerevisiae* (*Sc*) Prp19, is a conserved protein-only spliceosome subcomplex that has been isolated in both human and *S. cerevisiae* splicing extract systems (Tarn, Hsu et al. 1994, Ajuh, Kuster et al. 2000, Grote, Wolf et al. 2010). Stable binding of the NTC is required for the formation of a catalytically active spliceosome that is competent to precisely remove introns from precursor messenger RNA (pre-mRNA) to form mature message (mRNA). Although the exact NTC composition varies among eukaryotes, at least nine proteins are conserved in yeast and are often referred to as the core NTC complex. For both *S. cerevisiae* and *Schizosaccharomyces pombe* (*Sp*), these include the proteins: Prp19/Prp19 (*Sc*/*Sp*), Cef1/Cdc5, Prp46/Prp5, Clf1/Cwf4, Syf2/Cwf3, Cwc15/Cwf15, Isy1/Cwf12, Snt309/Cwf7, Cwc2/Cwf2 and *Sc* Ntc20 (Tarn, Hsu et al. 1994, Chen, Jan et al. 1998, Tsai, Chow et al. 1999, Ben-Yehuda, Dix et al. 2000, Chen, Tsai et al. 2001, Chen, Yu et al. 2002, Ohi and Gould 2002). In mammalian cells the NTC is called the hPrp19/CDC5L complex and is composed of a similar, but not identical, set of proteins.

These include: hPrp19 (Prp19/Prp19 (*Sc/Sp*)), CDC5L (Cef1/Cdc5), PRL1 (Prp46/Prp5), AD002 (Cwc15/Cwf15), SPF27 (Snt309/Cwf7), HSP73 (Ssa4/Sks2), and CTNNBL1 (*Sp* SPAC1952.06c) (Ajuh, Kuster et al. 2000, Grote, Wolf et al. 2010).

Although it is clear that the NTC is essential for the spliceosome to transition from an inactive to active complex (Chan and Cheng 2005), as well as ensuring both the fidelity and efficiency of the splicing reaction (Villa and Guthrie 2005, Pleiss, Whitworth et al. 2007); the molecular mechanism(s) of NTC function is not known. One model is that the NTC acts as a molecular scaffold supporting and/or facilitating essential RNA-RNA, RNA-protein, and protein-protein rearrangements that are required for the formation of a catalytically active spliceosome. In support of this model, a number of NTC components contain characteristic protein-protein interaction domains that include WD40 repeats, TPR (tetratricopeptide) repeats, and HAT (half-a-TPR) repeats. Additionally, two conserved components, Cef1/Cdc5 (*Sc*/*Sp*) and Cwc2/Cwf2 (*Sc*/*Sp*), contain Myb (myeloblastosis) repeats and a RRM (RNA recognition motif) respectively, serving as potential nucleic acid binding domains (Ohi, McCollum et al. 1994, McGrail, Krause et al. 2009). While the Zinc-finger (ZnF) and RRM in *Sc* Cwc2 and its mammalian homolog RBM22 cross-links directly to the U6 snRNA and pre-mRNA (McGrail, Krause et al. 2009, Rasche, Dybkov et al. 2012), the biochemical function of the Cef1/Cdc5 (*Sc*/*Sp*) Myb repeats has not been determined.

*S. pombe cdc5*$^{+}$ was first identified in a screen of fission yeast mutants defective for cell cycle progression (Nurse, Thuriaux et al. 1976) and subsequent studies have shown that *S. pombe* Cdc5 is an essential member of the spliceosome and functions in

pre-mRNA splicing (Burns, Ohi et al. 1999, McDonald, Ohi et al. 1999, Ohi, Link et al. 2002), a role conserved in other organisms (Burns, Ohi et al. 1999, Tsai, Chow et al. 1999, Ajuh, Kuster et al. 2000, Burns, Ohi et al. 2002, Zhou, Licklider et al. 2002, Liu, Gräub et al. 2003, Grote, Wolf et al. 2010). In addition to its essential function in pre-mRNA splicing, CDC5 proteins in various organisms have also been implicated in transcription (Lin, Yin et al. 2007), DNA damage response (Zhang, Kaur et al. 2005, Zhang, Kaur et al. 2009, Maréchal, Li et al. 2014, Wan and Huang 2014), mitotic spindle assembly (Hofmann, Tegha-Dunghu et al. 2013) and microRNA (miRNA) biogenesis (Zhang, Xie et al. 2013), although whether these diverse cellular activities are splicing dependent or independent has not been fully determined.

The N-terminus of Cdc5 family members contains two canonical Myb repeats (R1 and R2) and a third conserved domain (D3) previously classified as a Myb-like repeat (Ohi, Feoktistova et al. 1998) (MLR, also referred to as Cdc5-MLR3 and -MYB3 (Ohi, Feoktistova et al. 1998, Query and Konarska 2012)). Myb repeats are classically considered DNA binding motifs and multiple copies of these domains are often found in transcription factors (reviewed in (Prouse and Campbell 2012)). Structurally, the Myb domain is composed of three well-defined helices with the second and third helices adopting a fold similar to the canonical helix-turn-helix motif (HTH) (Ogata, Hojo et al. 1992). Despite its name, a Myb repeat is not an integrated tertiary motif in a larger domain, but rather an independently folded domain that is often found in multiple copies within proteins. Although the N-terminus of Cdc5 family members has been shown to interact with DNA *in vitro* (Ohi, McCollum et al. 1994, Hirayama and Shinozaki 1996,

Lei, Shen et al. 2000) and is required for cell viability (Ohi, McCollum et al. 1994, Tsai, Chow et al. 1999), the function of the Cdc5 N-terminus and its individual domains in pre-mRNA processing has yet to be determined. Unlike the N-terminus, the sequence of the Cdc5 C-terminus is not conserved and has no recognizable protein motifs (Ohi, Feoktistova et al. 1998); however, it interacts directly with several NTC core components (Ohi and Gould 2002, Grote, Wolf et al. 2010). The presence of nucleic acid binding domains in Cdc5 and its direct interaction with other core NTC members has led us to hypothesize that Cdc5 may facilitate NTC-mediated RNA-RNA and/or RNA-protein transitions by acting as a scaffold linking NTC components and RNAs, similar to what was observed with RRM containing RBM22/Cwc2 family members in humans and *S. cerevisiae* (McGrail, Krause et al. 2009, Rasche, Dybkov et al. 2012, Hogg, de Almeida et al. 2014).

To further characterize how the NTC stabilizes and activates the spliceosome, we investigated the role of the N-terminus of Cdc5 in cell function and its ability to bind RNA *in vitro*. Using a combination of yeast genetics and RNA binding assays we show that R1, R2 and D3 are all required for function and that the Cdc5 N-terminus binds RNA *in vitro*. Structural and biochemical analyses of Cdc5-D3 show that, unlike what has been predicted (Ohi, Feoktistova et al. 1998) this domain does not adopt a canonical Myb fold and that Cdc5-D3 preferentially binds double-stranded RNA *in vitro*. Our data demonstrate that the Cdc5 N-terminus (R1, R2, and D3) can function as an RNA binding platform and can directly interact with RNA structures found near the catalytic core of the spliceosome. Our results support a model where Cdc5, by interacting with both NTC

proteins and RNA, serves as an important scaffold that facilitates the conformational changes required for the formation of a catalytically active spliceosome.

**Materials and Methods**

**Strains, yeast methods, and molecular biology.** Strains and plasmids used in this study are listed in Tables 2-1 and 2-2. Yeast strains were grown in yeast extract (YE) media or Edinburgh minimal media (EMM) with appropriate supplements. A plasmid containing the *LEU2* marker (pIRT2) and the *ORF* of *cdc5*$^{+}$ and at least 500 base pairs (bp) of the 5′and 3′ flanking sequence was used to generate *cdc5*$^{\Delta R1}$, *cdc5*$^{\Delta R2}$, *cdc5*$^{\Delta D3}$, *cdc5*$^{\Delta L}$, *cdc5*$^{\Delta MID}$, and *cdc5*$^{\Delta D3::Myb}$ integration plasmids using QuickChange II (Agilent Technologies, Santa Clara, CA). All vector transformations were performed as previously described (Keeney and Boeke 1994). Both control and *cdc5* deletion plasmids were transformed into a diploid strain of *S. pombe*, *cdc5*$^{+}$/*cdc5*::*ura4*$^{+}$. Transformations were grown on minimal medium lacking leucine, adenine, and uracil. Colonies were grown in EMM (-nitrogen) to induce sporulation, and haploid cells were grown on minimal media (+ adenine) to select for haploid cells that were *cdc5*::*ura4*$^{+}$ and carried the pIRT2 plasmid. Stable integrants were selected based on resistance to 5-fluoroorotic acid (5-FOA) (Livesay, Collier et al. 2013) and the acquisition of the LEU$^{-}$ phenotype. Mutants were validated by whole-cell PCR with primers 5′ and 3′ of the *cdc5* gene. Deletions in pREP3X *cdc5*$^{+}$ (cDNA) vectors were generated as above and transformed to a haploid strain of *S. pombe*, *cdc5-TAP*. Transformants were grown on Minimal media lacking leucine and containing thiamine (30uM) for pREP3X repression or no thiamine for pREP3x induction. OD30 Lysate Western blots with anti-pSTAIR and anti-Cdc5 were

**Table 2-1.** *Schizosaccharomyces pombe* strains used in Chapter II

| Strain # | Genotype | Figure | Source |
|---|---|---|---|
| OHI001 | *ade6-M210 leu1-32 ura4-D18 h-* | 2-1 | K. Gould[a] |
| OHI018 | *nda3-km311 leu1-32 h+* | 2-1 | Toda et al. 1983 |
| OHI130 | *cdc5-tap::Kan$^r$ ade6-m210 leu1-32 ura4-D18 h+* | 2-2 | Adapted from Ohi et al. 2002 |
| OHI237 | *prp3-1* | 2-1 | Potashkin et al. 1989 |
| OHI269 | *cdc5-120 leu1-32 his3-D1 h+* | 2-1 | Nurse et al. 1976 |
| OHI332 | *cdc5$^{\Delta Mid}$ ade6-M216 leu1-32 ura4-D18 h-* | 2-1 | This work |
| OHI334 | *cdc5$^{\Delta L}$ ade6-M210 leu1-32 ura4-D18 h-* | 2-1 | This work |
| OHI379 | *cdc5+/cdc5::ura4+ ade6-M210/ade6-M216 leu1-32/leu1-32 ura4-D18/ura4-D18 h-/h+* | 2-1 | This work |
| OHI535 | *cdc5-120 prp3-1* | 2-1 | This Work |

[a]Vanderbilt University School of Medicine, Nashville, TN.

**Table 2-2.** Plasmids used in Chapter II

| Plasmid # | Plasmid | Insert | Figure | Source |
|---|---|---|---|---|
| OHI 698 | pMAL-c2x | MBP-MS2 Binding Protein | 2-8 | Zhou et al. 2002 |
| OHI 723 | pIRT2 | empty vector | 2-1 | Hindley et al. 1987 |
| OHI 843 | pIRT2 | *S.p. cdc5*$^{+}$(ORF)* | 2-1 | Ohi et al. 1994 |
| OHI 1135 | pIRT2 | *S.p. cdc5*$^{\Delta R1}$ (ORF)* | 2-1 | This work |
| OHI 1136 | pIRT2 | *S.p. cdc5*$^{\Delta R2}$ (ORF)* | 2-1 | This work |
| OHI 1137 | pIRT2 | *S.p. cdc5*$^{\Delta L}$ (ORF)* | 2-1 | This work |
| OHI 1148 | pIRT2 | *S.p. cdc5*$^{\Delta D3}$ (ORF)* | 2-1 | This work |
| OHI 1138 | pIRT2 | *S.p. cdc5*$^{\Delta Mid}$ (ORF)* | 2-1 | This work |
| OHI 1139 | pIRT2 | *S.p. cdc5*$^{D3::Myb}$ (ORF)* | 2-1 | This work |
| OHI 730 | pREP3X | empty vector | 2-2 | Foresburg, S.L. (1993) |
| OHI 1149 | pREP3X | *S.p. cdc5*$^{+}$(cDNA)$^{\#}$ | 2-2 | This work |
| OHI 1150 | pREP3X | *S.p. cdc5*$^{\Delta R1}$ (cDNA)$^{\#}$ | 2-2 | This work |
| OHI 1151 | pREP3X | *S.p. cdc5*$^{\Delta R2}$ (cDNA)$^{\#}$ | 2-2 | This work |
| OHI 1152 | pREP3X | *S.p. cdc5*$^{\Delta D3}$ (cDNA)$^{\#}$ | 2-2 | This work |
| OHI 1153 | pREP3X | *S.p. cdc5*$^{D3::Myb}$ (cDNA)$^{\#}$ | 2-2 | This work |
| OHI 1140 | pET15-b | *S.p. cdc5* (AA 155-214) | 2-3, 2-4, 2-5, 2-6,2-8, 2-9, 2-10, and 2-11 | This work |
| OHI 1141 | pET15-b | *S.p. cdc5* (AA 5-208$^{\Delta 111-146}$) | 2-8 & 2-9 | This work |

| | | | | |
|---|---|---|---|---|
| OHI 1142 | pET15-b | *S.p. cdc5* (AA 5-112) | 2-8 | This work |
| OHI 1143 | pET15-b | *S.p. cdc5* (AA 5-55) | 2-8 | This work |
| OHI 1144 | pET15-b | *S.p. cdc5* (AA 58-111) | 2-8 | This work |

* ORF = Open Reading Frame

[#] cDNA = complementary DNA

performed as previously described (Livesay, Collier et al. 2013). For spot assays, cells were grown to mid-log phase at 25°C and re-suspended in water to achieve an $OD_{595}$ of 0.3. 10-fold serial dilutions were made, and 2 µl of each dilution was plated on YE. Plates were incubated at the indicated temperatures for 3-5 days before imaging. Protparam (Wilkins, Gasteiger et al. 1999) was used to calculate pI's of protein domains.

**Protein Expression and Purification** Cdc5-R1-R2-D3$^{\Delta L}$ (amino acids (aa) 5-208$^{\Delta 111-146}$), Cdc5-R1-R2 (aa 5-112), Cdc5-R1 (aa 5-55), Cdc5-R2 (aa 58-111), Cdc5-D3 (aa 155-214) (*S. pombe* Cdc5, NP_593880) were cloned into pET-15b (NdeI/BamHI) (EMD Millipore, Darmstadt, Germany) and transformed into *Escherichia coli* Rosetta 2 (DE3) pLysS cells (EMD Millipore, Darmstadt, Germany). Cells were grown in Terrific broth (Invitrogen, Grand Island, NY) to an $OD_{595}$ of ~0.9 and cold shocked for 20 minutes on ice. Upon addition of 1 mM IPTG, the plasmids were over-expressed for 20h at 15°C. Cells were lysed in 25 mM MES (pH 6.0), 300 mM NaCl, 2.5mM Imidazole, 5% glycerol, 0.1% Triton X-100 and one SIGMA*FAST* protease tablet (Sigma-Aldrich, St. Louis, MO). Cdc5 constructs were purified using two 5 ml Histrap HP columns (GE Healthcare, Waukesha, WI) in 50 mM MES (pH 6.0), 500 mM NaCl, 5% glycerol and a 2.5-1,000 mM Imidazole linear gradient. After the Histrap column the protein fractions were concentrated and buffer exchanged into Heparin buffer A (10 mM Sodium Phosphate pH 7.0, 1 mM EDTA, and 5% glycerol) using a 3K Amicon Ultra-15 filter (Millipore, Billerica, MA). The pooled fractions were treated overnight at room temperature with RECOthrom (The Medicines Company, Parsippany, NJ) to cleave the $His_6$ tag. Cdc5 constructs were further purified using a Heparin Column (GE Healthcare,

Waukesha, WI) in 10 mM Sodium Phosphate (pH 7.0), 1 mM EDTA, 5% glycerol and a 0-1 M NaCl linear gradient. Gel filtration (Superdex 200, GE Healthcare, Waukesha, WI) in 25 mM MES (pH 6.0), 100 mM NaCl, and 1 mM EDTA was used for the final step of purification.

For NMR experiments *Sp* Cdc5-D3 was purified as above except cells were grown and expressed in M9 media supplemented with the appropriate isotopic label (either $^{15}N$ $NH_4Cl$ or $^{15}N$ $NH_4Cl$ and $^{13}C$ D-glucose) (Cambridge Isotopes, Andover, MA). 10% $D_2O$ was added to the final sample for all NMR experiments. For the $^{15}N$ Cdc5-D3 used in RNA titration experiments, S200-RNA buffer (25 mM MES (pH 6.0), 300 mM NaCl, 2 mM MgCl2 and 1 mM EDTA) was used for gel filtration. For $^{15}N$-Leucine and $^{15}N$-Histidine (Cambridge Isotopes, Andover, MA) specific labeling, M9 media was supplemented with the appropriate unlabeled amino acids as well as $^{15}N$ L-Leucine or $^{15}N$ L-Histidine (Cheng, Westler et al. 1995).

Maltose Binding Protein MS2 Binding Protein (MBP-MS2BP) was expressed and purified using both a MBP and Heparin column as described (Zhou, Sim et al. 2002). After elution from the heparin column fractions were pooled and concentrated using a 30K Amicon Ultra-15 filter (Millipore, Billerica, MA). Gel filtration (Superdex 200, GE Healthcare, Waukesha, WI) in 25 mM Tris (pH 7.4), 200 mM NaCl, 40 mM Maltose and 1 mM EDTA was used for the final step of purification. Samples were concentrated to approximately 1 mg/ml using a 30K Amicon Ultra-15 filter (Millipore, Billerica, MA) and stored at $-20^0$C. For the RNA pull-down experiment, samples were diluted in RNA

buffer (20 mM HEPES (pH 7.4), 100 mM NaCl, 2 mM $MgCl_2$, and 5% glycerol) to a concentration of ~0.15 mg/ml.

**Analytical Ultracentrifugation** Purified *Sp* Cdc5-D3 was run in an Optima XLI ultracentrifuge (Beckman Coulter, Brea, CA) equipped with a four-hole An-60 Ti rotor at 42,000 RPM at 4˚C. Samples were loaded into double-sector cells (path length of 1.2 cm) with charcoal-filled Epon centerpieces and sapphire windows. Sedfit (version 12.0)(Schuck 2000) was used to analyze velocity scans using every seven scans from a total of 360 scans. Approximate size distributions were determined for a confidence level of $p = 0.95$, a resolution of $n = 300$, and sedimentation coefficients between 0 and 15 S.

**NMR Spectroscopy** NMR experiments were performed at 25°C in a 3 mm NMR tube (Wilmad Lab Glass, Vineland, NJ). Four-channel Bruker AVIII 600 and 800 NMR spectrometers (Bruker, Billerica, MA) equipped with CPCQCI and CPTCI probes, respectively, and single axis pulsed-field gradients were used. The assignment of backbone resonances for Cdc5-D3 was completed using standard 2D sensitivity enhanced echo/antiecho $^{1}H$-$^{15}N$ heteronuclear single quantum coherence (HSQC) (Lewis, Keifer et al. 1992, Palmer, Cavanagh et al. 1992, Grzesiek and Bax 1993, Schleucher, Schwendinger et al. 1994), and 3D HNCO (Grzesiek and Bax 1992), HNCA (Grzesiek and Bax 1992), CBCA(CO)NH (Grzesiek and Bax 1993), HNCACB (Wittekind and Mueller 1993), HN(CA)CO (Clubb, Thanabal et al. 1992), HN(CO)CA (Grzesiek and Bax 1992), $^{15}N$-edited NOESY ($\tau = 120$ ms) (Davis, Keeler et al. 1992) experiments. The spectra were referenced to DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) at 0 ppm (parts per million).

**NMR analysis, TALOS+ and Chemical Shift Rosetta** Data were processed in Topspin 3.2 (Bruker, Billerica, MA) and analyzed with Sparky (T. D. Goddard and D. G. Kneller, University of California, San Francisco). Complete backbone $^{1}H$ and $^{15}N$ resonance assignments were obtained for all residues except amino acids E188 and the GSH residues remaining from the N-terminal His tag. The chemical shifts of H, N, Cα, Cβ and CO were analyzed with TALOS+ (Shen, Delaglio et al. 2009), a chemical shift index software that predicts secondary structure elements. An online version of Chemical Shift (CS)-Rosetta (https://condor.bmrb.wisc.edu/rosetta/) (Shen, Lange et al. 2008, Shen, Vernon et al. 2009) was also used to generate 3,000 models using the H, N, Cα, Cβ and CO backbone chemical shift data. The flexible C-terminus of D3 (aa 201-214) was removed for CS-Rosetta modeling to allow for model convergence.

**RNA and DNA Binding Assays** RNAs with a 5′ Biotin were ordered from Dharmacon (GE Healthcare, Waukesha, WI) (Table 2-3) and were 2′-bis(2-Acetoxyethoxy)methyl (ACE) deprotected according to the manufacturer's protocol. RNAs were further purified by ethanol precipitation and two 80% ethanol washes. RNA was heated to 94˚C for 10 minutes then placed on ice before use. DNAs with a 5′ Biotin were ordered from Operon (Eurofins Genomics, Huntsville, AL) (Table 2-3). DNAs were annealed by heating to 94˚C for 10 minutes and then slowly cooled to room temperature. Streptavidin agarose (Life Technologies, Grand Island, NY) blocked with Bovine Serum Albumin (BSA) (Sigma, St. Louis, MO) was incubated with 400µls of 50 µM RNA or DNA for 45 minutes. The RNA or DNA-resin was washed in 10 mM HEPES (pH 7.5), 100 mM NaCl and 10 mM $MgCl_2$ before the addition of ~0.15 mg/ml recombinant N-terminal domains

**Table 2-3.** RNA and DNA oligos used in Chapter II

| RNA # | RNA Name | RNA Sequence | Figure |
|---|---|---|---|
| RNA001 | U2 ssRNA | AAGUGUAG | 2-11 |
| RNA002 | U6 ssRNA | CAGAGAA | 2-11 |
| RNA003 | U6-ISL dsRNA | UGGCCCCUGCACAAGGAUGACA | 2-10 |
| RNA004 | MS2 dsRNA | CGUACACCAUCAGGGUACG | 2-11 |
| RNA005 | Bi-U2 ssRNA | Bi-AAGUGUAG | 2-8 |
| RNA006 | Bi-U6 ssRNA | Bi-CAGAGAA | 2-8 |
| RNA007 | Bi-U6-ISL dsRNA | Bi-UGGCCCCUGCACAAGGAUGACA | 2-8 |
| RNA008 | Bi-MS2 dsRNA | Bi-CGUACACCAUCAGGGUACG | 2-8 |

| | | | |
|---|---|---|---|
| DNA001 | Bi-U2 ssDNA | Bi-AAGTGTAG | 2-9 |
| DNA002 | Bi-U6 ssDNA | Bi-CAGAGAA | 2-9 |
| DNA003 | Bi-U6-ISL dsDNA | Bi-TGGCCCCTGCA | 2-9 |
| DNA004 | U6-ISL dsDNA | CAAGGATGACA | 2-9 |
| DNA005 | Bi-MS2 dsDNA | Bi-CGTACACCAT | 2-9 |
| DNA006 | MS2 dsDNA | CAGGGTACG | 2-9 |

of Cdc5 for 45 minutes at room temperature. New RNA or DNA-resin was made for each replicate experiment. The resin was then washed in RNA buffer (20 mM HEPES (pH 7.4), 100 mM NaCl, 2 mM $MgCl_2$, and 5% glycerol). For protein-RNA or DNA pull down experiments lithium dodecyl sulfate (LDS) sample buffer (Invitrogen, Carlsbad CA) was added directly to the resin that was then boiled except in the case of Cdc5-R1-R2, due the presence of a contaminant on the resin at the same molecular weight. Cdc5-R1-R2 was eluted from the RNA-resin using RNA buffer with 1 M NaCl. Samples were then treated with LDS and boiled as above. Samples were run on 4-12 % Bis-Tris PAGE gels (Invitrogen, Carlsbad CA) and stained with Colloidal Coomassie (Dyballa and Metzger 2009). All gels were quantified using ImageJ (Rasband 1997-2004). Average values were determined by comparing the intensity of each individual pull-down relative to the amount of protein loaded (relative to input %). Although adding excess RNA or DNA to the streptavidin agarose to saturate binding, we are unable to correct for the amount of RNA or DNA bound to the resin in these experiments. Graphs and statistics were generated using GraphPad Prism (version 5.0a).

**NMR RNA Titration** RNAs (Table 2-3) were ordered from Dharmacon (GE Healthcare) and were deprotected and washed per manufacturer's instructions. RNAs were re-suspended in S200-RNA buffer (25 mM MES (pH 6.0), 300 mM NaCl, 2 mM MgCl2 and 1 mM EDTA). For RNA titrations, Cdc5-D3 was at ~135 μM and the RNA concentrations were titrated at 1:0.25, 1:0.5, 1:1, 1:2, 1:3, and 1:5 protein to RNA molar ratios for U6-ISL dsRNA and examined at a 1:1 protein to RNA molar ratio for U2

ssRNA, U6 ssRNA, and MS2 dsRNA. Chemical shift differences were calculated using the following formula:

$$\Delta\delta_{\mathrm{Total}} = \sqrt{\left(\frac{\delta^{15}\mathrm{N_B} - \delta^{15}\mathrm{N_{B_0}}}{9.86204}\right)^2 + (\delta\mathrm{H_B} - \delta\mathrm{H_{B_0}})^2}$$

Where B equals the chemical shift at variable RNA concentrations and $B_0$ is the chemical shift of protein only. 9.86204 is the absolute ratio of the gyromagnetic constants of $^{1}H$ and $^{15}N$ (Martin-Tumasz, Reiter et al. 2010). Changes in chemical shifts were fit to a single site-binding curve using GraphPad Prism (version 5.0a) and the following formula:

$$\mathrm{Y} = \frac{\mathrm{Bmax} * \mathrm{X}}{(K_d + \mathrm{X})}$$

Where Bmax is the maximum binding and $K_d$ is the equilibrium binding constant(Barrett, Song et al. 2012).

**Circular Dichroism (CD)** Purified Cdc5-D3 was analyzed using a Jasco J-810 spectropolarimeter (Jasco Analytical Instruments, Easton, MD). Far-UV (ultraviolet) data were collected at a protein concentration of 0.15 mg/ml in a 1 mm quartz cuvette. Spectra were collected with an average time of 4 s for each point and a step size of 50 nm/min from 198 to 260 nm. Far-UV spectra were collected in quadruplicate and background-corrected against a buffer blank. Data were converted to mean residue ellipticity $[\theta]_m$ (degrees $cm^2dmol^{-1}$) using the following formula:

$$[\theta]_m = \frac{\theta}{(10 * l * c * n)}$$

Where $\theta$ is the measured ellipticity, $l$ is the cell path length in cm, $c$ is the molar concentration of protein in moles/liter, and $n$ is the number of amino acids.

Circular dichroism data was deposited in the PCDDB (PCDDB ID: 0004551000). Backbone resonance assignments have been deposited in the BMRB (accession no. 25084).

**Results**

**R1, R2 and D3 are essential for *S. pombe* Cdc5 function**

Although *cdc5* and *CEF1* deletion studies in *S. pombe* and *S. cerevisiae* have been reported (Ohi, McCollum et al. 1994, Tsai, Chow et al. 1999), most of the deletion mutants used in these analyses did not directly correspond to secondary structural elements making it difficult to determine the specific regions required for function. Using sequence alignments and secondary structure predictions, we identified five structural elements in the Cdc5 N-terminus, including the two Myb repeats (R1 and R2), a predicted Myb-like repeat (D3) (Ohi, Feoktistova et al. 1998), a non-structured loop (L) between R2 and D3, and a predicted a-helical region (MID) downstream from D3 (Fig. 2-1A). As a first step towards characterizing the role of the Cdc5 N-terminus in pre-mRNA splicing, we attempted to construct five *S. pombe* strains where the only copy of *cdc5* lacked one of the identified structural elements (Fig. 2-1A). Although strains containing only $cdc5^{\Delta L}$ or $cdc5^{\Delta MID}$ grew normally at all temperatures (Fig. 2-1B), we were unable to recover viable strains relying on only the expression of $cdc5^{\Delta R1}$, $cdc5^{\Delta R2}$, or $cdc5^{\Delta D3}$ (Fig. 2-1C). Importantly, the R1, R2, and D3 deletions did not destabilize Cdc5 as seen from Western blot analysis of wild-type cells overexpressing these mutants (Fig. 2-2). From these results we conclude that both of the canonical Myb repeats (R1 and R2), as well as Cdc5-D3, are essential for Cdc5 function, while the loop between R2 and D3 and the

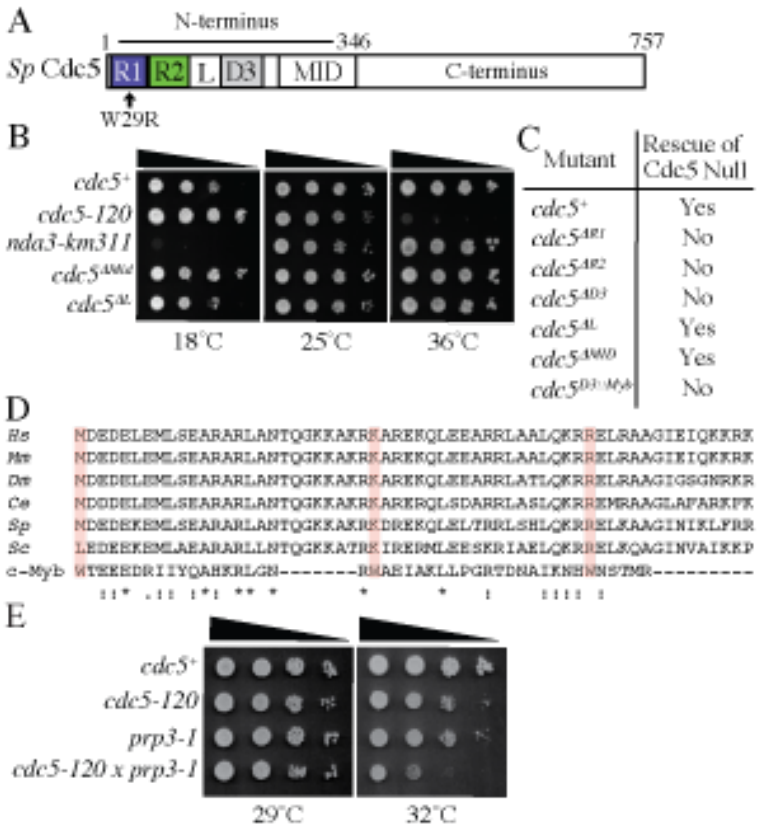



**Figure 2-1. The Canonical Myb repeats (R1, R2) and the Proposed Myb-like Repeat (D3) are Required for Function** (A) Domain architecture of *Sp* Cdc5. The Cdc5 N-terminus (aa 1-346) contains two canonical Myb repeats (R1, R2) and a third domain (D3) that is predicated to be a Myb-like repeat. R2 and D3 are separated by a loop region (L) predicted to be unstructured. A middle region of Cdc5 (MID) is predicted by PSI-PRED (Buchan et al. 2013) to be α-helical. Position of domains: R1 (aa 5-55); R2 (aa 58-111); Loop (aa 111-146), D3 (aa 155-214), and MID (aa 237-346). The Cdc5 C-terminus (aa 347-757) contains no predicted structural motifs. Black arrow marks the position of the mutation (W29R) found in temperature sensitive *cdc5-120*. (B) The loop and predicated α-helical regions of the Cdc5 N-terminus are not essential. Growth of serial dilutions of integrated $cdc5^{\Delta L}$ or $cdc5^{\Delta MID}$ cells grown at 18˚C, 25˚C and 36˚C. *cdc5-120* is a heat sensitive mutant, while *nda3-km11* is a cold sensitive mutant(100). (C) Summary of N-terminal deletions in *Sp* Cdc5. *cdc5* deletions were determined to be essential by assaying for ability to rescue the *cdc5* null. (D) Sequence alignment of Cdc5-D3 from yeast to humans compared with *Mus musculus* (*Mm*) c-Myb R3. *Hs, Homo sapiens*; *Dm, Drosophila melanogaster*; *Ce, Caenorhabditis elegans*; *Sp, Schizo-saccharomyces pombe*; *Sc, Saccharomyces cerevisiae*. The red boxes mark the location of the conserved tryptophan (W) residues in c-Myb and show that these tryptophans are not conserved in Cdc5-D3. "." represents weakly similar aa's, ":" represents strongly similar aa's, and "*" represents identical aa's. (E) *cdc5-120* and *prp3-1* (*Sp* cwf2) interact genetically. Growth of serial dilutions of wild-type, *cdc5-120*, *prp3-1*, and *cdc5-120 prp3-1* cells grown at 29˚C and 32˚C.

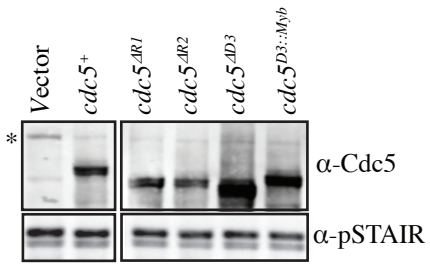



**Figure 2-2. Cdc5 Deletions are Stable** Western blot of a Cdc5-TAP strain overexpressing *cdc5*$^{+}$, *cdc5*$^{\Delta R1}$, *cdc5*$^{\Delta R2}$, *cdc5*$^{\Delta D3}$, and *cdc5*$^{\Delta D3::Myb}$ under control of the nmt+ promoter in pREP3X. Deletion constructs were overexpressed in the Cdc5-TAP background to easily distinguish between the endogenous and overexpressed protein. "*" marks the position of endogenous Cdc5-TAP. Top panels, anti-Cdc5 Western blot. Bottom panels, anti-pSTAIR Western blot.

MID region are not. Although it was previously demonstrated that mutations in both canonical Myb repeats R1 and R2 affect pre-mRNA splicing (Nurse, Thuriaux et al. 1976, Ohi, Feoktistova et al. 1998, Query and Konarska 2012), this is the first evidence that Cdc5-D3 is also required for Cdc5 function.

Myb repeats are approximately 50 amino acids in length with a tryptophan or tyrosine residue every 18-19 amino acids. The highly conserved tryptophan or tyrosine residues make up the hydrophobic core of a Myb repeat and are important for proper folding (Kanei-Ishii, Sarai et al. 1990, Saikumar, Murali et al. 1990, Ogata, Hojo et al. 1992). Structurally this domain is composed of three well-defined helices with the second and third helices adopting a fold similar to the canonical helix-turn-helix motif (HTH) found in many DNA binding proteins (Sauer, Yocum et al. 1982, Ohlendorf, Anderson et al. 1983, Ogata, Hojo et al. 1992). Cdc5-R1 and Cdc5-R2 adopt a canonical Myb-fold as determined by NMR analysis (PDB: 2DIM and 2DIN). While Cdc5-D3 has been predicted to be a Myb-like domain (Ohi, Feoktistova et al. 1998), the lack of highly conserved tryptophan residues usually found in Myb domains (Fig. 2-1D) suggests that Cdc5-D3 could adopt a different fold. To test whether the D3 region in Cdc5 could be replaced with a Myb fold, we replaced *cdc5-D3* with the sequence of Myb repeat 3 (R3) from *Mus musculus* (*Mm*) *c-MYB* (*cdc5*$^{D3::Myb}$) and asked whether this mutant could support cell function. The R3 Myb repeat from *Mm* c-Myb was chosen to replace Cdc5-D3 since it has been structurally characterized (Ogata, Hojo et al. 1992) and has a similar pI (isoelectric point) as Cdc5-D3 (10.4 and 10.7 respectively). As was seen with the *cdc5*$^{\Delta R1}$, *cdc5*$^{\Delta R2}$, and *cdc5*$^{\Delta D3}$ mutants, we were unable to recover a viable strain relying

solely on the domain swap mutant (Fig. 2-1C). The inability of a canonical Myb repeat to substitute for Cdc5-D3 in cells suggested that the functional surfaces present on Cdc5-D3 are essential for function and cannot be replaced with a structural Myb domain.

Recently another NTC component, *Sc* Cwc2 (*Sp* Cwf2/Prp3), was shown to interact directly with the U6 snRNA near the active site of the spliceosome (McGrail, Krause et al. 2009, Rasche, Dybkov et al. 2012, Li, Zhang et al. 2013). Interestingly both *Sc* Cef1 (*Sp* Cdc5) and *Sc* Cwc2 (*Sp* Cwf2/Prp3), the two NTC components with nucleic acid binding domains, interact directly with Prp19 (Ohi and Gould 2002, Vander Kooi, Ren et al. 2010), putting them in close physical proximity to each other and suggesting they could both interact with RNA near the catalytic core of the spliceosome. To investigate if there is a functional connection between Cdc5 and Cwf2/Prp3, we tested for a genetic interaction between *cdc5-120* and *prp3-1*. The *cdc5-120* mutation causes the amino acid substitution W29R in the R1 domain (Fig. 2-1A) (Ohi, Feoktistova et al. 1998), while *prp3-1* (Ohi and Gould 2002) causes the amino-acid substitution G123R in the ZnF domain, a domain that has been shown to cross-link to RNA in *S. cerevisiae* (Schmitzová, Rasche et al. 2012). Our analysis shows that *cdc5-120 prp3-1* cells are synthetically sick (Fig. 2-1E), suggesting that Cdc5 and Cwf2/Prp3 may function at similar stages of the splicing reaction.

**Secondary structure analysis of Cdc5-D3 shows this domain is composed of two α-helices and does not adopt a canonical Myb fold**

To examine the structural characteristics of Cdc5-D3, we expressed and purified recombinant Cdc5-D3 (aa 155-214) from *E. coli* (Fig. 2-3A). Analysis by sedimentation

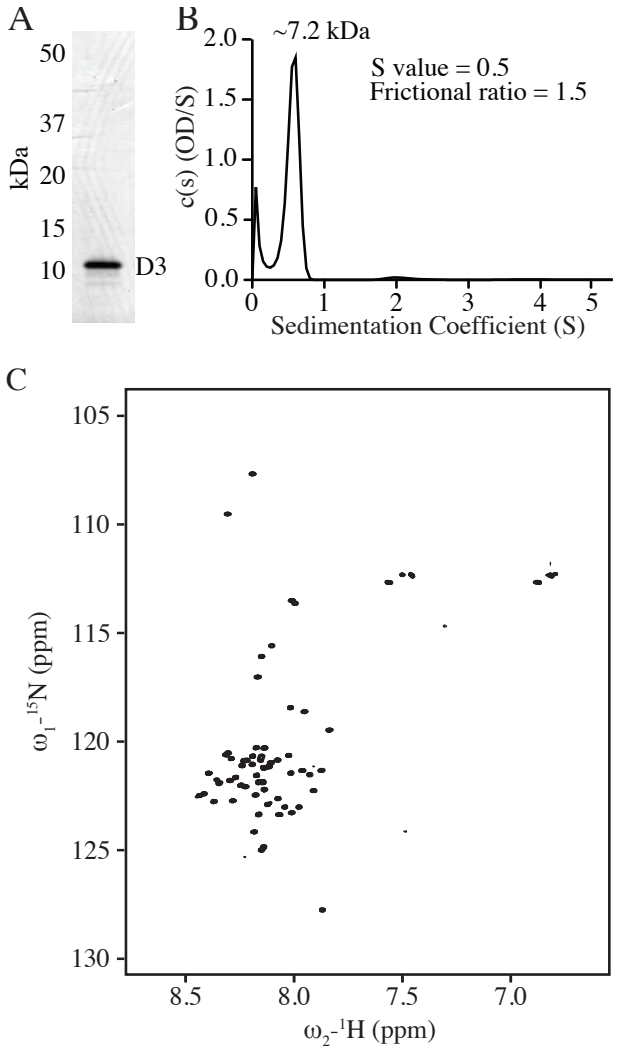



**Figure 2-3. Biophysical Characterization of Cdc5-D3.** (A) Coomassie stained SDS-PAGE of Cdc5-D3 (aa 155-214). Cdc5-D3 runs at a higher molecular weight than what is predicted because of its pI. (B) SVAU analysis of Cdc5-D3. The S value, frictional ratio, and determined molecular mass are given for the main peak which is ~93% of the sample. The RMSD is 0.004. (C) $^{15}$N-$^{1}$H HSQC spectra of Cdc5-D3. Peaks for 59 of the 63 (94%) expected residues were identified.

velocity analytical ultracentrifugation (SVAU) shows that Cdc5-D3 sediments as a monomer (s=0.5; predicted molecular mass, ~7.2kDa; root mean square deviation [RMSD] = 0.004) with a frictional ratio of 1.5 (Fig. 2-3B). Circular dichroism (CD) analysis of Cdc5-D3 using far-UV light was done to predict secondary structure. When comparing this spectrum to known spectra (Greenfield and Fasman 1969), Cdc5-D3 is primarily α-helical but contains random coil as indicated by the lower signal at 222nm (Fig. 2-4). To further examine the secondary structure of this domain, we $^{15}$N-labeled Cdc5-D3 and collected a two-dimensional (2D) $^{15}$N-$^{1}$H HSQC experiment using nuclear magnetic resonance (NMR) spectroscopy (Fig. 2-3C). In this spectrum, peaks were observed for ~94% of the expected amino acids in Cdc5-D3 (59 residues out of a total of 63, Fig. 2-5) making Cdc5-D3 amenable for NMR secondary structure analysis.

To determine the secondary structure of Cdc5-D3, we used a combination of NMR experiments and computational modeling. Using $^{13}$C- and $^{15}$N-labeled Cdc5-D3, we ran a series of two-dimensional (2D) and three-dimensional (3D) NMR experiments to determine the backbone connectivity of Cdc5-D3. Results from a $^{15}$N-NOESY-HSQC experiment showed there were two as opposed to three helical regions in Cdc5-D3 (Fig. 2-6A). We then used TALOS+ to predict the secondary structure of Cdc5-D3 using the NH, Cα, Cβ, CO, and N chemical shift data. This analysis confirmed that Cdc5-D3 contains two α-helical regions (Fig. 2-6A), rather than the three α-helices that would be expected for a Myb repeat (Fig. 2-7A-B). To generate a 3D model of Cdc5-D3, we used the backbone chemical shift data and Chemical-Shift Rosetta (CS-Rosetta) (Shen, Vernon et al. 2009, Shen and Bax 2012). For calculations of 3D models, the chemical shift values

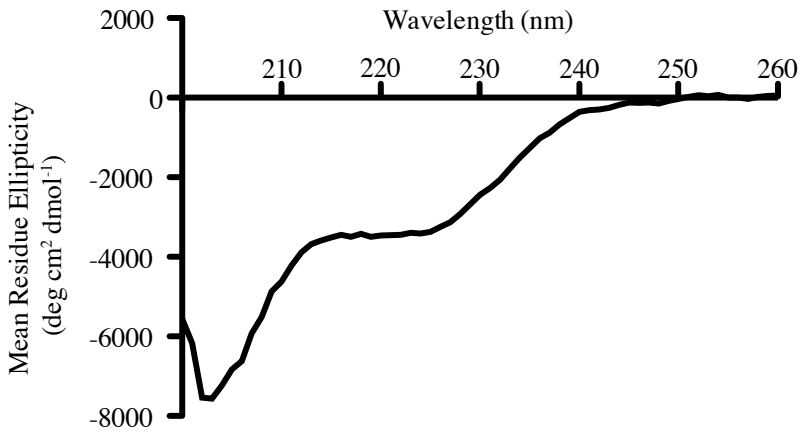



**Figure 2-4. Circular Dichroism Spectrum of Cdc5-D3** FAR-UV circular dichroism (CD) spectrum of Cdc5-D3 (aa 155-214).

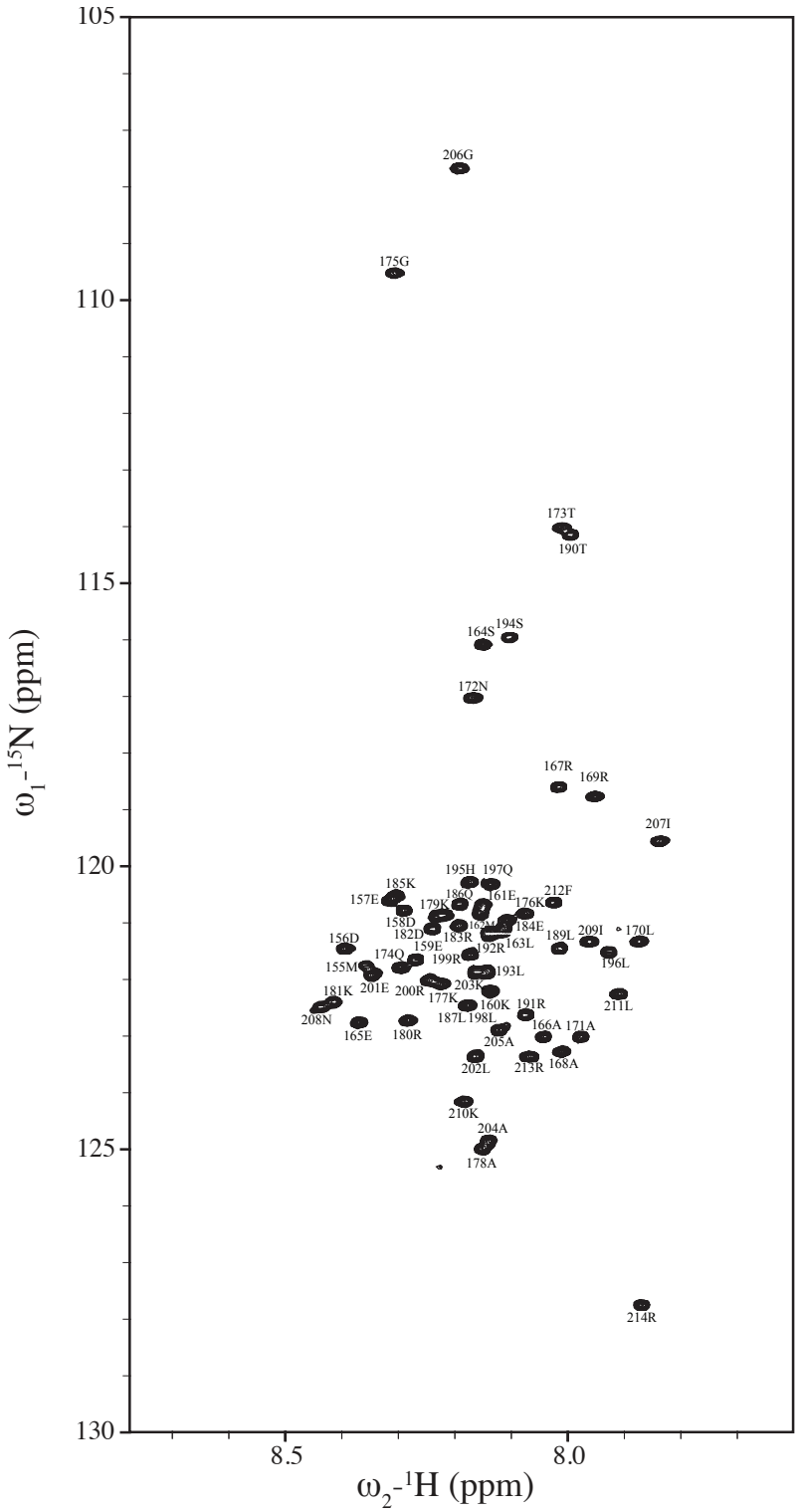



**Figure 2-5. Labeled 2D HSQC of Cdc5-D3** Each peak in the Cdc5-D3 HSQC is labeled with its corresponding amino acid assignment.

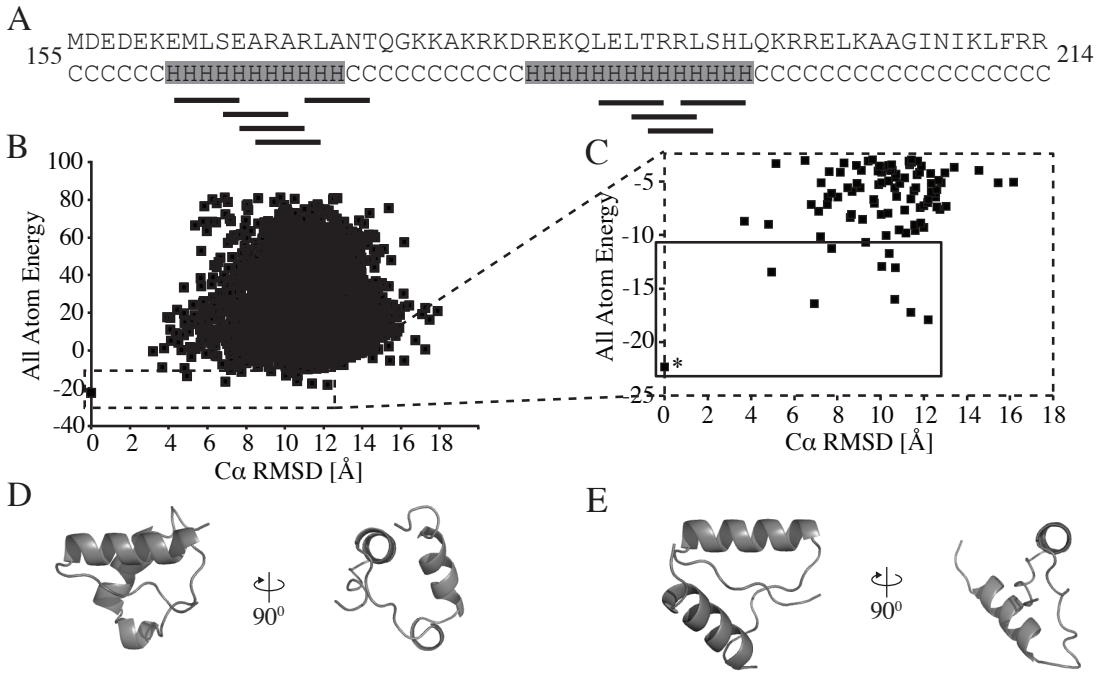



**Figure 2-6. Secondary Structure Analysis and 3D Modeling of Cdc5-D3 using Backbone Chemical Shifts** (A) Secondary structure analysis of Cdc5-D3 using TALOS+. Top row is the primary sequence for Cdc5-D3. The TALOS+ prediction is shown in the bottom row where "C" represents random coil and "H" represents helical. Gray highlights the predicted helical regions. The black lines correspond to the helical regions found using a $^{15}$N-NOESY-HSQC. Each bar represents an i,i$^{+4}$ or an i,i$^{-4}$ NOE. (B) Summary of 3,000 CS-Rosetta models of Cdc5-D3, comparing the all atom energy versus the Cα RMSD relative to the lowest energy model. The black dashed box corresponds to the 100 lowest energy structures. (C) Magnification of the 100 lowest energy structures from panel B. Solid black line encompasses the ten lowest energy structures. The "*" marks the lowest energy structure. (D and E) Two of the ten lowest energy models calculated by CS-Rosetta shown at 0˚ and 90˚. Panel D is the lowest energy model.

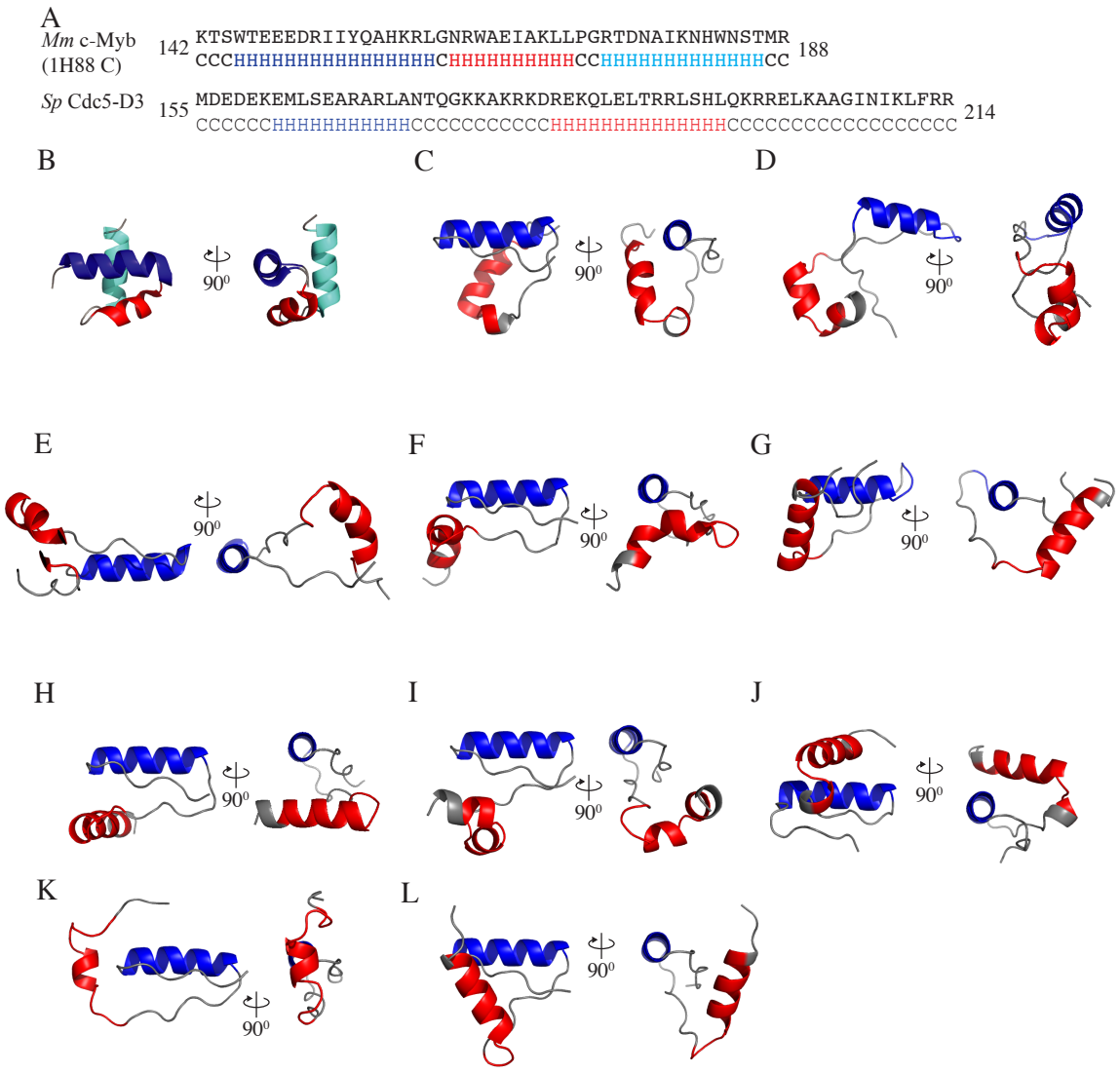



**Figure 2-7. Secondary Structure Analysis of Cdc5-D3** (A) Secondary structure of *Mm* c-Myb R3 compared to *Sp* Cdc5-D3. α-helix 1 colored blue, α-helix 2 colored red, and α-helix 3 (if applicable) colored cyan. (B) Solution structure of *Mm* c-Myb PDB: 1H88 chain C (aa 144-188). (C-L) The ten lowest energy models of Cdc5-D3 calculated using CS-Rosetta. Helices are colored according to the secondary structure from the NOESY and Talos+ experiments (Fig 2-6A). Helices colored as in Figure 2-7A. Structures are shown at 0˚ and 90˚.

for residues 201-214 were not included due to their predicted flexibility as calculated by TALOS+ and the lack of NOEs for these residues (Fig. 2-6A). CS-Rosetta generated 3,000 models o Cdc5-D3 and when these models were plotted with their all atom energy versus their Cα RMSD in angstroms (Å), the models converged on the lowest energy model (Fig. 2-6B-C). The ten lowest energy models contain two α-helices (Fig. 2-6D-E and 2-7C-L) and do not resemble a Myb fold (Fig. 2-7B). Strikingly, all of the lowest 100 models contained two rather than three α-helices, leading us to conclude that Cdc5-D3 does not adopt a canonical Myb fold and should not be referred to as a Myb-like repeat.

**The N-terminus of *Sp* Cdc5 binds RNA**

As an essential pre-mRNA splicing factor and a core component of the NTC, Cdc5 family members in *S. pombe*, *S. cerevisiae* and humans associate with the spliceosome starting with assembly of the B-complex and remain bound throughout the entire splicing reaction (Jurica, Licklider et al. 2002, Makarov, Makarova et al. 2002, Ohi, Link et al. 2002, Stevens, Ryan et al. 2002, Deckert, Hartmuth et al. 2006, Bessonov, Anokhina et al. 2008, Fabrizio, Dannenberg et al. 2009, Warkocki, Odenwalder et al. 2009, Bessonov, Anokhina et al. 2010, Lardelli, Thompson et al. 2010, Ren, McLean et al. 2011, Cvitkovic and Jurica 2013, Fourmann, Schmitzová et al. 2013). Because Cdc5 associates with the spliceosome, we hypothesize that the N-terminus may interact and bind to RNA. However, although the N-terminus of Cdc5 family members has been reported to bind DNA *in vitro* (Ohi, McCollum et al. 1994, Hirayama and Shinozaki 1996, Lei, Shen et al. 2000), the ability of Cdc5 to interact with RNA has not been reported. Cdc5, as a core member of the NTC, is required for formation of the

catalytic core(Chan and Cheng 2005), suggesting that Cdc5 may interact with core regions of the spliceosome required for catalysis. This possibility is supported by the ability of mutations in the *Sc* Cef1 N-terminus to suppress first and second step splicing defects of a variety of mutants, including mutations in introns and the U6 snRNA (Query and Konarska 2012). To test the ability of the Cdc5 N-terminus to bind RNA, we used well-characterized regions of the U2 and U6 snRNAs that are predicted to be near the catalytic core of the spliceosome (Madhani and Guthrie 1992, Lesser and Guthrie 1993, Sontheimer and Steitz 1993, Yean, Wuenschell et al. 2000) (Table 2-3). These RNAs include the double-stranded U6 snRNA inter-stem loop (U6-ISL), a single-stranded region of the U6 snRNA involved in 5′ splice-site selection (U6 ssRNA), and a single-stranded region of the U2 snRNA involved in branchpoint recognition (U2 ssRNA) (Madhani and Guthrie 1992, Sashital, Cornilescu et al. 2004, Burke, Sashital et al. 2012, Montemayor, Curran et al. 2014) (Fig. 2-12A). The MS2 RNA hairpin (MS2 dsRNA) from bacteriophage was also prepared to analyze the sequence specificity of dsRNA binding (Zhou, Sim et al. 2002). To test for a direct interaction with the Cdc5 N-terminus, biotinylated RNAs bound to streptavidin agarose beads were incubated with recombinant Cdc5-R1-R2-D3$^{\Delta L}$. The non-essential loop region (Fig. 2-1A-C) in this construct was deleted to generate a more stable protein. To detect binding, protein that remained bound to the resin after multiple washes was visualized by Coomassie staining (Fig. 2-8A) and the percentage of protein that remained bound as compared to the original input was quantified using the results from multiple binding assays (Fig. 2-8B). As seen in Figure 8A and B, Cdc5-R1-R2-D3$^{\Delta L}$ directly interacts with all the tested RNAs. The N-terminus

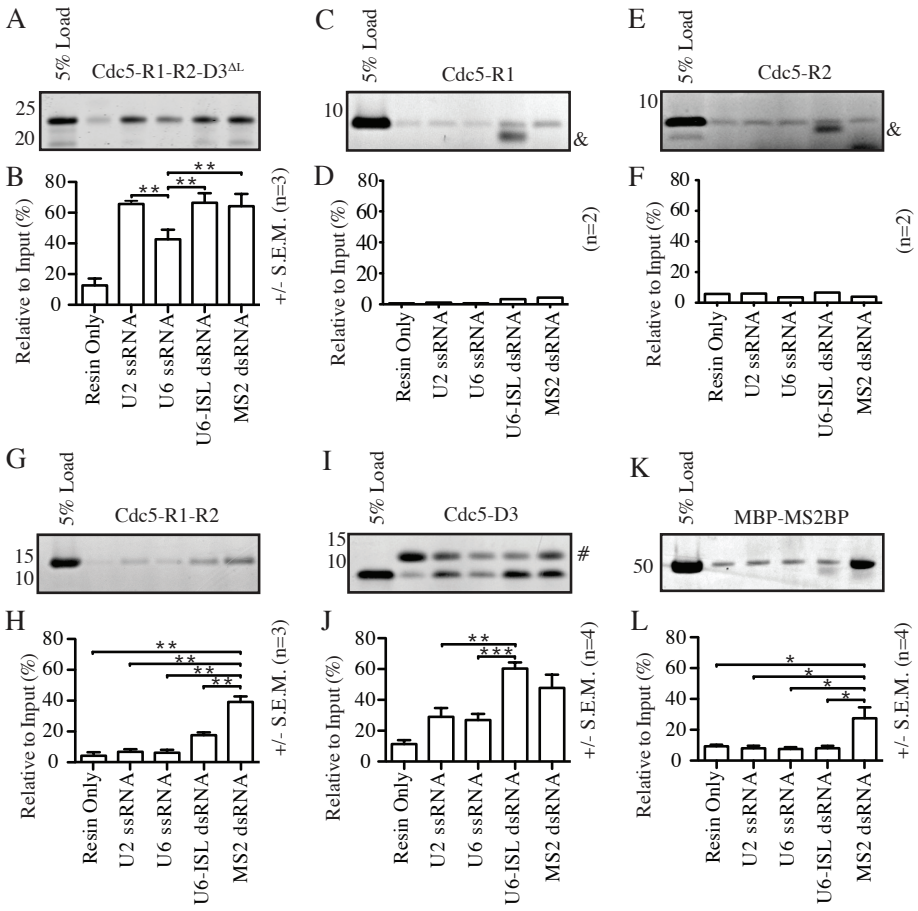



**Figure 2-8. The N-terminus of Cdc5 Binds RNA *In Vitro*** Streptavidin-RNA pull-down assays using the indicated recombinant protein. Panels A, C, E, G, I, and K are Coomassie-stained SDS-PAGE gels of a representative pull-down experiment. Panels B, D, F, H, J and L show the quantification of binding from multiple experiments. (A-B) Coomassie-stained SDS-PAGE gel of Cdc5 R1-R2-D3$^{\Delta L}$ bound to RNA agarose beads and quantification of three binding assays. (C-D) Coomassie-stained SDS-PAGE gel of Cdc5-R1 that binds to RNA agarose beads and average of two binding assays. "&" marks the position of a contamination band. (E-F) Coomassie-stained SDS-PAGE gel of Cdc5-R2 bound to RNA agarose beads and average of two binding assays. "&" marks the position of a contamination band . (G-H) Coomassie-stained SDS-PAGE gel of Cdc5-R1-R2 bound to RNA agarose beads and quantification of three binding assays. (I-J) Coomassie stained SDS-PAGE gel of Cdc5-D3 bound to RNA agarose beads and quantification of four binding assays. "#" represents contamination from the streptavidin resin. (K-L) Coomassie stained SDS-PAGE gel of MBP tagged MS2 binding protein (MBP-MS2BP) that binds to RNA agarose beads and quantification of four binding assays. SDS-PAGE gel quantification was done using ImageJ. Molecular weight markers shown to the left of each gel. Error bars and statistics were generated using GraphPad Prism (version 5.0a) where the error bars correspond to the standard error of the mean (S.E.M.) and the p-values are represented by "*" ($p < 0.05$ = *, $p < 0.01$ = ** and $p < 0.001$ = ***). 5'-biotinylated RNA sequences used in the RNA binding experiment can be found in Table 2-3.

of Cdc5 is composed of three essential domains, the two canonical Myb repeats (R1 and R2) and D3 (Fig. 2-1A and C). To determine if the canonical Cdc5 Myb domains (R1 and R2) can individually interact with RNA, we repeated the RNA binding experiment using recombinant Cdc5-R1 and Cdc5-R2 (Fig. 2-8C-F). In this assay neither R1 nor R2 binds RNA (Fig. 2-8C-F). However, since Myb repeats are often found as clustered groups (Prouse and Campbell 2012), we also investigated whether Cdc5-R1-R2 could bind RNA (Fig. 2-8G-H). While Cdc5-R1-R2 does bind RNA (Fig. 2-8G), we were surprised to find that when comparing the percentage of protein bound to the RNA (as compared to the initial input), only ~20% of Cdc5-R1-R2 versus over 60% of Cdc5-R1-R2-D3$^{\Delta L}$ was bound to the U6-ISL (Fig. 2-8B and H). Thus, although Cdc5-R1-R2 interacts with RNA, these two domains do not replicate the RNA binding seen with the entire Cdc5 N-terminus. Finally, since the N-terminus of Cdc5 family members have been reported to bind DNA (Ohi, McCollum et al. 1994, Hirayama and Shinozaki 1996, Lei, Shen et al. 2000), we also tested the ability of Cdc5-R1-R2-R3$^{\Delta L}$ to interact with DNA, which it does (Fig. 2-9A-B). Our RNA binding results suggests that the Cdc5-D3 region serves a more prominent role in facilitating RNA interactions than the canonical Myb repeats or, alternatively, all three domains are required for full RNA binding activity.

To discriminate between these possibilities, the binding experiment was repeated using Cdc5-D3 (Fig. 2-8I-J). Cdc5-D3 alone binds directly to all of the RNAs in the assay (Fig. 2-8I) with a significant binding preference for the U6-ISL (Fig. 2-8J). To test whether the spliceosomal RNAs bind non-specifically, we repeated the binding assays

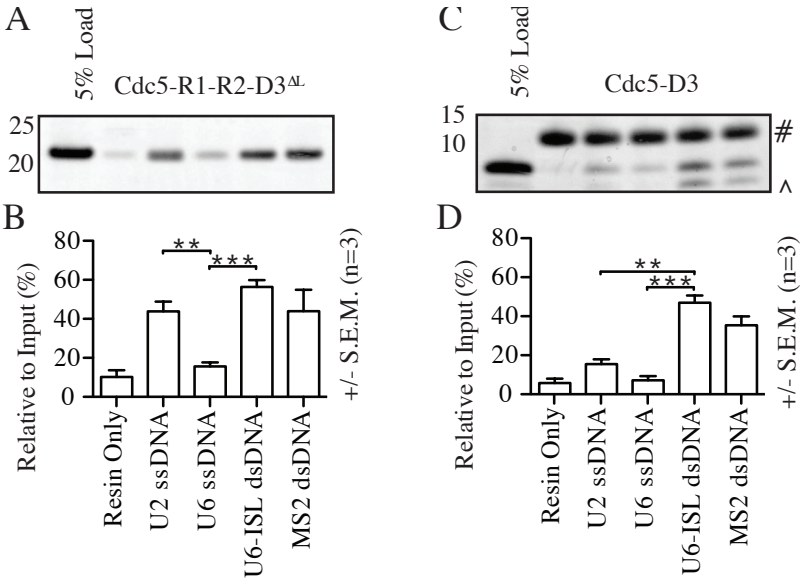



**Figure 2-9. The N-terminus of Cdc5 Binds DNA *In Vitro*** Streptavidin-DNA pull-down assays using the indicated recombinant protein. Panels A and C are Coomassie-stained SDS-PAGE gels of a representative pull-down experiment. Panels B and D show the quantification of binding from multiple experiments. (A-B) Coomassie-stained representative SDS-PAGE gel of Cdc5 R1-R2-D3$^{\Delta L}$ bound to DNA agarose beads and quantification of three binding assays. (C-D) Representative coomassie stained SDS-PAGE gel of Cdc5-D3 bound to DNA agarose beads and quantification of three binding assays. "#" marks contamination band from the streptavidin resin. "^" marks protein degradation product. SDS-PAGE gel quantification was done using ImageJ. Molecular weight markers shown to the left of each gel. Error bars and statistics were generated using GraphPad Prism (version 5.0a) where the error bars correspond to the standard error of the mean (S.E.M.) and the p-values are represented by "*" ($p < 0.05$ = *, $p < 0.01$ = ** and $p < 0.001$ = ***). 5'-biotinylated DNA sequences used in the DNA binding experiment can be found in Table 2-3.

using an N-terminally MBP tagged MS2 dsRNA binding protein (MBP-MS2BP) that interacts with the MS2 RNA hairpin loop (Zhou, Sim et al. 2002). While MPB-MS2BP bound to the MS2-dsRNA as expected, it did not bind to any of the spliceosomal RNAs (Fig. 2-8K-L). In addition, like Cdc5-R1-R2-R3$^{\Delta L}$, Cdc5-D3 also binds DNA (Fig. 2-9C-D). The results of these RNA binding assays demonstrate that the Cdc5 N-terminus binds to RNA directly and that, of the three essential domains, Cdc5-D3 binds RNA more robustly than the canonical Myb repeats alone.

**Cdc5-D3 preferentially binds double-stranded RNA**

To further characterize the interaction of Cdc5-D3 with RNA we performed several NMR RNA titration experiments using the U2 ssRNA, U6 ssRNA, U6-ISL dsRNA, MS2 dsRNA and $^{15}$N Cdc5-D3 (Figs. 2-10A and 2-11A-C). The preference for Cdc5-D3 binding to double-stranded RNAs detected in the *in vitro* binding assays (Fig. 2-8I-J) was confirmed through chemical shift perturbation by examining the HSQC spectra of $^{15}$N-labeled Cdc5-D3 after RNA was added at a 1:1 protein to RNA molar ratio (Figs. 2-10B and 2-11A-C). The interaction between the U6-ISL dsRNA and Cdc5-D3 was further characterized using a titration of RNA into $^{15}$N-labeled protein ranging from substoichiometric molar ratios of protein to RNA (1:0.25) to saturable concentrations (1:5). This RNA titration resulted in substantial chemical shift changes in the NMR spectra for a subset of Cdc5-D3 peaks (Fig. 2-10A), with the strongest shift differences localized to the last 16 amino acids (aa 199-214, Fig. 2-10B and E-F), a region predicted to be unstructured in our secondary structure calculations. The last 16 amino acids also showed strong shifts with the titration of MS2 dsRNA (Fig. 2-11C-E). Therefore it is

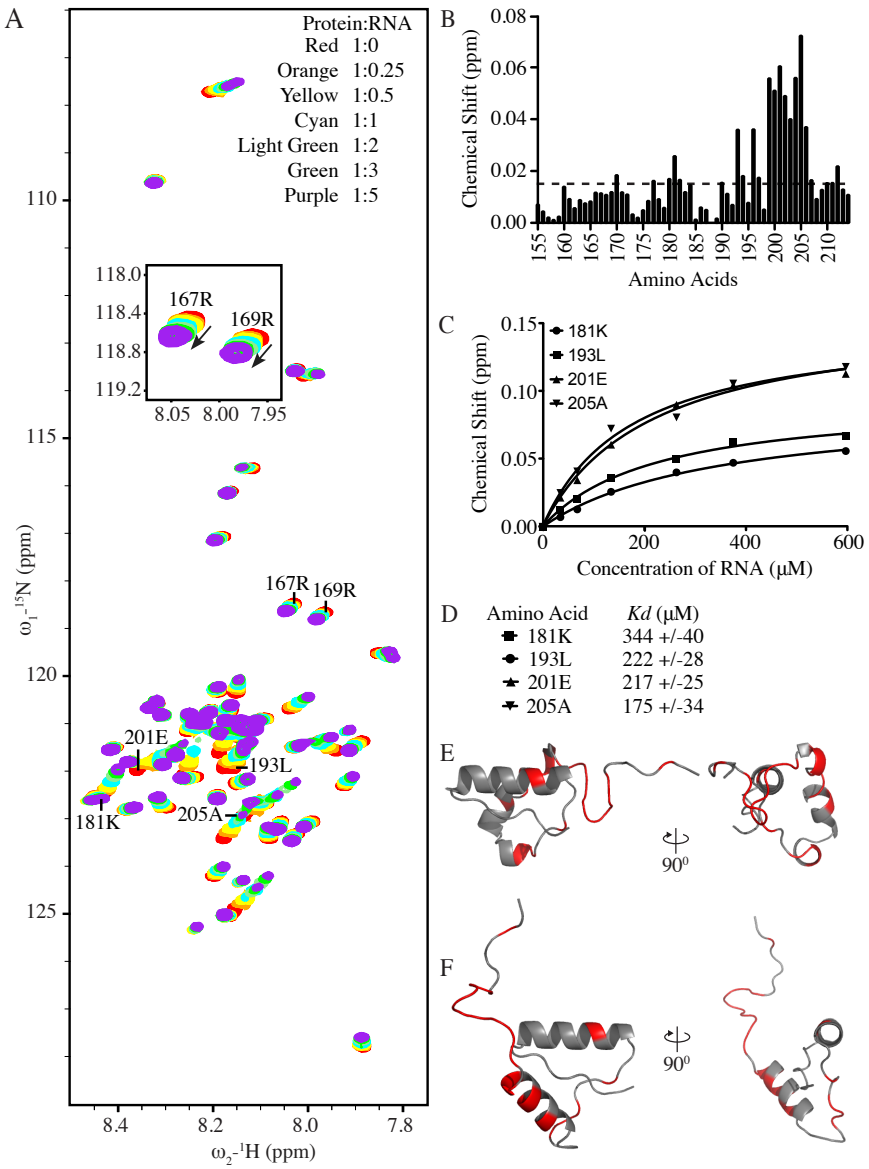



**Figure 2-10. Cdc5-D3 RNA Titration with U6-ISL dsRNA** (A) $^{15}$N-$^{1}$H HSQC spectra of Cdc5-D3 (~138μM *protein concentration changes with the addition of RNA) with varying concentrations of U6-ISL dsRNA (0, 35, 68, 134, 262, 373, and 593μM). The $^{15}$N-$^{1}$H HSQC inset shows amino acids 167R and 169R. The black arrows indicate the direction the peaks shift upon the addition of U6-ISL dsRNA. (B) Chemical shifts (Δ ppm) from $^{15}$N-$^{1}$H HSQC at a 1:1 molar ratio. The twenty strongest chemical shifts are above the dashed line. (C) Saturation curves of amino acids 181K, 193L, 201E and 205A are plotted and fit with a single-site binding curve. (D) Binding affinities (Kd) of 181K, 193L, 201E and 205A. (E-F) Chemical shifts from Cdc5-D3 and U6-ISL dsRNA NMR titration were plotted in red onto two of the ten lowest energy models of Cdc5-D3. Models have an added flexible C-terminal extension that was not included in the CS-Rosetta calculations due to lack of NOEs but do interact with RNA. Models shown at 0˚ and 90˚.

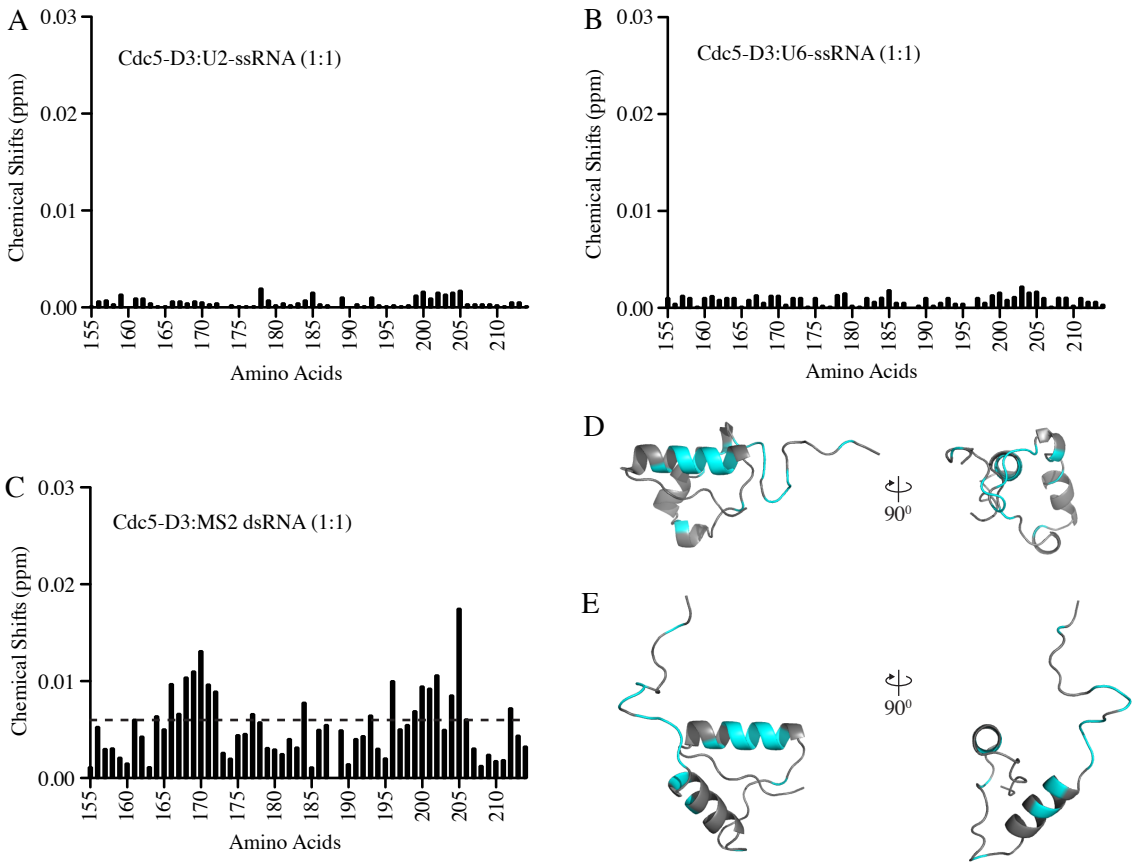



**Figure 2-11. Chemical Shift Analysis of Cdc5-D3 with ssRNA and MS2 dsRNA** All graphs are chemical shifts (Δppm) from $^{15}N$-$^{1}H$ HSQC of Cdc5-D3 at a 1:1 molar ratio with three different RNAs. (A) Cdc5-D3 titrated with U2 ssRNA. (B) Cdc5-D3 titrated with U6 ssRNA. (C) Cdc5-D3 titrated with MS2 dsRNA. The twenty strongest chemical shifts are above the dashed line. (D-E) Chemical shifts from Cdc5-D3 and MS2 dsRNA NMR titration were plotted in cyan onto two of the ten lowest energy models of Cdc5-D3. Models have an added flexible C-terminal extension that was not included in the CS-Rosetta calculations due to lack of NOEs but do interact with RNA. Models shown at 0˚ and 90˚.

possible that this flexible region of Cdc5-D3 may become ordered when bound to RNA. The chemical shift changes appear to saturate at high RNA concentrations and can be fit to a 1:1 binding model (Fig. 2-10C). Dissociation constants ($K_d$) from single site binding curves are ~210 μM (Fig. 2-10C-D). The locations of the Cdc5-D3 residues that interact with the U6-ISL in two of the ten lowest energy models are shown in Figure 2-10E and 10F. The RNA titration experiments confirm that Cdc5-D3 preferentially binds double-stranded RNAs.

**Discussion**

While the highly conserved N-terminus is essential for cell viability, the function of this portion of Cdc5 has not been characterized. The presence of Myb repeats in the N-terminus coupled with the ability of the C-terminus to directly interact with core NTC members (Ohi, McCollum et al. 1994, Tsai, Chow et al. 1999, Ohi and Gould 2002) led us to hypothesize that Cdc5 may facilitate NTC-mediated RNA-RNA and/or RNA-protein transitions by interacting with both RNA and protein in the spliceosome. Using a combination of yeast genetics and *in vitro* RNA binding assays we have shown that each of the R1, R2 and D3 domains in the Cdc5 N-terminus are essential and that Cdc5-R1-R2-D3$^{\Delta L}$ binds directly to regions within the U2 and U6 snRNAs. In addition, structural and biochemical analyses of Cdc5-D3 reveal that while this domain does not adopt a predicted Myb fold, it is able to interact preferentially with double-stranded RNAs suggesting that it may be a Myb-variant or a unique structural domain. Thus the N-terminus of Cdc5 can directly interact with multiple RNAs while the C-terminus contacts other core NTC components (Ohi and Gould 2002, Grote, Wolf et al. 2010), suggesting

that *in vivo* Cdc5 could tether NTC to RNA components within the spliceosome and provide an RNA binding platform that stabilizes the RNA-RNA and RNA-protein pre-mRNA splicing transitions.

Since Cdc5-D3 has a strong preference for U6-ISL and the MS2 dsRNAs despite the lack of sequence similarity between these RNAs, it seems likely that Cdc5-D3 might stabilize dsRNA conformations in a non-specific manner. However, an alternative explanation is that Cdc5-D3 binds with high specificity to an RNA that we did not directly test. Combined with our *in vivo* results showing that R1, R2 and D3 domains are all essential, we propose that Cdc5's N-terminus with an overall pI of 9.4 acts as a charged RNA binding platform that binds to dsRNAs found near the catalytic core of the spliceosome.

Our finding that all three Cdc5 nucleic acid binding domains (R1, R2 and D3) are essential suggests a potential model where these domains could act synergistically *in vivo*. The interaction of multiple Myb repeats with DNA is common in transcription factors (reviewed in (Prouse and Campbell 2012)). For example, the transcription factor c-Myb contains three Myb repeats (R1, R2, and R3) and it is the third helix in the R2 and R3 Myb domains that directly contacts the major groove of a specific DNA sequence (Ogata, Morikawa et al. 1994, Ogata, Morikawa et al. 1995). Therefore, using Myb containing proteins as an example, we speculate that the Cdc5 R1, R2 and D3 domains may bind RNA in a similar manner as c-Myb binds DNA, with the exception that the lack of the third canonical Myb repeat may make the Cdc5 N-terminus better suited for binding nucleic acid structures (*i.e.* dsRNA) versus binding specific nucleic acid

sequences. The ability of Cdc5 to interact with multiple RNA structures could provide a platform to stabilize the conformational changes that occur during the transition of the spliceosome from an inactive to activated complex.

The spliceosome is a dynamic macromolecular machine composed of both protein and RNA components. During the transition from the inactive spliceosomal B-complex to the activated $B^{act}$-complex, the U1 and U4 snRNAs dissociate allowing the U6 snRNA to change conformation by forming the catalytic U6-ISL and duplexes with both the U2 snRNA and the 5′ end of the intron via the ACAGAGA box (Madhani and Guthrie 1992). Although these RNA-RNA remodeling events are the major requirements for the formation of the spliceosome active site, additional events, including the stable association of the NTC, are required for catalysis to occur. The active site of the spliceosome is highly conserved from yeast to humans (Hausner, Giglio et al. 1990, Datta and Weiner 1991, Luukkonen and Séraphin 1998, Burke, Sashital et al. 2012) and is centered on a region of the U6 snRNA, which shares three similarities with Group II self-splicing introns, the AGC triad, $Mg^{2+}$ binding, and the U6-ISL (Reiter, Blad et al. 2004, Seetharaman, Eldho et al. 2006, Montemayor, Curran et al. 2014). Similarities between the U6 snRNA and the group II self-splicing introns (Fica, Mefford et al. 2014), as well as the splicing-like activity demonstrated by a protein-free RNA construct containing regions of the U6 and U2 snRNAs that base pair in the spliceosome (Valadkhan and Manley 2001, Valadkhan, Mohammadi et al. 2009), have led to the conclusion that the regions of U6 and U2 snRNAs make up essential components of the spliceosome active site. However, unlike the Group II self-splicing introns that use additional RNA domains

to support catalysis, the spliceosome has evolved to also use proteins to facilitate the splicing reaction. Two proteins, the highly conserved Prp8 (*Sp* Spp42) and *Hs* RBM22/*Sc* Cwc2 (*Sp* Cwf2/Prp3), have been shown to directly cross-link to both the U6 snRNA and the pre-mRNA within the active site of the spliceosome (Fig. 2-12) (McGrail, Krause et al. 2009, Rasche, Dybkov et al. 2012, Li, Zhang et al. 2013). The physical interaction between *Sc* Cwc2 and U6 snRNA places the NTC at the active site of the spliceosome and, interestingly, both *Sc* Cef1 (*Sp* Cdc5) and *Sc* Cwc2 (*Sp* Cwf2) bind directly to Prp19 (Ohi and Gould 2002, Vander Kooi, Ren et al. 2010) (Fig. 2-12B). Thus the only two NTC components that contain nucleic acid binding domains are likely in close physical proximity. However, although *Sp* Cwf2 family members have been shown to directly cross-link to the U6 snRNA *in vitro* (McGrail, Krause et al. 2009, Rasche, Dybkov et al. 2012, Schmitzová, Rasche et al. 2012), *Sp* Cdc5 family members have not. While there is no direct evidence that *Sp* Cdc5 interacts with RNA *in vivo*, genetic mutations in *Sc* Cef1 (*Sp* Cdc5) are able to suppress mutations in the U6 snRNA found in the active site of the spliceosome (Fig. 2-12A) (Query and Konarska 2012) and we have shown that there is a negative genetic interaction between *Sp cdc5* and *Sp cwf2*/*prp3* temperature sensitive alleles that generate point mutations in the RNA binding domains of each protein (Fig. 2-1E). The proximity of *Sc* Cwc2 to *Sc* Cef1 in the NTC (Ohi and Gould 2002), the capability of Cef1/Cdc5 (*Sc*/*Sp*) to suppress U6 snRNA mutations (Query and Konarska 2012), the ability of Cdc5 to bind RNA *in vitro*, and the negative genetic interaction between *Sp cdc5-120* and *Sp prp3-1* (*Sp cwf2*) makes it possible that Cdc5, as a core member of the NTC, is positioned near the active site of the spliceosome and can act as a

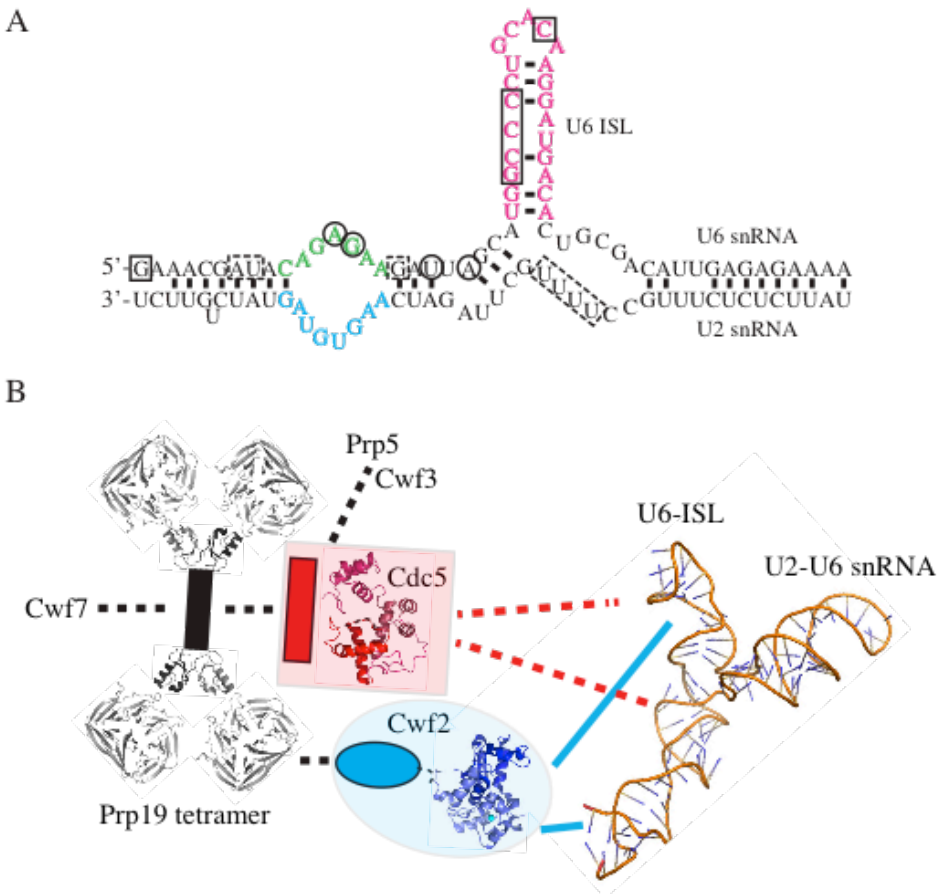



**Figure 2-12. Model of the U2 and U6 snRNAs and NTC Components in the Spliceosome B$^{act}$ Complex.** (A) A secondary structure model of U2 and U6 snRNA interactions in the *S. pombe* spliceosome active site as predicted from studies in *S. cerevisiae* (adapted from Burke et al. 2012). The snRNAs are highly conserved. The U6-ISL is pink, the U6 ssRNA is green, and the U2 ssRNA is blue. Dashed boxes are residues cross-linked to *Sc* Prp8 and solid boxes are residues crosslinked to *Sc* Cwc2. Circles mark the location of mutations in *Sc* U6 snRNA suppressed by mutations in the R1 domain of *Sc* Cef1. (B) Model of a subset of NTC components and the U2-U6 snRNA. The Prp19 tetramer binds directly to NTC components Cdc5 (*Sc* Cef1), Cwf7 (*Sc* Snt309), and Cwf2 (*Sc* Cwc2) (dashed lines). *Sc* Cwc2 crosslinks to regions of the U2 and U6 snRNAs. The C-terminus of Cdc5 binds directly to Prp5 (*Sc* Prp46) and Cwf3 (*Sc* Syf1), while the N-terminus binds RNA *in vitro*. Dashed lines indicate a direct *in vitro* interaction between components and solid lines represent *in vitro* protein-RNA crosslinking. Models are not to scale. Prp19 tetramer: light gray structures, WD40 repeats (PDB 3LVR); black and medium gray structures, U-box dimers (PDB 2BAY), and black rectangle, coiled-coil region. Cwc2: Blue structures, ZnF and RRM domain (PDB 3U1M), blue oval, the C-terminus. Cdc5: red structures, a model of the N-terminus, generated using Modeller 9v8 and *Hs*CDC5 R1 (PDB 2DIM), *Hs*CDC5 R2 (PDB 2DIN), and the lowest energy CS-Rosetta structure for Cdc5-D3; red box, the C-terminus. *Sc* U2-U6 snRNA (PDB 2LKR).

charged platform that helps stabilize the RNA-RNA rearrangements that occur during the transition from the B- to $B^{act}$-spliceosome (Fig. 2-12B).

Our findings suggest a model where the ability of the Cdc5 N-terminus to directly interact with a variety of RNA structures allows it to act as a binding scaffold that supports and/or facilitates the RNA-RNA and RNA-protein remodeling that occurs during the transition from the B- to $B^{act}$- spliceosomal complexes. Thus we propose that Cdc5 contributes to NTC function by playing an essential role in facilitating the conformational changes that occur during spliceosome activation. Our studies indicate that Cdc5 is biochemically positioned to play a pivotal role in pre-mRNA processing through its interaction with NTC components and RNA within the active site of the spliceosome.

# CHAPTER III

## Characterizing a mutant of *Schizosaccharomyces pombe* Prp16, a second step splicing factor

### Introduction

Pre-mRNA processing is catalyzed by the spliceosome, which is composed of the U1, U2, U4, U5 and U6 small nuclear ribonucleoproteins (snRNPs), the NineTeen Complex (NTC), and a number of other protein components. SnRNPs are protein-RNA complexes, which recognize conserved sequences of pre-mRNA, such as the 5′ and 3′ splice sites (SS) and the branchpoint (BP). The NTC is a protein-only complex required for spliceosome stabilization and activation. The snRNPs and NTC are subcomplexes of the spliceosome, which interact *de novo* with each pre-mRNA transcript forming the spliceosome. The spliceosome is a dynamic machine that undergoes structural rearrangements throughout the splicing cycle in order to efficiently process pre-mRNA. Eight conserved RNA helicases (*Sp* Uap56, *Sp* Prp11, *Sp* Prp28, *Sp* Brr2, *Sp* Cdc28, *Sp* Prp16, *Sp* Prp22, and *Sp* Prp43) drive the spliceosome transitions between distinct stages of the splicing cycle.

The current *in vitro* model for pre-mRNA splicing involves the step-wise association of the snRNPs and the NTC onto pre-mRNA, which are subsequently remodeled by RNA helicases. The U1 and U2 snRNPs first recognize the 5′SS and the BP, respectively, forming the A complex. The RNA helicases Uap56 and Prp11 promote A complex formation by removing protein factors from pre-mRNA and stabilizing the U2 snRNA allowing for the U2 snRNP to stably associate with the pre-mRNA. The B

complex is then formed after the addition of the tri-snRNP (U4, U5, and U6 snRNPs). In order for the spliceosome to become activated, the RNA helicases Prp28 and Brr2 act on the B complex to remove the U1 snRNP from the 5′SS and unwind the U4/U6 snRNAs. This helicase activity results in the loss of the U1 and U4 snRNPs from the B complex and with the addition of the NTC, the spliceosome transitions to the $B^{act}$ complex. The RNA helicase Cdc28 then interacts with the $B^{act}$ complex and removes components of the U2 snRNP allowing first step splicing components to interact with the BP. After Cdc28 activity, the spliceosome becomes the B* complex. The B* complex completes the first step of splicing, resulting in the formation of a free 5′ exon and a lariat intron. Upon completing the first step of splicing, the helicase Prp16 removes first step splicing factors and induces a structural change which brings the 3′SS into close proximity with the 5′SS for the second step of splicing, forming the C complex. When the second step of splicing is complete, the post-catalytic (P) complex is formed. The helicase Prp22 then removes the mRNA from the P complex forming the intron lariat spliceosome (ILS). Finally, the helicase Prp43 is recruited to the ILS where it releases the lariat RNA and displaces the remaining spliceosome components to be used for subsequent rounds of pre-mRNA splicing. Although the general mechanism of pre-mRNA splicing is known, the molecular details of how the spliceosome processes pre-mRNA and how these inter-molecular interactions change throughout the splicing cycle are incompletely understood.

The RNA helicases involved in spliceosome rearrangement are located at key positions driving the progression of the splicing cycle making RNA helicases a target for “trapping” and isolating large splicing complexes (Lardelli, Thompson et al. 2010).

Characterizing these "trapped" complexes allows for the molecular interactions within each stage of splicing to be elucidated. Spliceosomes purified in the presence of helicase mutants combined with experimental approaches such as electron microscopy (EM) provide insight into how the spliceosome is rearranged throughout the splicing cycle. For example, EM was used to characterize an *S. cerevisiae* mutant helicase in *Sc* Prp2 (*Sp* Cdc28) which was "trapped" both before and after the first step of splicing (Warkocki, Odenwalder et al. 2009). Although this is a promising method, to date, the structure of a helicase "trapped" spliceosome in *S. pombe* has yet to be determined.

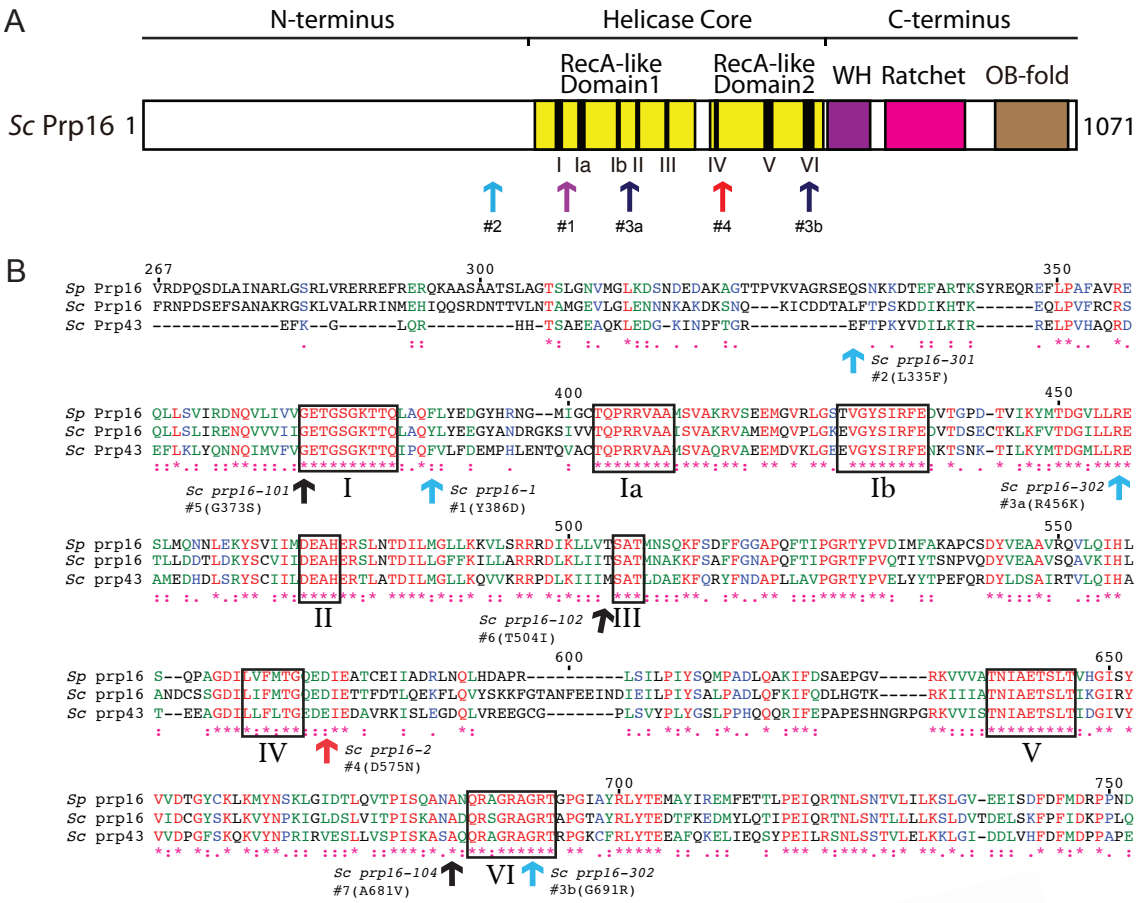
*S. pombe* is an ideal model system to study the various stages of splicing. Previous work in *S. pombe* has characterized a stable ILS complex purified from wild-type cells, and using EM the *S. pombe* ILS resembles the non-crosslinked human C complex (Jurica, Sousa et al. 2004, Ohi, Ren et al. 2007, Ilagan, Chalkley et al. 2013). In *S. pombe* no other large splicing complexes have been structurally characterized by EM, but there are several helicase mutants which are capable of "trapping" the spliceosome in *S. pombe* for EM studies. Helicase mutants in *S. pombe* are poorly characterized for their ability to "trap" the spliceosome at specific stages. However, several helicase mutants in *S. pombe* have characterized roles in splicing, such as *Sp* Brr2 (*spp41-1)*, *Sp* Prp11 (*prp11-1*), *Sp* Cdc28 (*cdcp8* and *prp8-1*), and *Sp* Prp16 (*prp14-1*/*prp16-1*), suggesting they are perturbing spliceosome function by "trapping" the spliceosome at a specific stage of splicing (Lundgren, Allan et al. 1996, Urushiyama, Tani et al. 1996, Schmidt, Richert et al. 1999). Whether or not the mutants of *S. pombe* RNA helicases can be used

to “trap” splicing complexes is still unknown, but these mutants do offer a potential way to characterize splicing complexes otherwise undetectable in the *S. pombe* model system.

*Sp* Prp16 is an RNA helicase of particular interest because it is well-characterized in *S. cerevisiae* and humans with a defined role in pre-mRNA processing (Schwer and Guthrie 1991, Zhou and Reed 1998). In general, Prp16 is a processive DEAH-box RNA helicase which has the ability to bind RNA duplexes *in vitro* and induce duplex unwinding (Wang, Wagner et al. 1998). In the spliceosome, homologs of *Sp* Prp16 have been shown to act after the first step of splicing, resulting in both the loss of first step splicing factors, *Sp* Cwf16 and *Sp* Cwf25 (*Sc* Yju2 and *Sc* Cwc25), and the structural rearrangement of the spliceosome, resulting in the protection of the 3′ SS from RNase H cleavage (Schwer and Guthrie 1991, Schwer and Guthrie 1992, Umen and Guthrie 1995, Tseng, Liu et al. 2011). In *S. cerevisiae*, a mutant of *Sc* Prp16 was used to isolate a second step splicing complex, supporting a role for Prp16 in transitioning the spliceosome from the B* to the C complex (Lardelli, Thompson et al. 2010). In addition to Prp16’s ATP dependent role, Prp16 also has an ATP independent role when an aberrant pre-mRNA transcript is present. Under these conditions, *Sc* Prp16 associates with the spliceosome prior to the first splicing reaction where it stabilizes *Sc* Cwc25 (*Sp* Cwf25) and proofreads 5′SS and BP mutants in pre-mRNA transcripts (Burgess and Guthrie 1993, Koodathingal, Novak et al. 2010, Tseng, Liu et al. 2011). Pre-mRNA proofreading by Prp16 enhances splicing fidelity by promoting the discard of aberrant transcripts (Koodathingal and Staley 2013). With Prp16’s well-characterized role in

humans and *S. cerevisiae* in transitioning the spliceosome from the B* to the C complex, we hypothesize that a mutant of *Sp* Prp16 will "trap" a second-step splicing complex.

Prp16 can be divided into three distinct regions: N-terminus, helicase core, and C-terminus (Fig. 3-1A). In *S. cerevisiae* Prp16 domain deletions as well as mutant analysis has defined which regions are required for function. For example, the N-terminal ~200 residues and the C-terminal ~100 residues are not essential for function *in vivo*; however, there are still regions of the N- and C-terminus, which are essential for function (Hotz and Schwer 1998, Wang, Wagner et al. 1998, Zhou and Reed 1998). The essential nature of the N-terminus of *Sc* Prp16 is due to its role in recruiting Prp16 to the spliceosome (Wang, Wagner et al. 1998). In addition to the essential N- and C-termini, site-specific mutations in key helicase core motifs result in lethality (Hotz and Schwer 1998). These site-specific mutations were used to determine which amino acids in the conserved sequence elements I, I, III, and VI are required for function (Fig. 3-1A). Conditional and suppressor alleles have also been isolated and used to further understand the function of *Sc* Prp16 (Fig. 3-1B). The suppressor alleles, although they have no phenotype on their own, are able to suppress A-C BP mutants (Burgess and Guthrie 1993). These suppressor mutants have a reduced rate of ATP hydrolysis (Fig. 3-1C), which is hypothesized by Burgess and Guthrie to allow for aberrantly formed lariat RNA intermediates to be corrected as apposed to being discarded (Burgess and Guthrie 1993). In addition to the suppressor mutants, both heat and cold sensitive conditional mutants have also been isolated. The conditional mutants when mapped to the crystal structure of the conserved helicase *Sc* Prp43 are located in the vicinity of the active site (Fig. 3-1A and D), although




C

| Summary of Sc Prp16 Activity | | Relative ATPase Activity* |
|---|---|---|
| **Cold Sensitive** | | |
| #1 *Sc prp16-1* | Y386D | 0.010 |
| #2 *Sc prp16-301* | L355F | N.D. |
| #3 *Sc prp16-302* | R456K<br>G691R | N.D. |
| **Heat Sensitive** | | |
| #4 *Sc prp16-2* | D575N | N.D. |
| **Suppressor Mutations** | | |
| #5 *Sc prp16-101* | G373S | 0.050 |
| #6 *Sc prp16-102* | T504I | 0.021 |
| #7 *Sc prp16-104* | A681V | 0.036 |

D

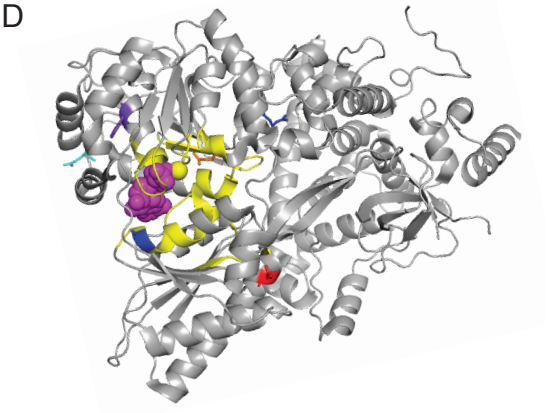


**Figure 3-1. *Sc* Prp16 Domain Architecture and Mutants** (A) Domain architecture of *Sc* Prp16. Colored arrows highlight RNA helicase mutants. (B) Sequence alignment of the helicase core domain, comparing *Sc* and *Sp* Prp16 with *Sc* Prp43. Numbering corresponds to *Sc* Prp16. Mutants are highlighed with arrows (red = heat sensitive and blue = cold sensitive) (C) *Sc* Prp16 helicase mutants and ATPase activity adapted from Schwer and Guthrie 1992Burgess and Guthrie 1993. (D) *Sc* Prp43 helicase core structure (PDB 2XAU) with *Sc* Prp16 helicase core mutatnts mapped in color (colors correspond to arrow in panel A).

none of the conditional mutants map to residues directly involved in ATP binding. Of the conditional alleles, *Sc prp16-1* is the most well-characterized mutant. When comparing the rate of ATP hydrolysis between *Sc* Prp16 and *Sc prp16-1* there is minimal *in vitro* NTP hydrolysis independent of the nucleotide used or whether RNA is present or not (Schwer and Guthrie 1992, Burgess and Guthrie 1993, Wang, Wagner et al. 1998) (Fig. 3-1C). The reduced rate of ATP hydrolysis also correlates with the inability for *Sc prp16-1* to unwind a U4/U6 snRNA duplex *in vitro* (Wang, Wagner et al. 1998). In *S. cerevisiae* there are distinct roles from the N-terminus, helicase core, and the C-terminus, where the helicase activity is required driving the progression of the splicing cycle. Although *Sc* Prp16 has been extensively studied, there is limited data on the function of *Sp* Prp16, and based on sequence homology *Sp* Prp16 is hypothesized to function similar to *Sc* Prp16.

In this study we characterized the *S. pombe prp16-1* mutant. Using a combination of yeast genetics and cell lysate analyses, we show that the *Sp prp16-1* helicase mutant results in a truncated protein due to a pre-mature stop codon, resulting in a protein with no helicase core or C-terminus. The *Sp prp16-1* mutant was found to have both a cold- and heat-sensitive phenotype. We further characterized this heat-sensitive phenotype and showed that in the presence of the *Sp prp16-1* mutant, a stable spliceosome can be isolated. Further, this complex contained RNAs and proteins associated with second step splicing complexes. Finally, EM analysis of the *Sp prp16-1* spliceosome shows a complex similar to the ILS. Taken together, these data demonstrate that an RNA helicase mutant in *S. pombe* can be used to isolate a second-step splicing complex.

## Materials and Methods

**Strains, yeast methods, and molecular biology.** Strains used in this study are listed in Table 3-1. Yeast strains were grown in liquid yeast extract (YE) or on YE agar plates. For TAP purifications cells were grown in 4x-YE. For spot assays, cells were grown to mid-log phase at 29°C and resuspended in water to achieve an optical density of 0.3 at 595 nm ($OD_{595}$). Tenfold serial dilutions were made, and 2 µl of each dilution was plated on YE. Plates were incubated at the indicated temperatures for 3 to 9 days before imaging.

**RNA Isolation, Northern Blotting, and RT-PCR.** Total RNA was prepared from cells by extraction with hot acidic phenol as described previously (Collart and Oliviero 2001). To visualize snRNAs from a Cdc5-TAP and anti-snRNA cap (anti-trimethylguanosine [$m_3G$]; Millipore) pulldowns, RNAs were resolved using 6% TBE-urea gels (Life Technologies). RNA was transferred to a Duralon-UV membrane (Agilent Technologies, Santa Clara, CA), UV cross-linked using energy setting 700 (UVC500 cross-linker; GE Healthcare, Piscataway, NJ), and detected by using [$\gamma$-$^{32}$P] ATP (PerkinElmer, Waltham, MA)-labeled oligonucleotides complementary to *S. pombe* U1, U2, U4, U5, and U6 (oligonucleotide sequences as published (Livesay, Collier et al. 2013)). Blots were exposed to phosphorimager screens overnight and visualized using a Typhoon FLA-7000IP instrument (GE Healthcare, Piscataway, NJ). For reverse transcription (RT)-PCR analysis, RNA was treated with DNase I (Life Technologies) and reverse transcribed according to the manufacturer's directions using random hexamers for priming (SuperScript; Life Technologies). The resulting cDNA was PCR amplified with *tbp1_a*

**Table 3-1.** *Schizosaccharomyces pombe* strains used in this study

| Strain # | Genotype | Figure | Source |
|---|---|---|---|
| OHI001 | *ade6-M210 leu1-32 ura4-D18 h-* | 3-2 | K. Gould[a] |
| OHI018 | *nda3-km311 leu1-32 h+* | 3-2 | Toda et al. 1983 |
| OHI130 | *cdc5-tap::Kan$^r$ ade6-m210 leu1-32 ura4-D18 h+* | 3-3, 3-5 | Adapted from Ohi et al. 2002 |
| OHI262 | *prp2-1 leu1-32 ura4-D18 h+* | 3-2 | Potashkin et al. 1976 |
| OHI269 | *cdc5-120 leu1-32 his3-D1 h+* | 3-2 | Nurse et al. 1976 |
| OHI420 | *prp16-1 (prp14-1)* | 3-2 | Urushiyama et al. 1996 |
| OHI665 | *prp16-1 cdc5-tap::Kan$^r$* | 3-3, 3-5 | This Work |

[a]Vanderbilt University School of Medicine, Nashville, TN.

(primers published in (Livesay, Collier et al. 2013)) for 32 cycles, run on a 2.5% gel and stained with ethidium bromide.

**Glycerol Gradients** For gradients from a Cdc5-TAP and from cell lysates, a 200-µl volume was layered onto a 10-30% glycerol gradient and centrifuged at 28,000 rpm at 4°C for 16 h in a SW55Ti rotor (Beckman). TAPs were performed as described previously (Gould, Ren et al. 2004). Fractions from the gradients were collected manually and resuspended in 4x LDS sample buffer (Life Technologies, Grand Island, NY). Parallel standard gradients contained thyroglobulin (19S) and catalase (11.35S) (HMW calibration kit; GE Healthcare, Piscataway, NJ).

**Western Blots** Poteins were resolved by a 4 to 12% Bis-Tris PAGE (and transferred by electroblotting to a PVDF membrane (Whatman, GE Healthcare, Piscataway, NJ). The primary antibody anti-Cdc5 (1/5,000) rabbit polyclonal antisera was used to detect Cdc5. The Cdc5 anti-body was detected by the secondary antibodies IRDye 800CW (LI-COR Biosciences, Lincoln, NE) (1/10,000 dilution) and visualized using an Odyssey scanner and software (LI-COR Biosciences).

**Mass Spectrometry** TAP elutions were trichloroacetic acid (TCA) precipitated, resolubilized in 8 M urea–100 mM Tris (pH 8.5), reduced, alkylated, and then diluted back to 2 M urea and digested overnight with trypsin as described previously (MacCoss, McDonald et al. 2002). Resulting peptides (corresponding to about 5% of the TAP eluate) were analyzed by a 70-min data-dependent liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis in the Vanderbilt Mass Spectrometry core. In brief, peptides were autosampled onto a 200-mm by 0.1-mm (Jupiter 3 micron, 300 Å) self-

packed analytical column coupled directly to an LTQ (ThermoFisher) using a nanoelectrospray source and resolved using an aqueous to organic gradient. A series of a full-scan mass spectrum followed by five data-dependent tandem mass spectra was collected throughout the run, and dynamic exclusion was enabled to minimize acquisition of redundant spectra. Tandem mass spectra were searched via Sequest against an *S. pombe* database (UniprotKB taxon 284812 reference proteome set) that also contained a reversed version for each of the entries (Yates, Eng et al. 1995). Identifications were filtered and collated at the protein level using Scaffold (Proteome Software).

**Electron Microscopy** Uranyl formate stained samples were prepared as described (Ohi, Li et al. 2004). In brief, 3 µl of sample was absorbed to a glow-discharged 400-mesh copper grid covered with carbon-coated collodion film. The grid was washed in two drops of water and then stained with two drops of uranyl formate (0.75%). Samples were imaged on a FEI Tecnai 200 kV electron microscope equipped with a field emission electron source (FEI, Hillsboro, OR) and 4K x 4K Gatan Ultrascan CCD.

**Results**

**Characterizing a mutant of the *S. pombe* helicase Prp16**

*S. pombe prp14-1* was isolated in a PRP screen where it was shown to be a cold sensitive (CS) mutant (Urushiyama, Tani et al. 1996). At the restrictive temperature (22ºC) *Sp prp14-1* was shown to be defective in pre-mRNA processing. In this study we determined *Sp prp14-1* to be a mutation in the *Sp prp16* gene and thus renamed this mutant to *Sp prp16-1* for clarity. The *Sp prp16-1* gene was PCR amplified and sequenced. Two nucleotide mutations were found, a G975A mutation resulting in a silent

mutation, K325K, and a C1630T mutation resulting in the introduction of a pre-mature stop codon, Q544stop. The *Sp prp16-1* Q544stop mutation is a severely truncated RNA helicase resulting in a Prp16 protein with only the non-conserved N-terminus and part of the first RecA-like domain found in the helicase core (Fig. 3-2A).

To address the impact of the *prp16-1* mutation in *S. pombe*, we used a spot assay to determine the relative growth of this mutant at different temperatures. We observed that wild-type, *cdc5-120*, *nda3-km311*, and *prp16-1* all grow at 29ºC (Fig. 3-2B middle panel). However, a temperature shift to either 18ºC or 36ºC results in a substantial decrease in the growth of the *Sp prp16-1* mutant similar to the CS phenotype of *nda3-km311* and the heat sensitive (HS) phenotype of *Sp cdc5-120* (Fig. 3-2B). These results indicate that the *Sp prp16-1* mutant has growth defects at both 18ºC and 36ºC. As the CS phenotype of *Sp prp16-1* has been previously characterized (Urushiyama, Tani et al. 1996), we began to investigate the HS phenotype of the *Sp prp16-1* mutant.

Prp16 is a well-characterized helicase, and its involvement in pre-mRNA splicing has been demonstrated in both *S. cerevisiae* and humans. To determine if the HS phenotype of the *Sp prp16-1* mutant results in a pre-mRNA splicing defect, we isolated total RNA from cells shifted to 36ºC for 4 hours. Using reverse transcriptase (RT)-PCR, the *Sp prp16-1* mutant showed an increase in *Sp tbp1* pre-mRNA transcript levels similar to a known pre-mRNA splicing mutant, *Sp prp2-1* (Fig. 3-2C). In addition to pre-mRNA analysis we next sought to determine the impact of the *Sp prp16-1* mutant on the spliceosome. Using Cdc5, a component of the NTC, as a marker for the spliceosome we ran *Sp prp16-1* cell lysate over a linear glycerol gradient. *Sp* Cdc5 was detected by

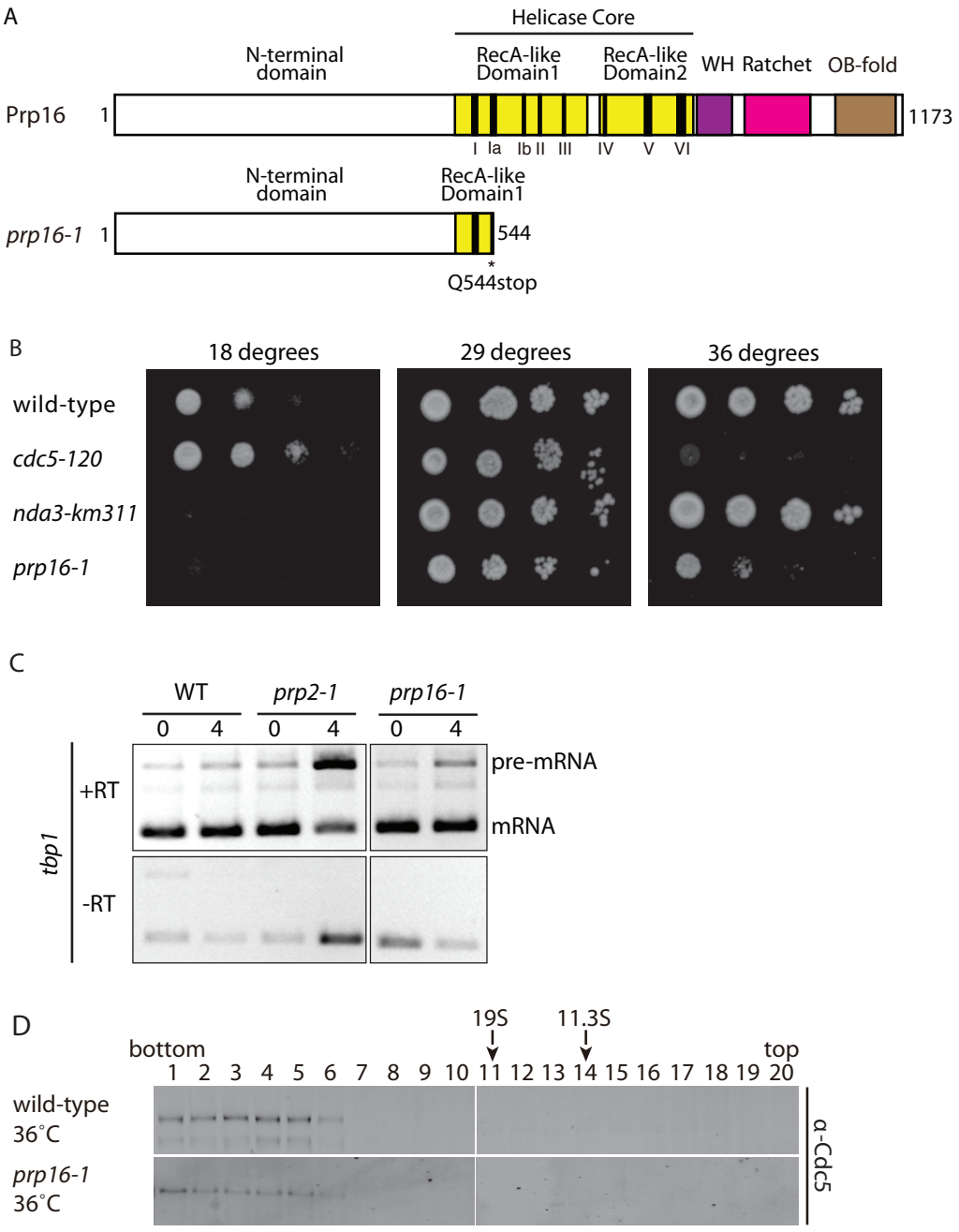



**Figure 3-2. Effect of the *prp16-1* Mutation on Cell Growth and Splicing** (A) Domain architecture of Prp16 and *prp16-1*. The "*" indicates the mutation in prp16-1. (B) Serial dilution spot assay at 18˚C, 29˚C, and 36˚C. (C) RT-PCR amplification of the *tbp1* RNA transcript from total RNA at 0 and 4 hours at 36˚C. The expected size for the pre-mRNA and mRNA transcripts are marked. (D) Western blot of a linear glycerol gradient with cell lysates shifted to 36˚C for 4 hours. The 19S thyroglobulin and 11.3S catalyase gradient markers are indicated above the gradient with arrows.

Western blot toward the bottom of the gradient in both wild-type and *Sp prp16-1* lysates (Fig. 3-2D). Despite the accumulation of pre-mRNA levels in the *Sp prp16-1* mutant at the restrictive temperature, the presence of a large splicing complex remains unchanged.

**The second step splicing complex in *S. pombe***

Mutations in the RNA helicases involved in pre-mRNA splicing can be useful tools for isolating the spliceosome at different stages of pre-mRNA processing. The RNA helicase mutants are hypothesized to be conditionally defective for RNA helicase activity and at extreme temperatures these helicase mutants would have reduced ATPase activity. For example, *Sc prp16-1* has reduced ATPase activity and *Sc prp16-302* was used to isolate a late splicing complex (Fig. 3-1B) (Burgess and Guthrie 1993, Lardelli, Thompson et al. 2010). We wanted to determine if the *Sp prp16-1* mutant, which is has no helicase domain, is capable of trapping a late splicing complex in *S. pombe* that is different from the previously characterized ILS purified from wild-type cells (Ohi, Ren et al. 2007). The spliceosome from *Sp prp16-1* cells was purified at the restrictive temperature using a tandem affinity purification (TAP) tag on Cdc5 (Cdc5-TAP). Upon purification of the Cdc5-TAP *prp16-1* complex, the presence of a stable splicing complex was assessed by a linear glycerol gradient (Fig. 3-3A). By Western blot, we detected Cdc5 toward the bottom of the gradient in both ILS and Cdc5-TAP *prp16-1* purifications. These data indicate that Cdc5 is associated with a stable splicing complex in the *Sp prp16-1* mutant.

As Cdc5 is associated with a stable spliceosome in the *Sp prp16-1* mutant strain at the restrictive temperature, we wanted to determine if this complex was also involved in

the first or second step of pre-mRNA splicing. We analyzed the snRNAs and a pre-mRNA transcript extracted from the Cdc5-TAP *prp16-1* purification and compared it to RNA extracted from the ILS to determine whether this complex is a first or second step splicing complex. Using a Northern blot and DNA probes for the five snRNAs, we found the Cdc5-TAP *prp16-1* complex contains the U2, U5 and U6 snRNAs similar to the ILS complex (Fig. 3-3B). We then wanted to use a model pre-mRNA transcript to determine whether the first or second step of splicing had occurred in the Cdc5-TAP *prp16-1* complex. Using RT-PCR we analyzed the state of the *cwf8* pre-mRNA transcript in the Cdc5-TAP *prp16-1* complex. We employed transcript-specific primers designed to amplify pre-mRNA, mRNA, and lariat RNA (Fig. 3-4), and found the ILS and Cdc5-TAP *prp16-1* complexes contain lariat RNA and mRNA without any detectable levels of unspliced pre-mRNA (Fig. 3-3C). RNA analysis of both the Cdc5-TAP *prp16-1* complex and the ILS show that each complex contains a similar RNA composition including the snRNAs (U2, U5, and U6), lariat RNA, and mRNA. The absence of pre-mRNA and the presence of the lariat RNA as well as mRNA indicates that the Cdc5-TAP *prp16-1* complex is a second step splicing complex.

In other organisms, mass spectrometry (MS) has been used to identify spliceosome protein components at distinct stages of pre-mRNA splicing. Interestingly, there are a number of protein components which are only required at specific stages of splicing. Using this biological feature to our advantage, we utilized a one-dimensional MS approach to determine the protein composition of the Cdc5-TAP *prp16-1* complex (Fig. 3-5B). Compared to the MS analysis of the ILS, the Cdc5-TAP *prp16-1* complex

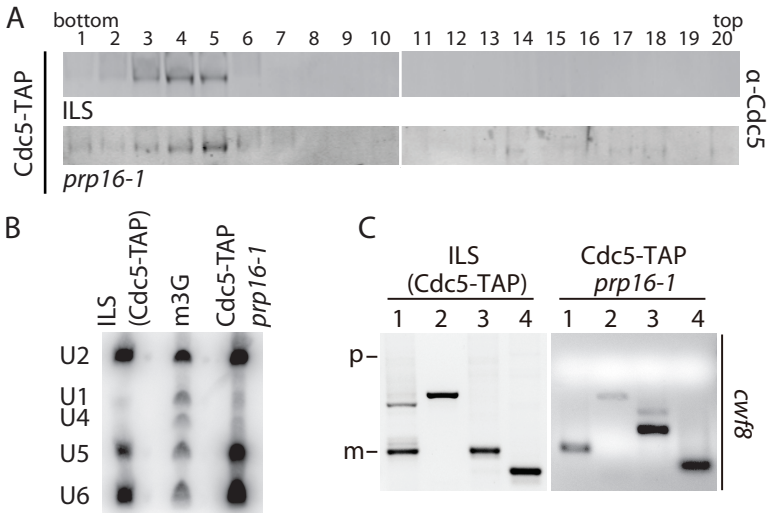



**Figure 3-3. Analysis of the Cdc5-TAP *prp16-1* Complex** (A) Anti-Cdc5 Western blot of a linear glycerol gradient with purified ILS (Cdc5-TAP) and Cdc5-TAP *prp16-1*. (B) Northern blot using $p^{32}$ labeled DNA oligos specific for the U1, U2, U4, U5 and U6 snRNA. m3G is a control for the five snRNAs in the spliceosome. (C) RT-PCR of the *cwf8* transcript from purified complexes. Numbers correspond to the reactions outlined in Figure 3-3. "p" represents the expected band size for pre-mRNA and "m" represents the expected band size for mRNA.

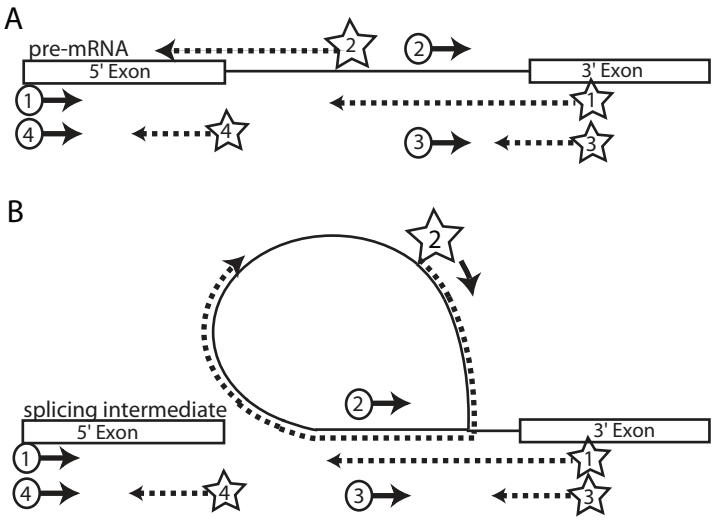



**Figure 3-4. Identifying RNAs Present in the *S. pombe* ILS and Cdc5-TAP *prp16-1* Complexes** Schematic of RT-PCR reactions used to characterize RNA extracted from both ILS and Cdc5-TAP *prp16-1* complexes shown in Figure 3-2. Exons are represented by boxes; introns are represented by solid lines. Positions of primers used in reactions 1–4 are shown in the diagram. Primers used for first-strand synthesis are marked by a star. These primers were used for subsequent amplifications. The arrows depict the direction of DNA synthesis. The dotted lines represent first-strand synthesis. Sequences of oligos used for these reactions are found in (Livesay et al. 2013). (A) pre-mRNA transcript found in the splicesosome prior to the first splicing reaction. (B) pre-mRNA transcript found in the spliceosome prior to the second step of splicing. This figure was adapted from Ohi et al, 2007.

contains components of the SF3A and SF3B complexes as well as Brr2, a U5 snRNP component; however, only a few peptides from each of the components were detected (Fig. 3-5A and B). In contrast, the Cdc5-TAP *prp16-1* complex does not contain detectable amounts of the RNA helicase *Sp* Cdc28, the components of the retention in splicing (RES) complex, or Mug161, which modulates debranching in *S. cerevisiae.* From the RNA and protein analyses of the Cdc5-TAP *prp16-1* complex, this complex is similar to the ILS complex in RNA composition but contains a unique set of detectable protein components by one-dimensional MS.

**EM analysis of the *prp16-1* second step splicing complex**

Since the protein composition between the Cdc5-TAP *prp16-1* complex and the ILS is different, we wanted to determine if these differences could be detected in 2D EM projections of purified complexes. Using negative stain EM, we compared the structural features of the Cdc5-TAP *prp16-1* complex with the ILS (Fig. 3-5C and D). Similar to the glycerol gradients, a large complex was present in both purifications. After selecting ~7000 particles from each purification, we rotationally and translationally aligned these particles generating 2D projections. Analysis of a single 2D projection from each complex indicates that there are three clear globular domains in both complexes (Fig. 3-5C and D, insets), which have been previously identified in the 3D structure of the ILS (Ohi, Ren et al. 2007). The 2D projections have clear similarities in the "foot-like" domain (Fig. 3-5C and D, green arrow) and in the "ridge" domain (Fig. 3-5C and D, black arrow). In contrast the "head-like" domain shows clear differences (Fig. 3-5C and D, white arrow). Taken together, these second step splicing complexes have similar RNA

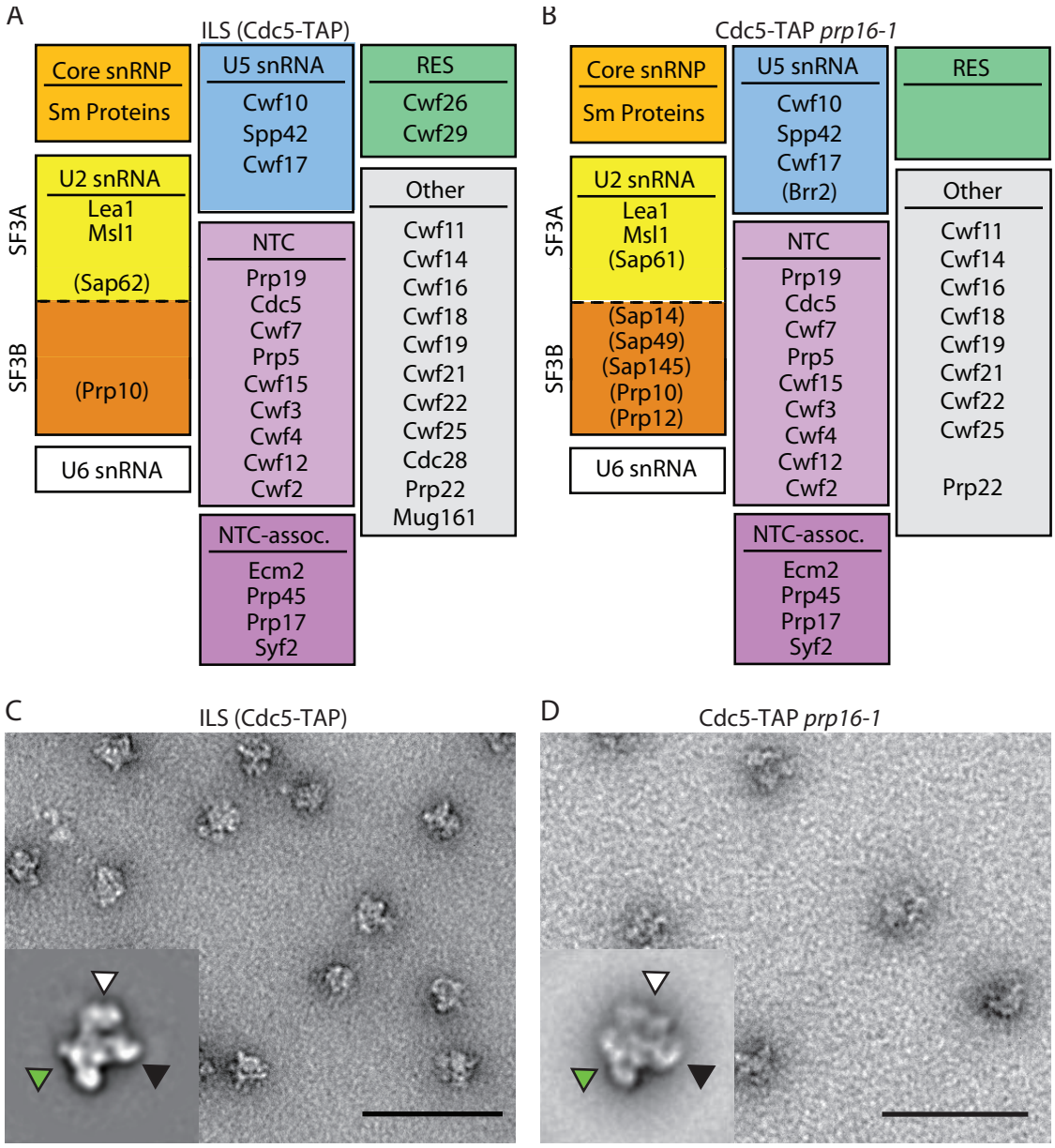



**Figure 3-5. MS and EM Analysis of ILS (Cdc5-TAP) and Cdc5-TAP *prp16-1* Complexes** (A) MS table from two replicates of the ILS (Cdc5-TAP). (B) MS table from two replicates of the Cdc5-TAP *prp16-1* complex. MS tables in panels A and B were generated using the criteria that a protein was present in both replicates with at least two peptides per replicate. Proteins outlined with parentheses are present in low abundance. (C) Negative stain EM analysis of the ILS (Cdc5-TAP). Inset is a 2D projection of the ILS. Arrows in the inset signify the three domains in the 2D projection (Scale bar = 50nm). (D) Negaitve stain EM analysis of the Cdc5-TAP *prp16-1* complex. Inset is a 2D projection of the Cdc5-TAP *prp16-1* complex. Arrows in the inset signify the three domains in the 2D projection (Scale bar = 50nm).

and protein composition with protein differences most likely found in the "head-like" domain.

**Discussion**

Here we characterized *prp16-1,* an RNA helicase mutant in *S. pombe* involved in pre-mRNA splicing. Previous work with *Sp prp16-1* showed that this mutant has a CS growth and a CS splicing defect. We further characterized this mutant and showed that the *Sp prp16-1* mutation is the result of a pre-mature stop codon resulting in a truncated protein containing only the non-conserved N-terminus. In addition to the previously characterized CS phenotype, we showed that the *Sp prp16-1* mutant results in a HS phenotype also resulting in a splicing defect. Despite the HS splicing defect, we were able to show a large splicing complex is present in cell lysate. These results suggest a functional *Sp* Prp16 is required at extreme temperatures and that at the permissive temperature the helicase and C-terminal functions of *Sp* Prp16 are not essential for function. These results are interesting as the helicase function of Prp16 is thought to be essential in all organisms. It is possible that another helicase such as *Sp* Prp22 is capable of rescuing the *Sc prp16*-1 at the restrictive temperature or that the helicase function of *Sp* Prp16 is not required at the permissive temperature. Using the helicase deficient *Sp prp16-1* mutant a stable second step spliceosome may be trapped, similar to what is seen using mutants in the *Sc prp16* helicase domain. Using this HS phenotype of *Sp prp16-1*, we wanted to determine if the Cdc5-TAP *prp16-1* spliceosome was a second step splicing complex similar to the ILS. When comparing both complexes on a linear gradient, they both peak toward the bottom of the gradient suggesting they have the same overall shape

and size. SnRNA analysis further supports the similarities between these two complexes as both spliceosomes contain the same snRNAs. Further analysis of RNA extracted from these spliceosomes shows that the *Sp cwf8* transcript has no detectable amount of pre-mRNA in either complex. The absence of pre-mRNA suggests the Cdc5-TAP *prp16-1* complex, like the ILS, is a second step splicing complex.

Although the general size and shape as well as the RNA composition between the Cdc5-TAP *prp16-1* and the ILS are similar, MS analysis highlights differences in their protein composition. In the Cdc5-TAP *prp16-1* complex, *Sp* Brr2 and components of the SF3A and SF3B complex were detected, although in low abundance. Detecting *Sp* Brr2 in the Cdc5-TAP *prp16-1* complex is interesting as it is typically not detected in the *S. pombe* ILS and the lack of *Sp* Brr2 in the ILS is hypothesized to be partially responsible for why the ILS is stable and not recycled (Chen, Shulha et al. 2014). These results suggest that *Sp* Brr2 is not stably associated with spliceosome in *S. pombe* after the first step of splicing. In *S. cerevisiae* the SF3A and SF3B complexes are thought to be loosely associated after the first step of splicing. However in humans, the association of the SF3A and SF3B complexes seems to be more dynamic as these complexes are detected in both the first and second step splicing complexes (Bessonov, Anokhina et al. 2010, Ilagan, Chalkley et al. 2013). Alternatively, the presence of the SF3A and SF3B complexes in the Cdc5-TAP *prp16-1* spliceosome may represent a C complex with loosely associated SF3A and SF3B complexes like in *S. cerevisiae* (Lardelli, Thompson et al. 2010). In contrast, the presence of the SF3A and SF3B complexes may represent the P complex found in humans (Ilagan, Chalkley et al. 2013). In either case, the low abundance of both

Brr2 and the SF3A and SF3B complexes in the Cdc5-TAP *prp16-1* spliceosome indicates a heterogeneous spliceosome. The slight differences in protein components between the ILS and the Cdc5-TAP *prp16-1* complexes suggest these complexes are slightly different mixtures of second step splicing complexes.Using EM we wanted to determine if the Cdc5-TAP *prp16-1* spliceosome had any detectable difference from the ILS. Using negative stain EM we were able to detect similar 2D projections from both purifications. Using the naming from the structure of the ILS, the "foot-like" and "ridge" domains look similar. Strikingly the "head-like" domain is well ordered in the ILS compared to the "head-like" domain of the Cdc5-TAP *prp16-1* complex. The differences in the "head-like" domain can represent a slightly different structure or the differences could indicate that the "head-like" region of the Cdc5-TAP *prp16-1* complex is heterogeneous.

In summary, we have characterized a mutant of the Prp16 helicase in *S. pombe.* We have shown this mutation results in a truncated protein and a CS and HS growth phenotype. The *Sp prp16-1* mutant has a splicing defect, although a stable spliceosome can be purified. The *Sp prp16-1* spliceosome resembles the ILS in both snRNA and pre-mRNA content, indicating that the *Sp prp16-1* spliceosome is a second-step splicing complex. MS and EM analysis of the *Sp prp16-1* spliceosome show that this complex is slightly different than the ILS potentially representing a new *S. pombe* spliceosome which can be used to provide further insight into the molecular details of how the spliceosome functions.

# CHAPTER IV

## Discussion

This work focuses on characterizing the structure and function of the *S. pombe* spliceosome. Chapter II centers on elucidating the function of the N-terminus of *Sp* Cdc5 and its potential role in the active site of the spliceosome. Chapter III is concentrated on describing the function of the *Sp* Prp16 helicase mutant and using this mutant a second step splicing complex was isolated from *S. pombe*. These studies have advanced the current understanding of spliceosome function and have laid the groundwork for future studies.

### NTC Characterization

In Chapter II the structure and function of N-terminal domains of *Sp* Cdc5 were investigated. In this study, *Sp* Cdc5-D3 was found to be structurally distinct from the canonical Myb repeats R1 and R2. Further investigation showed that the N-terminus of Cdc5 binds RNA *in vitro*, preferentially dsRNA. Finally a synthetic sick interaction between mutants of the N-terminus of *Sp* Cdc5 and *Sp* Cwf2, an RNA binding protein in the active site of the spliceosome, suggest the N-terminus of *Sp* Cdc5 may bind RNA *in vivo* within the active site of the spliceosome. These findings support a model where in the active site of the spliceosome *Sp* Cdc5, in concert with other RNA binding proteins, acts to stabilize and activate the spliceosome (Fig. 4-1A).

A potential role for the N-terminus of *Sp* Cdc5 in the active site fits well with previously published literature. For example, mutations in the Myb repeats of *Sp* Cdc5 and its *S. cerevisiae* homolog, *Sc* Cef1, result in a temperature-sensitive cell cycle arrest

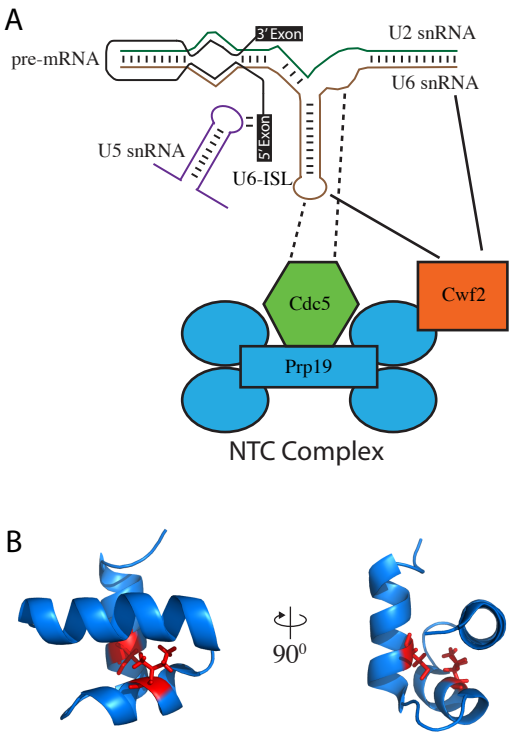



**Figure 4-1. Model of NTC Protein-RNA Interactions and Mapping of *Sc* Cef1 Mutations** (A) Model of a subset of NTC components and the secondary structure of the active site RNA in the *S. pombe* spliceosome. The U6 snRNA is in brown, the U2 snRNA is in green, and the U5 snRNA is in purple. The Prp19 tetramer in blue binds directly to NTC components Cdc5 (*Sc* Cef1) and Cwf2 (*Sc* Cwc2). *Sc* Cwc2 cross-links to regions of the U2 and U6 snRNAs. Dashed lines indicate a direct *in vitro* interaction between components and solid lines represent *in vitro* protein-RNA crosslinking. Models are not to scale. (B) The structure of *Hs* Cdc5 R1 (PDB 2DIM) is shown in blue at 0˚ and 90˚. The amino acids in red correspond to the *Sc* Cef1 mutants identified by Query and Konarska, V36R and S48R (or K). These mutations map to the core of *Hs* Cdc5 R1.

(Ohi, McCollum et al. 1994, Ohi, Feoktistova et al. 1998, Burns, Ohi et al. 2002). Further showing Cdc5's importance in splicing, work in *S. cerevisiae* shows that *Sc* Cef1 is required for the first and second steps of splicing (Burns, Ohi et al. 1999, Query and Konarska 2012). Recent work by Query and Konarska using a copper sensitivity assay show that silent mutations in the *Sc* Cef1 R1 act to suppress intron mutations, U6 snRNA mutations, and the splicing defect caused by the deletion of *Sc* Prp17 (Query and Konarska 2012). Interestingly, this study also suggests that the *Sc* Cef1 R1 mutations, V36R and S48R (or –K), can activate alternative splice sites. They hypothesized that the charged mutations in R1 form new, stabilizing interactions resulting in alternative splice site selection. My results from Chapter II support an alternative hypothesis to the activation of alternative splice sites. The alternative hypothesis is that charged mutations in the *Sp* Cdc5/*Sc* Cef1 disrupt the tertiary structure of the Myb repeats rendering them inactive and without their RNA-RNA stabilizing effects, pre-mRNA processing becomes "sloppy", resulting in poor splice site selection. Supporting this hypothesis, the suppressor mutations identified by Query and Konarska are located in the hydrophobic core of R1 (Fig. 4-1B), suggesting that these mutations disrupt the tertiary fold of the Myb repeat instead of forming new binding sites. Combined with results from Chapter II, the inactivation/unfolding of R1 would also result in the loss of R2 function as both domains in *Sp* Cdc5 bind RNA cooperatively *in vitro*. With the loss of both Myb repeats RNA-RNA interactions would become destabilized resulting in aberrant splicing. Taken together these studies suggest *Sp* Cdc5/*Sc* Cef1's role in pre-mRNA processing is to

stabilize RNA-RNA interactions within the active site of the spliceosome to promote the fidelity of splicing.

Although *Sp* Cdc5 and its homologs have clear roles in splicing (Ohi, Feoktistova et al. 1998, Burns, Ohi et al. 1999, McDonald, Ohi et al. 1999, Tsai, Chow et al. 1999), negative genetic interactions with components located at the core of the spliceosome (Query and Konarska 2012, Livesay, Collier et al. 2013), and this study showing *Sp* Cdc5 can bind RNA *in vitro*, there are no direct *in vivo* studies describing a target RNA for *Sp* Cdc5 or its homologs within the spliceosome. **Chapter II leaves the open question of what RNA(s) interacts with the N-terminus of *Sp* Cdc5 *in vivo***. Future experiments should be focused on describing the protein-RNA interactions with Cdc5 and active site RNA. To accomplish this goal, RNA crosslinking should be used to detect protein-RNA interactions with Cdc5 *in vivo* and this can be done using ultraviolet (UV) crosslinking with endogenous RNA or exogenously added photoactivatable nucleosides (Urlaub, Hartmuth et al. 2000, Ascano, Hafner et al. 2012). Mapping protein-RNA contacts with the N-terminus of Cdc5 (R1, R2, and D3) and the spliceosome active site RNA will be an important step forward for understanding the molecular details on how Cdc5 contributes to RNA stabilization in the active site of the spliceosome.

The ability for the N-terminus of Cdc5 to bind RNA *in vitro* has implications for understanding how the NTC directly stabilizes active site RNA. Several NTC components have domains capable of binding RNA in addition to *Sp* Cdc5. For example, homologs of the NTC component, *Sp* Cwf2 (*Sc* Cwc2 and *Hs* RBM22), have been shown to bind active site RNA *in vivo* throughout multiple stages of splicing (McGrail, Krause

et al. 2009, Rasche, Dybkov et al. 2012). These studies with homologs of *Sp* Cwf2 used UV crosslinking to show direct protein-RNA interactions with the U6-ISL, the U6 snRNA and the pre-mRNA intron near the 5′SS. Due to *Sp* Cwf2 homologs binding the active site, these reports suggest *Sp* Cwf2 homologs are directly responsible for the stabilization of RNA-RNA interactions by the NTC. However, these reports fail to consider other potential RNA binding proteins in the NTC. In Chapter II, *Sp* Cdc5 is clearly capable of binding RNA *in vitro* and in combination with the genetic interactions with *Sp* Cwf2 and previously reported genetic interactions (Query and Konarska 2012, Livesay, Collier et al. 2013), *Sp* Cdc5 has a clear role in the active site of the spliceosome. In addition to *Sp* Cwf2 and *Sp* Cdc5, other NTC components may play a role in RNA stabilization by the NTC. For example the HAT (half-a-TPR) repeats found in both *Sp* Cwf3 and *Sp* Cwf4 can potentially bind RNA (Hammani, Cook et al. 2012) as well as the WD40 repeats which are found in *Sp* Prp19 and *Sp* Prp5 (Lunde, Moore et al. 2007, Lau, Bachorik et al. 2009, Tycowski, Shu et al. 2009). Further supporting the hypothesis that multiple NTC components are needed to stabilize active site RNA, depletion of the NTC in *S. cerevisiae* extracts results in the destabilization of the U6 snRNA with the 5′SS as well as the U5 snRNA with the 5′SS (Chan and Cheng 2005). The crosslinking data with homologs of *Sp* Cwf2 can directly account for U6 snRNA/5′SS stabilization only. At least one other NTC component is required to stabilize the U5 snRNA/5′SS interaction, again supporting the hypothesis that multiple NTC components are needed to stabilize RNA in the active site of the spliceosome. Previous work with *Sp* Cwf2 homologs and this work with the N-terminus of *Sp* Cdc5 are just the

tip of the iceberg in understanding how the NTC functions to stabilize RNA-RNA and protein-RNA interactions within the spliceosome. Further analyses of protein-RNA interactions with NTC components are needed to answer the question: **which NTC components can bind RNA and contribute to the stabilization of the active site of the spliceosome**?

**Macromolecular Characterization of the Spliceosome**

Within the confines of the spliceosome the NTC not only interacts with RNA but it must also interact with other spliceosome subcomplexes. The NTC and spliceosome research in general needs to be focused on the question of **how the spliceosome is assembled and how the inter-complex interactions change throughout the splicing cycle.** Understanding these inter-complex interactions within the spliceosome will provide mechanistic details on how the spliceosome is assembled to accomplish pre-mRNA splicing.

Toward the goal of understanding how the spliceosome is structurally organized, in Chapter III the structure of the spliceosome was investigated in the presence of a mutant RNA helicase, *Sp* Prp16. In summary, *S. pombe prp16-1* was shown to have growth and splicing defects at 36ºC. Further analysis of this temperature sensitive mutant showed that at 36ºC the RNA and protein composition as well as a 2D EM projection of a large splicing complex resembles a second step splicing complex similar to the *S. pombe* ILS. Analysis of the *prp16-1* second step splicing complex has identified a new spliceosome in *S. pombe* albeit similar to the ILS.

In Chapter III, the isolation of a second step splicing complex using a mutant of *Sp* Prp16 is consistent with previously published literature from homologs of *Sp* Prp16. The hypothesis in Chapter III is that a mutant of *Sp* Prp16 would "trap" a new second step splicing complex in *S. pombe*. This hypothesis is based on the well-characterized role of *Sc* Prp16 and *Hs* Prp16. Supporting this hypothesis, *Sc* Prp16 acts after the first step of splicing to promote the rearrangement of the active site of the spliceosome resulting in protection of the 3′SS from RNase H digestion and the formation of the C complex (Schwer and Guthrie 1991). *Sc* Prp16 also functions to remove the proteins *Sc* Cwc25 (*Sp* Cwc25) and *Sc* Yju2 (*Sp* Cwf16) although by MS these proteins are still detected in the *S. cerevisiae* C complex trapped on a mutant pre-mRNA (Fabrizio, Dannenberg et al. 2009). Finally, a mutant of *Sc* Prp16, *Sc prp16-302,* has been used to trap a second step splicing complex (Lardelli, Thompson et al. 2010). The *Sp prp16-1* spliceosome purified in Chapter III is consistent with second step splicing complexes purified in *S. cerevisiae*. Although there are variations between the *S. pombe* and *S. cerevisiae* complexes, these disparities are also found in different purifications from *S. cerevisiae* and may reflect different purification and isolation conditions between complexes (Lardelli, Thompson et al. 2010). Taken together, the isolation of a second step splicing complex in *S. pombe* using a mutant of the RNA helicase, *Sp* Prp16, is consistent with previously published work in both *S. cerevisiae* and humans.

The *S. pombe prp16-1* second step spliceosome is a step forward in describing new splicing complexes in *S. pombe* and understanding how the spliceosome is structurally organized. MS analysis suggests this new *S. pombe* spliceosome represents a

heterogeneous mixture of second step splicing complexes with the SF3A and SF3B complexes being found in low abundance. The low abundance of these factors is also seen by MS in *S. cerevisiae* C complexes (Fabrizio, Dannenberg et al. 2009). In order for the *Sp prp16-1* to be structurally characterized, a more homogeneous complex is needed. One way to increase particle homogeneity of this complex is to use GraFix, a linear glycerol gradient with protein crosslinking, to keep the complex from disassociating after purification. GraFix has been used to purify a more homogeneous human spliceosome (Stark 2010). In the human model system there are two 3D structures of the C complex with clear discrepancies between the structures (Jurica, Sousa et al. 2004, Golas, Sander et al. 2010). These discrepancies were resolved when 2D EM projections of human C complexes were compared with and without crosslinking (Ilagan, Chalkley et al. 2013). Results from the human C complex clearly show that protein crosslinking can facilitate complex purification by preserving a more homogeneous sample. With further sample optimization the 3D structure of the *S. pombe prp16-1* spliceosome can be determined.

In Chapter III, a mutant RNA helicase was used to trap a second step splicing complex, but this study leaves the **question of whether other *S. pombe* helicase mutants can be used to isolate splicing complexes for structural characterization**. Consistent with other organisms, *S. pombe* has several RNA helicases involved in splicing but how mutants of these *S. pombe* helicases affect the structure of the spliceosome has yet to be determined. Expanding beyond the second step of splicing, mutants of RNA helicases *Sp* Prp28 and *Sp* Cdc28 are interesting candidates as they are positioned in the splicing cycle to "trap" first step splicing complexes. Structural

comparison of first and second step spliceosomes purified from RNA helicase mutants will provide details on how the spliceosome changes throughout the splicing cycle and understanding these changes is critical for fully comprehending how the spliceosome functions.

In Chapter III, the structure of a second step splicing complex in *S. pombe* was characterized. Although this complex is heterogeneous, using EM there are clear differences in the "head-like" domain when compared to the previously characterized *S. pombe* ILS. The isolation of a second step splicing complex in Chapter III now allows for the structures of the spliceosome in *S. pombe* to be compared using EM. Future work will focus on characterizing the 3D EM structure of the *prp16-1* spliceosome with the goal of comparing this structure to the previously described 3D structure of the ILS (Ohi, Ren et al. 2007). The comparison of second step splicing complexes in *S. pombe* in addition to structures of first step splicing complexes, potentially isolated using mutants of RNA helicases, will provide details on how pre-mRNA processing is facilitated by the spliceosome and with increasing structural resolution of these complexes more details of pre-mRNA processing will be elucidated (Fig 4-2). For example, the current EM model of the *S. pombe* ILS is at 29Å and at this resolution the general shape and size of a structure can be determined (Ohi, Ren et al. 2007). Although no mechanistic details can be determined at this resolution, several techniques including difference mapping and domain localization can be used to understand gross changes in the spliceosome when comparing spliceosome structures at different stages of splicing. With an increase in resolution to ~12Å, RNA within the spliceosome can be distinguished from protein and

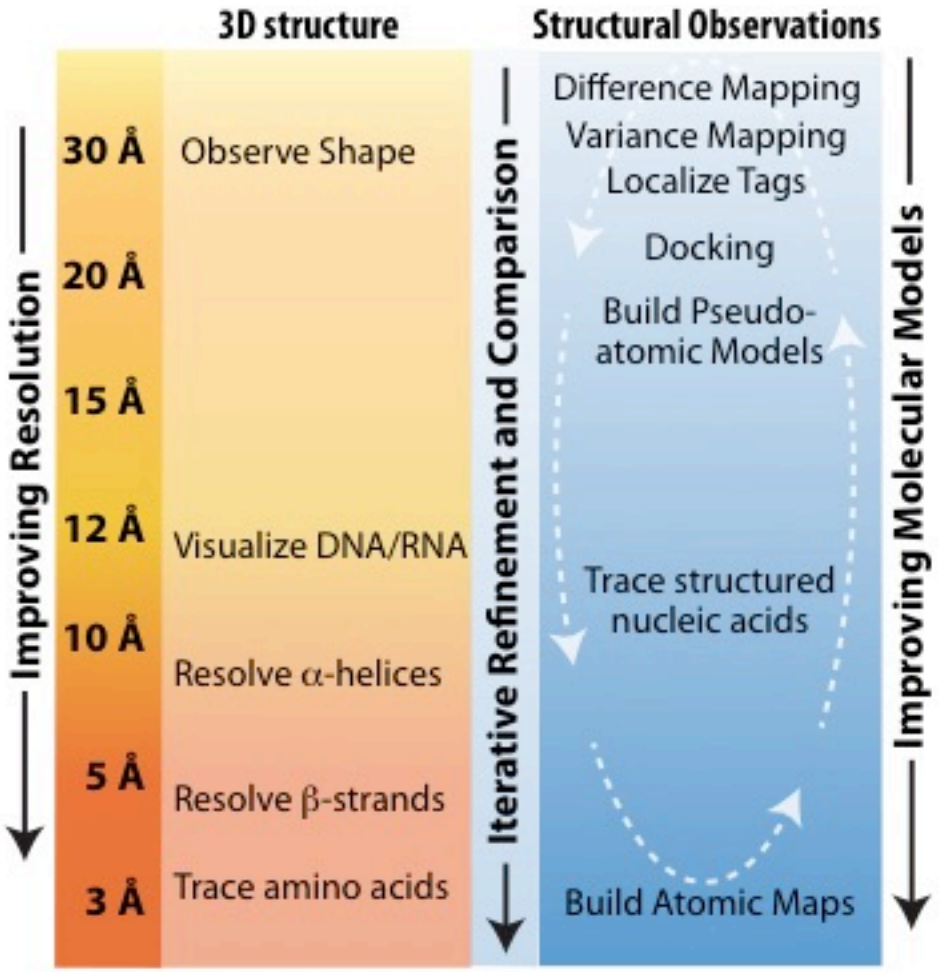



**Figure 4-2. Schematic of Structural Observations that can be made at Different Resolutions** Increasingly detailed molecular models of macromolecules can be generated when intermediate-resolution structures are combined with approaches that localize individual components in 3D maps, dock available high-resolution X-ray structures of individual components into lower resolution EM densities, and map protein-protein interaction networks. The use of hybrid approaches has been a important for building pseudo-atomic models of structurally challenging complexes. (Figure courtesy of M. Ohi)

the general location of the active site of the spliceosome can be determined. The tertiary structure of the RNA can then be compared with that of Group II introns and used to potentially support the spliceosome as a ribozyme hypothesis. The protein components of the spliceosome can start to be resolved with a further increase in resolution with α-helices being resolved at ~10Å and β-strands being resolved at ~5Å. The 5Å-10Å resolution range is particularly useful for generating pseudo-atomic models of the spliceosome. These pseudo-atomic models will provide an unprecedented level of detail on how the spliceosome is assembled and functions to process pre-mRNA. Finally at a resolution of ~3Å the amino acid backbone can be traced and at this resolution atomic models of the spliceosome can be made answering questions of how the spliceosome functions and piecing together decades of genetic and functional data on the spliceosome. Determining high resolution structures of multiple splicing complexes will allow for mechanistic details to be determined on how proteins and RNA are rearranged throughout the splicing cycle.

Processing pre-mRNAs by the spliceosome is a basic biological property required for life in eukaryotic organisms. Since pre-mRNA splicing is essential to life, perturbations in this process through mutations in pre-mRNAs or in splicing factors can often result in disease (López-Bigas, Audit et al. 2005, Ward and Cooper 2009, Douglas and Wood 2011, Xiong, Alipanahi et al. 2015). A simplified model system is often used to understand the mechanism by which these mutations cause disease. Previous and ongoing work in *S. pombe* shows that this simplified system correlates well in humans, with similar human disease mutations being found in *S. pombe* and splicing-specific

drugs in humans affecting the same processes in *S. pombe*. For example, in humans, mutations in the *Hs SF3B1* gene are often found associated with chronic lymphocytic leukemia (CLL) (Hahn and Scott 2011, Lu, Lu et al. 2011). The disease correlative mutations in *Hs* SF3B1 are grouped in the HEAT (Huntingtin, EF3, PP2A, and Tor1) repeats. *Hs* SF3B1 is conserved in *S. pombe* (*Sp* Prp10) and interestingly there is a temperature sensitive mutant of *Sp* Prp10 which has a mutation in the HEAT repeats (Habara, Urushiyama et al. 1998). With similar mutations between *Hs* SF3B1 and *Sp* Prp10 affecting cell function, what is learned by studying splicing in *S. pombe* can be correlated with human disease. Further supporting this idea is evidence that the drug, spliceostatin A, inhibits the *S. pombe* spliceosome. In both human and *S. pombe* model systems spliceostatin A binds to components of the SF3B complex and promotes pre-mRNA accumulation in the nucleus (Kaida, Motoyoshi et al. 2007, Lo, Kaida et al. 2007). Similar splicing protein mutations and drug responses in the *S. pombe* and human model systems suggest that what is learned in *S. pombe* will correlate with the human model system.

The study of the spliceosome centers on understanding a basic biological function of eukaryotic organisms. The ultimate goal of this research is to understand how the essential function of the spliceosome affects human health and disease. Studying the spliceosome under non-disease conditions will provide insight into how the spliceosome is perturbed due to mutations in pre-mRNA and splicing components. For example, understanding how different regions of a pre-mRNA effect splicing has led to the generation of an algorithm which scores how strongly genetic variants affect RNA

splicing (Xiong, Alipanahi et al. 2015). 1393 unique pre-mRNA sequence features and known RNA binding proteins were used to train an algorithm and predict with relatively accuracy the effects of known single nucleotide variants on human disease. Although this algorithm was relatively successful, it cannot take into account all of the effects the spliceosome as the **detailed mechanisms of pre-mRNA processing by the spliceosome are not known**. With further characterization of the mechanisms of pre-mRNA splicing by the spliceosome, this algorithm or algorithms like it will be able to more accurately assess the effects of mutations in pre-mRNAs and the spliceosome.

In summary, this work characterizes the structure and function of the *S. pombe* spliceosome. This work furthers splicing research in *S. pombe* by characterizing *Sp* Cdc5 and *Sp* Prp16. The N-terminus of *Sp* Cdc5 was shown to bind RNA which has implications for how the NTC stabilizes and activates the spliceosome. A mutant of *Sp* Prp16, an RNA helicase, was used to isolate a new second step splicing complex in *S. pombe.* This work is a step forward in understanding the mechanisms behind pre-mRNA processing by the spliceosome by focusing on characterizing protein-RNA interactions in the active site of the spliceosome. Although this work progresses our understanding of spliceosome function, several questions remain, such as; (1) What RNA(s) does *Sp* Cdc5 bind *in vivo*? (2) Which NTC components can bind RNA and contribute to the stabilization of the active site of the spliceosome? (3) How is the spliceosome structurally organized? (4) Can other *S. pombe* helicase mutants be used to isolate splicing complexes for structural characterization? (5) What is the detailed mechanism of pre-mRNA processing by the spliceosome?

## PUBLISHED MANUSCRIPTS

1. Johnson, A. E., **Collier, S. E.**, Ohi, M. D., and Gould, K. L. (2012) Fission yeast Dma1 requires RING domain dimerization for its ubiquitin ligase activity and mitotic checkpoint function, In *J Biol Chem*, pp 25741-25748.
2. Chambers, M. G., Pyburn, T. M., González-Rivera, C., **Collier, S. E.**, Eli, I., Yip, C. K., Takizawa, Y., Lacy, D. B., Cover, T. L., and Ohi, M. D. (2013) Structural analysis of the oligomeric states of Helicobacter pylori VacA toxin, *J Mol Biol 425*, 524-535.
3. Folkmann, A. W., **Collier, S. E.**, Zhan, X., Aditi, Ohi, M. D., and Wente, S. R. (2013) Gle1 functions during mRNA export in an oligomeric complex that is altered in human disease, In *Cell*, pp 582-593.
4. Livesay, S. B., **Collier, S. E.**, Bitton, D. A., Bähler, J., and Ohi, M. D. (2013) Structural and functional characterization of the N terminus of Schizosaccharomyces pombe Cwf10, *Eukaryot Cell 12*, 1472-1489.
5. Sturgill, E. G., Das, D. K., Takizawa, Y., Shin, Y., **Collier, S. E.**, Ohi, M. D., Hwang, W., Lang, M. J., and Ohi, R. (2014) Kinesin-12 Kif15 targets kinetochore fibers through an intrinsic two-step mechanism, *Curr Biol 24*, 2307-2313.
6. **Collier, S. E.**, Voehler, M., Peng, D., Ohi, R., Gould, K. L., Reiter, N. J., and Ohi, M. D. (2014) Structural and functional insights into the N-terminus of Schizosaccharomyces pombe Cdc5, *Biochemistry 53*, 6439-6451.
7. Gao, M., Thomson, T. C., Creed, T. M., Tu, S., Loganathan, S. N., Jackson, C. A., McCluskey, P., Lin, Y., **Collier, S. E.**, Weng, Z., Lasko, P., Ohi, M. D., and Arkov, A. L. (2015) Glycolytic enzymes localize to ribonucleoprotein granules in Drosophila germ cells, bind Tudor and protect from transposable elements, *EMBO Rep*.

## REFERENCES


Abelson, J. (2008). "Is the spliceosome a ribonucleoprotein enzyme?" Nat Struct Mol Biol **15**(12): 1235-1237.

Ajuh, P., B. Kuster, K. Panov, J. C. Zomerdijk, M. Mann and A. I. Lamond (2000). "Functional analysis of the human CDC5L complex and identification of its components by mass spectrometry." EMBO J **19**(23): 6569-6581.

Albers, M., A. Diment, M. Muraru, C. S. Russell and J. D. Beggs (2003). Identification and characterization of Prp45p and Prp46p, essential pre-mRNA splicing factors. RNA. **9:** 138-150.

Ambrósio, D. L., N. Badjatia and A. Günzl (2015). "The spliceosomal PRP19 complex of trypanosomes." Mol Microbiol **95**(5): 885-901.

Aravind, L., H. Watanabe, D. J. Lipman and E. V. Koonin (2000). "Lineage-specific loss and divergence of functionally linked genes in eukaryotes." Proc Natl Acad Sci U S A **97**(21): 11319-11324.

Aronova, A., D. Bacíková, L. B. Crotti, D. S. Horowitz and B. Schwer (2007). "Functional interactions between Prp8, Prp18, Slu7, and U5 snRNA during the second step of pre-mRNA splicing." RNA **13**(9): 1437-1444.

Ascano, M., M. Hafner, P. Cekan, S. Gerstberger and T. Tuschl (2012). "Identification of RNA-protein interaction networks using PAR-CLIP." Wiley Interdiscip Rev RNA **3**(2): 159-177.

Azubel, M., S. G. Wolf, J. Sperling and R. Sperling (2004). Three-dimensional structure of the native spliceosome by cryo-electron microscopy. Mol Cell. **15:** 833-839.

Barrett, P. J., Y. Song, W. D. Van Horn, E. J. Hustedt, J. M. Schafer, A. Hadziselimovic, A. J. Beel and C. R. Sanders (2012). "The amyloid precursor protein has a flexible transmembrane domain and binds cholesterol." Science **336**(6085): 1168-1171.

Bartels, C., C. Klatt, R. Lührmann and P. Fabrizio (2002). "The ribosomal translocase homologue Snu114p is involved in unwinding U4/U6 RNA during activation of the spliceosome." EMBO Rep **3**(9): 875-880.

Behzadnia, N., M. M. Golas, K. Hartmuth, B. Sander, B. Kastner, J. Deckert, P. Dube, C. L. Will, H. Urlaub, H. Stark and R. Lührmann (2007). Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. EMBO J. **26:** 1737-1748.

Ben-Yehuda, S., I. Dix, C. S. Russell, M. McGarvey, J. D. Beggs and M. Kupiec (2000). "Genetic and physical interactions between factors involved in both cell cycle progression and pre-mRNA splicing in Saccharomyces cerevisiae." Genetics **156**(4): 1503-1517.

Bentley, D. (2002). "The mRNA assembly line: transcription and processing machines in the same factory." Curr Opin Cell Biol **14**(3): 336-342.

Bessonov, S., M. Anokhina, A. Krasauskas, M. M. Golas, B. Sander, C. L. Will, H. Urlaub, H. Stark and R. Lührmann (2010). "Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis." RNA **16**(12): 2384-2403.

Bessonov, S., M. Anokhina, C. L. Will, H. Urlaub and R. Lührmann (2008). "Isolation of an active step I spliceosome and composition of its RNP core." Nature **452**(7189): 846-850.

Boehringer, D., E. M. Makarov, B. Sander, O. V. Makarova, B. Kastner, R. Lührmann and H. Stark (2004). "Three-dimensional structure of a pre-catalytic human spliceosomal complex B." Nat Struct Mol Biol **11**(5): 463-468.

Bonomi, S., S. Gallo, M. Catillo, D. Pignataro, G. Biamonti and C. Ghigna (2013). "Oncogenic alternative splicing switches: role in cancer progression and prospects for therapy." Int J Cell Biol **2013**: 962038.

Boon, K. L., T. Auchynnikava, G. Edwalds-Gilbert, J. D. Barrass, A. P. Droop, C. Dez and J. D. Beggs (2006). "Yeast ntr1/spp382 mediates prp43 function in postspliceosomes." Mol Cell Biol **26**(16): 6016-6023.

Boon, K. L., R. J. Grainger, P. Ehsani, J. D. Barrass, T. Auchynnikava, C. F. Inglehearn and J. D. Beggs (2007). "prp8 mutations that cause human retinitis pigmentosa lead to a U5 snRNP maturation defect in yeast." Nat Struct Mol Biol **14**(11): 1077-1083.

Bottner, C. A., H. Schmidt, S. Vogel, M. Michele and N. F. Käufer (2005). Multiple genetic and biochemical interactions of Brr2, Prp8, Prp31, Prp1 and Prp4 kinase suggest a function in the control of the activation of spliceosomes in Schizosaccharomyces pombe. Curr Genet. **48:** 151-161.

Braun, V., M. Mehlig, M. Moos, M. Rupnik, B. Kalt, D. E. Mahony and C. von Eichel-Streiber (2000). "A chimeric ribozyme in clostridium difficile combines features of group I introns and insertion elements." Mol Microbiol **36**(6): 1447-1459.

Brenner, T. J. and C. Guthrie (2005). "Genetic analysis reveals a role for the C terminus of the Saccharomyces cerevisiae GTPase Snu114 during spliceosome activation." Genetics **170**(3): 1063-1080.

Brenner, T. J. and C. Guthrie (2006). "Assembly of Snu114 into U5 snRNP requires Prp8 and a functional GTPase domain." RNA **12**(5): 862-871.

Brown, G. G., C. Colas des Francs-Small and O. Ostersetzer-Biran (2014). "Group II intron splicing factors in plant mitochondria." Front Plant Sci **5**: 35.

Burgess, S., J. R. Couto and C. Guthrie (1990). "A putative ATP binding protein influences the fidelity of branchpoint recognition in yeast splicing." Cell **60**(5): 705-717.

Burgess, S. M. and C. Guthrie (1993). "A mechanism to enhance mRNA splicing fidelity: the RNA-dependent ATPase Prp16 governs usage of a discard pathway for aberrant lariat intermediates." Cell **73**(7): 1377-1391.

Burke, J. E., D. G. Sashital, X. Zuo, Y.-X. Wang and S. E. Butcher (2012). "Structure of the yeast U2/U6 snRNA complex." RNA **18**(4): 673-683.

Burns, C. G., R. Ohi, A. R. Krainer and K. L. Gould (1999). Evidence that Myb-related CDC5 proteins are required for pre-mRNA splicing. Proc Natl Acad Sci USA. **96:** 13789-13794.

Burns, C. G., R. Ohi, S. Mehta, E. T. O'Toole, M. Winey, T. A. Clark, C. W. Sugnet, M. Ares and K. L. Gould (2002). Removal of a single alpha-tubulin gene intron suppresses cell cycle arrest phenotypes of splicing factor mutations in Saccharomyces cerevisiae. Mol Cell Biol. **22:** 801-815.

Busch, A. and K. J. Hertel (2012). "Evolution of SR protein and hnRNP splicing regulatory factors." Wiley Interdiscip Rev RNA **3**(1): 1-12.

Butcher, S. E. (2009). "The spliceosome as ribozyme hypothesis takes a second step." Proc Natl Acad Sci U S A **106**(30): 12211-12212.

Chan, S.-P. and S.-C. Cheng (2005). "The Prp19-associated complex is required for specifying interactions of U5 and U6 with pre-mRNA during spliceosome activation." J Biol Chem **280**(35): 31190-31199.

.

Chen, C.-H., W.-C. Yu, T. Y. Tsao, L.-Y. Wang, H.-R. Chen, J.-Y. Lin, W.-Y. Tsai and S.-C. Cheng (2002). "Functional and physical interactions between components of the Prp19p-associated complex." Nucleic Acids Res **30**(4): 1029-1037.

Chen, C. H., W. Y. Tsai, H. R. Chen, C. H. Wang and S. C. Cheng (2001). "Identification and characterization of two novel components of the Prp19p-associated complex, Ntc30p and Ntc20p." J Biol Chem **276**(1): 488-494.

Chen, H. R., S. P. Jan, T. Y. Tsao, Y. J. Sheu, J. Banroques and S. C. Cheng (1998). "Snt309p, a component of the Prp19p-associated complex that interacts with Prp19p and associates with the spliceosome simultaneously with or immediately after dissociation of U4 in the same manner as Prp19p." Mol Cell Biol **18**(4): 2196-2204.

Chen, J. Y., L. Stands, J. P. Staley, R. R. Jackups, L. J. Latus and T. H. Chang (2001). "Specific alterations of U1-C protein or U1 small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD box splicing factor." Mol Cell **7**(1): 227-232.

Chen, W. and M. J. Moore (2014). "The spliceosome: disorder and dynamics defined." Curr Opin Struct Biol **24**: 141-149.

Chen, W., H. P. Shulha, A. Ashar-Patel, J. Yan, K. M. Green, C. C. Query, N. Rhind, Z. Weng and M. J. Moore (2014). "Endogenous U2·U5·U6 snRNA complexes in S. pombe are intron lariat spliceosomes." RNA **20**(3): 308-320.

Cheng, H., W. M. Westler, B. Xia, B. H. Oh and J. L. Markley (1995). "Protein expression, selective isotopic labeling, and analysis of hyperfine-shifted NMR signals of Anabaena 7120 vegetative [2Fe-2S]ferredoxin." Arch Biochem Biophys **316**(1): 619-634.

Chu, V. T., Q. Liu, M. Podar, P. S. Perlman and A. M. Pyle (1998). "More than one way to splice an RNA: branching without a bulge and splicing without branching in group II introns." RNA **4**(10): 1186-1202.

Clubb, R. T., V. Thanabal and G. Wagner (1992). "A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue 1HN, 15N, and 13C′ chemical shifts in15N-13C-labelled proteins." J. Magn. Reson. **97**: 213-217.

Collart, M. A. and S. Oliviero (2001). "Preparation of yeast RNA." Curr Protoc Mol Biol **Chapter 13**: Unit13.12.

Collins, C. A. and C. Guthrie (2000). "The question remains: is the spliceosome a ribozyme?" Nat Struct Biol **7**(10): 850-854.

Company, M., J. Arenas and J. Abelson (1991). "Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes." Nature **349**(6309): 487-493.

Consortium, I. H. G. S. (2004). "Finishing the euchromatic sequence of the human genome." Nature **431**(7011): 931-945.

Cordin, O. and J. D. Beggs (2013). "RNA helicases in splicing." RNA Biol **10**(1): 83-95.

Cordin, O., D. Hahn, R. Alexander, A. Gautam, C. Saveanu, J. D. Barrass and J. D. Beggs (2014). "Brr2p carboxy-terminal Sec63 domain modulates Prp16 splicing RNA helicase." Nucleic Acids Res **42**(22): 13897-13910.

Cvitkovic, I. and M. S. Jurica (2013). "Spliceosome database: a tool for tracking components of the spliceosome." Nucleic Acids Res **41**(Database issue): D132-141.

Datta, B. and A. M. Weiner (1991). "Genetic evidence for base pairing between U2 and U6 snRNA in mammalian mRNA splicing." Nature **352**(6338): 821-824.

Davis, A. L., J. Keeler, E. D. Laue and D. Moskau (1992). "Experiments for recording pure-absorption heteronuclear correlation spectra using pulsed field gradients." J. Magn. Reson. **98**: 207-216.

de Andrade, H. H., E. K. Marques, A. C. Schenberg and J. A. Henriques (1989). "The PSO4 gene is responsible for an error-prone recombinational DNA repair pathway in Saccharomyces cerevisiae." Mol Gen Genet **217**(2-3): 419-426.

Deckert, J., K. Hartmuth, D. Boehringer, N. Behzadnia, C. L. Will, B. Kastner, H. Stark, H. Urlaub and R. Luhrmann (2006). "Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions." Mol Cell Biol **26**(14): 5528-5543.

Douglas, A. G. and M. J. Wood (2011). "RNA splicing: disease and therapy." Brief Funct Genomics **10**(3): 151-164.

Dyballa, N. and S. Metzger (2009). "Fast and sensitive colloidal coomassie G-250 staining for proteins in polyacrylamide gels." J Vis Exp(30).

Emerick, V. L. and S. A. Woodson (1994). "Fingerprinting the folding of a group I precursor RNA." Proc Natl Acad Sci U S A **91**(21): 9675-9679.

Fabrizio, P., J. Dannenberg, P. Dube, B. Kastner, H. Stark, H. Urlaub and R. Luhrmann (2009). "The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome." Mol Cell **36**(4): 593-608.

Fabrizio, P., J. Dannenberg, P. Dube, B. Kastner, H. Stark, H. Urlaub and R. Lührmann (2009). "The Evolutionarily Conserved Core Design of the Catalytic Activation Step of the Yeast Spliceosome." Mol Cell **36**(4): 593-608.

Fabrizio, P., B. Laggerbauer, J. Lauber, W. S. Lane and R. Lührmann (1997). "An evolutionarily conserved U5 snRNP-specific protein is a GTP-binding factor closely related to the ribosomal translocase EF-2." EMBO J **16**(13): 4092-4106.

Faustino, N. A. and T. A. Cooper (2003). "Pre-mRNA splicing and human disease." Genes Dev **17**(4): 419-437.

Fica, S. M., M. A. Mefford, J. A. Piccirilli and J. P. Staley (2014). "Evidence for a group II intron-like catalytic triplex in the spliceosome." Nat Struct Mol Biol **21**(5): 464-471.

Fica, S. M., N. Tuttle, T. Novak, N. S. Li, J. Lu, P. Koodathingal, Q. Dai, J. P. Staley and J. A. Piccirilli (2013). "RNA catalyses nuclear pre-mRNA splicing." Nature **503**(7475): 229-234.

Fleckner, J., M. Zhang, J. Valcárcel and M. R. Green (1997). "U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction." Genes Dev **11**(14): 1864-1872.

Fourmann, J. B., J. Schmitzová, H. Christian, H. Urlaub, R. Ficner, K. L. Boon, P. Fabrizio and R. Lührmann (2013). "Dissection of the factor requirements for spliceosome disassembly and the elucidation of its dissociation products using a purified splicing system." Genes Dev **27**(4): 413-428.

Galej, W. P., T. H. Nguyen, A. J. Newman and K. Nagai (2014). "Structural studies of the spliceosome: zooming into the heart of the machine." Curr Opin Struct Biol **25**: 57-66.

Galej, W. P., C. Oubridge, A. J. Newman and K. Nagai (2013). "Crystal structure of Prp8 reveals active site cavity of the spliceosome." Nature **493**(7434): 638-643.

Golas, M. M., B. Sander, S. Bessonov, M. Grote, E. Wolf, B. Kastner, H. Stark and R. Lührmann (2010). 3D cryo-EM structure of an active step I spliceosome and localization of its catalytic core. Mol Cell. **40:** 927-938.

Golden, B. L., A. R. Gooding, E. R. Podell and T. R. Cech (1998). "A preorganized active site in the crystal structure of the Tetrahymena ribozyme." Science **282**(5387): 259-264.

Gordon, P. M., E. J. Sontheimer and J. A. Piccirilli (2000). "Metal ion catalysis during the exon-ligation step of nuclear pre-mRNA splicing: extending the parallels between the spliceosome and group II introns." RNA **6**(2): 199-205.

Gould, K. L., L. Ren, A. S. Feoktistova, J. L. Jennings and A. J. Link (2004). "Tandem affinity purification and identification of protein complex components." Methods **33**(3): 239-244.

Greenfield, N. and G. D. Fasman (1969). "Computed circular dichroism spectra for the evaluation of protein conformation." Biochemistry **8**(10): 4108-4116.

Grote, M., E. Wolf, C. L. Will, I. Lemm, D. E. Agafonov, A. Schomburg, W. Fischle, H. Urlaub and R. Luhrmann (2010). "Molecular architecture of the human Prp19/CDC5L complex." Mol Cell Biol **30**(9): 2105-2119.

Grote, M., E. Wolf, C. L. Will, I. Lemm, D. E. Agafonov, A. Schomburg, W. Fischle, H. Urlaub and R. Lührmann (2010). "The Molecular Architecture of the Human Prp19/CDC5L Complex." Mol Cell Biol **30**(9): 2105–2119.

Grzesiek, S. and A. Bax (1992). "An Efficient Experiment for SequentialBackboneAssignment of Medium-Sized Isotopically Enriched Protein." Journal of Magnetic Resonance **99**: 201-207.

Grzesiek, S. and A. Bax (1993). "Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins." J Biomol NMR **3**(2): 185-204.

Grzesiek, S. and A. Bax (1993). "The importance of not saturating water in protein NMR. Application to sensitivity enhancement and NOE measurements." J. Am. Chem. Soc. **115**(26): 12593-12594.

Gómez, E. B., V. T. Angeles and S. L. Forsburg (2005). "A screen for Schizosaccharomyces pombe mutants defective in rereplication identifies new alleles of rad4+, cut9+ and psf2+." Genetics **169**(1): 77-89.

Habara, Y., S. Urushiyama, T. Tani and Y. Ohshima (1998). "The fission yeast prp10(+) gene involved in pre-mRNA splicing encodes a homologue of highly conserved splicing factor, SAP155." Nucleic Acids Res **26**(24): 5662-5669.

Hage, R., L. Tung, H. Du, L. Stands, M. Rosbash and T. H. Chang (2009). "A targeted bypass screen identifies Ynl187p, Prp42p, Snu71p, and Cbp80p for stable U1 snRNP/Pre-mRNA interaction." Mol Cell Biol **29**(14): 3941-3952.

Hahn, C. N. and H. S. Scott (2011). Spliceosome mutations in hematopoietic malignancies. Nat Genet. **44:** 9-10.

Hahn, D., G. Kudla, D. Tollervey and J. D. Beggs (2012). "Brr2p-mediated conformational rearrangements in the spliceosome during activation and substrate repositioning." Genes Dev **26**(21): 2408-2421.

Hammani, K., W. B. Cook and A. Barkan (2012). "RNA binding and RNA remodeling activities of the half-a-tetratricopeptide (HAT) protein HCF107 underlie its effects on gene expression." Proc Natl Acad Sci U S A **109**(15): 5651-5656.

Harris-Kerr, C. L., M. Zhang and C. L. Peebles (1993). "The phylogenetically predicted base-pairing interaction between alpha and alpha' is required for group II splicing in vitro." Proc Natl Acad Sci U S A **90**(22): 10658-10662.

Hartwell, L. H. (1967). Macromolecule synthesis in temperature-sensitive mutants of yeast. J Bacteriol. **93:** 1662-1670.

Hartwell, L. H. (1978). "Cell division from a genetic perspective." J Cell Biol **77**(3): 627-637.

Hausner, G., M. Hafez and D. R. Edgell (2014). "Bacterial group I introns: mobile RNA catalysts." Mob DNA **5**(1): 8.

Hausner, T. P., L. M. Giglio and A. M. Weiner (1990). "Evidence for base-pairing between mammalian U2 and U6 small nuclear ribonucleoprotein particles." Genes Dev **4**(12A): 2146-2156.

Henriques, J. A., E. J. Vicente, K. V. Leandro da Silva and A. C. Schenberg (1989). "PSO4: a novel gene involved in error-prone repair in Saccharomyces cerevisiae." Mutat Res **218**(2): 111-124.

Herold, N., C. L. Will, E. Wolf, B. Kastner, H. Urlaub and R. Lührmann (2009). Conservation of the protein composition and electron microscopy structure of Drosophila melanogaster and human spliceosomal complexes. Mol Cell Biol. **29:** 281-301.

Higgs, P. G. and N. Lehman (2015). "The RNA World: molecular cooperation at the origins of life." Nat Rev Genet **16**(1): 7-17.

Hirayama, T. and K. Shinozaki (1996). "A cdc5+ homolog of a higher plant, Arabidopsis thaliana." Proc Natl Acad Sci USA **93**(23): 13371-13376.

Hofmann, J. C., A. Husedzinovic and O. J. Gruss (2010). The function of spliceosome components in open mitosis. Nucleus (Austin, Tex). **1:** 447-459.

Hofmann, J. C., J. Tegha-Dunghu, S. Dräger, C. L. Will, R. Lührmann and O. J. Gruss (2013). "The Prp19 complex directly functions in mitotic spindle assembly." PLoS One **8**(9): e74851.

Hogg, R., R. A. de Almeida, J. P. Ruckshanthi and R. T. O'Keefe (2014). "Remodeling of U2-U6 snRNA helix I during pre-mRNA splicing by Prp16 and the NineTeen Complex protein Cwc2." Nucleic Acids Res **42**(12): 8008-8023.

Hogg, R., J. C. McGrail and R. T. O'Keefe (2010). "The function of the NineTeen Complex (NTC) in regulating spliceosome conformations and fidelity during pre-mRNA splicing." Biochem Soc Trans **38**(4): 1110-1115.

Hoskins, A. A. and M. J. Moore (2012). The spliceosome: a flexible, reversible macromolecular machine. Trends Biochem Sci.

Hotz, H. R. and B. Schwer (1998). "Mutational analysis of the yeast DEAH-box splicing factor Prp16." Genetics **149**(2): 807-815.

Huang, T., J. Vilardell and C. C. Query (2002). "Pre-spliceosome formation in S.pombe requires a stable complex of SF1-U2AF(59)-U2AF(23)." EMBO J **21**(20): 5516-5526.

Huppler, A., L. J. Nikstad, A. M. Allmann, D. A. Brow and S. E. Butcher (2002). "Metal binding and base ionization in the U6 RNA intramolecular stem-loop structure." Nat Struct Biol **9**(6): 431-435.

Häcker, I., B. Sander, M. M. Golas, E. Wolf, E. Karagöz, B. Kastner, H. Stark, P. Fabrizio and R. Lührmann (2008). "Localization of Prp8, Brr2, Snu114 and U4/U6 proteins in the yeast tri-snRNP by electron microscopy." Nat Struct Mol Biol **15**(11): 1206-1212.

Ilagan, J. O., R. J. Chalkley, A. L. Burlingame and M. S. Jurica (2013). "Rearrangements within human spliceosomes captured after exon ligation." RNA **19**(3): 400-412.

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

Jhanwar, S. C. (2014). "Genetic and epigenetic pathways in myelodysplastic syndromes: A brief overview." Adv Biol Regul.

Jurica, M. S., L. J. Licklider, S. R. Gygi, N. Grigorieff and M. J. Moore (2002). "Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis." RNA **8**(4): 426-439.

Jurica, M. S., D. Sousa, M. J. Moore and N. Grigorieff (2004). Three-dimensional structure of C complex spliceosomes by electron microscopy. Nat Struct Mol Biol. **11:** 265-269.

Kaida, D., H. Motoyoshi, E. Tashiro, T. Nojima, M. Hagiwara, K. Ishigami, H. Watanabe, T. Kitahara, T. Yoshida, H. Nakajima, T. Tani, S. Horinouchi and M. Yoshida (2007). "Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA." Nat Chem Biol **3**(9): 576-583.

Kanei-Ishii, C., A. Sarai, T. Sawazaki, H. Nakagoshi, D. N. He, K. Ogata, Y. Nishimura and S. Ishii (1990). "The tryptophan cluster: a hypothetical structure of the DNA-binding domain of the myb protooncogene product." J Biol Chem **265**(32): 19990-19995.

Keeney, J. B. and J. D. Boeke (1994). "Efficient targeted integration at leu1-32 and ura4-294 in Schizosaccharomyces pombe." Genetics **136**(3): 849-856.

Kim, D. H. and J. J. Rossi (1999). "The first ATPase domain of the yeast 246-kDa protein is required for in vivo unwinding of the U4/U6 duplex." RNA **5**(7): 959-971.

Kim, S. H. and R. J. Lin (1993). "Pre-mRNA splicing within an assembled yeast spliceosome requires an RNA-dependent ATPase and ATP hydrolysis." Proc Natl Acad Sci U S A **90**(3): 888-892.

Kim, S. H. and R. J. Lin (1996). "Spliceosome activation by PRP2 ATPase prior to the first transesterification reaction of pre-mRNA splicing." Mol Cell Biol **16**(12): 6810-6819.

Kistler, A. L. and C. Guthrie (2001). "Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for sub2, an essential spliceosomal ATPase." Genes Dev **15**(1): 42-49.

Koodathingal, P., T. Novak, J. A. Piccirilli and J. P. Staley (2010). "The DEAH box ATPases Prp16 and Prp43 cooperate to proofread 5' splice site cleavage during pre-mRNA splicing." Mol Cell **39**(3): 385-395.

Koodathingal, P. and J. P. Staley (2013). "Splicing fidelity: DEAD/H-box ATPases as molecular clocks." RNA Biol **10**(7): 1073-1079.

Kuhn, A. N. and N. F. Käufer (2003). "Pre-mRNA splicing in Schizosaccharomyces pombe: regulatory role of a kinase conserved from fission yeast to mammals." Curr Genet **42**(5): 241-251.

Käufer, N. F. and J. Potashkin (2000). "Analysis of the splicing machinery in fission yeast: a comparison with budding yeast and mammals." Nucleic Acids Res **28**(16): 3003-3010.

Kück, U., I. Godehardt and U. Schmidt (1990). "A self-splicing group II intron in the mitochondrial large subunit rRNA (LSUrRNA) gene of the eukaryotic alga Scenedesmus obliquus." Nucleic Acids Res **18**(9): 2691-2697.

Laggerbauer, B., T. Achsel and R. Lührmann (1998). "The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplices in vitro." Proc Natl Acad Sci U S A **95**(8): 4188-4192.

Lambowitz, A. M. and S. Zimmerly (2011). "Group II introns: mobile ribozymes that invade DNA." Cold Spring Harb Perspect Biol **3**(8): a003616.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng,

A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blöcker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos,

M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y. J. Chen, J. Szustakowki and I. H. G. S. Consortium (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.

Lardelli, R. M., J. X. Thompson, J. R. Yates and S. W. Stevens (2010). "Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing." RNA **16**(3): 516-528.

Lau, C. K., J. L. Bachorik and G. Dreyfuss (2009). "Gemin5-snRNA interaction reveals an RNA binding function for WD repeat domains." Nat Struct Mol Biol **16**(5): 486-491.

Lei, X. H., X. Shen, X. Q. Xu and H. S. Bernstein (2000). "Human Cdc5, a regulator of mitotic entry, can act as a site-specific DNA binding protein." J Cell Sci **113 Pt 24**: 4523-4531.

Lesser, C. F. and C. Guthrie (1993). "Mutations in U6 snRNA that alter splice site specificity: implications for the active site." Science **262**(5142): 1982-1988.

Lewis, K. E., P. Keifer and T. Saarinen (1992). "Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity." J. Am. Chem. Soc. **114**: 10663-10665.

Li, X., W. Zhang, T. Xu, J. Ramsey, L. Zhang, R. Hill, K. C. Hansen, J. R. Hesselberth and R. Zhao (2013). "Comprehensive in vivo RNA-binding site analyses reveal a role of Prp8 in spliceosomal assembly." Nucleic Acids Res **41**(6): 3805-3818.

Lin, Z., K. Yin, D. Zhu, Z. Chen, H. Gu and L. J. Qu (2007). "AtCDC5 regulates the G2 to M transition of the cell cycle and is critical for the function of Arabidopsis shoot apical meristem." Cell Res **17**(9): 815-828.

Liu, H. L. and S. C. Cheng (2012). "The interaction of Prp2 with a defined region of the intron is required for the first splicing reaction." Mol Cell Biol **32**(24): 5056-5066.

Liu, L., R. Gräub, M. Hlaing, C. L. Epting, C. W. Turck, X.-Q. Xu, L. Zhang and H. S. Bernstein (2003). "Distinct domains of human CDC5 direct its nuclear import and association with the spliceosome." Cell Biochem Biophys **39**(2): 119-132.

Liu, S., R. Rauhut, H. P. Vornlocher and R. Lührmann (2006). "The network of protein-protein interactions within the human U4/U6.U5 tri-snRNP." RNA **12**(7): 1418-1430.

Livesay, S. B., S. E. Collier, D. A. Bitton, J. Bähler and M. D. Ohi (2013). "Structural and functional characterization of the N terminus of Schizosaccharomyces pombe Cwf10." Eukaryotic Cell **12**(11): 1472-1489.

Livesay, S. B., S. E. Collier, D. A. Bitton, J. Bähler and M. D. Ohi (2013). "Structural and functional characterization of the N terminus of Schizosaccharomyces pombe Cwf10." Eukaryot Cell **12**(11): 1472-1489.

Lo, C. W., D. Kaida, S. Nishimura, A. Matsuyama, Y. Yashiroda, H. Taoka, K. Ishigami, H. Watanabe, H. Nakajima, T. Tani, S. Horinouchi and M. Yoshida (2007). "Inhibition of splicing and nuclear retention of pre-mRNA by spliceostatin A in fission yeast." Biochem Biophys Res Commun **364**(3): 573-577.

Lu, P. L., G. F. Lu, C. Y. Yan, L. Wang, W. Q. Li and P. Yin (2011). Structure of the mRNA splicing complex component Cwc2: insights into RNA-recognition. Biochem J.

Lu, X. and R. J. Legerski (2007). The Prp19/Pso4 core complex undergoes ubiquitylation and structural alterations in response to DNA damage. Biochem Biophys Res Commun. **354:** 968-974.

Lunde, B. M., C. Moore and G. Varani (2007). "RNA-binding proteins: modular design for efficient function." Nat Rev Mol Cell Biol **8**(6): 479-490.

Lundgren, K., S. Allan, S. Urushiyama, T. Tani, Y. Ohshima, D. Frendewey and D. Beach (1996). "A connection between pre-mRNA splicing and the cell cycle in fission yeast: cdc28+ is allelic with prp8+ and encodes an RNA-dependent ATPase/helicase." Mol Biol Cell **7**(7): 1083-1094.

Luukkonen, B. G. and B. Séraphin (1998). "Genetic interaction between U6 snRNA and the first intron nucleotide in Saccharomyces cerevisiae." RNA **4**(2): 167-180.

López-Bigas, N., B. Audit, C. Ouzounis, G. Parra and R. Guigó (2005). "Are splicing mutations the most frequent cause of hereditary disease?" FEBS Lett **579**(9): 1900-1903.

MacCoss, M. J., W. H. McDonald, A. Saraf, R. Sadygov, J. M. Clark, J. J. Tasto, K. L. Gould, D. Wolters, M. Washburn, A. Weiss, J. I. Clark and J. R. Yates (2002). "Shotgun identification of protein modifications from protein complexes and lens tissue." Proc Natl Acad Sci U S A **99**(12): 7900-7905.

Madhani, H. D. and C. Guthrie (1992). "A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome." Cell **71**(5): 803-817.

Maeder, C., A. K. Kutach and C. Guthrie (2009). "ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C terminus of Prp8." Nat Struct Mol Biol **16**(1): 42-48.

Makarov, E. M., O. V. Makarova, H. Urlaub, M. Gentzel, C. L. Will, M. Wilm and R. Lührmann (2002). "Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome." Science **298**(5601): 2205-2208.

Martin-Tumasz, S., N. J. Reiter, D. A. Brow and S. E. Butcher (2010). "Structure and functional implications of a complex containing a segment of U6 RNA bound by a domain of Prp24." RNA **16**(4): 792-804.

Maréchal, A., J. M. Li, X. Y. Ji, C. S. Wu, S. A. Yazinski, H. D. Nguyen, S. Liu, A. E. Jiménez, J. Jin and L. Zou (2014). "PRP19 transforms into a sensor of RPA-ssDNA after DNA damage and drives ATR activation via a ubiquitin-mediated circuitry." Mol Cell **53**(2): 235-246.

Mayas, R. M., H. Maita and J. P. Staley (2006). Exon ligation is proofread by the DExD/H-box ATPase Prp22p. Nat Struct Mol Biol. **13:** 482-490.

McDonald, W. H., R. Ohi, N. Smelkova, D. Frendewey and K. L. Gould (1999). Myb-related fission yeast cdc5p is a component of a 40S snRNP-containing complex and is essential for pre-mRNA splicing. Mol Cell Biol. **19:** 5352-5362.

McGrail, J., A. Krause and R. O'Keefe (2009). The RNA binding protein Cwc2 interacts directly with the U6 snRNA to link the nineteen complex to the spliceosome during pre-mRNA splicing. Nucleic Acids Res.

McGrail, J. C., A. Krause and R. T. O'Keefe (2009). "The RNA binding protein Cwc2 interacts directly with the U6 snRNA to link the nineteen complex to the spliceosome during pre-mRNA splicing." Nucleic Acids Res **37**(13): 4205-4217.

Montemayor, E. J., E. C. Curran, H. H. Liao, K. L. Andrews, C. N. Treba, S. E. Butcher and D. A. Brow (2014). "Core structure of the U6 small nuclear ribonucleoprotein at 1.7-A resolution." Nat Struct Mol Biol **21**(6): 544-551.

Mozaffari-Jovin, S., K. F. Santos, H. H. Hsiao, C. L. Will, H. Urlaub, M. C. Wahl and R. Lührmann (2012). "The Prp8 RNase H-like domain inhibits Brr2-mediated U4/U6 snRNA unwinding by blocking Brr2 loading onto the U4 snRNA." Genes Dev **26**(21): 2422-2434.

Nielsen, K. H. and J. P. Staley (2012). "Spliceosome activation: U4 is the path, stem I is the goal, and Prp8 is the keeper. Let's cheer for the ATPase Brr2!" Genes Dev **26**(22): 2461-2467.

Nordgård, O., T. Andersen and O. S. Gabrielsen (2004). "Selective inhibition of c-Myb DNA-binding by RNA polymers." BMC Biochem **5**: 15.

Nurse, P. (1975). "Genetic control of cell size at cell division in yeast." Nature **256**(5518): 547-551.

Nurse, P., P. Thuriaux and K. Nasmyth (1976). "Genetic control of the cell division cycle in the fission yeast Schizosaccharomyces pombe." Mol Gen Genet **146**(2): 167-178.

Ogata, K., H. Hojo, S. Aimoto, T. Nakai, H. Nakamura, A. Sarai, S. Ishii and Y. Nishimura (1992). "Solution structure of a DNA-binding unit of Myb: a helix-turn-helix-related motif with conserved tryptophans forming a hydrophobic core." Proc Natl Acad Sci U S A **89**(14): 6428-6432.

Ogata, K., S. Morikawa, H. Nakamura, H. Hojo, S. Yoshimura, R. Zhang, S. Aimoto, Y. Ametani, Z. Hirata and A. Sarai (1995). "Comparison of the free and DNA-complexed forms of the DNA-binding domain from c-Myb." Nat Struct Biol **2**(4): 309-320.

Ogata, K., S. Morikawa, H. Nakamura, A. Sekikawa, T. Inoue, H. Kanai, A. Sarai, S. Ishii and Y. Nishimura (1994). "Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices." Cell **79**(4): 639-648.

Ohi, M., Y. Li, Y. Cheng and T. Walz (2004). "Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy." Biol Proced Online **6**: 23-34.

Ohi, M. D. and K. L. Gould (2002). "Characterization of interactions among the Cef1p-Prp19p-associated splicing complex." RNA **8**(6): 798-815.

Ohi, M. D., A. J. Link, L. Ren, J. L. Jennings, W. H. McDonald and K. L. Gould (2002). "Proteomics analysis reveals stable multiprotein complexes in both fission and budding yeasts containing Myb-related Cdc5p/Cef1p, novel pre-mRNA splicing factors, and snRNAs." Mol Cell Biol **22**(7): 2011-2024.

Ohi, M. D., L. Ren, J. S. Wall, K. L. Gould and T. Walz (2007). Structural characterization of the fission yeast U5.U2/U6 spliceosome complex. Proc Natl Acad Sci USA. **104:** 3195-3200.

Ohi, M. D., C. W. Vander Kooi, J. A. Rosenberg, W. J. Chazin and K. L. Gould (2003). Structural insights into the U-box, a domain associated with multi-ubiquitination. Nat Struct Biol. **10:** 250-255.

Ohi, M. D., C. W. Vander Kooi, J. A. Rosenberg, L. Ren, J. P. Hirsch, W. J. Chazin, T. Walz and K. L. Gould (2005). Structural and functional analysis of essential pre-mRNA splicing factor Prp19p. Mol Cell Biol. **25:** 451-460.

Ohi, R., A. Feoktistova, S. McCann, V. Valentine, A. T. Look, J. S. Lipsick and K. L. Gould (1998). "Myb-related Schizosaccharomyces pombe cdc5p is structurally and functionally conserved in eukaryotes." Mol Cell Biol **18**(7): 4097-4108.

Ohi, R., D. McCollum, B. Hirani, G. J. Den Haese, X. Zhang, J. D. Burke, K. Turner and K. L. Gould (1994). "The Schizosaccharomyces pombe cdc5+ gene encodes an essential protein with homology to c-Myb." EMBO J **13**(2): 471-483.

Ohlendorf, D. H., W. F. Anderson and B. W. Matthews (1983). "Many gene-regulatory proteins appear to have a similar alpha-helical fold that binds DNA and evolved from a common precursor." J Mol Evol **19**(2): 109-114.

Oka, Y., H. Varmark, K. Vitting-Seerup, P. Beli, J. Waage, A. Hakobyan, M. Mistrik, C. Choudhary, M. Rohde, S. Bekker-Jensen and N. Mailand (2014). "UBL5 is essential for pre-mRNA splicing and sister chromatid cohesion in human cells." EMBO Rep **15**(9): 956-964.

Palmer, A., J. Cavanagh, R. Byrd and M. Rance (1992). "Sensitivity improvement in 3-dimensional heteronuclear correlation NMR spectroscopy." J Magn Reson **96**(2): 416-424.

Perriman, R. and M. Ares (2010). "Invariant U2 snRNA nucleotides form a stem loop to recognize the intron early in splicing." Mol Cell **38**(3): 416-427.

Perriman, R., I. Barta, G. K. Voeltz, J. Abelson and M. Ares (2003). "ATP requirement for Prp5p function is determined by Cus2p and the structure of U2 small nuclear RNA." Proc Natl Acad Sci U S A **100**(24): 13857-13862.

Perriman, R. J. and M. Ares (2007). "Rearrangement of competing U2 RNA helices within the spliceosome promotes multiple steps in splicing." Genes Dev **21**(7): 811-820.

Pleiss, J. A., G. B. Whitworth, M. Bergkessel and C. Guthrie (2007). "Transcript specificity in yeast pre-mRNA splicing revealed by mutations in core spliceosomal components." PLoS Biol **5**(4): e90.

Podar, M., V. T. Chu, A. M. Pyle and P. S. Perlman (1998). "Group II intron splicing in vivo by first-step hydrolysis." Nature **391**(6670): 915-918.

Preker, P. J. and W. Keller (1998). "The HAT helix, a repetitive motif implicated in RNA processing." Trends Biochem Sci **23**(1): 15-16.

Prouse, M. B. and M. M. Campbell (2012). "The interaction between MYB proteins and their target DNA binding sites." Biochim Biophys Acta **1819**(1): 67-77.

Query, C. C. and M. M. Konarska (2012). "CEF1/CDC5 alleles modulate transitions between catalytic conformations of the spliceosome." RNA **18**(5): 1001-1013.

Raghunathan, P. L. and C. Guthrie (1998). "RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2." Curr Biol **8**(15): 847-855.

Rasband, W. S. (1997-2004). ImageJ. http://imagej.nih.gov/ij/, U. S. National Institutes of Health, Bethesda, Maryland, USA.

Rasche, N., O. Dybkov, J. Schmitzova, B. Akyildiz, P. Fabrizio and R. Luhrmann (2012). "Cwc2 and its human homologue RBM22 promote an active conformation of the spliceosome catalytic centre." EMBO J **31**(6): 1591-1604.

Rasche, N., O. Dybkov, J. Schmitzová, B. Akyildiz, P. Fabrizio and R. Lührmann (2012). "Cwc2 and its human homologue RBM22 promote an active conformation of the spliceosome catalytic centre." EMBO J **31**(6): 1591-1604.

Reiter, N. J., H. Blad, F. Abildgaard and S. E. Butcher (2004). "Dynamics in the U6 RNA intramolecular stem-loop: a base flipping conformational change." Biochemistry **43**(43): 13739-13747.

Ren, L., J. R. McLean, T. R. Hazbun, S. Fields, C. Vander Kooi, M. D. Ohi and K. L. Gould (2011). Systematic two-hybrid and comparative proteomic analyses reveal novel yeast pre-mRNA splicing factors connected to Prp19. PLoS ONE. **6:** e16719.

Roy, B., L. M. Haupt and L. R. Griffiths (2013). "Review: Alternative Splicing (AS) of Genes As An Approach for Generating Protein Complexity." Curr Genomics **14**(3): 182-194.

Russell, P. and P. Nurse (1986). "Schizosaccharomyces pombe and Saccharomyces cerevisiae: a look at yeasts divided." Cell **45**(6): 781-782.

Russell, R., I. Jarmoskaite and A. M. Lambowitz (2013). "Toward a molecular understanding of RNA remodeling by DEAD-box proteins." RNA Biol **10**(1): 44-55.

Sabatinos, S. A. and S. L. Forsburg (2010). "Molecular genetics of Schizosaccharomyces pombe." Methods Enzymol **470**: 759-795.

Saikumar, P., R. Murali and E. P. Reddy (1990). "Role of tryptophan repeats and flanking amino acids in Myb-DNA interactions." Proc Natl Acad Sci U S A **87**(21): 8452-8456.

Saldanha, R., G. Mohr, M. Belfort and A. M. Lambowitz (1993). "Group I and group II introns." FASEB J **7**(1): 15-24.

Sander, B., M. M. Golas, E. M. Makarov, H. Brahms, B. Kastner, R. Lührmann and H. Stark (2006). Organization of core spliceosomal components U5 snRNA loop I and U4/U6 Di-snRNP within U4/U6.U5 Tri-snRNP as revealed by electron cryomicroscopy. Mol Cell. **24:** 267-278.

Santos, K. F., S. M. Jovin, G. Weber, V. Pena, R. Lührmann and M. C. Wahl (2012). "Structural basis for functional cooperation between tandem helicase cassettes in Brr2-mediated remodeling of the spliceosome." Proc Natl Acad Sci U S A **109**(43): 17418-17423.

Sarkar, N. (1997). "Polyadenylation of mRNA in prokaryotes." Annu Rev Biochem **66**: 173-197.

Sashital, D. G., G. Cornilescu, C. J. McManus, D. A. Brow and S. E. Butcher (2004). "U2-U6 RNA folding reveals a group II intron-like domain and a four-helix junction." Nat Struct Mol Biol **11**(12): 1237-1242.

Sauer, R. T., R. R. Yocum, R. F. Doolittle, M. Lewis and C. O. Pabo (1982). "Homology among DNA-binding proteins suggests use of a conserved super-secondary structure." Nature **298**(5873): 447-451.

Schleucher, J., M. Schwendinger, M. Sattler, P. Schmidt, O. Schedletzky, S. J. Glaser, O. W. Sørensen and C. Griesinger (1994). "A general enhancement scheme in heteronuclear

multidimensional NMR employing pulsed field gradients." J Biomol NMR **4**(2): 301-306.

Schmidt, C., C. Lenz, M. Grote, R. Lührmann and H. Urlaub (2010). Determination of protein stoichiometry within protein complexes using absolute quantification and multiple reaction monitoring. Anal Chem. **82:** 2784-2796.

Schmidt, H., K. Richert, R. A. Drakas and N. F. Käufer (1999). "spp42, identified as a classical suppressor of prp4-73, which encodes a kinase involved in pre-mRNA splicing in fission yeast, is a homologue of the splicing factor Prp8p." Genetics **153**(3): 1183-1191.

Schmitzová, J., N. Rasche, O. Dybkov, K. Kramer, P. Fabrizio, H. Urlaub, R. Lührmann and V. Pena (2012). "Crystal structure of Cwc2 reveals a novel architecture of a multipartite RNA-binding protein." EMBO J **31**(9): 2222-2234.

Schneider, S., E. Campodonico and B. Schwer (2004). "Motifs IV and V in the DEAH box splicing factor Prp22 are important for RNA unwinding, and helicase-defective Prp22 mutants are suppressed by Prp8." J Biol Chem **279**(10): 8617-8626.

Schuck, P. (2000). "Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling." Biophys J **78**(3): 1606-1619.

Schwer, B. (2008). "A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release." Mol Cell **30**(6): 743-754.

Schwer, B. and C. Guthrie (1991). "PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome." Nature **349**(6309): 494-499.

Schwer, B. and C. Guthrie (1992). "A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis." EMBO J **11**(13): 5033-5039.

Schwer, B. and C. Guthrie (1992). "A dominant negative mutation in a spliceosomal ATPase affects ATP hydrolysis but not binding to the spliceosome." Mol Cell Biol **12**(8): 3540-3547.

Seetharaman, M., N. V. Eldho, R. A. Padgett and K. T. Dayie (2006). "Structure of a self-splicing group II intron catalytic effector domain 5: parallels with spliceosomal U6 RNA." RNA **12**(2): 235-247.

Semlow, D. R. and J. P. Staley (2012). "Staying on message: ensuring fidelity in pre-mRNA splicing." Trends Biochem Sci **37**(7): 263-273.

Shatkin, A. J. (1976). "Capping of eucaryotic mRNAs." Cell **9**(4 PT 2): 645-653.

Shen, H., X. Zheng, J. Shen, L. Zhang, R. Zhao and M. R. Green (2008). "Distinct activities of the DExD/H-box splicing factor hUAP56 facilitate stepwise assembly of the spliceosome." Genes Dev **22**(13): 1796-1803.

Shen, Y. and A. Bax (2012). "Identification of helix capping and b-turn motifs from NMR chemical shifts." J Biomol NMR **52**(3): 211-232.

Shen, Y., F. Delaglio, G. Cornilescu and A. Bax (2009). "TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts." J Biomol NMR **44**(4): 213-223.

Shen, Y., O. Lange, F. Delaglio, P. Rossi, J. M. Aramini, G. Liu, A. Eletsky, Y. Wu, K. K. Singarapu, A. Lemak, A. Ignatchenko, C. H. Arrowsmith, T. Szyperski, G. T.

Montelione, D. Baker and A. Bax (2008). "Consistent blind protein structure generation from NMR chemical shift data." Proc Natl Acad Sci U S A **105**(12): 4685-4690.

Shen, Y., R. Vernon, D. Baker and A. Bax (2009). "De novo protein structure generation from incomplete chemical shift assignments." J Biomol NMR **43**(2): 63-78.

Silverman, E., G. Edwalds-Gilbert and R. J. Lin (2003). "DExD/H-box proteins and their partners: helping RNA helicases unwind." Gene **312**: 1-16.

Song, E. J., S. L. Werner, J. Neubauer, F. Stegmeier, J. Aspden, D. Rio, J. W. Harper, S. J. Elledge, M. W. Kirschner and M. Rape (2010). The Prp19 complex and the Usp4Sart3 deubiquitinating enzyme control reversible ubiquitination at the spliceosome. Genes Dev. **24:** 1434-1447.

Sontheimer, E. J. and J. A. Steitz (1993). "The U5 and U6 small nuclear RNAs as active site components of the spliceosome." Science **262**(5142): 1989-1996.

Spingola, M., L. Grate, D. Haussler and M. Ares (1999). "Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae." RNA **5**(2): 221-234.

Staley, J. P. and C. Guthrie (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell. **92:** 315-326.

Staley, J. P. and C. Guthrie (1999). "An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p." Mol Cell **3**(1): 55-64.

Stark, H. (2010). GraFix: stabilization of fragile macromolecular complexes for single particle cryo-EM. Meth Enzymol. **481:** 109-126.

Stevens, S. W., D. E. Ryan, H. Y. Ge, R. E. Moore, M. K. Young, T. D. Lee and J. Abelson (2002). "Composition and functional characterization of the yeast spliceosomal penta-snRNP." Mol Cell **9**(1): 31-44.

Sundaramoorthy, S., M. D. Vázquez-Novelle, S. Lekomtsev, M. Howell and M. Petronczki (2014). "Functional genomics identifies a requirement of pre-mRNA splicing factors for sister chromatid cohesion." EMBO J **33**(22): 2623-2642.

Tahirov, T. H., K. Sato, E. Ichikawa-Iwata, M. Sasaki, T. Inoue-Bungo, M. Shiina, K. Kimura, S. Takata, A. Fujikawa, H. Morii, T. Kumasaka, M. Yamamoto, S. Ishii and K. Ogata (2002). "Mechanism of c-Myb-C/EBP beta cooperation from separated sites on a promoter." Cell **108**(1): 57-70.

Tanikawa, J., T. Yasukawa, M. Enari, K. Ogata, Y. Nishimura, S. Ishii and A. Sarai (1993). "Recognition of specific DNA sequences by the c-myb protooncogene product: role of three repeat units in the DNA-binding domain." Proc Natl Acad Sci U S A **90**(20): 9320-9324.

Tanner, N. K., O. Cordin, J. Banroques, M. Doère and P. Linder (2003). "The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis." Mol Cell **11**(1): 127-138.

Tarn, W. Y., C. H. Hsu, K. T. Huang, H. R. Chen, H. Y. Kao, K. R. Lee and S. C. Cheng (1994). "Functional association of essential splicing factor(s) with PRP19 in a protein complex." EMBO J **13**(10): 2421-2431.

Tarn, W. Y., K. R. Lee and S. C. Cheng (1993). "Yeast precursor mRNA processing protein PRP19 associates with the spliceosome concomitant with or just after dissociation of U4 small nuclear RNA." Proc Natl Acad Sci USA **90**(22): 10821-10825.

Teigelkamp, S., M. McGarvey, M. Plumpton and J. D. Beggs (1994). "The splicing factor PRP2, a putative RNA helicase, interacts directly with pre-mRNA." EMBO J **13**(4): 888-897.

Toor, N., K. S. Keating, S. D. Taylor and A. M. Pyle (2008). "Crystal structure of a self-spliced group II intron." Science **320**(5872): 77-82.

Tsai, R. T., R. H. Fu, F. L. Yeh, C. K. Tseng, Y. C. Lin, Y. H. Huang and S. C. Cheng (2005). "Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2." Genes Dev **19**(24): 2991-3003.

Tsai, W. Y., Y. T. Chow, H. R. Chen, K. T. Huang, R. I. Hong, S. P. Jan, N. Y. Kuo, T. Y. Tsao, C. H. Chen and S. C. Cheng (1999). "Cef1p is a component of the Prp19p-associated complex and essential for pre-mRNA splicing." J Biol Chem **274**(14): 9455-9462.

Tseng, C. K. and S. C. Cheng (2008). "Both catalytic steps of nuclear pre-mRNA splicing are reversible." Science **320**(5884): 1782-1784.

Tseng, C. K., H. L. Liu and S. C. Cheng (2011). "DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps." RNA **17**(1): 145-154.

Tycowski, K. T., M. D. Shu, A. Kukoyi and J. A. Steitz (2009). "A conserved WD40 protein binds the Cajal body localization signal of scaRNP particles." Mol Cell **34**(1): 47-57.

Umen, J. G. and C. Guthrie (1995). "Prp16p, Slu7p, and Prp8p interact with the 3' splice site in two distinct stages during the second catalytic step of pre-mRNA splicing." RNA **1**(6): 584-597.

Urlaub, H., K. Hartmuth, S. Kostka, G. Grelle and R. Lührmann (2000). A general approach for identification of RNA-protein cross-linking sites within native human spliceosomal small nuclear ribonucleoproteins (snRNPs). Analysis of RNA-protein contacts in native U1 and U4/U6.U5 snRNPs. J Biol Chem. **275:** 41458-41468.

Urushiyama, S., T. Tani and Y. Ohshima (1996). "Isolation of novel pre-mRNA splicing mutants of Schizosaccharomyces pombe." Mol Gen Genet **253**(1-2): 118-127.

Valadkhan, S. and J. L. Manley (2001). "Splicing-related catalysis by protein-free snRNAs." Nature **413**(6857): 701-707.

Valadkhan, S. and J. L. Manley (2003). "Characterization of the catalytic activity of U2 and U6 snRNAs." RNA **9**(7): 892-904.

Valadkhan, S., A. Mohammadi, Y. Jaladat and S. Geisler (2009). "Protein-free small nuclear RNAs catalyze a two-step splicing reaction." Proc Natl Acad Sci U S A **106**(29): 11901-11906.

Valdivia, H. H. (2007). "One gene, many proteins: alternative splicing of the ryanodine receptor gene adds novel functions to an already complex channel protein." Circ Res **100**(6): 761-763.

van der Feltz, C., K. Anthony, A. Brilot and D. A. Pomeranz Krummel (2012). "Architecture of the spliceosome." Biochemistry **51**(16): 3321-3333.

van der Lelij, P., R. R. Stocsits, R. Ladurner, G. Petzold, E. Kreidl, B. Koch, J. Schmitz, B. Neumann, J. Ellenberg and J. M. Peters (2014). "SNW1 enables sister chromatid cohesion by mediating the splicing of sororin and APC2 pre-mRNAs." EMBO J **33**(22): 2643-2658.

van Nues, R. W. and J. D. Beggs (2001). "Functional contacts with a range of splicing proteins suggest a central role for Brr2p in the dynamic control of the order of events in spliceosomes of Saccharomyces cerevisiae." Genetics **157**(4): 1451-1467.

Vander Kooi, C. W., M. D. Ohi, J. A. Rosenberg, M. L. Oldham, M. E. Newcomer, K. L. Gould and W. J. Chazin (2006). The Prp19 U-box crystal structure suggests a common dimeric architecture for a class of oligomeric E3 ubiquitin ligases. Biochemistry. **45:** 121-130.

Vander Kooi, C. W., L. Ren, P. Xu, M. D. Ohi, K. L. Gould and W. J. Chazin (2010). "The Prp19 WD40 domain contains a conserved protein interaction region essential for its function." Structure **18**(5): 584-593.

Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R.

Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y. H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigó, M. J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y. H. Chiang, M. Coyne, C. Dahlke, A.

Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh and X. Zhu (2001). "The sequence of the human genome." Science **291**(5507): 1304-1351.

Villa, T. and C. Guthrie (2005). "The Isy1p component of the NineTeen complex interacts with the ATPase Prp16p to regulate the fidelity of pre-mRNA splicing." Genes Dev **19**(16): 1894-1904.

Wahl, M. C., C. L. Will and R. Lührmann (2009). The spliceosome: design principles of a dynamic RNP machine. Cell. **136:** 701-718.

Wan, L. and J. Huang (2014). "The PSO4 protein complex associates with replication protein A (RPA) and modulates the activation of ataxia telangiectasia-mutated and Rad3-related (ATR)." J Biol Chem **289**(10): 6619-6626.

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

Ward, A. and T. Cooper (2009). The pathobiology of splicing. J Pathol.

Warkocki, Z., P. Odenwalder, J. Schmitzova, F. Platzmann, H. Stark, H. Urlaub, R. Ficner, P. Fabrizio and R. Luhrmann (2009). "Reconstitution of both steps of

Saccharomyces cerevisiae splicing with purified spliceosomal components." Nat Struct Mol Biol **16**(12): 1237-1243.

Watrin, E., M. Demidova, T. Watrin, Z. Hu and C. Prigent (2014). "Sororin pre-mRNA splicing is required for proper sister chromatid cohesion in human cells." EMBO Rep **15**(9): 948-955.

Webb, C. J., C. M. Romfo, W. J. van Heeckeren and J. A. Wise (2005). "Exonic splicing enhancers in fission yeast: functional conservation demonstrates an early evolutionary origin." Genes Dev **19**(2): 242-254.

Wilkins, M. R., E. Gasteiger, A. Bairoch, J. C. Sanchez, K. L. Williams, R. D. Appel and D. F. Hochstrasser (1999). "Protein identification and analysis tools in the ExPASy server." Methods Mol Biol **112**: 531-552.

Will, C. L. and R. Luhrmann (2011). "Spliceosome structure and function." Cold Spring Harb Perspect Biol **3**(7).

Will, C. L. and R. Lührmann (2001). "Spliceosomal UsnRNP biogenesis, structure and function." Curr Opin Cell Biol **13**(3): 290-301.

Wittekind, M. and L. Mueller (1993). "HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha- and Beta-Carbon Resonances in Proteins." J. Magn. Reson. B **101**(2): 201-205.

Wolf, E., B. Kastner, J. Deckert, C. Merz, H. Stark and R. Lührmann (2009). Exon, intron and splice site locations in the spliceosomal B complex. EMBO J.

Wood, V., R. Gwilliam, M. A. Rajandream, M. Lyne, R. Lyne, A. Stewart, J. Sgouros, N. Peat, J. Hayles, S. Baker, D. Basham, S. Bowman, K. Brooks, D. Brown, S. Brown, T.

Chillingworth, C. Churcher, M. Collins, R. Connor, A. Cronin, P. Davis, T. Feltwell, A. Fraser, S. Gentles, A. Goble, N. Hamlin, D. Harris, J. Hidalgo, G. Hodgson, S. Holroyd, T. Hornsby, S. Howarth, E. J. Huckle, S. Hunt, K. Jagels, K. James, L. Jones, M. Jones, S. Leather, S. McDonald, J. McLean, P. Mooney, S. Moule, K. Mungall, L. Murphy, D. Niblett, C. Odell, K. Oliver, S. O'Neil, D. Pearson, M. A. Quail, E. Rabbinowitsch, K. Rutherford, S. Rutter, D. Saunders, K. Seeger, S. Sharp, J. Skelton, M. Simmonds, R. Squares, S. Squares, K. Stevens, K. Taylor, R. G. Taylor, A. Tivey, S. Walsh, T. Warren, S. Whitehead, J. Woodward, G. Volckaert, R. Aert, J. Robben, B. Grymonprez, I. Weltjens, E. Vanstreels, M. Rieger, M. Schäfer, S. Müller-Auer, C. Gabel, M. Fuchs, A. Düsterhöft, C. Fritzc, E. Holzer, D. Moestl, H. Hilbert, K. Borzym, I. Langer, A. Beck, H. Lehrach, R. Reinhardt, T. M. Pohl, P. Eger, W. Zimmermann, H. Wedler, R. Wambutt, B. Purnelle, A. Goffeau, E. Cadieu, S. Dréano, S. Gloux, V. Lelaure, S. Mottier, F. Galibert, S. J. Aves, Z. Xiang, C. Hunt, K. Moore, S. M. Hurst, M. Lucas, M. Rochet, C. Gaillardin, V. A. Tallada, A. Garzon, G. Thode, R. R. Daga, L. Cruzado, J. Jimenez, M. Sánchez, F. del Rey, J. Benito, A. Domínguez, J. L. Revuelta, S. Moreno, J. Armstrong, S. L. Forsburg, L. Cerutti, T. Lowe, W. R. McCombie, I. Paulsen, J. Potashkin, G. V. Shpakovski, D. Ussery, B. G. Barrell, P. Nurse and L. Cerrutti (2002). "The genome sequence of Schizosaccharomyces pombe." Nature **415**(6874): 871-880.

Xiong, H. Y., B. Alipanahi, L. J. Lee, H. Bretschneider, D. Merico, R. K. Yuen, Y. Hua, S. Gueroussov, H. S. Najafabadi, T. R. Hughes, Q. Morris, Y. Barash, A. R. Krainer, N. Jojic, S. W. Scherer, B. J. Blencowe and B. J. Frey (2015). "RNA splicing. The human

splicing code reveals new insights into the genetic determinants of disease." Science **347**(6218): 1254806.

Xu, Y. Z., C. M. Newnham, S. Kameoka, T. Huang, M. M. Konarska and C. C. Query (2004). "Prp5 bridges U1 and U2 snRNPs and enables stable U2 snRNP association with intron RNA." EMBO J **23**(2): 376-385.

Xu, Y. Z. and C. C. Query (2007). "Competition between the ATPase Prp5 and branch region-U2 snRNA pairing modulates the fidelity of spliceosome assembly." Mol Cell **28**(5): 838-849.

Yates, J. R., J. K. Eng, A. L. McCormack and D. Schieltz (1995). "Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database." Anal Chem **67**(8): 1426-1436.

Yean, S. L., G. Wuenschell, J. Termini and R. J. Lin (2000). Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. Nature. **408:** 881-884.

Zaug, A. J., M. M. McEvoy and T. R. Cech (1993). "Self-splicing of the group I intron from Anabaena pre-tRNA: requirement for base-pairing of the exons in the anticodon stem." Biochemistry **32**(31): 7946-7953.

Zhang, M. and M. R. Green (2001). "Identification and characterization of yUAP/Sub2p, a yeast homolog of the essential human pre-mRNA splicing factor hUAP56." Genes Dev **15**(1): 30-35.

Zhang, N., R. Kaur, S. Akhter and R. J. Legerski (2009). "Cdc5L interacts with ATR and is required for the S-phase cell-cycle checkpoint." EMBO Rep **10**(9): 1029-1035.

Zhang, N., R. Kaur, X. Lu, X. Shen, L. Li and R. J. Legerski (2005). "The Pso4 mRNA splicing and DNA repair complex interacts with WRN for processing of DNA interstrand cross-links." J Biol Chem **280**(49): 40559-40567.

Zhang, S., M. Xie, G. Ren and B. Yu (2013). "CDC5, a DNA binding protein, positively regulates posttranscriptional processing and/or transcription of primary microRNA transcripts." Proc Natl Acad Sci U S A **110**(43): 17588-17593.

Zhang, Y. and M. J. Leibowitz (2001). "Folding of the group I intron ribozyme from the 26S rRNA gene of Candida albicans." Nucleic Acids Res **29**(12): 2644-2653.

Zhao, C., D. L. Bellur, S. Lu, F. Zhao, M. A. Grassi, S. J. Bowne, L. S. Sullivan, S. P. Daiger, L. J. Chen, C. P. Pang, K. Zhao, J. P. Staley and C. Larsson (2009). "Autosomal-dominant retinitis pigmentosa caused by a mutation in SNRNP200, a gene required for unwinding of U4/U6 snRNAs." Am J Hum Genet **85**(5): 617-627.

Zhou, Z., L. J. Licklider, S. P. Gygi and R. Reed (2002). "Comprehensive proteomic analysis of the human spliceosome." Nature **419**(6903): 182-185.

Zhou, Z. and R. Reed (1998). "Human homologs of yeast prp16 and prp17 reveal conservation of the mechanism for catalytic step II of pre-mRNA splicing." EMBO J **17**(7): 2095-2106.

Zhou, Z., J. Sim, J. Griffith and R. Reed (2002). "Purification and electron microscopic visualization of functional human spliceosomes." Proc Natl Acad Sci U S A **99**(19): 12203-12207.