Spliceosomal GTPase Cwf10: Roles in Splicing *in vivo* and Characterization of the N-terminal Extension (NTE) Domain

By

Stephen Brent Livesay

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

December, 2013

Nashville, Tennessee

Approved:

Susan R. Wente, Ph.D.

Laura A. Lee, M.D., Ph.D.

Ronald B. Emeson, Ph.D.

James G. Patton, Ph.D.

Melanie D. Ohi, Ph.D.

ACKNOWLEDGEMENTS

This project has involved a lot of perseverance from myself and the help of many people, some for only a short time, and others throughout the entire process. Acknowledging my beginnings, I would like to thank those who got me interested in molecular biology and the processes of the cell, namely Dr. Alicia Schaffner at Cedarville University, Dr. Alison John at Oxford University, and Drs. T. J. Murphy and Chad Kitchen at Emory University. I would also like to thank Drs. Florent Elefteriou and Xiangli Yang of Vanderbilt University for developing my scientific critical thinking skills while I was a technician in Florent's lab. I would like to thank the Vanderbilt BRET office and especially Drs. Roger Chalkley and Jim Patton for their tireless work for me and the other first-year Ph.D. students.

I also am saying "thank you" to those who have made my journey more interesting, enjoyable, and meaningful. My friends and counselors have helped me through difficulties, and I want to especially acknowledge Ms. Diane Majors for encouraging me to finish my doctoral training at a critical juncture. The simple smiles of my fellow lab members and those I would see in the hallways every day were extremely welcome and uplifting. My parents were instrumental in helping me find a quiet, restful place to stay while studying in Nashville, and their care is appreciated.

In my thesis lab, I was supported by generous funding from the National Institutes of Health through grant T32 CA009582 to myself and grant DP2OD004483 to my advisor and mentor, Dr. Melanie D. Ohi. I would like to thank Mel for interesting me in the world of single-particle electron microscopy and for allowing me to take a place in

her lab. One of the highlights of my training was attending the Practical Course on Three-dimensional Cryo Electron Microscopy of Single Particles at the University of Vermont, and I thank Mel for supporting me in that adventure. I have learned so much working for and alongside her. I received expert training and advice from numerous other scientists, and would like to especially acknowledge Drs. Janel McLean and Dewight Williams for patient explanation of techniques and Ms. Liping Ren for being supportive of my project and success in many ways.

My thesis advisory committee provided expert guidance for the completion of my first-author manuscript and for my education in research as a whole. Thank you to Drs. Jim Patton, Laurie Lee, and Ron Emeson. My two committee chairs, Drs. Trish Labosky and Susan Wente, went above and beyond the call of duty to help me through some difficult spots in my progress, which I gratefully acknowledge.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Introduction to pre-mRNA splicing

All organisms have genomes that are transcribed by polymerases. In eukaryotes, the resulting pre-messenger RNA (pre-mRNA) may contain intervening sequences (introns) that must be removed so that the remaining sequences (exons) can be joined together and form the mature mRNA. The process of intron removal and exon ligation is termed pre-mRNA splicing. Splicing is a highly regulated step in gene expression, extensively coupled to preceding and following steps in the mRNA's maturation (reviewed in Montes, Becerra et al. 2012 and Pawlicki and Steitz 2010). When we consider a single intron surrounded by two exons (exon 1 and exon 2), the splicing chemistry depends on three sequence features within the intron: the 5' splice site (SS), the branch point, and the 3' SS. In the first reaction step, the 2' hydroxyl of the branch point adenosine attacks the 5' SS phosphate, releasing the sequence 5' of the intron (exon 1), and forming a lariat intermediate. In the second step, the 3' hydroxyl of exon 1 attacks the phosphate at the 3' SS, joining exon 1 with the sequence downstream of the intron, termed exon 2. The ligated exons are included in the mRNA to be translated by the ribosome, while the intron, in lariat form, is discarded.

Because splicing alters the mRNA sequence, fidelity of this process is of great importance to proper gene expression. In an average human gene, a sequence of 28,000 nucleotides contains nine relatively short exons (120 nucleotides each) alternatively

interspersed with eight introns of between 100 and 100,000 nucleotides each (Lander, Linton et al. 2001). The choice of splice sites can be changed by mutation of a single nucleotide, and many genetic diseases are caused by such single nucleotide changes (reviewed in Singh and Cooper 2012). Transcripts in higher eukaryotes often undergo differential patterns of splicing depending on cell type, stage of development, and other signals, a process termed alternative splicing (Pan, Shai et al. 2008, Wang, Sandberg et al. 2008). In this process, the normal, or constitutive, pattern of intron/exon boundaries is modified in some fashion, leading to such transcript changes as exon exclusion, exon length modifications, and intron inclusion (Maniatis and Tasic 2002). Alternative splicing thus functions to increase the number of potential proteins that can be coded by a single gene.

Spliceosome machinery

The regulation of splice site choice and facilitation of splicing chemistry are performed by a ubiquitous eukaryotic machine termed the spliceosome. Our current knowledge of splicing factors and their interactions within the spliceosome comes from many decades of research. The first genetic means of defining splicing factors was a screen in 1967 by L. Hartwell, who isolated *Saccharomyces cerevisiae* temperature sensitive (ts) mutants with RNA synthesis defects (Hartwell 1967). M. Rosbash and colleagues later demonstrated that some of Hartwell's RNA synthesis mutants had defects in splicing (reviewed in Ruby and Abelson 1991), and termed the complementation groups *PRP* (precursor RNA processing). Using *in vitro* splicing assays, developed in the 1980s (Cheng, Newman et al. 1990 and references therein), groups

confirmed many *PRP* factors have direct involvement in the splicing reaction. As additional splicing factors continued to be discovered genetically using *S. cerevisiae*, screens for splicing mutants in the fission yeast *Schizosaccharomyces pombe* led to the definition of *PRP* groups in that organism (Potashkin, Li et al. 1989, Urushiyama, Tani et al. 1996). Early biochemical work using HeLa cells led to the hypothesis that structural RNA molecules termed small nuclear RNAs (snRNAs) were required for splicing (Yang, Lerner et al. 1981). These uridine-rich snRNAs include U1, U2, U4, U5, and U6. The same types of experiments were used to isolate each snRNA and its associating protein components, together termed the small nuclear ribonucleoproteins (snRNPs) (for example, Lerner and Steitz 1979, Bringmann, Appel et al. 1984, Bach, Winkelmann et al. 1989, Behrens, Tyc et al. 1993). The U1, U2, U4, and U5 snRNA are packaged with a common set of stabilizing proteins, termed the Sm ring, while the U6 snRNA is stabilized by a similar structure, termed the Lsm ring. The snRNA, Sm or Lsm ring, and other proteins specific to each snRNA together organize into a snRNP.

HeLa cell *in vitro* splicing assays revealed that each snRNP is only required for a portion of the splicing reaction (Bindereif and Green 1987). The *in vitro* splicing assays from both budding yeast and HeLa cells led to a model of spliceosome assembly and catalysis defined by ordered binding and dissociation of snRNPs. In this model, spliceosome assembly begins with the recognition of the 5' SS, branch point, and polypyrimidine tract downstream of the branch point by the U1 snRNP, the SF1 protein, and the U2 auxilary factor (U2AF), respectively. Afterwards, in an ATP-dependent step, the U2 snRNA basepairs with the branch point, forming Complex A (Figure 1-1). The subsequent engagement of the U5.U4/U6 tri-snRNP at Complex B triggers the unwinding

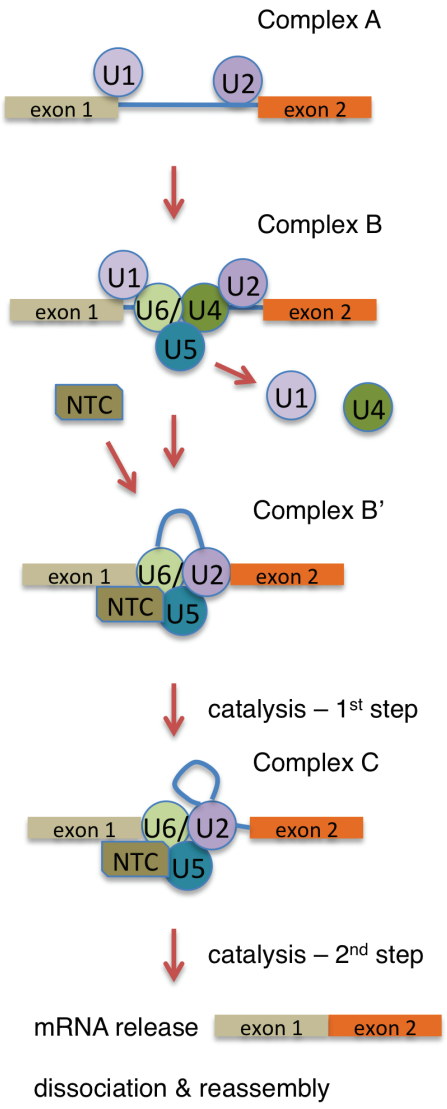


FIGURE 1-1

Schematic model of a single round of pre-mRNA splicing
U1, U2, U4, U5, and U6 represent snRNPs. NTC stands for the Prp **n**ine**t**een **c**omplex.

of the U4/U6 snRNA duplex that is then replaced with the U2/U6 snRNA duplex. Furthermore the U1 snRNA base pairing at the 5' splice site is disrupted and exchanged for base pairing between the 5' splice site and the U6 snRNA. The subsequent release of the U1 and U4 snRNPs marks the transition from the inactive to active spliceosome, termed Complex B'. The activated Complex B' catalyzes the first transesterification reaction, and Complex C catalyzes the second transesterification reaction. The post-catalytic complex then releases the mRNA and disassembles.

In addition to the snRNP specific components, the spliceosome contains hundreds of additional proteins that associate with the complex at different points in the reaction. While many of these pre-mRNA splicing factors join and leave the spliceosome singly, there is one group of proteins that appears to join as a complex, termed the NTC for its founding member (Prp **n**ine**t**een **c**omplex) (Tarn, Hsu et al. 1994, Tsai, Chow et al. 1999, Agafonov, Deckert et al. 2011). The NTC binds following Complex B formation and is required for spliceosome activation (Chan, Kao et al. 2003, Makarova, Makarov et al. 2004). The advent of mass spectrometry for identification of proteins in purifications was used with *in vitro* spliceosomes stalled at various stages of assembly to determine the full complement of proteins present at each stage and revealed that the protein composition of the spliceosome changes dramatically during a splice cycle (Hartmuth, Urlaub et al. 2002, Jurica, Licklider et al. 2002, Makarov, Makarova et al. 2002, Makarova, Makarov et al. 2004, Deckert, Hartmuth et al. 2006, Behzadnia, Golas et al. 2007, Bessonov, Anokhina et al. 2008, Fabrizio, Dannenberg et al. 2009). Particularly exciting was the application of a novel semiquantiative proteomic analysis method to show that many spliceosome

proteins associate sub-stoichiometrically (Agafonov, Deckert et al. 2011), indicating that such proteins are not required to splice every substrate.

The composition of the active site and the identity of the molecules that accomplish actual catalysis at Complexes B' and C are still unclear. RNPs, which catalyze many cellular reactions besides splicing, can contain an RNA and/or protein-rich active site. In true ribozymes, the RNA catalyzes the reaction, and RNP proteins play supporting roles, while protein-catalyzed RNPs use only proteins for catalysis. Genetic and biochemical analyses have clearly demonstrated that both proteins and snRNAs are required for the splicing reaction in the context of the spliceosome; however, the spliceosome's status as a ribozyme is unclear. Our current understanding of RNA/RNA pairing in the active site is illustrated in Figure 1-2. The U2 and U6 snRNAs are extensively base-paired to each other, and each strand contacts the substrate – U2 contacts the branch point and U6 contacts the 5' SS. The U5 snRNA is also present and is thought to align the two exons for second step catalysis (Sontheimer and Steitz 1993). *In vitro*, a system containing only portions of U6, U2, and the pre-mRNA was able to catalyze a reaction similar to the first step of splicing (Valadkhan and Manley 2001). There are a handful of proteins present near the active site, including U5 snRNP proteins Prp8, Brr2, and Snu114. Regarding proteins that may catalyze splicing, Prp8, a component of the U5 snRNP, can be crosslinked to the 5' SS, branch point, and 3' SS prior to catalysis (reviewed in Grainger and Beggs 2005). A recent crystal structure of a large portion of Prp8 revealed that it forms a cavity large enough to contain the self-splicing RNA structures of a group II intron, which undergoes the same two-step chemistry as spliceosome-processed pre-mRNA (Galej, Oubridge et al. 2013). High-

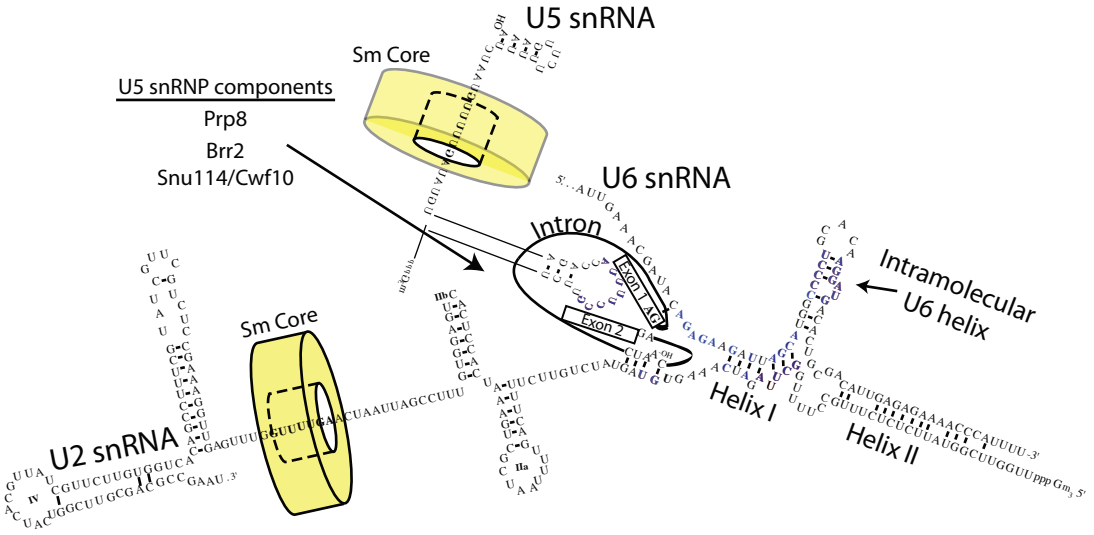


FIGURE 1-2

Model of two-dimensional RNA/RNA interactions in the active site

The sequences of *S. pombe* U2, U5, and U6 are shown. U5 loop 1 is shown contacting the pre-mRNA exons, while the U2 and U6 sequences, extensively base-paired, contact the branch point and 5' splice site, respectively. Nucleic acids in blue are known to be involved in the active site.

resolution structural data of the active site would be required, however, to fully visualize the proteins, RNA, or both that participate in catalysis.

Structural characterization of splicing complexes

Structural characterization on the spliceosome has lagged behind genetic and biochemical analyses due to certain peculiarities of the complex. Superficially, the spliceosome appears to be similar in complexity to another large RNP, the ribosome, which has been successfully structurally modeled through both electron microscopy (EM) and x-ray crystallography (Taylor, Nilsson et al. 2007, Ben-Shem, Jenner et al. 2010). However, the mechanisms of function are quite different between the two complexes. Once the ribosome is in elongation mode, it has pre-formed active sites that can function outside of the context of the entire complex, while the spliceosome and its active sites form *de novo* on the substrate for each round of splicing. The compositional heterogeneity of the spliceosome at each splicing intermediate and the sub-stoichiometric association of many spliceosome proteins are additional challenges unique to the spliceosome. Interactions observed in, for example, the pre-engagement U5 snRNP are unlikely to be entirely preserved in the catalytically active spliceosome due to the many actions of enzymes during spliceosome assembly (see section Facilitation of Spliceosome Dynamics below). Many spliceosome proteins are predicted to be unstructured out of the context of their binding partners (Korneta, Magnus et al. 2012, Coelho Ribeiro, Espinosa et al. 2013), which makes characterization of these smaller units by crystallization difficult. Finally, spliceosomes are large complexes with dozens of unique proteins, making reconstitution from purified individual components prohibitive.

Current structural data on the spliceosome has been extensively reviewed elsewhere (Zhang, Li et al. 2013) and is composed mostly of high-resolution crystallography and NMR structures of the domains of individual pre-RNA splicing factors and of low-resolution single-particle EM structures of entire splicing complexes. Some of the most informative high-resolution data so far include two crystal structures of the human U1 snRNP at 5.5 Å (Pomeranz Krummel, Oubridge et al. 2009) and 4.4 Å (Weber, Trowitzsch et al. 2010) resolution. These structures are complex, including the snRNA, the entire Sm ring, and the three U1-specific proteins. These studies enhanced our understanding of the Sm ring/snRNA interaction and revealed U1-70K's capacity to link the snRNA with another U1 specific protein through discrete domains. Additionally, a model of the U4 snRNP core at 3.6 Å resolution showed direct one-to-one interactions between the seven nucleotides of the U4 snRNA "Sm site heptad" and the seven Sm proteins taking place in the central canal of the Sm ring (Leung, Nagai et al. 2011).

Regarding spliceosome proteins, several domains of U5-snRNP protein Prp8 have been crystallized (Pena, Liu et al. 2007, Zhang, Shen et al. 2007, Pena, Rozov et al. 2008, Ritchie, Schellenberg et al. 2008, Yang, Zhang et al. 2008). The largest structure, with ~1000 residues of Prp8 in complex with assembly factor Aar2 at 1.9 Å, revealed the presence of Prp8 domains similar to bacterial and fungal Group II intron-encoded proteins (IEPs) (Galej, Oubridge et al. 2013). Because the splicing reaction is similar to a Group II intron self-splicing event, and because the IEPs act upon the Group II intron following splicing, the structural model revealed a remarkable adaptation of these ancient proteins for use in the spliceosome catalytic core. A protein domain successfully crystallized and analyzed by NMR is the RNA-recognition motif (RRM), present in

multiple spliceosomal proteins (reviewed in Zhang, Li et al. 2013), and able to bind either canonically to RNA or, by adaptations, to other proteins. And finally, multiple RNA helicases, which drive conformational changes and proofread within the spliceosome, have been crystallized, including Prp43 (He, Andersen et al. 2010, Walbott, Mouffok et al. 2010), which has high sequence similarity to other spliceosomal helicases including Prp2, Prp16, and Prp22. Also, a large portion of Brr2, the only helicase that is an integral member of a snRNP (U5), was crystallized at 2.7 Å (Santos, Jovin et al. 2012). Brr2 is a 250-kDa protein that contains two tandem helicase cassettes, and the structure revealed the interaction of the catalytically active N-terminal cassette with the C-terminal cassette, which cannot hydrolyze ATP but contacts binding partners and regulates N-terminal cassette activity.

Because of the challenges of crystallizing large complexes, EM has been the preferred method to interrogate the structural organization of entire spliceosomes. However, the low-resolution structures obtained thus far have been useful mainly in defining the general shapes of the complexes. Currently there are three-dimensional structures for the yeast U5.U4/U6 tri-snRNP (Hacker, Sander et al. 2008), yeast U6 snRNP (Karaduman, Dube et al. 2008), the human U5.U4/U6 tri-snRNP (21 Å resolution), and its components, U5 snRNP (26-32 Å resolution) and U4/U6 snRNPs (40 Å resolution) (Sander, Golas et al. 2006). The two tri-snRNP papers, with included labeling studies, satisfyingly demonstrated the organization of the tri-snRNP, with U5-snRNP proteins Prp8 and Snu114 present in the main body, and the U4/U6 di-snRNP and Brr2 forming two projections both upwards from the main body. Concerning spliceosome intermediates, there are 3D structural models of human complex A at 40-50 Å resolution

(Behzadnia, Golas et al. 2007), human BΔU1 at 40 Å resolution (Boehringer, Makarov et al. 2004), human C at 20-29 Å resolution (Golas, Sander et al. 2010), and a much smaller human C structure at 30 Å resolution (Jurica, Sousa et al. 2004). Disparities between the two C complex structures have not been fully explained; however, the fixation protocol used to prepare spliceosomes in (Golas, Sander et al. 2010) appears to account for at least some of the differences (Ilagan, Chalkley et al. 2013). Figure 1-3 shows the splicing intermediate structures, as well as the human tri-snRNP structure (Sander, Golas et al. 2006) at proper scale relative to each other. These structures highlight the huge conformational and size changes the spliceosome undergoes throughout a single splicing cycle. An additional cryo-EM structure of an endogenous *S. pombe* spliceosome at 29 Å resolution (Ohi, Ren et al. 2007) is strikingly similar in size and shape to the smaller human C complex (Jurica, Sousa et al. 2004), suggesting high conservation of overall spliceosome structure from fission yeast to humans.

In addition to the challenges to structural determination outlined at the beginning of this section, single-particle EM analysis of spliceosomes requires overcoming additional spliceosome-specific issues, including conformational heterogeneity and lack of symmetry. One goal of our lab, and others, is to improve the sample quality of spliceosomes through genetic and biochemical manipulation as well as purification strategies. Improved samples, along with the continual advances in analysis software and processing power, promise to expand the information EM can reveal about splicing complexes.

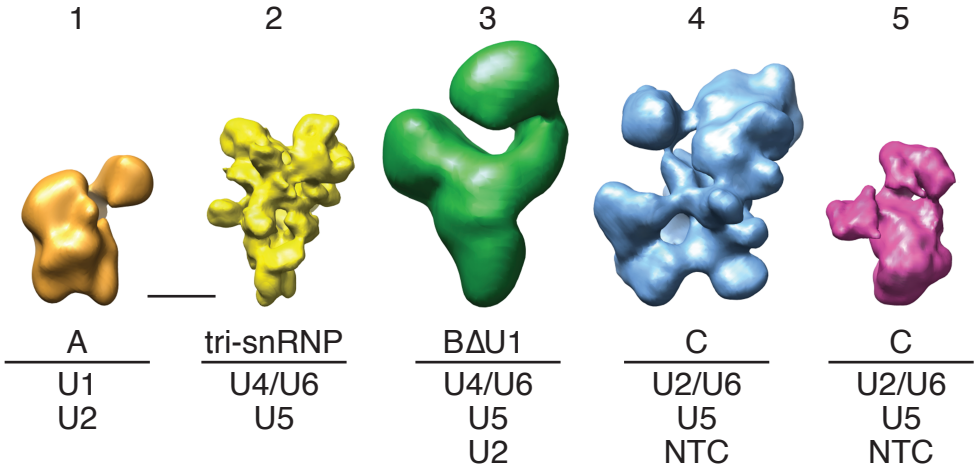


FIGURE 1-3

Comparison of current 3D EM structures of human spliceosome intermediates
The spliceosome stage and major components are listed below each structure. 1) Behzadnia, et al., 2007, 2) Sander, et al., 2006, 3) Boehringer, et al., 2004, 4) Golas, et al., 2010, 5) Jurica, et al., 2004. Scale bar (between structures 1 and 2) indicates 100 Å.

Facilitation of spliceosome dynamics

Because of the massive conformational and compositional changes that happen between intermediates in the splicing cycle, we would like to know how these changes are facilitated. The list of spliceosome-associated proteins contains a number of candidates with enzymatic activity, including evolutionarily conserved DEAD-box (DExD/H) ATPase helicases, a ubiquitin ligase, kinases, phosphatases, and a GTPase (Staley and Guthrie 1998, Will and Lührmann 2011, van der Feltz, Anthony et al. 2012). The GTPase, Snu114/Cwf10, will be discussed at length in a following section. The eight spliceosome-associated helicases are the best-studied transition-related enzymes, and they function at discrete stages of spliceosome assembly, activation, and disassembly (Staley and Guthrie 1998). Helicases generally act by either binding dsRNA and unwinding the strands locally, or by translocating along a single RNA strand until a duplex is encountered and displaced (Jankowsky, Gross et al. 2000, Yang, Del Campo et al. 2007). DExD/H box helicases can also disrupt protein/RNA interactions (Jankowsky, Gross et al. 2001). Thus, it is proposed that the spliceosome helicases unwind RNA duplexes or displace RNA-bound proteins while translocating along the RNA. Spliceosome helicases also function at splicing fidelity checkpoints by delaying splicing progression of sub-optimal substrates such that a discard pathway is used (Koodathingal, Novak et al. 2010, Mayas, Maita et al. 2010). While helicase function is absolutely required for some spliceosome transitions *in vitro*, many of the helicase targets are unknown (Cordin, Hahn et al. 2012).

The spliceosome contains at least one known ubiquitin ligase, Prp19, which ubiquitylates the U4 snRNP protein Prp3, increasing its affinity for U5 protein Prp8 and

encouraging assembly of the U5.U4/U6 tri-snRNP (Song, Werner et al. 2010). Deubiquitylation of Prp3 during spliceosome activation is presumed to encourage ejection of U4 and tri-snRNP-specific proteins from the spliceosome during activation. A separate set of experiments demonstrated similar roles for ubiquitylation in tri-snRNP integrity and spliceosome activation, although the targets were not identified (Bellare, Small et al. 2008).

Kinases and phosphatases are also required for spliceosome transitions, although it is likely that some of these events are not conserved in budding yeast, which lacks a gene encoding the spliceosome-specific kinase Prp4. Kinase activity by human Prp4, however, is required for Complex B formation (Schneider, Hsiao et al. 2010). Furthermore, Prp28 phosphorylation by SRPK2 likely plays a role in Complex B formation (Mathew, Hartmuth et al. 2008). Phosphatases PP1 and PP2A are required for the first-to-second step catalytic transition in human splicing (Shi, Reddy et al. 2006). These results, and the increasing number of phosphorylated human splicing factors being reported (Agafonov, Deckert et al. 2011), suggest phosphorylation may be a master regulator of spliceosome function.

Functions and components of the U5 snRNP

The U5 snRNP is a substrate and/or driver in many of the splice cycle transitions, mostly due to the U5-specific proteins. Before engaging the spliceosome, the U5 snRNP incorporates in to the U5.U4/U6 tri-snRNP. Following splice site selection in Complex A, the tri-snRNP joins the spliceosome in Complex B. The U5 snRNP here undergoes major changes, as some U5-specific proteins are ejected from the spliceosome, the NTC joins,

and the mature active site, likely in the vicinity of U5 snRNP protein Prp8, is formed (Galej, Oubridge et al. 2013, Schellenberg, Wu et al. 2013). The U5 snRNP core and at least proteins Prp8, Brr2, Snu114, and U5-40 remain with the spliceosome through both steps of catalysis and mRNA release (Fabrizio, Dannenberg et al. 2009, Agafonov, Deckert et al. 2011, Fourmann, Schmitzova et al. 2013). The U5 snRNA, specifically loop 1 (L1), is required for tethering of the two exons prior to their ligation in catalytic step two (O'Keefe and Newman 1998). Another element of the U5 snRNA, internal loop 1 (IL1), acts as an assembly platform for U5 snRNP proteins Brr2, Prp8, and Snu114 (Nancollis, Ruckshanthi et al. 2013). In addition to potentially housing the active site, Prp8 can be mutated to affect both steps of catalysis, and the protein physically contacts the 5' and 3' splice sites, the intron's branch point sequence, and the U5 and U6 snRNAs (Dix, Russell et al. 1998, Grainger and Beggs 2005, Turner, Norman et al. 2006). Brr2, an RNA helicase, is required for spliceosome remodeling events, specifically the disruption of U4/U6 interactions (Maeder, Kutach et al. 2009) and the release of U6 from U2 in spliceosome disassembly (Small, Leggett et al. 2006). Snu114 contains a GTPase domain and, *in vitro*, is required for Brr2 function, modulating Brr2's ATPase activity (Small, Leggett et al. 2006, Maeder, Kutach et al. 2009). These three proteins, Prp8, Brr2, and Snu114, interact amongst each other both physically and genetically (Brenner and Guthrie 2005, Boon, Norman et al. 2006, Liu, Rauhut et al. 2006, Zhang, Xu et al. 2009, Mozaffari-Jovin, Wandersleben et al. 2013), suggesting that these proteins coordinate to control certain spliceosome transitions.

Structure and function of the spliceosomal GTPase

The GTPase Snu114 appears to be a ribosomal protein re-engineered for use in the spliceosome. *S. cerevisiae* Snu114 shares 27% sequence identity with *S. cerevisiae* Eft1/EF2, a ribosomal translocase (Fabrizio, Laggerbauer et al. 1997). EF2 functions in eukaryotic translation similarly to its better-studied bacterial homolog, EF-G. During the elongation phase of translation, EF-G favors tRNA translocation by driving a conformational change in the ribosome (reviewed in Rodnina and Wintermeyer 2011), and GTP hydrolysis is required for optimal translocation efficiency (Rodnina, Savelsbergh et al. 1997). While EF2/EF-G's interactions and binding partners within the ribosome are known, and include both RNAs and other proteins (Spahn, Gomez-Lorenzo et al. 2004, Taylor, Nilsson et al. 2007, Sengupta, Nilsson et al. 2008), the exact sequence of events – how GTP hydrolysis leads to the ribosome's conformational changes – is unknown. Inter-domain movement, however, appears to be important, as crosslinking of domains I and V inactivates the protein's translocation function (Peske, Matassova et al. 2000), and it is known that GTP hydrolysis causes movement of domain IV relative to domains I and II (Stark, Rodnina et al. 2000, Jørgensen, Ortiz et al. 2003).

We made use of crystal structures of EF2 (Jørgensen, Ortiz et al. 2003) to model Snu114's homolog in *Schizosaccharomyces pombe*, Cwf10 (Figure 1-4A). The domain map of Cwf10 (Figure 1-4B) is colored to match the 3D representation in Figure 1-4A. In addition to the EF2-like portion, Cwf10/Snu114 contains an N-terminal extension (NTE) that is not present in EF2. The NTE has approximately 120 amino acids and, at least based on primary sequence, is not homologous to any known domain or motif.

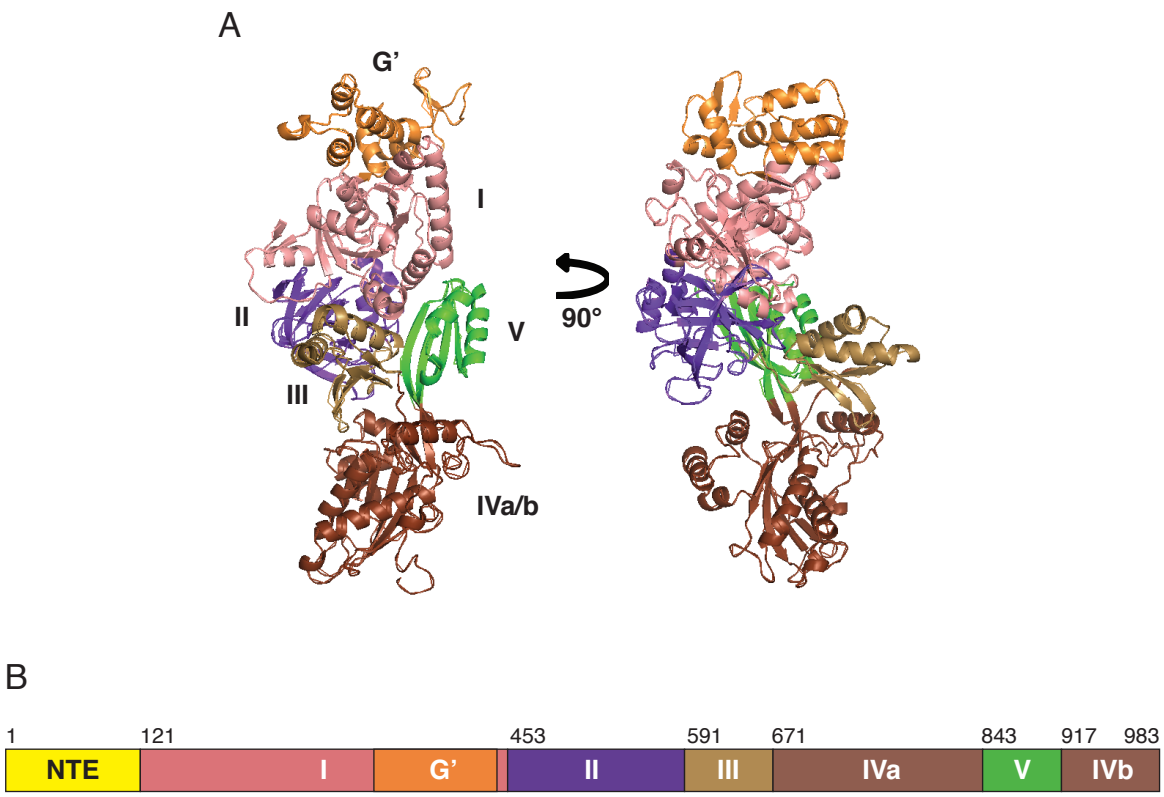


FIGURE 1-4

Structural model of spliceosomal GTPase Cwf10

A) Cwf10 three-dimensional model, as threaded on to the determined structure of EF2 (PDB 1N0V). B) Linear representation of the domains of Cwf10, as defined by homology to EF2. Domains are colored as in A. The 120 amino acid N-terminal extension (NTE) is not conserved in EF2 and is not modeled.

What can our knowledge of EF2/EF-G predict about Snu114's mechanisms? Snu114 likely retains the capacity to bind both proteins and RNA. Indeed Snu114 is able to crosslink to the pre-mRNA polypyrimidine tract (Chiara, Palandjian et al. 1997) and U5 snRNA IL1 (Dix, Russell et al. 1998). The GTPase activity of Snu114 is predicted to be required for optimal function; however, evidence suggests that nucleotide binding status, but not hydrolysis, is important for Snu114's role in U4/U6 unwinding and U2/U6 unwinding *in vitro* (Small, Leggett et al. 2006). Finally, inter-domain movement is likely important. Large Snu114 conformational changes could facilitate such transitions as the release of tri-snRNP proteins during activation, and the integration of U5 snRNP in the U5.U4/U6 tri-snRNP. Besides what has been inferred about Snu114/Cwf10 from its homology to EF2, Snu114 is known to function in both spliceosome activation and disassembly (Bartels, Klatt et al. 2002, Brenner and Guthrie 2006, Small, Leggett et al. 2006), and its presence is required for the integrity of the U5 snRNP and tri-snRNP (Bartels, Klatt et al. 2002, Brenner and Guthrie 2006). Additionally, several studies implicate Snu114 in the second step of catalysis (Fabrizio, Laggerbauer et al. 1997, Liu, Laggerbauer et al. 1997, Frazer, Lovell et al. 2009). Thus, Snu114/Cwf10 is important throughout the splice cycle; unfortunately its mechanisms beyond nucleotide binding are elusive. A spliceosome protein that may act as a guanine nucleotide exchange factor (GEF) is unknown. Additionally, the local movements Snu114 may drive with its neighboring proteins and RNAs are completely unknown. Finally, the ways in which GTP binding, relative domain movement, and NTE coupling to the EF2-like domains influence each other and how much they each contribute to the protein's functions are unknown.

Summary

A testament to the process's complexity, spliceosome-catalyzed splicing is still not well understood in several respects after over three decades of research. While the molecular players – the structural RNAs and proteins – are well-catalogued, we have only an intermediate understanding of the post-translational modifications important for splicing and the functions of the individual components. Knowledge is perhaps most lacking in the mechanisms involved in each spliceosome transition, and, relatedly, the spatial organization of the complex at multiple intermediate stages. In this work, I have manipulated *S. pombe* Cwf10, the sole spliceosomal GTPase, to examine its functions and how its mutation influences splicing complexes. In chapter II, I will characterize the NTE domain of Cwf10, and in chapter III, I will characterize ts *cwf10* mutants generated to improve homogeneity of splicing complexes for structural characterization.

CHAPTER II

STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF THE N-TERMINUS OF SPLICEOSOMAL GTPase CWF10

S. Brent Livesay[1], Scott E. Collier[1], Danny A. Bitton[2], Jürg Bähler[2], and Melanie D. Ohi[1]

[1]Department of Cell and Developmental Biology, Vanderbilt University Medical Center, Nashville, TN 37232.

[2]Department of Genetics, Evolution and Environment, University College London, London, WC1E 6BT, UK.

A slightly modified version of this chapter has been accepted for publication in *Eukaryotic Cell*.

Author contributions

Scott Collier performed all biophysical experiments and analyzed the data from those experiments. Danny Bitton performed all analyses of the RNA-seq data. Brent Livesay performed all other experiments.

Introduction

The spliceosome is composed of several structural RNAs and dozens of protein components which undergo an amazing procession of changes during the course of one

splice cycle (reviewed in Will and Lührmann 2011). Snu114/Cwf10 is a member of the U5 snRNP, contains a GTPase domain, and functions during multiple spliceosome transitions (Fabrizio, Laggerbauer et al. 1997, Achsel, Ahrens et al. 1998, Bartels, Klatt et al. 2002, Ohi, Link et al. 2002, Brenner and Guthrie 2006, Small, Leggett et al. 2006). The Snu114 family of proteins shares homology with the eukaryotic translation elongation factor EF2 (Fabrizio, Laggerbauer et al. 1997) but differs significantly from EF2 in that they contain a conserved N-terminal extension (NTE) (Figure 2-1A). The NTE is approximately 120 amino acids long and is rich in acidic residues, with 39% of the first 56 residues being aspartate or glutamate in *Schizosaccharomyces pombe* Cwf10 (Figure 2-1B). The human and *S. pombe* NTEs are 43% identical, a much larger percentage than the human and *Saccharomyces cerevisiae* NTEs (26% identical), suggesting that studies in *S. pombe* may also yield relevant insights into the activity of the human NTE.

Functional data implicate the NTE in spliceosome activation and catalysis; however, very little is known about the structure and binding partners of the NTE. In human *in vitro* splicing assays, addition of antibodies against the U5-116K NTE partially blocks the second step of splicing (Fabrizio, Laggerbauer et al. 1997). Removal of *Sc* Snu114's NTE (*Sc snu114*Δ*N*) causes a temperature sensitive (ts) phenotype *in vivo*, while *in vitro* splicing extracts prepared from *Sc snu114*Δ*N* cells are unable to efficiently unwind the U4/U6 snRNAs prior to the first catalytic step of pre-mRNA splicing (Bartels, Klatt et al. 2002). These extracts also contain destabilized U5 snRNPs (Bartels, Klatt et al. 2002). Finally, the *Sc snu114*Δ*N* allele shows synthetic lethal and synthetic sick interactions with mutations in the U5 snRNA loops 1 and IL1, as well as with U6 snRNA

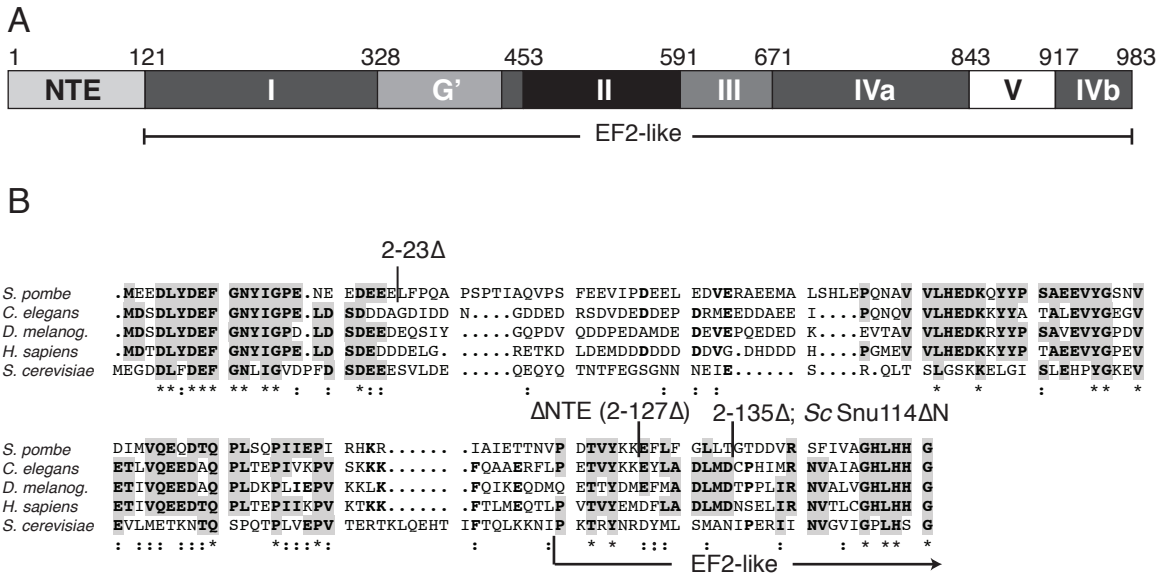


FIGURE 2-1

The N-terminal extension is conserved in Snu114/Cwf10 family members
A) Domain map of *S. pombe* Cwf10, as defined by three-dimensional modeling (MODELLER; Sali and Blundell, 1993) of *S. pombe* Cwf10 on to the crystal structure of *S. cerevisiae* EF2 (PDB 1N0V). Domains are named as defined by homology with EF2 and residue numbers listed above domains refer to *S. pombe* Cwf10. The NTE domain is not present in EF2. B) CLUSTALW (Thompson et al., 1994) sequence alignment of the N-termini of Snu114/Cwf10 homologs. Residues that are identical in at least three of the five homologs are in bold. Clusters of identical residues are highlighted in gray. N-terminal boundaries of the various Cwf10 truncations and the *S. cerevisiae* Snu114ΔN truncation are noted above the alignment. "*" identical, ":" similar. Arrow marks the transition from the NTE to EF2-like domain I, as assigned in A.

alleles that disrupt U2/U6 base pairing (Frazer, Lovell et al. 2009). These results led the authors to speculate that the *Sc* Snu114 NTE may be involved in facilitating U5 and U6 snRNA interactions near the 5' splice site. Interestingly, no physical interactions have been determined for the NTE in any organism, leaving open the question of how the NTE is spatially oriented within the spliceosome. Additionally, the only structural information about the NTE is a bioinformatics prediction that the NTE of human U5-116K is composed of two disordered regions of equal length (Korneta, Magnus et al. 2012).

In this study we characterize the *S. pombe* Cwf10 NTE both *in vivo* and biochemically. We show that although the NTE is not essential in *S. pombe*, deleting this region leads to a general splicing defect at all temperatures. We define a small region of the NTE required for efficient splicing and demonstrate the presence of both structural order and disorder within the NTE. Finally, we show that when the NTE is overexpressed *in vivo* it stably associates with a protein complex similar to the *S. pombe* U5.U2/U6 spliceosomal complex, and rescues the splicing defect caused by deletion of the NTE from endogenous *cwf10*$^{+}$. Taken together, these findings suggest that the NTE is a semi-ordered domain that has the ability to function *in trans* to the EF2-like portion of Cwf10.

Results

Deleting the *Sp* Cwf10 NTE impairs splicing *in vivo*

To more closely examine the *in vivo* role of the Cwf10 NTE in pre-mRNA splicing, we prepared an *S. pombe cwf10* construct corresponding to *S. cerevisiae snu114ΔN* (Bartels, Klatt et al. 2002). Although *cwf10 2-135Δ* aligns with *snu114ΔN*

(Figure 2-1B), the *S. pombe* allele did not support growth at either 25°C or 32°C when integrated at the endogenous locus (data not shown), while the *S. cerevisiae* allele is viable at temperatures below 40°C (Frazer, Lovell et al. 2009). Upon closer examination of the alignment of Snu114 family members with EF2, it appears that Cwf10 2-135Δ lacks residues corresponding to domain I (Figure 2-1B) that may affect the folding of this essential domain. Therefore, we deleted eight fewer amino acids and successfully integrated *cwf10 2-127Δ* (hereafter *cwf10-ΔNTE*) at the endogenous locus. This result suggests that some of the defects seen in the *S. cerevisiae snu114ΔN* strain may partially result from a defect in the folding of domain I of the EF2-like region of Snu114.

We further characterized *cwf10-ΔNTE* to examine the NTE's contribution to *in vivo* splicing. The *cwf10-ΔNTE* strain formed slightly smaller colonies on solid media at all temperatures tested (Figure 2-2A) and exhibited slow growth in liquid culture at 25°C and 36°C compared with wild-type cells (Figure 2-2B and data not shown). Conversely, overexpression of the NTE (*cwf10 1-135* or *1-107*) under the high-strength *nmt* (no message in thiamine) promoter did not impair cell viability, indicating its expression is not dominant negative (data not shown). RT-PCR analysis of *tbp1_a*, an mRNA intron highly sensitive to splicing defects (Habara, Urushiyama et al. 2001), indicated that *cwf10-ΔNTE* cells grown at 25°C accumulate unspliced transcript (Figure 2-2C). Thus, removal of the Cwf10 NTE reduces splicing efficiency *in vivo*.

In our experience, yeast spliceosome mutants often have a lower steady-state mutant protein level which correlates with the splicing defect. To determine whether the deletion of the Cwf10 NTE simply decreases the amount of Cwf10 in cells, thus reducing splicing efficiency, antibodies were generated against Cwf10 amino acid residues 673-

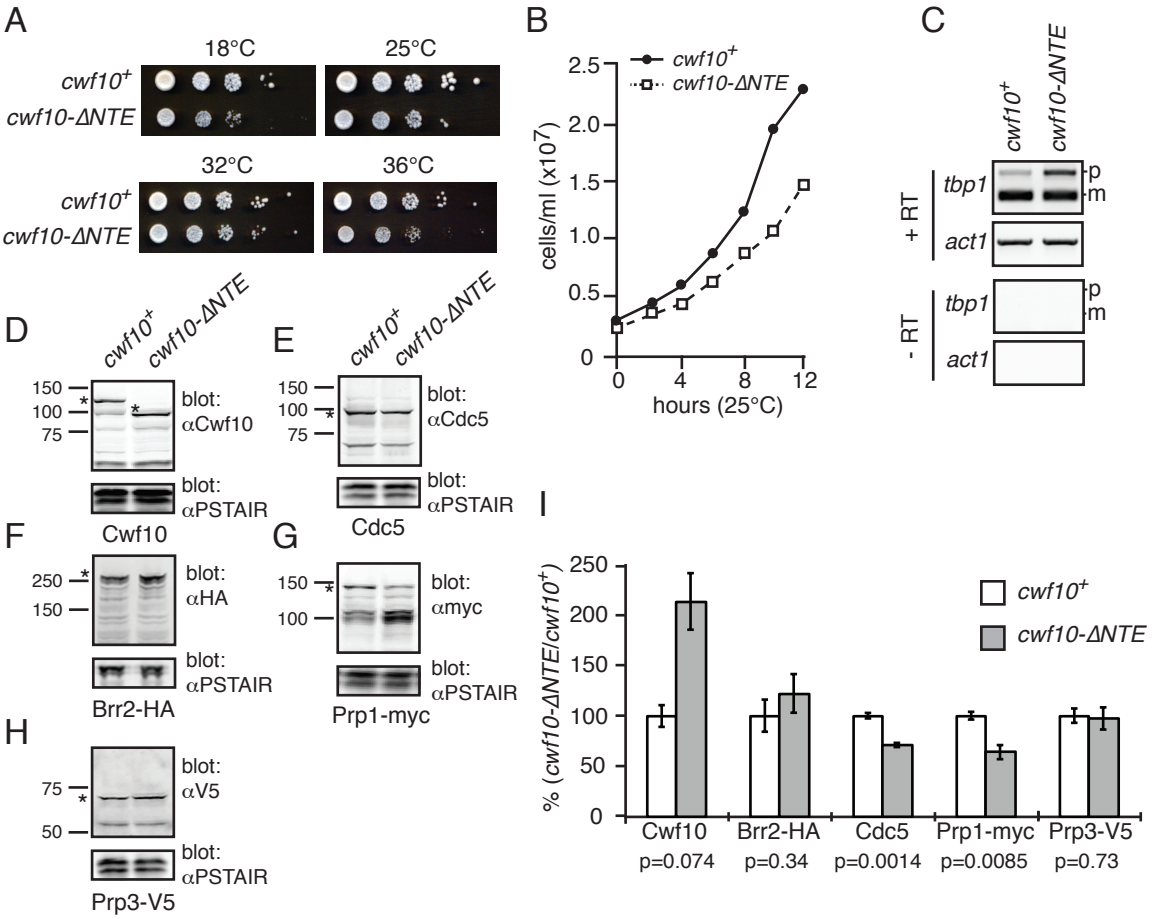


FIGURE 2-2

Characterization of the *S. pombe cwf10-ΔNTE* allele

A) Growth of serial dilutions of wild-type (*cwf10+*) and *cwf10-ΔNTE*. Serial dilutions of equal numbers of cells were spotted on solid YE media and grown at 18°C, 25°C, 32°C, and 36°C. B) Growth of *cwf10+* and *cwf10-ΔNTE* cells in liquid YE at 25°C. Samples were collected at 2 hour intervals to determine cell number. C) The NTE is required for efficient pre-mRNA splicing. RT-PCR was used to characterize RNA extracted from *cwf10+* and *cwf10-ΔNTE* cells. Total RNA was isolated from cells grown at 25°C. RT-PCR was performed with random hexamer primers. Portions of the *tbp1* (containing introns) and *act1* (intronless) transcripts were amplified from the RT-PCR products. Mature (m) and pre-mRNA (p) forms are labeled. No bands were detected in any RT-PCR samples in which the reverse transcriptase enzyme was omitted (-RT). D-H) Immunoblots demonstrating relative protein levels of pre-mRNA splicing factors in either *cwf10+* or *cwf10-ΔNTE* backgrounds. "*" marks the position of each protein on the blot. D-E) Lysates from *cwf10+* and *cwf10-ΔNTE* cells were probed with anti-PSTAIR and either anti-Cwf10 (D) and anti-Cdc5 (E) antibodies. F) Lysates from *brr2-HA* and *brr2-HA cwf10-ΔNTE* cells probed with anti-HA and anti-PSTAIR. G) Lysates from *prp1-myc* and *prp1-myc cwf10-ΔNTE* cells probed with anti-myc and anti-PSTAIR. H) Lysates from *prp3-V5* and *prp3-V5 cwf10-ΔNTE* cells probed with anti-V5 and anti-PSTAIR. Anti-PSTAIR detects Cdc2. Numbers to the left of each blot indicate molecular weight in kilodaltons. I) Comparison of protein levels of each pre-mRNA splicing factor in *cwf10-ΔNTE* cell lysates relative to the *cwf10+* value (100%), normalized to Cdc2, quantitated from Western analysis. Values were calculated using three biological replicates. Error bars indicate standard error of the mean. The P value was determined using the Student's t test.

983 (domains IV and V). Immunoblotting of *S. pombe* lysate with α-Cwf10 antiserum detected a single protein at the anticipated molecular weight of ~116 kDa (Figure 2-2D, *cwf10*$^{+}$ lane). In *cwf10-ΔNTE* lysates the antiserum detected a band ~20 kDa smaller than the wild-type protein, consistent with the predicted molecular weight of the deletion (Figure 2-2D). Using quantitative Western blotting, we found that levels of Cwf10 were consistently higher in *cwf10-ΔNTE* cells than in wild-type cells when quantified against the loading control Cdc2 (Cdk1) (Figure 2-2I). These results show that low levels of Cwf10 do not cause the splicing defect in *cwf10-*Δ*NTE* cells. To examine whether levels of other splicing components might be affected in this background, we analyzed the protein levels of *S. pombe* Cdc5 (homolog of *S. cerevisiae* Cef1) (Figure 2-2E), a core component of the nineteen complex (NTC) (Tsai, Chow et al. 1999, Ohi and Gould 2002), Brr2 (homolog of *S. cerevisiae* Brr2) (Figure 2-2F), a core U5 snRNP member (Achsel, Ahrens et al. 1998), as well as Prp1 (homolog of *S. cerevisiae* Prp6) and Prp3 (homolog of *S. cerevisiae* Prp3) (Figure 2-2G and H), both members of the U4/U6/U5 tri-snRNP and B-complex (Stevens and Abelson 1999, Fabrizio, Dannenberg et al. 2009). In the *cwf10-ΔNTE* background, levels of Brr2 and Prp3 were similar to levels seen in wild-type cells; however, both Cdc5 and Prp1 levels were reduced to ~70% and ~65%, respectively (Figure 2-2I). Since the *prp1* transcript does not contain introns, we attribute the lower level of these proteins to reduced protein stability rather than a general reduction in transcript levels.

RNA-seq demonstrates a general splicing defect in *cwf10-ΔNTE* cells

Previously it has been demonstrated that mutations in core spliceosome proteins affect splicing of introns to different degrees (Romfo, Lakhe-Reddy et al. 1999, Pleiss, Whitworth et al. 2007, Sridharan, Heimiller et al. 2011). That is, one mutant can have a profile of inefficiently spliced introns that is significantly different from the profile of another mutant. To examine the extent of splicing changes in *cwf10-ΔNTE* cells, we performed transcriptome analysis. RNA was extracted from *cwf10*$^{+}$ and *cwf10-ΔNTE* cells grown at 25°C, and two replicates for each genetic background were sequenced on the Illumina HiSeq platform. Two methods were employed to evaluate the effect of NTE deletion on splicing efficiency using the RNA sequencing data. A Cochran-Mantel-Haenszel (CMH) test assessed the significance of the change in the ratio of spliced to unspliced transcripts between *cwf10-ΔNTE* and *cwf10*$^{+}$ (Materials and Methods), while DESeq (Anders and Huber 2010) revealed differentially expressed introns between the two conditions. Splicing efficiency displayed a nearly-global shift when *cwf10-ΔNTE* and *cwf10*$^{+}$ were compared; this can be seen as a shift of most points toward the right of the diagonal in Figure 2-3A. Splicing efficiency was highly reproducible between replicates, with a tight distribution along the diagonal (Figure 2-4). This suggests an extensive splicing defect in the NTE-deleted strain that affected most introns. Using the splicing efficiency (SE) approach under a stringent threshold (q-value <0.05; fold-change >2), we identified 1,193 introns (out of 5,361 possible introns) whose splicing efficiency was significantly compromised in the mutant. Similarly, using DESeq, 883 introns were more highly expressed in *cwf10-ΔNTE* relative to *cwf10*$^{+}$, indicating that their splicing was significantly less efficient in the mutant strain (adjusted p-value <0.05 and fold-change

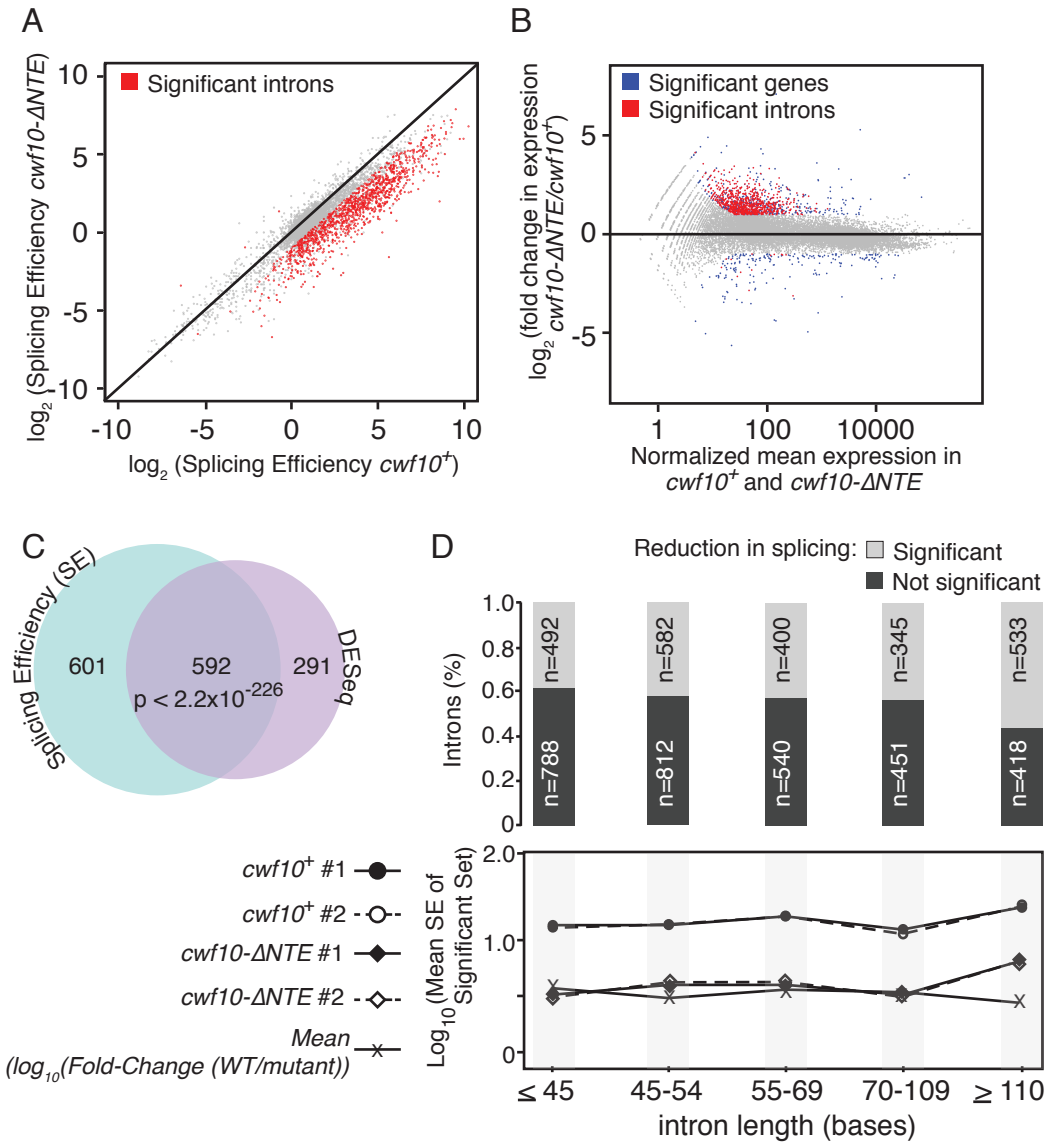


FIGURE 2-3

RNA-seq data reveal a nearly-global splicing defect in *cwf10-ΔNTE* cells

A) A global reduction in splicing efficiency in the *cwf10-ΔNTE* strain (Y-axis; $\log_2$ scale) compared to wild-type (X-axis; $\log_2$ scale). Each point represents an intron (5,361 in total), while red dots represent the 1,193 introns for which splicing efficiency was significantly compromised in the mutant compared to wild-type (CMH test, q-value <0.05, absolute fold change cutoff >2). Note that this graph compares one replicate of each genotype. B) Differentially expressed genes (blue) and introns (red) identified by DESeq (adjusted p-value <0.05 and absolute fold-change >2; X-axis: normalized mean expression; Y-axis log2 fold-change). A total of 21,193 genes were interrogated, of which 883 introns were up-regulated in the mutant. (C) Venn Diagram illustrating the overlap between the two methods used to assess the extent of the splicing defect (SE-light blue; DESeq-light pink). The probability of observing the same number of overlaps or more by chance assuming a hypergeometric distribution was determined by "phyper" function in R. (D) Top panel: The proportion of significant and non-significant introns within a given intron-length bin (2,352 introns identified by SE or DESeq, with q-value or adjusted p-value <0.05, respectively; for a given intron, splicing efficiency was compromised in both replicates; light grey: significant; dark grey: non-significant; exact number are as indicated). Bottom panel: The splicing efficiency of the significant intron set as a function of their length (straight and dashed lines with circles: technical replicates of wild-type; straight and dashed lines with diamonds: technical replicates of mutant; line with X's: mean $\log_{10}$ of SE fold-change mutant/wild-type).

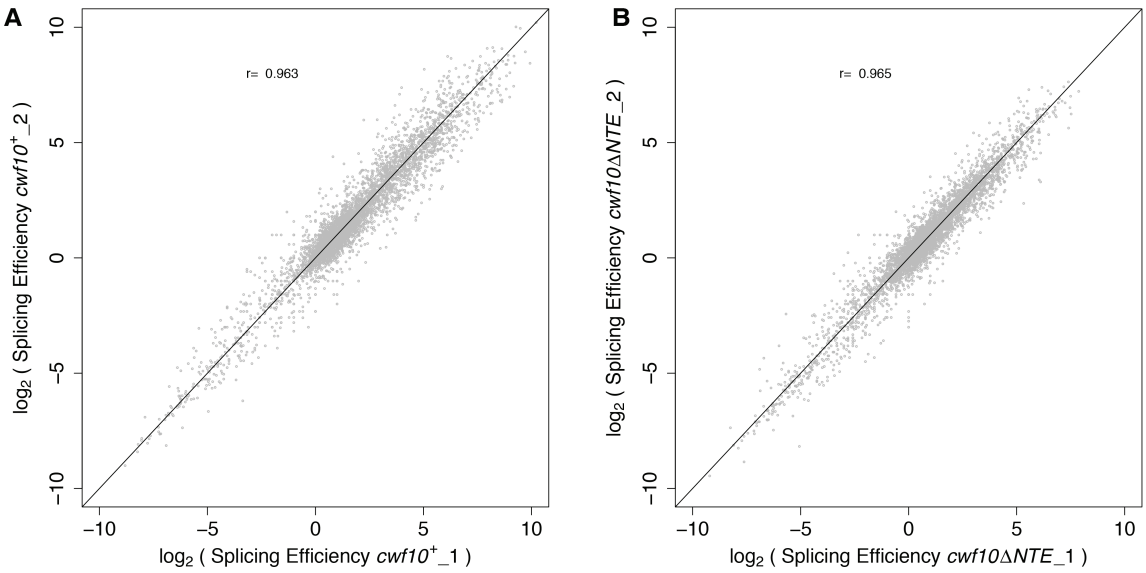


FIGURE 2-4

Reproducibility of splicing efficiency in replicates
Each point represents an intron (5,361 in total; 'r' Pearson's correlation coefficient; $\log_2$ scale). A) Two wild-type replicates. B) Two *cwf10-ΔNTE* replicates.

>2; Figure 2-3B). There was a good correspondence between the two methods ($p < 2.2e-226$, hypergeometric test; Figure 2-3C), which was even stronger when no fold-change cutoff was applied ($p < 2.7e-249$; data not shown). Overall, the splicing of 2,352 introns (approximately 44% of all introns) was significantly compromised by the NTE deletion, as indicated by at least one method without the fold-change cutoff. The effect of the NTE deletion was not dependent on intron length, and neither was the magnitude of the reduction in splicing efficiency (Figure 2-3D, top and bottom panels, respectively). Moreover, the effect of the mutation was not dependent on intron GC content, branch-site sequence, 5' intron sequence, or on the position of the intron within the transcript (data not shown). For a small set of introns (17 total), splicing efficiency was significantly improved by the deletion, but there was no agreement between the two methods under these criteria (14 found by DESeq and 3 by SE). However, when the magnitude of the fold-change was ignored, the splicing efficiency of 223 introns was significantly improved (209 by DESeq and 24 by SE, of which ten were identified by both methods; $p < 3.9e-10$). We did not find any clear sequence features that could explain why a small percentage of introns were either spliced with the same or better efficiency in the NTE deletion background. Taken together, the RNA-seq data indicate that the NTE of Cwf10 is required for efficient splicing of most *S. pombe* introns.

The lack of Cwf10-NTE changes spliceosome sedimentation patterns

In *in vitro* splicing assays, deletion of the *Sc* Snu114 NTE inhibits U4/U6 snRNA unwinding, thus impairing activation of the spliceosome (Bartels, Klatt et al. 2002). Since *S. pombe* does not have a robust *in vitro* splicing extract system, likely due to the stability

of the U5.U2/U6 spliceosomal complex (Huang, Vilardell et al. 2002, Ohi, Ren et al. 2007), we could not directly test whether *Sp* Cwf10-NTE plays a similar role in fission yeast. However, we postulated that if the *S. pombe* Cwf10-NTE shares a similar function to its *S. cerevisiae* ortholog, cells that lack the Cwf10-NTE might have altered sedimentation patterns of spliceosomal complexes and/or snRNAs. Thus, to test the effect of deleting Cwf10-NTE *in vivo*, we compared the sedimentation patterns of Cwf10, Cdc5 (*Sc* Cef1), and the five snRNAs in sucrose gradients using native lysates made from either wild-type or *cwf10*Δ-*NTE* cells (Figure 2-5). It has been previously shown that in lysates from asynchronously growing wild-type *S. pombe* cells, the majority of spliceosomes sediment as a stable ~37S U5.U2/U6 complex (McDonald, Ohi et al. 1999, Huang, Vilardell et al. 2002, Ohi, Link et al. 2002, Ohi, Ren et al. 2007), although less abundant U5/U6/U4/U2/U1 and U5/U6/U4/U2 complexes have also been characterized (Carnahan, Feoktistova et al. 2005, Lützelberger, Bottner et al. 2009). Unlike what is found in other organisms, *S. pombe* lysates lack detectable quantities of a pre-assembled U5.U4/U6 tri-snRNP (Huang, Vilardell et al. 2002).

As expected, in wild-type lysates a majority of Cwf10 sediments at an ~37S peak corresponding to the sedimentation pattern of the U5.U2/U6 complex (Figure 2-5A and B, fraction 9, (Ohi, Link et al. 2002)); however, the sedimentation pattern of Cwf10-ΔNTE in *cwf10*-Δ*NTE* lysates changes, with a portion of Cwf10-ΔNTE shifting to higher molecular mass fractions (Figure 2-5A and B, fractions 11 and 12). The sedimentation pattern of Cdc5, a core NTC component, does not appear grossly altered in lysates from *cwf10*-Δ*NTE* cells (Figure 2-5C and D). We next compared the sedimentation profiles of spliceosomal snRNAs in both wild-type and *cwf10*-Δ*NTE* lysates. As expected in the

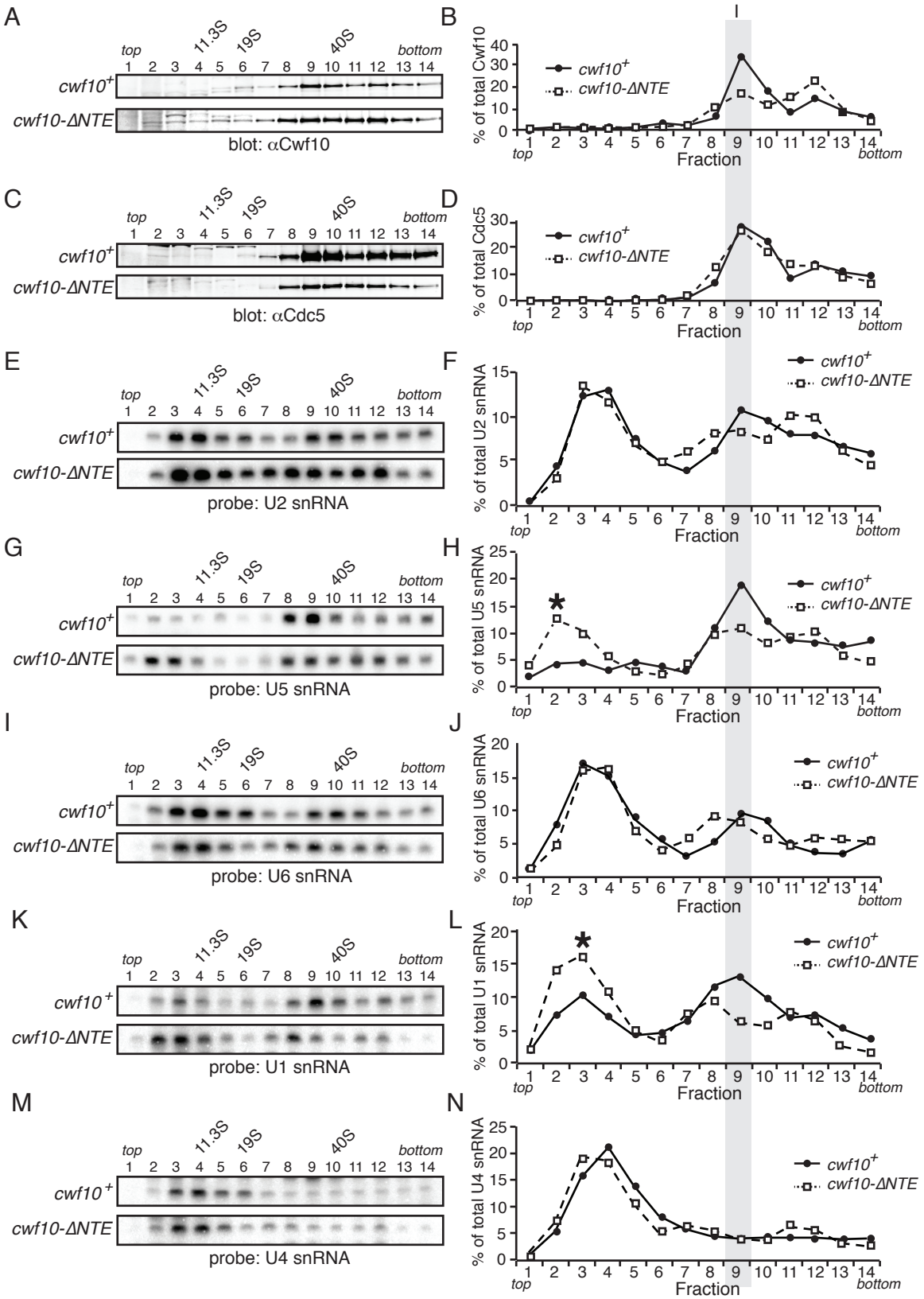

A
B
C
D
E
F
G
H
I
J
K
L
M
N
top
bottom
11.3S
19S
40S
1 2 3 4 5 6 7 8 9 10 11 12 13 14
cwf10+
cwf10-ΔNTE
blot: αCwf10
blot: αCdc5
probe: U2 snRNA
probe: U5 snRNA
probe: U6 snRNA
probe: U1 snRNA
probe: U4 snRNA
% of total Cwf10
% of total Cdc5
% of total U2 snRNA
% of total U5 snRNA
% of total U6 snRNA
% of total U1 snRNA
% of total U4 snRNA
Fraction
*

FIGURE 2-5 (from previous page)

Sedimentation of spliceosome components is altered in *cwf10-ΔNTE* cells
Lysates from *cwf10+* and *cwf10-ΔNTE* cells grown at 25°C were run on 10-30% sucrose gradients. Fractions were collected and either TCA precipitated for immunoblot analysis or RNA was extracted for Northern analysis. A and C) Immunoblots of fractions collected from 10-30% sucrose gradients of lysates from *cwf10+ prp1-myc* and *cwf10-ΔNTE prp1-myc* using anti-Cwf10 (A) and anti-Cdc5 (C). B and D) Intensities of the bands of the anti-Cwf10 blot (B) and anti-Cdc5 (D) were quantified and plotted as a percentage of the sum of the signal from all fractions. The bands were quantified by near infrared detection and Odyssey software (Licor). E, G, I, K, and M) RNA was isolated from fractions collected from 10-30% sucrose gradients of *cwf10+* and *cwf10-ΔNTE* lysates. Blots were probed with ^{32}P-labeled oligonucleotides complementary to the *S. pombe* U2 (E), U5 (G), U6 (I), U1 (K), and U4 (M) snRNAs. F, H, J, L, and N) Intensities of the bands of the U2 (F), U5 (H), U6 (J), U1 (L), U4 (N) probes were quantified and plotted as a percentage of the sum of the signal from all fractions. The bands were quantified by phosphor-imaging and ImageQuant TL 8.1 software (GE Healthcare). The "*" in H and L highlights the peak of the U5 and U1 snRNAs found in a small molecular weight fraction in *cwf10-ΔNTE*. Column I indicates where the U5.U2/U6 spliceosomal complex sediments in the gradients. The migration of FAS (40S), thyroglobulin (19S), and catalase (11.3S) collected from parallel gradients is indicated.

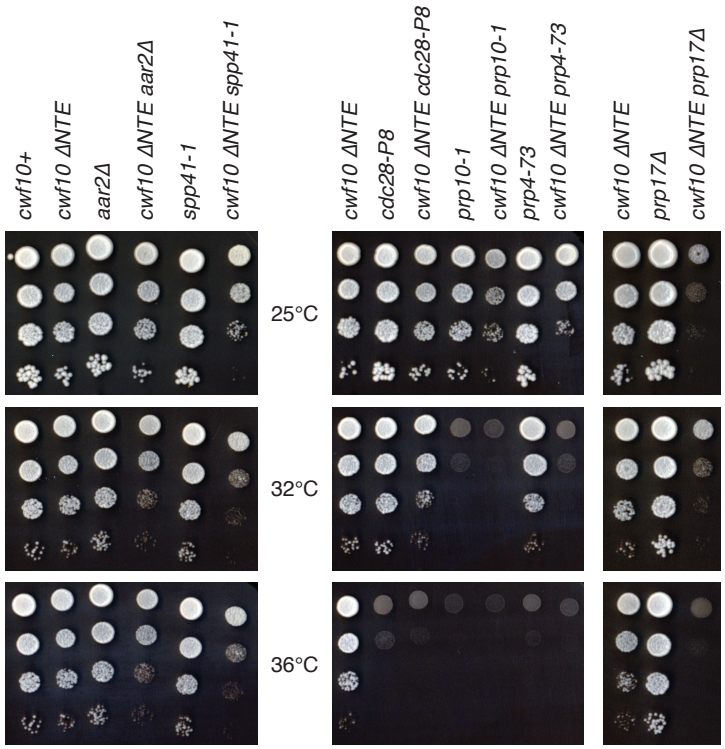
wild-type background, the U2, U5, and U6 snRNAs co-sedimented at ~37S as described before (McDonald, Ohi et al. 1999, Huang, Vilardell et al. 2002, Ohi, Link et al. 2002, Ohi, Ren et al. 2007) (Figure 2-5 E-J). The sedimentation pattern of U4, U5, and U6 snRNAs from wild-type cells (Figure 2-5 G-J and M-N) are consistent with previous reports that *S. pombe* appears to lack an independently sedimenting U5.U4/U6 tri-snRNP (Huang, Vilardell et al. 2002). Another apparent difference between fission yeast and other organisms is the abundance of the U2 snRNP at 12S (Figure 2-5E and F), which most likely represents the core U2 snRNP rather than the larger 17S complex that contains the SF3a and SF3b sub-complexes (Behrens, Galisson et al. 1993, Behrens, Tyc et al. 1993, Brosi, Hauri et al. 1993, Pauling, McPheeters et al. 2000). However, when comparing wild-type and *cwf10*-Δ*NTE* lysates, the sedimentation patterns of all the snRNAs were altered to some degree (Figure 2-5 E-N). The differences include a small but consistent shift of the U2, U4, U5, and U6 snRNAs to higher molecular mass fractions (>40S, fractions 11-13, Figure 2-5 E-J and M-N), as well as, increases in the amount of the U5 and U1 snRNAs at lower molecular mass fractions (<11.3S, Figure 2-5 G-H and K-L). Thus, from these analyses we conclude that deleting the *Sp* Cwf10 NTE does alter the distribution of spliceosomal complexes *in vivo*, although in a complex pattern with shifts to both higher and lower molecular masses.

cwf10-ΔNTE synthetically interacts with factors predicted to be involved in spliceosome activation

As another approach to investigate whether *S. pombe* NTE function can be correlated to a specific step of the pre-mRNA splicing reaction or stage of spliceosomal

organization, we tested for genetic interactions between *cwf10-ΔNTE* and eight previously characterized *S. pombe* pre-mRNA splicing alleles (Figure 2-6). Our analysis revealed that *cwf10-ΔNTE* is synthetic lethal when combined with *prp1-4* (Urushiyama, Tani et al. 1996) (homolog of *S. cerevisiae PRP6)*, *spp42-1* (Schmidt, Richert et al. 1999) (homolog of *Sc PRP8*), and *cdc5-120* (Nurse, Thuriaux et al. 1976) (homolog of *Sc CEF1*). *cwf10-ΔNTE* has an extreme synthetic sick phenotype in combination with *prp17Δ* (Sapra, Arava et al. 2004). When *cwf10-ΔNTE* is combined with either *prp4-73* (Rosenberg, Alahari et al. 1991) (homolog of human PRPF4B) or *spp41-1* (Schmidt, Richert et al. 1999) (homolog of *Sc BRR2*), the cells are synthetically sick. Conversely, synthetic interactions are weak or non-existent between *cwf10-ΔNTE* and *aar2Δ* (Kim, Hayles et al. 2010) (homolog of *Sc AAR2*), *prp10-1* (Urushiyama, Tani et al. 1996) (homolog of *Sc HSH155*), or *cdc28-P8* (Nasmyth and Nurse 1981) (homolog of *Sc PRP2*). Two of the three synthetic lethal interactions, *prp1-4* and *spp42-1*, are with genes previously shown to be important for spliceosome activation (Kuhn, Li et al. 1999, Lützelberger, Bottner et al. 2009). The third synthetic lethal interaction, *cdc5-120*, is interesting because Cdc5 is likely involved in spliceosome activation as a component of the NTC (Chan, Kao et al. 2003) and in modulating the transition between first and second step catalysis (Query and Konarska 2012). The most obvious synthetic sick phenotype is in combination with deletion of *prp17*, a second catalytic step factor (Jones, Frank et al. 1995). We conclude that the *S. pombe* Cwf10 NTE is likely important for a specific stage in spliceosome activation, and possibly also for a later stage(s) in the splicing reaction.

A



B

cwf10-ΔNTE crossed with:	Synthetic Lethal	Growth 25˚C	Growth 36˚C	homolog	Spliceosomal Complex
prp1-4	**Yes**	-	-	*Sc* Prp6	U5 snRNP
spp42-1	**Yes**	-	-	*Sc* Prp8	U5 snRNP
cdc5-120	**Yes**	-	-	*Sc* Cef1	NTC
prp4-73	No	-/+	-	hPRPF4B	Kinase
prp17Δ	No	-/+	-/+	*Sc* Prp17	NTC
prp10-1	No	+	-	*Sc* Hsh155	U2 snRNP
cdc28-P8	No	+	-	*Sc* Prp2	B^{act} complex
spp41-1	No	-/+	-/+	*Sc* Brr2	U5 snRNP
aar2Δ	No	+	-/+	*Sc* Aar2	U5 snRNP

- = does not grow -/+ = slow growth compared to single mutant + = grows

FIGURE 2-6

Genetic interactions of *cwf10-ΔNTE* with alleles of other pre-mRNA splicing factors
A) Serial dilutions of single and double mutants grown at 25°C, 32°C, and 36°C were used to examine growth. B) Table summarizing results of crossing *cwf10-ΔNTE* with other pre-mRNA splicing factor mutants. The column labeled “Growth” indicates growth of the double mutant compared with the single mutant allele in a wild-type (*cwf10+*) background at 25°C and 36°C. “-“ = no growth, “-/+” = slow growth, “+” = growth, “Sc” = S. cerevisiae; “h” = human.

To further address whether the NTE may be involved in facilitating the transition from an inactive to activated spliceosome, we attempted to analyze the protein composition of spliceosomal complexes purified from either a *cwf10*-Δ*NTE* or wild-type background. Purification of tagged Cwf10 was our first choice, due to changes in sedimentation of this protein in wild-type and mutant backgrounds (Figure 2-5A and B). However, we were not able to tag endogenous *cwf10*-Δ*NTE* with a C-terminal TAP tag. Purification of the Prp1-TAP complex, which is associated with a group of proteins similar to Complex B (Carnahan, Feoktistova et al. 2005), failed due to extremely low levels of the Prp1-TAP fusion protein in *cwf10*-Δ*NTE prp1-TAP* lysates. Therefore, we purified a

U5.U2/U6 spliceosomal complex from *cdc5-TAP* cells in either a *cwf10*-Δ*NTE* or wild-type background at 75 mM salt and identified associating proteins using one-dimensional (1D) liquid chromatography-tandem mass spectrometry (LC-MS/MS). Cdc5 sedimented similarly in the wild-type and mutant *cwf10* backgrounds (Figure 2-5C and D), but we reasoned that using lower salt (75 mM) in the purification would preserve weaker interactions that may be unique to one of the two backgrounds. From analysis of peptides in two biological replicates of each genetic background, the U5.U2/U6 complex appeared largely similar in *cwf10*$^{+}$ and *cwf10*-Δ*NTE* (Table 2-1), suggesting that, although Cwf10 is highly represented in the U5.U2/U6 complex, the NTE domain is not required to maintain most associations within that complex.

	S. pombe	ORF number	Cdc5-TAP				*S. cerevisiae*	*H. sapiens*
			cwf10-ΔNTE		wild-type			
			I	II	I	II		
Core snRNP	Smb1	SPAC26A3.08	3	2	12	11	Smb1	SMB/B'
	Smd1	SPAC27D7.07c	4	3	8	8	Smd1	SMD1
	Smd2	SPAC2C4.03c	5	6	14	13	Smd2	SMD2
	Smd3	SPBC19C2.14	---	3	5	14	Smd3	SMD3
	Sme1	SPBC11G11.06c	2	1	6	4	Sme1	SME1
	Smf1	SPBC3E7.14	5	8	6	10	Smf1	SMF1
	Smg1	SPBC4B4.05	1	3	10	8	Smg2	SMG1
U2	Lea1	SPBC1861.08c	11	11	28	27	Lea1	U2A'
	Msl1	SPBC8D2.09c	3	2	2	6	Msl1	U2B"
	Prp2/Mis11	SPBC146.07	---	---	---	4	Mud2	U2AF65
	Sap61	SPBC36.09	3	2	2	3	Prp9	SF3A60
	Sap114	SPAC22A12.09c	1	1	1	3	Prp21	SF3A1
	Ini1	SPAC23H3.02c	---	---	1	2	Rds3	SF3b14b
	Sap145	SPAC22F8.10c	---	2	1	1	Cus1	SF3B150
	Prp10	SPAC27F1.09c	3	2	3	8	Hsh155	SF3B160
	Prp12	SPAPJ698.03c	6	9	5	3	Rse1	SF3B130
U5	Cwf10-ΔNTE	SPBC215.12	34	37	---	---	Snu114	U5-116
	Cwf10	SPBC215.12	---	---	100	73	Snu114	U5-116
	Brr2	SPAC9.03c	---	---	4	11	Brr2	U5-200
	Spp42	SPAC4F8.12c	59	78	186	171	Prp8	U5-220
	Cwf17/Spf38	SPBC1289.11	14	14	41	38	UNK	U5-40
NTC core	Cdc5	SPAC644.12	31	29	92	71	Cef1	CDC5
	Cwf2	SPAC3A12.11c	5	7	37	38	Cwc2	RBM22
	Cwf7	SPBC28F2.04c	10	14	34	13	Snt309	SPF27
	Cwf15	SPBC337.06c	2	3	13	11	Cwc15	AD002
	Prp5	SPBP22H7.07	19	19	51	57	Prp46	PRL1
	Prp19	SPAC29A4.08c	25	24	75	48	Prp19	PRP19
NTC-associated	Cwf3	SPBC211.02c	23	29	47	56	Syf1	SYF1
	Cwf4	SPBC31F10.11c	18	23	67	48	Clf1	CRN1
	Cwf5/Ecm2	SPCC550.02c	12	17	52	42	Ecm2/Slt11	RBM22
	Cwf12	SPBC32F12.05c	6	3	16	12	Isy1	ISY1
	Prp45	SPCC188.11	13	19	28	30	Prp45	SKIP
	Prp17	SPBC6B1.10	14	14	38	45	Prp17	hPRP17
	Syf2	SPBC3E7.13c	6	10	19	25	Syf2	SYF2
Other	Cwf11	SPBC646.02	34	37	115	97	UNK	hCWF11
	Cwf14	SPBC24C6.11	7	9	16	13	Cwc14	G10
	Cwf16	SPAC9.13c	2	3	1	8	Yju2	CCDC94
	Cwf18	SPCP1E11.07c	7	6	12	13	UNK	MGC23918
	Cwf19	SPAC30D11.09	18	16	37	47	UNK	hCWF19L2
	Cwf21	SPAC4A8.09c	3	1	6	7	Cwc21	SRm300
	Cwf22	SPBC13E7.01	7	9	15	18	Cwc22	hCWC22
	Cwf25	SPBC146.05c	---	1	---	10	Cwc25	hCWF25
	Cwf26	SPCC1620.10	1	1	3	1	Bud13	BUD13
	Cyp1	SPAC57A10.03	4	3	7	7	UNK	PPIL1
	Mug161	SPAC1F3.09	---	---	8	9	YGR093W	CWF19L1
	Prp16	SPBC1711.17	---	---	---	6	Prp16	PRP16
	Cdc28	SPBC19C2.01	---	---	---	5	Prp2	DHX16
	Prp43	SPBC16H5.10c	---	---	9	2	Prp43	hPRP43
	Prp22	SPAC10F6.02c	---	3	13	21	Prp22	hPRP22
	Saf4	SPBC18H10.10c	1	3	11	11	UNK	CCDC130
	Prp40/Usp104	SPAC4D7.13	1	1	---	3	Prp40	PRPF40A
	Usp105	SPBC4B4.09	1	1	2	1	Prp39	UNK
	Bpb1	SPCC962.06c	1	---	---	4	Msl5	SF1
	Rsd1	SPAC19G12.07c	4	3	---	11	UNK	RNPC2
	Srp1	SPAC29A4.06c	---	---	1	2	UNK	SRFS2
	Dsk1	SPBC530.14c	---	---	5	5	Sky1	SRPK1
	Srp2	SPAC16.02c	2	----	1	7	Npl3	UNK

	Sum3	SPCC1795.11	3	1	1	4	Ded1	DDX3

Table 2-1. Splicing proteins co-purifying with Cdc5-TAP in either a *cwf10*-Δ*NTE* or wild-type background using 75 mM NaCl. Two biological replicates were analyzed using one-dimensional (1D) liquid chromatography-tandem mass spectrometry (LC-MS/MS) per genotype (I and II). Peptide counts are included for each replicate. UNK indicates that an ortholog is unknown or not present. "---", no peptides identified.

The Cwf10 NTE is partially folded, but contains regions of disordered coiled-coil

The structural characteristics of the NTE were of interest to us for several reasons. First, the only published data on NTE structure is a bioinformatics predication that the region is unfolded (Korneta, Magnus et al. 2012). Second, the NTE does not carry strong homology to known protein domains or to primary sequences in other proteins (data not shown). To more carefully examine the structural characteristics of this domain, we used DISOPRED (Ward, McGuffin et al. 2004), a program that predicts structure disorder, to precisely map potential regions of intrinsic disorder in the Cwf10 NTE. This analysis showed that a majority of the first N-terminal 60 amino acids are strongly predicted to lack secondary structure. The C-terminal 75 amino acids are predicted to be structurally ordered (Figure 2-7A), perhaps as β-strands as suggested by analysis with PSIPRED version 3.3 (Jones 1999, Bryson, McGuffin et al. 2005), a program that predicts secondary structure (data not shown). To experimentally test this model we expressed and purified recombinant Cwf10(1-135)His_6 from *E. coli* (Figure 2-8A). Analysis by sedimentation analytical ultracentrifugation (SVAU) shows that the Cwf10 NTE sediments as a monomer (s = 0.8, predicted molecular mass of ~19 kDa, rsmd = 0.09) with a frictional ratio of 2.13 (Figure 2-7B). Cwf10(1-135)His_6 was then analyzed by circular dichroism (CD) spectrometry using near UV wavelengths, which can be used to detect tertiary structure. Analysis of this spectrum shows a strong signal between 260-290 nm, suggesting that there are some aromatic residues found in a folded environment (Figure 2-7C). As expected, this signal is no longer seen when the protein is completely denatured in 6 M guanidine-HCl (Figure 2-7D). The NTE contains six tyrosines, two in the first 12 amino acids and the remaining four between residues 78 and 125 (Figure 2-

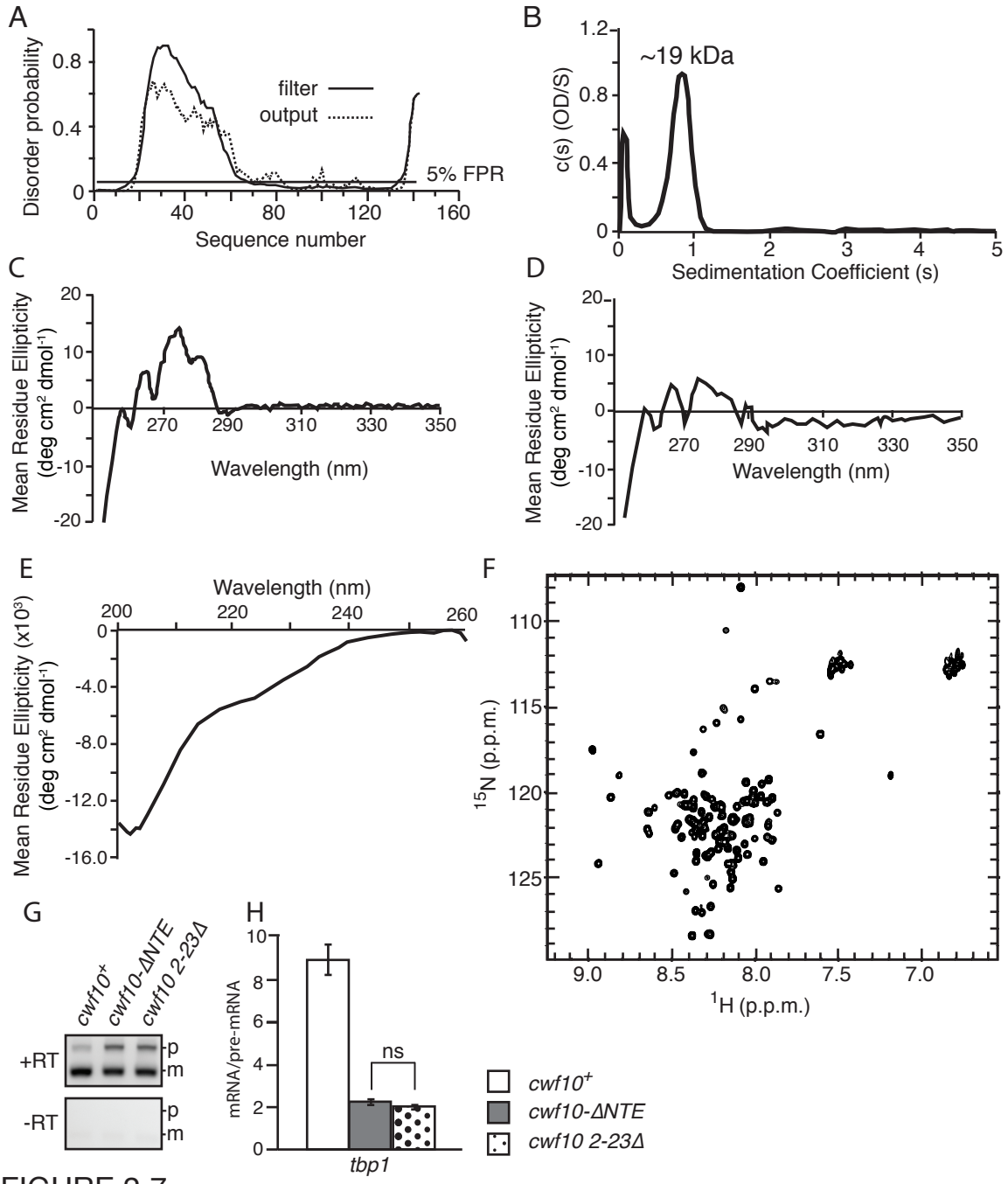


FIGURE 2-7

Biophysical characterization shows the Cwf10 NTE contains regions of order and disorder

A) Residues 18-61 of the NTE are predicted to be intrinsically disordered in solution using DISOPRED (Ward, et al., 2004), a program that predicts structure disorder. The false positive rate (FPR) was set at 5%. B) Cwf10(1-135)His_6 is monomeric by SVAU. The calculated s [c(s)] is plotted versus the sedimentation coefficient (s) for Cwf10(1-135)His_6. The sedimentation velocity profiles were fitted to a continuous sedimentation distribution. C) Near-UV CD spectrum of Cwf10(1-135)His_6. D) Near-UV CD spectrum of Cwf10(1-135)His_6 in 6 M guanidine-HCl. E) Far-UV CD spectrum of Cwf10(1-135)His_6. F) ^{15}N-^{1}H HSQC spectra of Cwf10(1-135)His_6. G, H) The first 23 amino acids of the Cwf10-NTE are required for efficient splicing. RNA was prepared from *cwf10+*, *cwf10-ΔNTE*, and *cwf10 2-23Δ* cells grown at 25°C and RT-PCR was done using random hexamer primers. A portion of the *tbp1* transcript was amplified from the RT-PCR products and ratios of mRNA/pre-mRNA were calculated. Quantification represents the average of three biological replicates per genotype. Error bars indicate standard error of the mean.

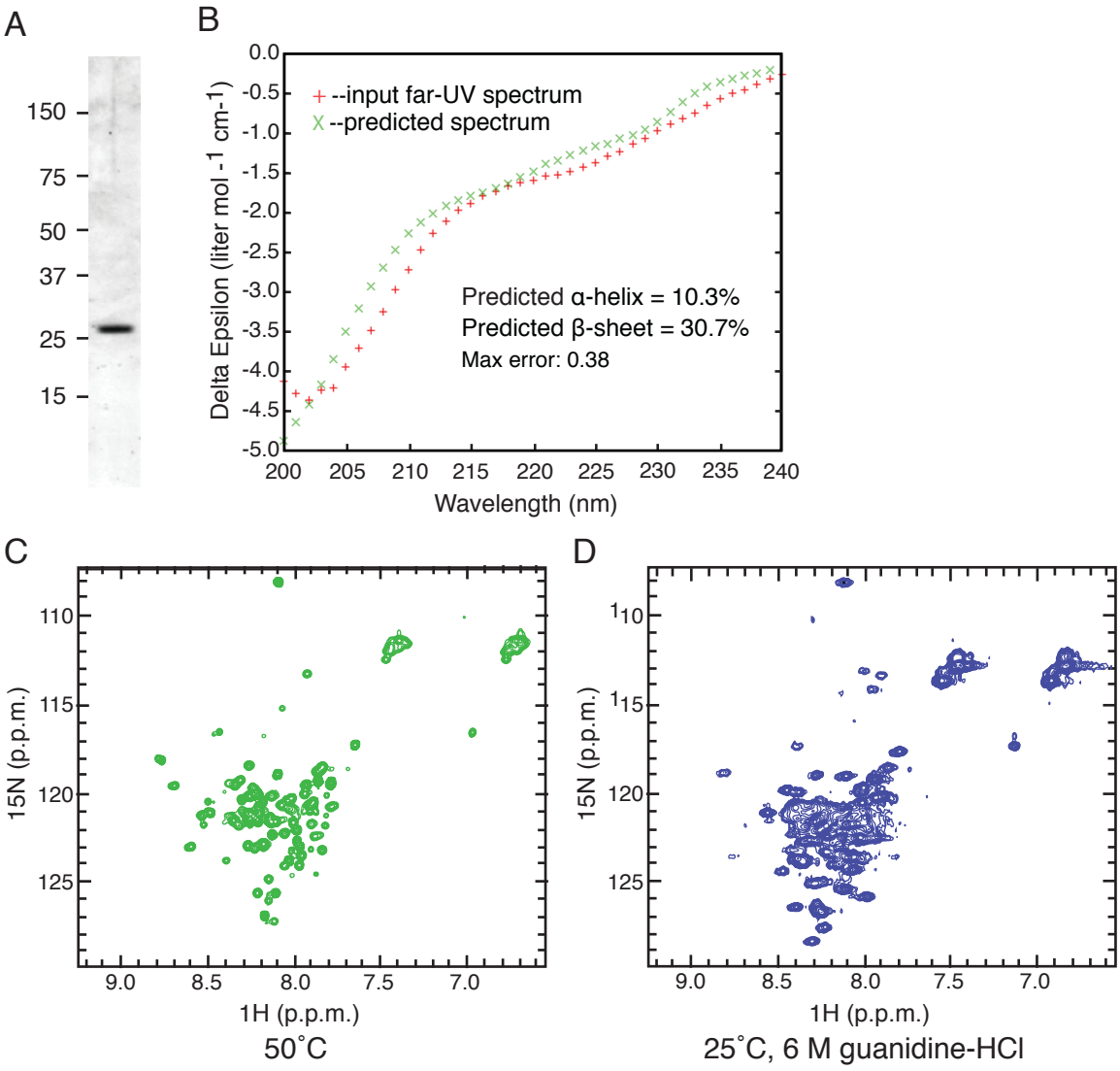


FIGURE 2-8

Analysis of recombinant Cwf10(1-135)His$_6$

A) Coomassie stained SDS-PAGE gel of recombinant Cwf10(1-135)His$_6$. Molecular weight markers in kilodaltons are shown on left. B) Predicted secondary structure calculated from the Far-UV spectrum seen in Fig. 2-7E using the program K2D2 (Perez-Iratxeta and Andrade-Navarro, 2008). C) ^{15}N-^{1}H HSQC spectrum of Cwf10(1-135)His$_6$ at 50°C. D) ^{15}N-^{1}H HSQC spectrum of Cwf10(1-135)His$_6$ in 6 M guanidine-HCl.

1B). Interestingly, these regions both correspond with predicted regions of order (Figure 2-7A). Next the protein was analyzed by CD spectrometry using far ultra-violet (UV) wavelengths (Figure 2-7E), which can be used to predict secondary structure. Analysis of this spectra using the program K2D2 (Perez-Iratxeta and Andrade-Navarro 2008) predicts that Cwf10(1-135)His_6 in solution is composed of ~10% α-helix and ~31% β-sheet, leaving over 50% of the NTE likely disordered (Figure 2-8B).

To more carefully examine the tertiary structure of the NTE, we ^{15}N labeled Cwf10(1-135)His_6 and collected a two-dimensional ^{15}N-^{1}H HSQC (Heteronuclear Single-Quantum Correlation) experiment using nuclear magnetic resonance (NMR) spectroscopy. By this analysis we found that the Cwf10-NTE contains 103 well-dispersed resonances (out of a total of 141 residues) (Figure 2-7F). When Cwf10(1-135)His_6 is either heated to 50°C or treated with 6 M guanidine-HCl in order to cause complete denaturation of any folded domain(s), the resonance peaks collapse and are no longer dispersed (Figure 2-8C and D). Well-dispersed resonances in HSQC spectra result from the variable environment of the amines in a folded protein. Thus, the NMR analysis confirms both the computational and CD analyses that the Cwf10 NTE contains regions of disorder, most likely in the first 60 amino acids (Figure 2-7A), as well as regions that adopt a well-ordered secondary structure.

Interestingly, amino acids 1-23 of the NTE are 70% identical between *S. pombe* and human orthologs (Figure 2-1B), and most of these residues (amino acids 1-17) are not predicted to be disordered (Figure 2-7A). To test if this patch of conserved residues is required for NTE function, we replaced the chromosomal copy of *cwf10*$^{+}$ with *cwf10 2-23Δ* to observe whether this smaller truncation would also cause a pre-mRNA splicing

defect. RNAs from wild-type, *cwf10 2-23Δ*, and *cwf10-ΔNTE* cells were extracted and RT-PCR performed (Figure 2-7G). Measures were taken to improve the quantitative nature of the PCR (Marone, Mozzetti et al. 2001), including reverse transcribing similar amounts of RNA, reducing PCR cycles, and quantifying against an inherent internal control (the spliced and unspliced forms are amplified in the same reaction). The ratio of mature to premature signal for the *tbp1_a* intron was almost identical between the two truncations (*cwf10-ΔNTE* 2.1 ± 0.2 vs. *cwf10 2-23Δ* 2.0 ± 0.3), while the ratio in the wild-type strain was 9.7 ± 1.5 (Figure 2-7H). This suggests that the conserved, extreme N-terminus (residues 1-23) is required for NTE function.

The Cwf10 NTE co-purifies with splicing factors

Spliceosomal binding partners of the NTE have not been determined in any organism. We attempted to screen interactions by yeast two-hybrid; however, the assay was hindered by the high self-activation of the GAL binding domain when fused to the acidic NTE. Therefore, we fused *cwf10 2-135* to the N-terminal tandem affinity purification (TAP) tag in a pREP41 NTAP vector (Tasto, Carnahan et al. 2001). We overexpressed *NTAP-NTE* in *cwf10-ΔNTE* cells, performed a two-step TAP purification at 150 mM NaCl, and analyzed the eluate using negative stain electron microscopy (EM). Although the purification was clearly dilute, we did see particles that were reminiscent of negative stain images of the *S. pombe* U5.U2/U6 complex (Figure 2-9A) (Ohi, Link et al. 2002, Ohi, Ren et al. 2007), suggesting that the Cwf10 NTE is able to interact with spliceosomal complexes on its own. To further characterize the composition of the NTAP-NTE purification, we analyzed the protein content of the eluate using 1D LC-

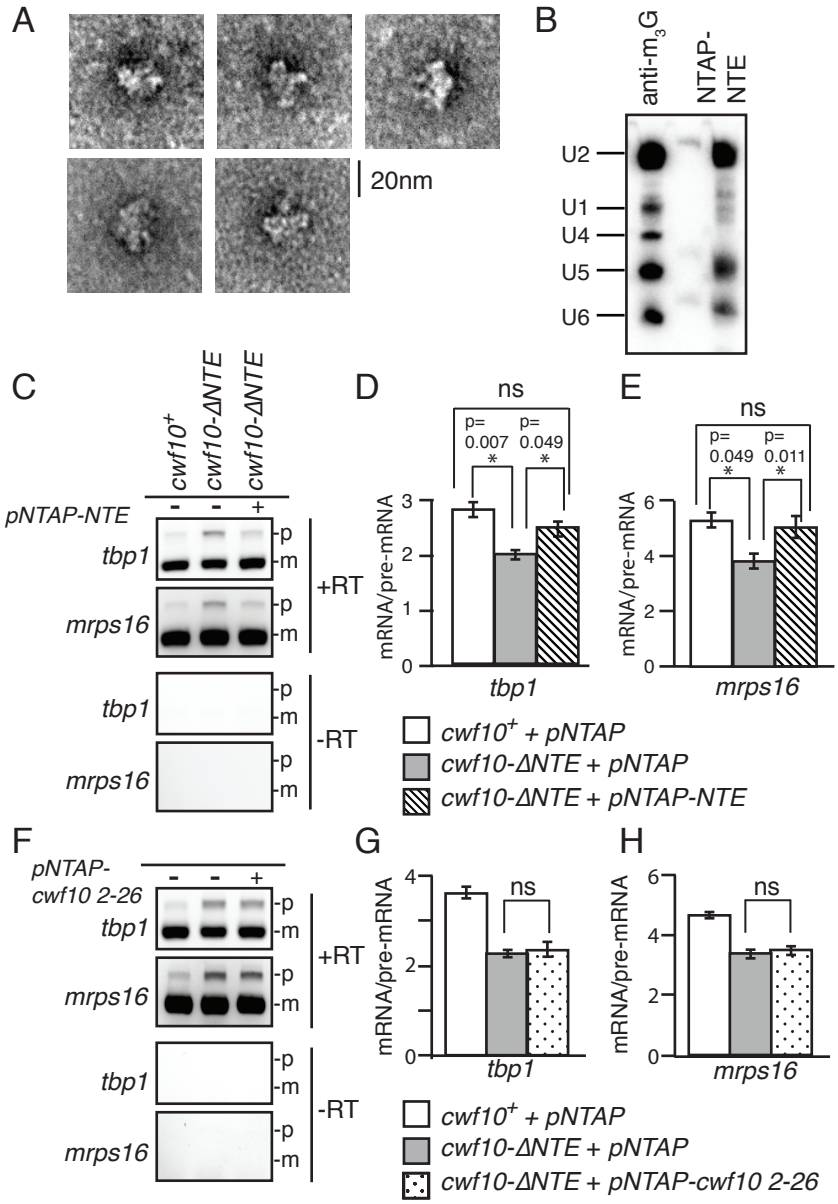


FIGURE 2-9

Cwf10 NTE can incorporate into spliceosomal complexes and partially rescue the splicing defect in *cwf10-ΔNTE* cells

A) Gallery of particles present in NTAP-NTE purifications from *cwf10-ΔNTE* cells as seen by negative stain EM. Scale bar is 20 nm. B) NTAP-NTE associates with the U2, U5, and U6 snRNAs. RNA was isolated from NTAP-NTE purified from *cwf10-ΔNTE* cells and anti-trimethylguanosine (m3G) purified from wild-type cells. Blots were probed with ^{32}P-labeled oligonucleotides complimentary to the *S. pombe* U1, U2, U4, U5 and U6 snRNAs. C) Overexpression of the NTEdomain in *cwf10-ΔNTE* cells rescues splicing defects. Total RNA was isolated from cells grown at 32°C. RT-PCR was performed with random hexamer primers. Portions of the *tbp1* and *mrps16* transcripts were amplified from the RT-PCR products. Mature (m) and pre-mRNA (p) forms are indicated. RT = reverse transcriptase enzyme. D and E) Quantification of mRNA/pre-mRNA ratios for three biological replicates per genotype for RT-PCR analysis of *tbp1* (D) and *mrps16* (E) transcripts. Error bars indicate standard error of the mean. F) Overexpression of *cwf10 2-23* in *cwf10-ΔNTE* cells does not rescue splicing defects. Total RNA was isolated from cells grown at 32°C. RT-PCR was performed with random hexamer primers. Portions of the *tbp1* and *mrps16* transcripts were amplified from the RT-PCR products. G and H) Quantification of mRNA/pre-mRNA ratios for three biological replicates per genotype for RT-PCR analysis of *tbp1* (G) and *mrps16* (H) transcripts.

MS/MS. Peptide counts for spliceosomal proteins found in each LC-MS/MS run are included in Table 2-2 to provide a semi-quantitative indication of protein amounts. After heat shock proteins (data not shown), the next-highest group of proteins identified in the purification comprised pre-mRNA splicing factors that are similar to the *S. pombe* U5.U2/U6 spliceosomal complex (Ohi, Link et al. 2002, Ren, McLean et al. 2011) (Table 2-2). Highly represented in this group are the Sm proteins, components of the U5 snRNP, and components of and related to the hPrp19/Cdc5L complex. Importantly, the EF2-like portion of Cwf10 (from the genomic *cwf10-ΔNTE allele*) was also highly represented, confirming the capacity of Cwf10-ΔNTE to incorporate in to higher-order complexes, as seen by sucrose gradient sedimentation (Figure 2-5A). The presence of relatively high counts of the small, 27kDa U2 snRNP protein Lea1 indicates at least the partial presence of the U2 snRNP.

To examine how closely the NTAP-NTE purification matched endogenous U5.U2/U6 complexes, we purified Cdc5-TAP (homolog of *Sc* Cef1), which associates with the *Sp* U5.U2/U6 complex (Ohi, Link et al. 2002), and performed duplicate 1D MS analyses under the same conditions used to analyze NTAP-NTE. The compositions of Cdc5-TAP and NTAP-NTE are almost identical (Table 2-2). As one further characterization of the NTAP-NTE complex, we analyzed the snRNA content of the NTAP-NTE purification from *cwf10-ΔNTE* cells. As seen in Cdc5-TAP (Ohi, Link et al. 2002), NTAP-NTE is associated with the U2, U5, and U6 snRNAs (Figure 2-9B). While it is possible that a small amount of U1 snRNA is present in the purification (Figure 2-9B), since no U1 snRNP components were found in the 1D LC-MS/MS analysis (Table 2-2) this interaction would have to be sub-stoichiometric. Thus, the NTAP-NTE is able to

	S. pombe	ORF number	TAP-NTE *cwf10-ΔNTE*		Cdc5-TAP wild-type		*S. cerevisiae*	*H. sapiens*
			I	II	I	II		
Core snRNP	Smb1	SPAC26A3.08	4	3	16	6	Smb1	SMB/B'
	Smd1	SPAC27D7.07c	1	3	12	4	Smd1	SMD1
	Smd2	SPAC2C4.03c	6	3	15	5	Smd2	SMD2
	Smd3	SPBC19C2.14	3	1	15	5	Smd3	SMD3
	Sme1	SPBC11G11.06c	1	1	18	8	Sme1	SME1
	Smf1	SPBC3E7.14	5	3	11	6	Smf1	SMF1
	Smg1	SPBC4B4.05	---	3	22	9	Smg2	SMG1
U2	Lea1	SPBC1861.08c	9	3	25	14	Lea1	U2A'
	Msl1	SPBC8D2.09c	1	---	10	5	Msl1	U2B"
	Prp10	SPAC27F1.09c	---	---	8	---	Hsh155	SF3B160
U5	Cwf10-ΔNTE	SPBC215.12	60	42	---	---	Snu114	U5-116
	Cwf10	SPBC215.12	---	---	116	54	Snu114	U5-116
	Brr2	SPAC9.03c	---	---	4	---	Brr2	U5-200
	Spp42	SPAC4F8.12c	57	33	221	89	Prp8	U5-220
	Cwf17/Spf38	SPBC1289.11	9	4	53	21	UNK	U5-40
NTC core	Cdc5	SPAC644.12	34	13	99	50	Cef1	CDC5
	Cwf2	SPAC3A12.11c	13	5	52	20	Cwc2	RBM22
	Cwf7	SPBC28F2.04c	10	6	42	13	Snt309	SPF27
	Cwf15	SPBC337.06c	6	1	23	13	Cwc15	AD002
	Prp5	SPBP22H7.07	14	8	74	26	Prp46	PRL1
	Prp19	SPAC29A4.08c	18	17	93	43	Prp19	PRP19
NTC-associated	Cwf3	SPBC211.02c	26	9	87	36	Syf1	SYF1
	Cwf4	SPBC31F10.11c	21	11	63	35	Clf1	CRN1
	Cwf5/Ecm2	SPCC550.02c	12	10	64	21	Ecm2/Slt11	RBM22
	Cwf12	SPBC32F12.05c	6	6	25	8	Isy1	ISY1
	Prp45	SPCC188.11	3	3	52	23	Prp45	SKIP
	Prp17	SPBC6B1.10	17	15	51	27	Prp17	hPRP17
	Syf2	SPBC3E7.13c	10	5	16	15	Syf2	SYF2
Other	Cwf11	SPBC646.02	36	17	142	55	UNK	hCWF11
	Cwf14	SPBC24C6.11	8	4	14	8	Cwc14	G10
	Cwf18	SPCP1E11.07c	6	5	13	9	UNK	MGC23918
	Cwf19	SPAC30D11.09	11	3	63	24	UNK	hCWF19L2
	Cwf21	SPAC4A8.09c	1	---	8	1	Cwc21	SRm300
	Cwf22	SPBC13E7.01	2	1	20	6	Cwc22	hCWC22
	Cyp1	SPAC57A10.03	5	1	12	4	UNK	PPIL1
	Mug161	SPAC1F3.09	---	---	8	1	YGR093W	CWF19L1
	Prp16	SPBC1711.17	---	---	---	---	Prp16	PRP16
	Prp43	SPBC16H5.10c	---	---	3	3	Prp43	hPRP43
	Prp22	SPAC10F6.02c	---	---	6	---	Prp22	hPRP22
	Saf4	SPBC18H10.10c	3	3	13	4	UNK	CCDC130
	Sum3	SPCC1795.11	7	1	1	3	Ded1	DDX3
	unnamed	SPAC20H4.09	---	---	5	3	UNK	DHX35

Table 2-2. Splicing proteins co-purifying with NTAP-NTE in a *cwf10-ΔNTE* background and Cdc5-TAP in a wild-type background. Two biological replicates were analyzed using one-dimensional (1D) liquid chromatography-tandem mass spectrometry (LC-MS/MS) per genotype (I and II). Peptide counts are included for each replicate. UNK indicates that an ortholog is unknown or not present. "---", no peptides identified.

bind to a surface(s) of the U5.U2/U6 core without being covalently linked to the EF2-like portion of Cwf10 (residues 128-983). Additionally, this suggests that the protein(s) and/or RNA(s) that create the binding surface(s) for the Cwf10 NTE are likely present in the purified complex of 32 detected spliceosome components.

Notably, peptides for Brr2, a core U5 snRNP component, were inconsistently found in the Cdc5-TAP purifications at both 75 mM and 150 mM salt and were not found in either of the Cdc5-TAP or NTAP-NTE purifications from *cwf10-ΔNTE* cells (Table 2-2). To more closely examine this result, we asked whether NTAP-NTE interacts with Brr2 by co-immunoprecipitation. To this end, either *NTAP* or *NTAP-NTE* was overexpressed in a *brr2-3XHA* strain. Immunoprecipitations were performed using IgG-Sepharose beads followed by immunoblotting with anti-HA antibodies to detect Brr2-HA. Using this approach, NTAP-NTE pulled down Brr2-HA, while NTAP alone did not (Figure 2-10). However, one caveat is that this experiment does not replicate the two-step TAP protocol that was used for purifying sample for the mass spectrometry analysis. Thus, in aggregate these results show that Brr2 is likely present at only sub-stoichiometric levels in both NTAP-NTE and Cdc5-TAP purifications, with one explanation being that it is not as tightly associated with the U5.U2/U6 complex as other U5 snRNP components.

Because Cwf10-NTE co-purified with spliceosomal components, we next asked whether the domain could restore pre-mRNA splicing efficiency in *cwf10-ΔNTE*, *i.e.* function *in trans*. To test this, we performed semi-quantitative RT-PCR on RNA extracted from *cwf10-ΔNTE* cells over-expressing either NTAP or NTAP-NTE and looked at the splicing efficiency of two introns. Analysis of the *tbp1_a* and *mrps16_b*

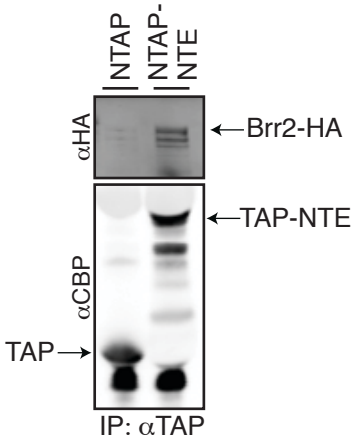


FIGURE 2-10

NTAP-NTE co-immunoprecipitates Brr2-HA
pREP41 NTAP or pREP41 NTAP-NTE were expressed in *cwf10-ΔNTE brr2-HA* cells. NTAP and NTAP-NTE were immunoprecipitated with IgG-Sepharose beads and then immunoblotted with anti-HA antibodies. Positions of proteins on the immunoblot are labeled.

introns revealed that overexpression of the NTAP-NTE in the *cwf10-ΔNTE* background improved the mature/pre-mature ratio over the NTAP control by ~25% and 30% respectively, returning splicing efficiency close to wild-type levels (Figure 2-9 C-E). Although only semi-quantitative, the statistical significance of these data supports a model in which the NTE can incorporate into spliceosomal complexes independent of any covalent connection with the C-terminal EF2-like portion of Cwf10, partially rescuing the splicing deficiency seen in *cwf10-*Δ*NTE* cells.

Given the previous result that *cwf10 2-127*Δ *(cwf10-*Δ*NTE*) and *cwf10 2-23*Δ have almost identical accumulation of the *tbp1_a* intron (Figure 2-7G and H), we wondered whether overexpressing the first 26 amino acids of the NTE alone would be sufficient to restore the pre-mRNA splicing efficiency of *cwf10-*Δ*NTE*. Therefore, we repeated the RT-PCR experiment, using RNA extracted from *cwf10-ΔNTE* cells over-expressing either NTAP or NTAP-Cwf10 2-26. Although Western blotting confirmed the expression of the NTAP-Cwf10 2-26 protein (data not shown), the smaller region failed to complement deletion of the entire NTE (Figure 2-9 F-H). Thus, the conserved region of amino acids 2-23 is necessary for splicing but not sufficient to complement *cwf10-ΔNTE* in *trans*.

Discussion

Deletion of the NTE and its effects on transcripts

In this study, we have investigated the function and structural characteristics of the Cwf10 NTE. Although this domain is not essential in fission yeast, deletion of the

NTE causes a general splicing defect, a change in the sedimentation patterns of *in vivo* spliceosomal complexes, and synthetic lethal and synthetic sick interactions with mutant alleles of other pre-mRNA splicing factors.

Because *cwf10*-Δ*NTE* cells show robust growth at all temperatures, we were surprised to detect a pre-mRNA splicing defect and postulated that the lack of the NTE domain may affect the splicing of a specific subset of pre-mRNAs rather than cause a global pre-mRNA splicing defect. To test if this was the case, we used deep sequencing analysis to comprehensively determine which transcripts are affected by this mutation. Interestingly, most introns show a tendency toward reduced splicing efficiency (all points to the right of the diagonal in Figure 2-3A), and the reduced splicing is statistically significant for about 44% of total introns. It is not surprising, then, that there is no obvious type of transcript (i.e. specific splice sites, number of introns, and/or size of introns) that is specifically affected. Thus, the splicing defect in *cwf10*-Δ*NTE* cells appears to be global in nature and widespread in scope. This suggests that the NTE is important for a general (not transcript-specific) process in the splicing reaction.

Because functions have not been assigned for any sub-region of the NTE, we removed only the first twenty-three amino acids from the NTE and found a similar splicing defect between *cwf10*-Δ*NTE* and *cwf10 2-23*Δ. We speculate that loss of this small, highly-conserved, acidic region undermines the NTE's capacity to bridge interactions or support conformational rearrangements in the spliceosome critical for the NTE's function (see "Model" section below).

The NTE's roles in the splicing cycle

A previous study using *Sc Snu114ΔN* demonstrated a role for the Snu114/Cwf10 NTE in U4/U6 snRNA unwinding (Bartels, Klatt et al. 2002). However, a study using human splicing extracts and antibodies directed against the hU5-116K NTE suggests that the NTE may also be involved in the first-to-second step transition (Fabrizio, Laggerbauer et al. 1997). This NTE function, uncovered in the human system, is further supported by negative genetic interactions (Frazer, Lovell et al. 2009) between *Sc Snu114ΔN* and both a substitution at snRNA U6-A59, which inhibits the second step of splicing (Fabrizio and Abelson 1990), and several alleles of the U5 loop 1, which helps position the exons for ligation in the second step (Newman and Norman 1992, O'Keefe and Newman 1998).

Our genetic findings are consistent with a role for the NTE in both spliceosome activation and second step catalysis. *cwf10-ΔNTE* interacts with several alleles of genes known to be involved in spliceosome activation (Figure 2-6), including *spp42-1* (*Sc PRP8*), *prp1-4* (*Sc PRP6*), and *cdc5-120* (*Sc CEF1*). Additionally, the *cdc5-120* mutation, with which *cwf10-ΔNTE* is synthetically lethal, may play an additional role in second step chemistry. First, *cdc5-120* is a point mutation in one of the two conserved Myb-repeats (Ohi, Feoktistova et al. 1998), a region important for first step to second step modulation in ortholog *Sc* Cef1 (Query and Konarska 2012). Second, *cdc5-120* is lethal with *S. pombe prp17Δ* (Ren, McLean et al. 2011), a known second-step splicing factor in *S. cerevisiae* (Jones, Frank et al. 1995). Additionally, a role in the second step is supported by the strong negative interaction between *cwf10-ΔNTE* and *prp17Δ* (Figure 2-6A). The fact that the NTE does not immunoprecipitate a pre-activation spliceosome (Table 2-2),

but rather a complex similar to U5.U2/U6, could indicate the NTE becomes stably "locked" to the spliceosome following activation and is positioned near the catalytic core to help modulate the first to second step transition.

Alterations in sizes of snRNA-containing complexes

Analysis of snRNAs in cells lacking the NTE shows that loss of this domain changes the sedimentation patterns of spliceosomal complexes. Although all the snRNA sedimentation patterns show some degree of change in the Δ*NTE* background (Figure 2-5 E-N), for both the U1 and U5 snRNAs there is a shift away from higher molecular weight fractions into lower molecular mass fractions (<11.3S, Figure 2-5H and L). For the U5 snRNA, this shift is too high in the gradient (<11.3S, Figure 2-5H) to contain the full complement of U5-snRNP specific proteins, as both human and *S. cerevisiae* U5 snRNPs sediment at 15-20S when bound to the U5-specific proteins (Bach, Winkelmann et al. 1989, Neubauer, Gottschalk et al. 1997). Similar results with the *Sc snu114*Δ*N* allele led the authors to speculate that the NTE is required for U5 snRNP stability (Bartels, Klatt et al. 2002), and our results are consistent with that hypothesis. For the U1 snRNA, it is unclear whether this lower molecular weight sedimenting fraction represents free U1 snRNA or the snRNA complexed with at least some of the U1 snRNP proteins.

Sucrose gradient analysis also revealed that a small fraction of the U2, U5, U6, U4 snRNAs, as well as Cwf10, shifted to high molecular mass fractions that are more pronounced in *cwf10*-Δ*NTE* (Figure 2-5B, F, H, J, and N). The presence of this peak at fractions 11/12 could suggest that pre-activation spliceosomes are stalled and

accumulating. Alternatively, this peak could represent a different multi-snRNP complex or aggregates of spliceosome components that are not able to organize properly.

Although the sedimentation data support a model in which the *in vivo cwf10-ΔNTE* splicing complexes that are slow to activate or complete catalysis may be accumulating or aggregating, other possible fates for stalled spliceosomes include degradation by the proteasome and disassembly. The lower steady-state levels of Cdc5 and Prp1-myc (Figure 2-2I), which sediment mainly as part of large complexes in *cwf10-ΔNTE* (Figure 2-5C and D and data not shown), could favor the hypothesis that these slow-to-splice complexes are degraded.

The NTE's structure

Although the NTE was predicted to be intrinsically unstructured in solution (Korneta, Magnus et al. 2012), our modeling and biophysical analysis unexpectedly showed that approximately half of NTE residues are in a folded environment in solution. Because the DISOPRED program (Figure 2-7A) predicted disorder to exist mostly in the N-terminus (AA 18-61) of the NTE, and the ordered amino acids to exist at the extreme N-terminus (AA 1-17) and the C-terminus (62-120), it is tempting to begin thinking of the NTE as three sub-regions with separate characteristics. The extreme N-terminus, which is not predicted to be disordered, is required for function, since deleting 23 conserved residues in the N-terminus fully recapitulates the splicing defect seen when the entire NTE is deleted (Figure 2-7G and H). However, this small region cannot complement the *cwf10* mutant lacking the NTE *in trans*, suggesting that the C-terminal NTE regions are also important. The middle region of the NTE, AA 18-61, which is

predicted to be disordered, contains many acidic residues. The acidic charge of the NTE is likely essential since we were unable to integrate a *cwf10* mutant that had all negatively-charged residues in AA 1-61 replaced with alanine residues (data not shown). The proposed third region of the NTE, amino acids 62-120, is predicted to be structured and links the N-terminal NTE regions to the "EF2-like" portion of the protein. Further studies are needed to delineate the boundaries of order and disorder within the NTE, as well as to determine how these regions may interact with each other.

The unstructured region(s) of the NTE may become folded in the context of the dynamic spliceosome when it comes into contact with a binding partner(s). This idea has precedent in the pre-mRNA splicing machinery. A 70 amino acid intrinsically disordered region of NTC-associated protein hSKIP undergoes a disordered-to-ordered transition upon binding to cyclophilin PPIL1 (Wang, Zhang et al. 2010). Similarly, a 31 amino acid, predominantly random coil region of human U4/U6-60K adopts structure when bound to U4/U6-20K (another cyclophilin) (Reidt, Wahl et al. 2003). However, the disordered regions of hSKIP or U4/U6-60K do not share the same high acidic content of the proposed disordered middle region of the NTE, AA 18-61. The determination of binding partners for the NTE will pave the way for experiments to examine whether order-to-disorder transitions occur within this domain.

Potential binding environment in the spliceosome

When overexpressing *NTAP-NTE* (Cwf10 residues 2-135) in *cwf10-ΔNTE* cells, we found this fragment has the capacity to incorporate into a spliceosomal complex that is similar in composition to the *S. pombe* U5.U2/U6 complex. From this observation, we

conclude that the Cwf10-NTE recognizes and interacts with other spliceosomal components independently of its polypeptide linkage to the main Cwf10 "EF2-like" portion. What does the NTE bind to in the context of the late-stage spliceosome? Likely, the binding site does not involve Brr2, because it is not present at stoichiometric amounts in the purification (Table 2-2), although other U5-specific proteins Spp42 (*Sc* Prp8), Spf38 (human U5-40), and the EF2-like portion of Cwf10 are highly represented. Therefore, the binding partner(s) is likely one of the above U5-specific proteins, the U5 Sm core, U2 snRNP protein Lea1, or one or more of the NTC and NTC-related proteins. Additionally, the specific genetic interactions of *Sc snu114ΔN* with snRNA alleles (Frazer, Lovell et al. 2009) may also support an orientation of the NTE near U5 loop 1 and U6.

Importantly, NTE binding partners may change as protein-protein, protein-RNA, and RNA-RNA rearrangements occur during the splicing reaction. The combination of structured and unstructured regions within the NTE may allow the domain to bridge transient conformational states adopted by the spliceosome as splicing occurs. However, regardless of exact mechanism, the presence of the Cwf10 NTE *in trans* is sufficient to partially rescue the splicing defect seen in *cwf10-ΔNTE* cells. There are other spliceosome proteins that can function when their domains are expressed separately, highlighting the multiple interactions between spliceosome components. Examples include the large, 280kDa *Sc* Prp8 (Boon, Norman et al. 2006) and first-step factor *Sc* Yju2 (Chiang and Cheng 2013).

Model of Cwf10 NTE interactions in the spliceosome

Based on our data and the work of others outlined above, we propose a model of NTE function that includes (a) stabilization of the U5 snRNP structure and (b) facilitation of specific conformations among U5 snRNP components or between the U5 snRNP and other spliceosome proteins/RNAs. The NTE is capable of independently incorporating into the spliceosome and likely makes contact with more than one splicing component through its predicted structured and unstructured regions. Although we were not able to determine specific binding partners for the Cwf10 NTE, these partners must be in the vicinity of the EF2-like portion of Cwf10. It is tempting to speculate the these contacts promote cohesion within the U5 snRNP, and may transmit a signal or form a stabilizing bridge, first within the context of the pre-activation Complex B, then near the catalytic center of Complex C.

CHAPTER III

RANDOM MUTAGENESIS OF SPLICEOSOMAL GTPase CWF10

Introduction

While great advances have been made in the past decades toward deciphering the pre-mRNA splicing process, much remains to be learned, particularly about the specific interactions that exist among the spliceosome's numerous polypeptides and snRNAs. To this end, structural studies have determined atomic-level structures of individual domains of pre-mRNA splicing factors using x-ray crystallography and defined the general shapes of spliceosome intermediates through single-particle electron microscopy (EM) (reviewed in Zhang, Li et al. 2013). Indeed, one such intermediate, the U5.U2/U6 complex from *S. pombe*, was characterized using EM in our lab (Ohi, Ren et al. 2007). However, the resolution achieved in these EM density maps needs to be improved from the current ~30 Å to better than 10 Å in order to begin observing specific interactions among components.

In this study, we investigate whether more homogenous spliceosome populations can be purified from *S. pombe*, with the aim of achieving higher-resolution EM models of these spliceosomes. Our approach was to purify complexes from cells that have a block in splicing progression. Heat- and cold-sensitive yeast harboring mutations in spliceosome components often block *in vitro* splicing reactions at a particular step, and recently such mutants have been used to biochemically and structurally analyze spliceosome intermediates (Warkocki, Odenwälder et al. 2009, Lardelli, Thompson et al.

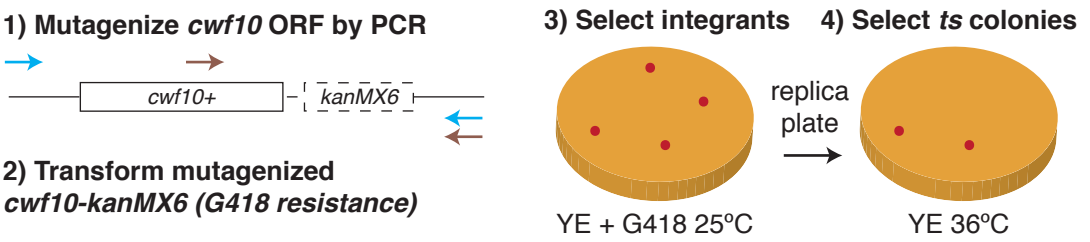
2010). Toward this end, we used a random mutagenesis strategy to produce multiple *cwf10* mutant strains. Cwf10 is the EF2-like GTPase component of the U5 snRNP that, as discussed in detail in Chapter I, occupies a position near the splicing catalytic site and has the capacity to favor or inhibit certain structural transitions through its nucleotide-binding status. We herein characterize the cellular and splicing phenotypes of the mutant strains and biochemically analyze changes in the spliceosome complexes present in cell lysates. These data suggest that our *cwf10* mutant strains contain destabilized multi-snRNP complexes, as marked by the pre-mRNA splicing factors Cdc5 and Prp1, and thus are not useful tools for isolating complexes for EM analysis.

Results

Nine ts alleles are isolated by random mutagenesis

In order to isolate *cwf10* ts mutants, we employed a random mutagenesis strategy of *cwf10*$^{+}$*::kanMX6* PCR fragments (Figure 3-1A). The fragments were amplified using error-prone PCR and transformed into yeast carrying wild-type *cwf10*$^{+}$. Integrants were isolated from YE G418 plates at 25°C, and ts colonies were identified by their failure to grow on YE at 36°C. All together, nine distinct ts alleles were isolated (*cwf10-2* through *cwf10-10*). Each mutated ORF contained on average 6.0 base pair changes, which caused 4.1 amino acid changes. Three different mutagenesis conditions were used to produce these nine alleles, and we wondered whether the mutation rate was responsive to the error-prone PCR conditions used. Three (3) mutants were isolated by using 15 cycles of error-prone PCR of the entire gene (3,093 bp; blue arrows Figure 3-1A), and these

A



B

Observed frequently (7-11 times)	Observed rarely (0-2 times)
A—G	A—C
A—T	C—A
G—A	C—G
	C—T
T—A	
T—C	G—C
	G—T
	T—G

C

Name	NTE	Domain I	Domain II	Domain III	Domain IVa/b	Domain V
Cwf10-1[a]		C323Y				
Cwf10-2		V161I, K172R, S362L		M596K	N711I	S879T
Cwf10-3		L266P	E525D		D752V	
Cwf10-4					G716E, D941G	E848K, D854N, G878V
Cwf10-5					T682A, R726G, L804Q, D923A, V980E	
Cwf10-6					E795V	C894G, V902E
Cwf10-7				H640L, E665G	S689T, L983S	G895E
Cwf10-8			G517C	I593T	C676R, M700V	
Cwf10-9					E703V, G818D	S893Y
Cwf10-10	E116D	M230K				Y882F

[a] allele described in Bayne et al. (2008)

D

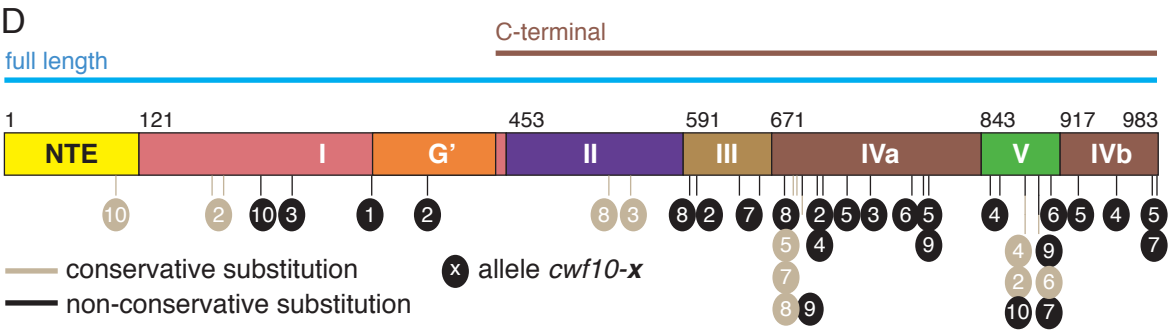


FIGURE 3-1

***cwf10* ts alleles generated by random mutagenesis contain multiple mutations**

A) A kanMX6 cassette (dashed lines and box) conferring G418 resistance was introduced 3' of the cwf10 gene at its endogenous locus. Following integration, the cwf10-kanMX6 sequence was amplified by PCR with primers 500 bases upstream and downstream (blue arrows). This PCR amplicon served as a template to mutagenize the full length (blue arrows) or C-terminal (red arrows) cwf10 ORF at random sites using Taq polymerase (see Methods). Mutagenized PCR was transformed in to yeast with an unmodified cwf10+ locus, and integrants were selected on G418 media at 25°C. After replica plating to rich media at 36°C, temperature sensitive (ts) strains were isolated. B) Table grouping nucleotide substitutions found in alleles cwf10-2 through -10 by their frequency of occurrance. C) Table listing the exact amino acid substitutions found in each cwf10 allele grouped by domain. D) Linear representation of the domains of Cwf10, as defined by homology to EF2. Mutations are represented as vertical lines underneath the position of the amino acid, with the Cwf10 allele carrying the mutation (-1 through -10) encircled below. Conservative amino acid substitutions are indicated in peach, while non-conservative substitutions are indicated in black. The G' subdomain is AA 328-435. "N" indicates the N-terminal domain. The two mutagenesis target fragments - full length and C-terminal - are depicted above the domains.

mutants had an average of 2.3 mutated bases/kb. Two (2) mutants were isolated from 8 cycles of PCR covering the entire gene, with an average of 1.8 mutated bases/kb. The remaining four (4) mutants were isolated from transformation of 15 cycles of the 3' end of the gene (1,100 bp; red arrows, Figure 3-1A), and these mutants had an average of 5.0 mutated bases/kb. These results indicate a mutation rate dependent on the PCR conditions and a trend toward more mutations with more error-prone PCR cycles. The specific base pair changes were also analyzed to determine the biases present in this protocol. Specifically, five of the sixteen possible base pair changes were over-represented, while seven of the sixteen occurred infrequently (Figure 3-1B). While these data may be influenced by the requirement that the set of mutations resulted in a viable allele in order to have been sequenced, the information may still be useful when applying this method to other genes to determine error-prone PCR parameters and predict base changes.

The amino acid substitutions associated with each *cwf10* ORF are listed in Figure 3-1C. We created a visual representation of these substitutions indicating (a) whether the substitution was conservative or non-conservative and (b) in which domain the mutation occurred (Figure 3-1D). The mutated amino acids are spread throughout all domains of the protein, although there is a clear preference for accumulation of mutations toward the C-terminus (domains IV and V). This is likely due in part to the four mutants that were generated by mutating only the C-terminus (*cwf10-4* through -*7*; red line, Figure 3-1D). A previously described ts point mutant, *cwf10-1* (Bayne, Portoso et al. 2008), is also diagrammed in Figures 3-1C and 3-1D and is phenotypically analyzed in following sections.

cwf10 mutants display defects in growth, splicing, and silencing of heterochromatin transcription

As expected from alleles harboring different mutations, the mutants displayed a range of growth rates on solid media at 18°C and 32°C, with all displaying cell death at 36°C (Figure 3-2A). The ts phenotype was further investigated in a subset of alleles by quantifying cell growth in liquid cultures. All mutants displayed negligible growth compared with wild-type following placement in 36°C YE media (Figure 3-2B). We wanted to confirm that the ts phenotypes were truly caused by mutations in the *cwf10* ORFs. To investigate this, we expressed a high-copy vector containing the *cwf10* ORF in the context of ~500 bp native upstream and downstream chromosomal sequence, and expressed the vector in several mutants, including *cwf10-1* and three mutants isolated from the random mutagenesis screen. The wild-type pIRT2 $cwf10^{+}$ plasmid rescued growth of every *cwf10* mutant, but not of a *prp4* mutant (Figure 3-2C), indicating that these mutations are recessive and can be complemented by the $cwf10^{+}$ chromosomal sequence. *S. pombe* ts mutants can arrest in a variety of cell shapes and sizes, which often correlate with an interpretation of cell cycle defects (Fantes 1977, Potashkin, Kim et al. 1998). In every case examined, the *cwf10* ts mutants arrested as long cells, rarely with more than one septum, indicating a block in mitotic progression and subsequent cell septation (Figure 3-2D).

We next wanted to determine whether these cells had splicing defects. The presence of an increased splicing defect upon shift to 36°C would support the idea that cell death in these mutants is caused by a defect in the splicing machinery. RNA was extracted from each mutant grown at 25°C and 36°C, and semi-quantative reverse

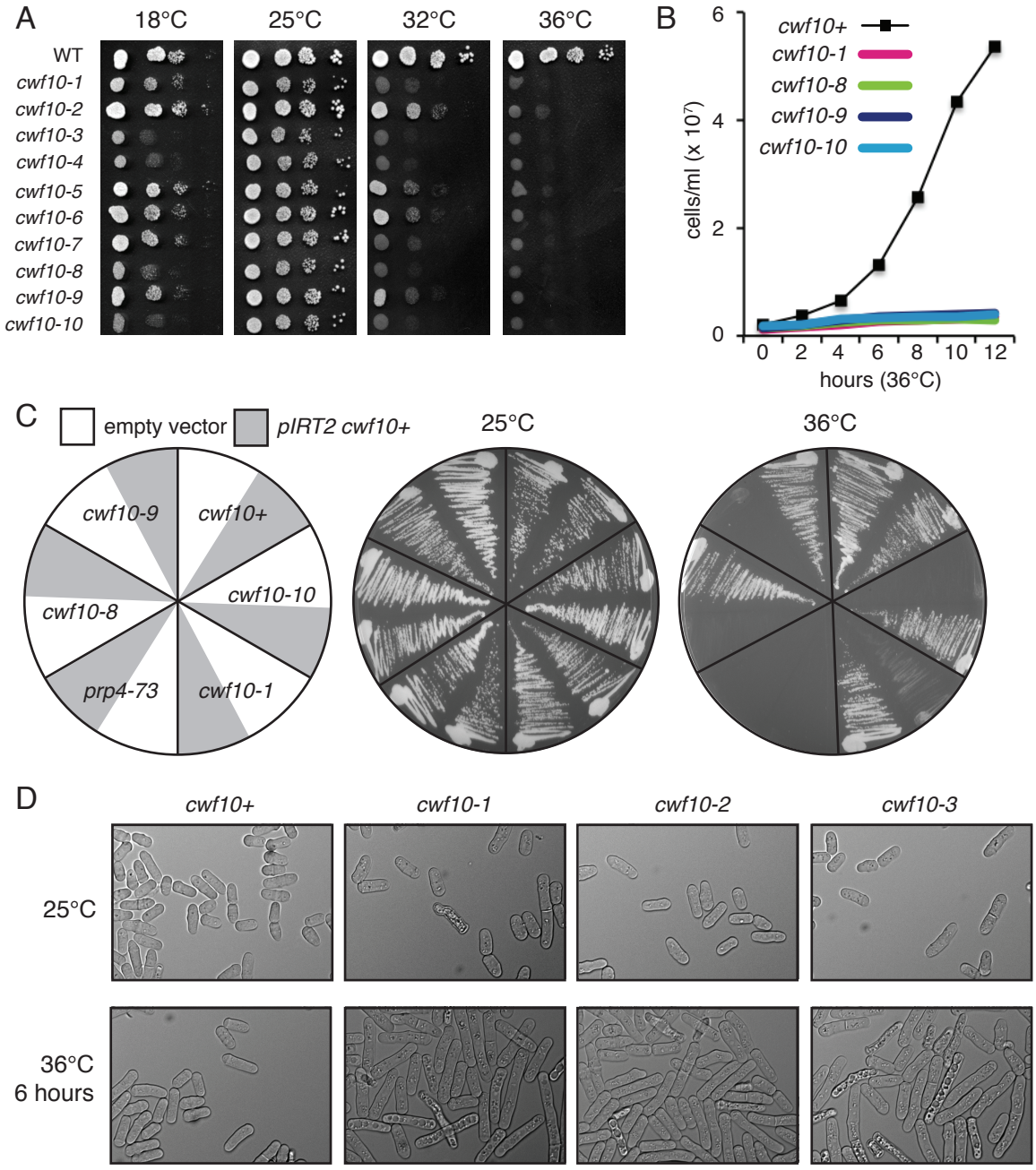


FIGURE 3-2

***cwf10* ts mutants exhibit cell growth and morphology defects**

A) The strains indicated were serially diluted and spotted on YE at the indicated temperatures to assay temperature-specific growth defects. B) Cells of the indicated strains were grown at 25°C and filtered in to 36°C rich media at an approximate concentration of $1.5x10^6$ cells/mL. Cells were counted every hour. C) The strains indicated were transformed with either pIRT2 or pIRT2 containing the cwf10 genomic sequence. Transformants were struck to MAU plates at the indicated temperatures. D) The indicated strains were grown in YE media at the indicated temperatures and imaged with brightfield microscopy.

transcriptase-PCR (RT-PCR) analysis of the first *tbp1* intron was performed (Figure 3-3A). ts alleles displayed splicing defects at 25°C, and the defect was exacerbated by the 36°C temperature shift. Interestingly, however, unlike in the temperature-shifted *prp2-1* and *prp4-73* mutants, for the *cwf10* mutants the mature (spliced) band remains strong at 36°C (Figure 3-3A). Other introns besides *tbp1_a* were not tested.

The connection between the splicing machinery and RNAi processes has been appreciated only recently (Dumesic, Natarajan et al. 2013, Tabach, Billi et al. 2013), and we wanted to investigate this connection using our *cwf10* mutants. Cwf10 has previously been shown to co-isolate with RNAi processing factors, and the *cwf10-1* allele alleviates silencing of genes located in the centromeric heterochromatin (Bayne, Portoso et al. 2008), which, in fission yeast, is maintained in part by RNAi (Buhler and Moazed 2007, Grewal and Jia 2007). We asked if our *cwf10* alleles also had gene silencing defects, using RT-PCR amplification of the *cen-dg* heterochromatic transcripts, which are normally silenced. Indeed, at the permissive temperature of 25°C, all *cwf10* alleles displayed accumulation of *cen-dg* RNAs in abundance similar to deletion of the known chromatin silencing factor *swi6* (Figure 3-3B). This defect is not observed in splicing factor mutants *prp2-1* and *prp4-73*.

cwf10 mutations cause changes to spliceosomal complexes

Because our ultimate goal was to use these ts alleles to isolate a homogenous splicing complex, we began biochemically probing spliceosomes in these mutants. Alleles *cwf10-1*, *-8*, *-9*, and *-10* were chosen to investigate further because their non-conservative amino acid substitutions lie in different domains of Cwf10 (Figures 3-1C

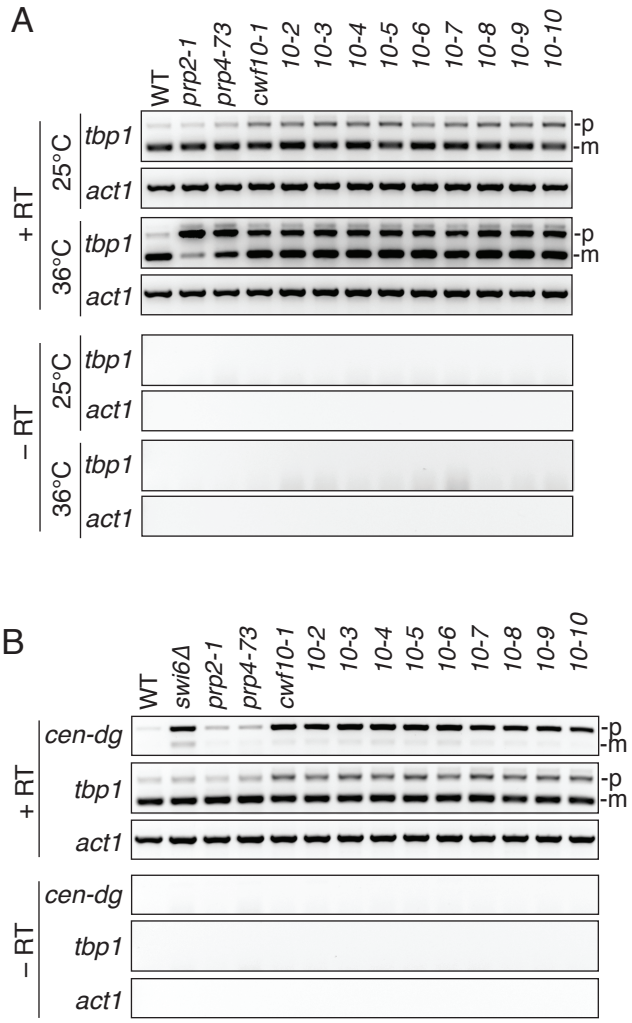


FIGURE 3-3

***cwf10* ts mutants exhibit splicing defects and transcription from centromeric heterochromatin**

A) Total RNA was isolated from cells grown at 25°C (permissive temperature) or grown at 25°C and shifted to 36°C for six hours (restrictive temperature). RT-PCR was performed with random hexamer primers. From the RT-PCR reactions, portions of the tbp1 (containing introns) and act1 (intronless) transcripts were amplified by PCR. Mature (m) and pre-mRNA (p) forms are indicated. No bands were detected in any RT-PCR samples in which the reverse transcriptase enzyme was omitted (-RT). B) Accumulation of centromeric transcripts in cwf10 mutants. Total RNA was isolated from cells grown at 25°C. RT-PCR was performed with an oligo(dT)12–18 primer, and tbp1 (containing introns), cen-dg (centromeric), or act1 (intronless) transcripts were amplified from RT-PCR products. Mature (m) and pre-mRNA (p) forms are indicated. No bands were detected in any RT-PCR samples in which the reverse transcriptase enzyme was omitted (-RT).

and 3-1D). Cwf10-1 contains a substitution in domain I, Cwf10-8 in domains III and IVa, Cwf10-9 in domains IVa and V, and Cwf10-10 in domain V. Because mutations in different *S. cerevisiae* Snu114 domains have different biochemical effects on spliceosomes (Brenner and Guthrie 2006), we reasoned that these different alleles would possibly impair splicing progression uniquely.

We first investigated the density gradient sedimentation of two marker spliceosome proteins, Prp1 and Cdc5. Prp1 associates with a group of proteins that resembles a pre-activation spliceosome (Carnahan, Feoktistova et al. 2005), while Cdc5 immunoprecipitates exclusively U5, U2, and U6 snRNAs and proteins reminiscent of a post-activation spliceosome (Ohi, Link et al. 2002). Although we have never observed the sedimentation pattern of Cdc5 to change in any spliceosome mutant background ((Ohi, Link et al. 2002) and data not shown), Prp1-myc readily redistributes to <19S in a two hour *prp2-1* arrest (data not shown). The mutant *cwf10* strains were arrested for six hours at 36°C before loading the lysates on density gradients. We observed that the sedimentation pattern of Cdc5 was relatively unchanged in the mutants (Figure 3-4A, right blots). Prp1-myc did not shift to the top of the gradient as in *prp2-1*; however, the sedimentation pattern >37S appeared to be altered among the tested alleles (Figure 3-4A, left blots). These differences were difficult to quantify because the Western blot signal from Prp1-myc in the mutants was very low. We used lysates to calculate the levels of Prp1-myc and Cdc5 against levels of housekeeping protein Asp1, using near-infrared Western blotting (Figure 3-4B). Interestingly, the marker proteins were reduced 47-91% in each *cwf10* mutant shifted to 36°C for 6 hours. The reduced protein level of Cdc5 and its similar sedimentation profile in all mutant strains indicated that the remaining Cdc5

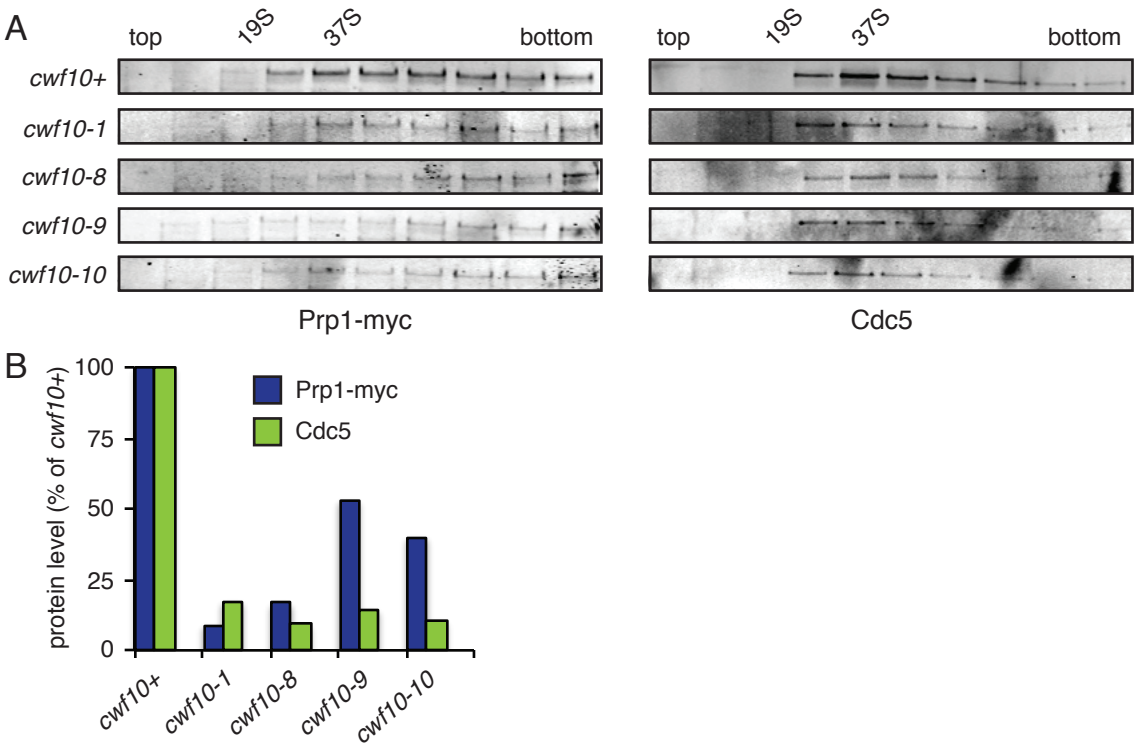


FIGURE 3-4

Spliceosome complex marker proteins Prp1 and Cdc5 become less abundant in *cwf10* ts arrests
A) Lysates from strains shifted to 36ºC for six hours were run over 10-30% sucrose gradients. Every other fraction was blotted for the indicated proteins. 19S and 37S were determined by the sedimentation of thyroglobulin and Cdc5, respectively. B) Lysates from strains shifted to 36ºC for six hours were blotted for the indicated proteins and quantified against each lysate's amount of housekeeping protein Asp1. The graph represents a single biological/technical replicate.

protein remained bound to its normal 37S complex. Likewise, the reduced protein level of Prp1-myc and possible shifts in its sedimentation indicated Prp1-myc-containing complexes are reduced in number, and these complexes may be altered in composition or conformation.

To further clarify the behavior of multi-snRNP complexes in the *cwf10* mutants, we ran lysates over density gradients and isolated the RNA from each fraction. Blotting for the U5 snRNA, which participates in both pre- and post-activation spliceosomes, we observed that U5 no longer sediments mainly in a ~40S peak (Figure 3-5A, fraction 9, and quantified in Figure 3-5D), but rather sediments below 11.3S in the *cwf10-1* and *-9* alleles. The sedimentation of U5 in the *prp2-1* arrest clearly shows the presence of a <11.3S peak (fraction 2) and a ~19S peak (fraction 5). The 19S peak is large enough to contain the U5-specific proteins, including Cwf10, while the <11.3S peak likely contains only the U5 snRNA with or without the Sm ring (Bach, Winkelmann et al. 1989, Neubauer, Gottschalk et al. 1997). This result indicates that these *cwf10* ts mutants cause (a) U5 snRNA to dissociate from high molecular weight complexes and (b) Cwf10 and other U5-specific proteins to dissociate from the U5 snRNA. Additional blotting for the U2 and U6 snRNAs (Figures 3-5B and 3-5C) and quantification in Figures 3-5E and 3-5F shows the redistribution of U2 from the peak at fraction 9 to two separate patterns in *cwf10-1* and *cwf10-9*, as well as the almost complete removal of U6 from high molecular weight fractions in these mutants. This, along with the reduced protein levels of Cdc5 (Figure 3-4B) suggests that U2, U5, and U6 no longer associate as a multi-snRNP complex at high levels and that splicing defects in these strains are likely due, at least in part, to lack of assembled splicing complexes.

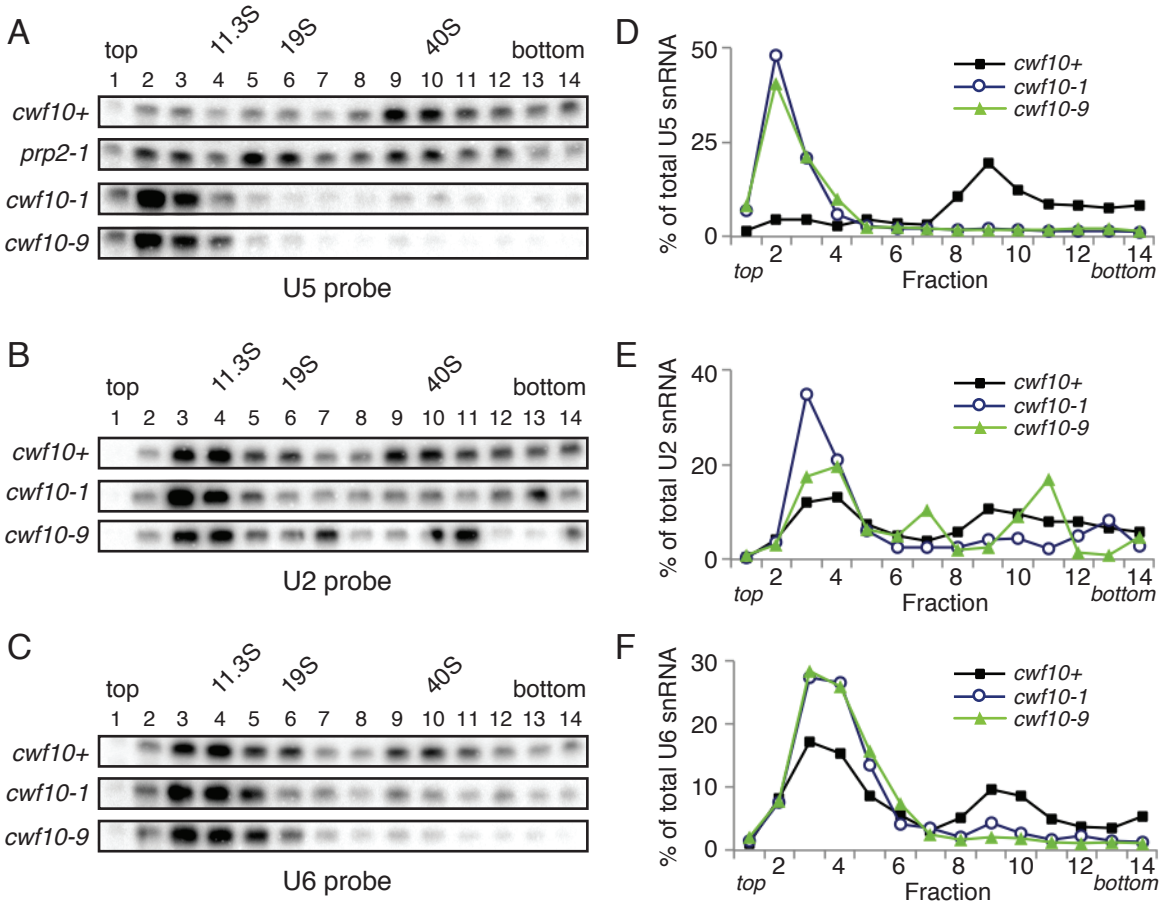


FIGURE 3-5

High molecular weight snRNA-containing complexes become less abundant in *cwf10* ts arrests
A) Lysates from strains grown at 25ºC (*cwf10+*), shifted to 36ºC for two hours (*prp2-1*), or shifted to 36ºC for six hours (*cwf10-1* and *-9*) were run over 10-30% sucrose gradients and RNA extracted from each fraction. Northern blot analysis using a U5 probe is shown. 11.3S, 19S, and 40S were determined by the sedimentation of catalase, thyroglobulin, and *S. cerevisiae* Fas2-V5, respectively. B) Samples are as in A, except *prp2-1* is not included. Band shown is the U2 snRNA. C) Samples are as in A, except *prp2-1* is not included. Band shown is the U6 snRNA. D) Band intensities from A were quantified and plotted as a percentage of the sum of the signal from all fractions. The bands were quantified by phosphorimaging and ImageQuant TL 8.1 software (GE Healthcare). E) Band intensities from B were quantified and plotted as a percentage of the sum of the signal from all fractions. F) Band intensities from C were quantified and plotted as a percentage of the sum of the signal from all fractions.

cwf10 mutations cause reduced levels of Cwf10 protein

Finally, we asked what happens to the stability of the mutant Cwf10 proteins. Using polyclonal Cwf10 antiserum developed for this project (characterized in Figure 2-2D), the levels of Cwf10 in all the mutants was determined to be significantly lower than wild-type, even at the permissive temperature of 25°C (Figure 3-6). Although it is not clear whether the proteins are inherently unstable, or if they are degraded in response to altered interactions with the spliceosome, these data led us to consider the possibility that the splicing phenotypes of the *cwf10* ts mutants were all attributable to destabilization of the Cwf10 protein structure, rather than to alterations in specific domains of the protein. Even if there were unique domain-specific effects being exerted on spliceosomes, the low levels of the Cdc5 and Prp1 marker proteins indicated that it would be difficult to isolate large quantities of these complexes to analyze by single-particle EM.

Discussion

In this study, random mutagenesis by PCR was used to generate ts alleles of *cwf10*. We generated a panel of nine mutants with different sets of mutations; however, it is unclear whether the mutations generally destabilize the protein or whether they inhibit splicing by mechanisms unique to each mutant. We have used this method to generate ts strains with other targeted spliceosome mutations, including *cwf5 S166P* (Livesay and Ohi, unpublished observation) and multiple *cdc28* mutants (Takizawa and Ohi, unpublished observations), although the same question remains of how the mutations exert the splicing defect. The mutagenesis method is not suitable for every gene, as

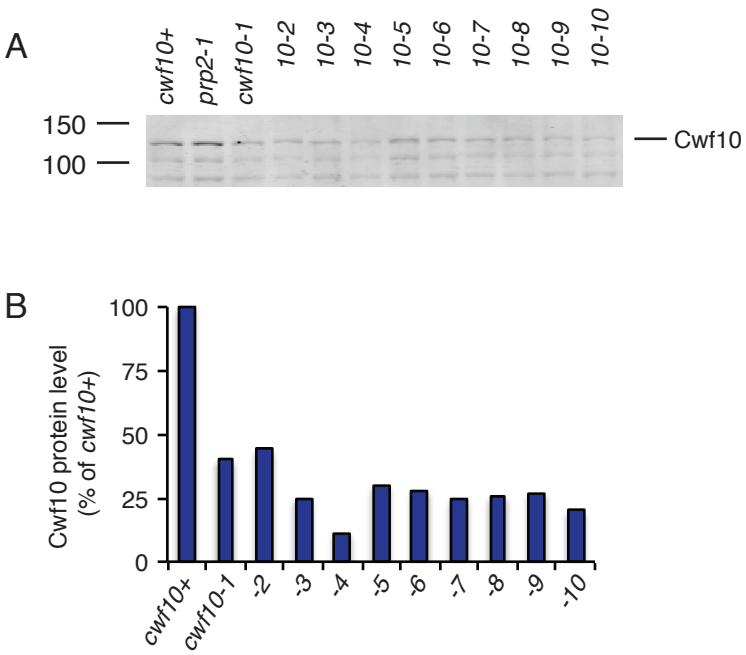


FIGURE 3-6

***cwf10* ts mutants have reduced levels of Cwf10 at the permissive temperature**
A) Representative blot of lysates from strains grown at the permissive temperature (25ºC) blotted for Cwf10. B) Quantification of Cwf10 against each lysate's amount of housekeeping protein Asp1. The graph represents the average of two technical replicates.

demonstrated by our multiple unsuccessful attempts to generate a *prp19* ts strain. Another caveat of this approach is that the mutants must have properties that allow them to be screened by replica plating, such as sensitivity to temperature or a drug. It is notable that all of the ts mutants we isolated (*cwf10-2* through *-10*) contained multiple mutations. Since it has been demonstrated that a single mutation ts *cwf10* strain can be generated (*cwf10-1*, recovered from EMS-treated cells) (Ekwall, Cranston et al. 1999), we hypothesize that parameters of the mutagenesis PCR reaction could be modified to favor recovery of single-mutation alleles.

cen-dg transcript accumulation in *cwf10* mutant cells (Figure 3-3B), and previous work on *cwf10*'s role in RNAi processing (Bayne, Portoso et al. 2008) beg the question of how Cwf10 protein is required for these processes. The current understanding of spliceosome interaction with RNAi machinery places the two processes in competition (Dumesic, Natarajan et al. 2013). That is, weak splicing substrates, which do not undergo the kinetic mechanisms of forward splicing as quickly as strong substrates, are more likely to be processed by the RNAi machinery. When the spliceosome is stalled on the substrate, the RNAi machinery has opportunity to access and process the transcript. It is not immediately apparent, then, how a splicing mutant that, presumably, would cause additional stalling in splicing, would cause inefficient RNAi processing. One idea is that Cwf10 is involved in the association of the RNAi machinery with the spliceosome. If Cwf10 is not able to recruit the RNAi machinery efficiently, then the transcript will not be processed by the RNAi machinery, even if splicing is stalled. Another idea is that the spliceosome does not actually form correctly on the transcript, thus not recruiting the RNAi machinery. We see that the U5 snRNP is largely disassembled in the mutant

(Figure 3-5A), possibly indicating that Complex B cannot form from Complex A already assembled on the transcript, because U5 snRNP and thus the U5.U4/U6 tri-snRNP are not available to bind.

It is intriguing to note how the snRNAs have dramatically altered sedimentation patterns in the *cwf10* mutant arrests while the spliceosome proteins Cdc5 and Prp1 have relatively unaltered sedimentation patterns (Figure 3-4A). Upon temperature shift of the *cwf10* mutants, several of the snRNA signals shift mainly to low molecular weight fractions (Figure 3-5 D-F), while Cdc5 and Prp1-myc remain mostly visible in higher molecular weight fractions. One explanation is that Cdc5 and Prp1 are somehow degraded when their complexes are disassembled, while the snRNAs (probably in complex with the Sm/Lsm proteins) remain stable. We also considered that transcripts of *cdc5* and *prp1* are not spliced properly, and thus less protein is produced; however, the *prp1* transcript contains no introns. It is possible that Cdc5 and Prp1 are vulnerable to degradation as unfolded proteins when they are not in complexes; another intriguing possibility is that a specific component(s) of a stalled spliceosome is targeted by ubiquitination for proteasomal degradation. While ubiquitination is a known process in the splice cycle (Bellare, Small et al. 2008, Song, Werner et al. 2010), no targeted proteasomal degradation process has been described.

Based on the low levels of Prp1-myc and Cdc5 protein in arrested *cwf10* mutant cells, we propose that these mutants are not good choices for isolating large quantities of homogenous multi-snRNP complexes for structural analysis. There are some other possible ways to use *cwf10* for structural analysis discussed below. One is the isolation of *cwf10* alleles that are not temperature sensitive. *cwf10-ΔNTE* cells, for example, grow

well at a range of temperatures and have high steady-state protein levels of Cwf10 ΔNTE (see Chapter II). Although that mutant's levels of Prp1-myc and Cdc5 are reduced, similar mutations may exist that slow splice cycle progression through a certain transition, yet maintain levels of that complex. Another possibility is overexpression of *cwf10* alleles. Preliminary results with overexpressed pREP3X *cwf10 E842I* (OHI plasmid 863) indicate that the allele causes cells to die, even though chromosomal *cwf10*$^{+}$ is also expressed. This dominant negative phenotype could indicate that Cwf10 E842I sequesters other spliceosome components in a stalled complex, reducing their availability to bind with wild-type Cwf10. It would be interesting to TAP-tag this overexpressed allele and examine co-purifying complexes for protein composition, snRNA composition, and types of splicing intermediates.

CHAPTER IV

CONCLUSIONS AND FUTURE DIRECTIONS

Snu114/Cwf10 plays several roles in pre-mRNA splicing as a U5 snRNP component and the sole spliceosomal GTPase. In this work, I have examined *Schizosaccharomyces pombe* Cwf10 through genetic, biochemical, and biophysical means to better understand its functions and mechanisms of action.

Chapter highlights

In chapter II, I explored the functional and structural characteristics of the N-terminal extension (NTE) domain of Cwf10. Deletion of the NTE led to a global splicing defect, suggesting the NTE is required for a general, not transcript-specific, process in the splicing reaction. The Cwf10 ΔNTE protein was stably expressed and retained the capacity to associate with high molecular weight complexes; however, there was an increased presence of Cwf10 ΔNTE over wild-type Cwf10 in some particularly large complexes of unknown composition. The NTE deletion also caused an increase in U5 and U1 snRNAs found in low molecular weight fractions, probably mono-snRNPs. Our genetic findings were consistent with roles for the NTE in spliceosome activation and second-step catalysis. The biophysical characteristics of recombinant NTE led us to propose a model by which, in solution, the NTE contains a short, structured segment followed by two larger segments, the first disordered and the second ordered. The NTE, expressed as a single domain, co-immunoprecipitated U5.U2/U6 spliceosome almost

exclusively, suggesting the domain is stably bound when associating with this spliceosome stage. The overexpressed NTE complemented the splicing defect of *cwf10-*Δ*NTE*, suggesting that the NTE can efficiently localize to the spliceosome and function even when the domain is not fused to the rest of Cwf10. We conclude that the NTE is important for multiple phases of the splice cycle and may be composed of structurally-distinct regions with high affinity for distinct binding partners, causing the NTE to stably incorporate into the catalytic-stage spliceosome.

In chapter III, I applied a PCR-based mutagenesis strategy to create genomic mutants of *cwf10*. Nine ts alleles were recovered, all with unique mutations. The mutations caused amino acid changes covering every domain of the protein. The mutants (*cwf10-2* through *cwf10-10*) generally displayed splicing defects at both the restrictive and permissive temperatures, and they displayed heterochromatin-silencing defects at the permissive temperature. The mutants caused levels of multi-snRNP protein markers Cdc5 and Prp1 to decrease, and snRNA sedimentation patterns in the mutants shifted, providing clues to the mechanism of the ts alleles' effects on splicing. Finally, even at the permissive temperature, the mutant Cwf10 proteins appeared unstable, with lower steady-state levels as determined by Western blot. It is likely that the mutated forms of Cwf10 are unstable and do not support efficient function. Multi-snRNP complexes, marked by Cdc5 and Prp1, became much less abundant in the mutants, possibly because spliceosomes formed inefficiently and the individual components were vulnerable to degradation. We conclude that these *cwf10* ts mutants are unlikely to be useful for purifying multi-snRNP structures for analysis by single-particle EM.

Taken together, these studies highlight the functions of Cwf10, advancing the field's understanding of Cwf10 NTE structure, global importance for splicing, and binding partners, and providing a clearer understanding of genetic means in *S. pombe* for preparing homogenous spliceosome purifications. This work paves the way for several paths of future investigation.

Addressing Cwf10's position within the catalytic core

Snu114/Cwf10 has several connections to spliceosome catalysis. As discussed in Chapter I, Cwf10 has genetic and biochemical interactions with Prp8, which is near, and may form part of, the active site. Additionally, human Cwf10 and Prp8, under chaotropic salt conditions that disrupt most U5 snRNP interactions, remain stably bound, suggesting that the shared interfaces are extensive (Achsel, Ahrens et al. 1998). This work and the work of others (Fabrizio, Laggerbauer et al. 1997, Liu, Laggerbauer et al. 1997, Frazer, Lovell et al. 2009) supports a role for Cwf10 in the second step of splicing, indicating the protein must be in a position to affect the catalytic core, directly or indirectly. While yeast two-hybrid assays have been helpful in determining interactions between specific regions of Cwf10 and Prp8 (for example, Boon, Norman et al. 2006 and Hegele, Kamburov et al. 2012), there are currently no co-crystal structures of Cwf10/Prp8 interfaces, as there are for Brr2/Prp8 regions (Mozaffari-Jovin, Wandersleben et al. 2013). Thus the three-dimensional orientation of Cwf10 in relation to Prp8 is unclear. An extensive genetic analysis of *S. cerevisiae* Cwf10 alleles with snRNA alleles (Frazer, Lovell et al. 2009) was interpreted to support the orientation of Cwf10 domain I near exon 1 prior to the first catalytic step. Human Cwf10 can be specifically crosslinked to

exogenous hairpins inserted between the branch point and 3' SS that block the second step (Liu, Laggerbauer et al. 1997). These data are consistent with Cwf10 being located near the substrate and having an active role in locating the 3' SS before second step catalysis.

While the relative positions of the Cwf10 domains are expected to be similar to those of homolog EF2, the NTE is not present in EF2. Thus, the relative orientation of the NTE to the other Cwf10 domains and/or other binding partners is unknown. However, the NTE may be oriented toward the 3' SS of the pre-mRNA, as two studies (Fabrizio, Laggerbauer et al. 1997, Frazer, Lovell et al. 2009) as well as my genetic data (Figure 2-6) have implicated the NTE domain in facilitating the second step.

Crystallization of the entire catalytic core of the spliceosome would provide amazing structural insight about, among many things, the orientation of Cwf10. Difficulties lie, however, in purifying sufficient quantities of the core and isolating a homogenous core complex from assembled spliceosomes. An alternative approach, building the core from recombinant components, is perhaps more prohibitive, as the components could not organize themselves into a functional active site without undergoing spliceosome assembly. Additional yeast two-hybrid, co-IP, and RNA binding assays of Cwf10 with candidate proteins and RNAs in the spliceosome may provide tools to better understand Cwf10's positioning in the spliceosome during catalysis.

Understanding mechanisms of NTE's functions in U5 snRNP stability and spliceosome activation

Apart from the NTE's positioning prior to second step catalysis, the NTE may contact different binding partners and be oriented differently in the U5 snRNP and the U5.U4/U6 tri-snRNPs. This is suggested by the NTE's additional functions in U5 snRNP stability (Bartels, Klatt et al. 2002 and Figure 2-5G and H) and in activation of the spliceosome (transition from Complex B to Complex B') (Bartels, Klatt et al. 2002).

The NTE could function actively in U5 snRNP stability, playing a role in the assembly or recycling of the U5 snRNP, or it could function passively, preserving important interactions that are disrupted by the NTE's absence. As no enzymatic activity has been reported for the NTE, an active role is less likely. *de novo* U5 snRNP formation involves the binding of Prp8 and Cwf10 to the cytoplasmic protein Aar2 (Boon, Grainger et al. 2007, Weber, Cristao et al. 2011). The *cwf10-ΔNTE aar2Δ* double mutant grew fairly well at all temperatures (Figure 2-6A), indicating that the NTE's function in U5 snRNP stability may occur after the initial assembly with Aar2. Structural data of the NTE's positioning in the U5 snRNP will be essential in understanding how the NTE stabilizes the snRNP.

The NTE's function in spliceosome activation is not defined specifically enough to hypothesize a mechanism of action. It could facilitate an early step, such as U4/U6 unwinding, or a later step, such as the helicase Prp2-mediated remodeling of the Complex B^{act} to Complex B' (Complex B^{act} is a recently described intermediate between Complex B and Complex B') (Warkocki, Odenwälder et al. 2009). Our genetic analysis of the *cwf10-ΔNTE* allele provides some clues as to the timeframe. *cwf10-ΔNTE* is synthetically

sick with *prp4-73*, an allele of the splicing kinase. Prp4 kinase has been shown to phosphorylate Prp1 (homolog of human U5-102) in *S. pombe* and human systems and to phosphorylate PRP31 in the human system (Schwelnus, Richert et al. 2001, Schneider, Hsiao et al. 2010). Prp1 and PRP31 are both components of the U5.U4/U6 tri-snRNP and leave the spliceosome prior to Complex B^{act} formation (Fabrizio, Dannenberg et al. 2009). The *cwf10-ΔNTE prp4-73* interaction suggests the NTE may act in an early stage of activation. Additional evidence for this comes from the synthetic lethality of *cwf10-ΔNTE* with *prp1-4*. Prp1 is thought to be the major "bridge" connecting the U5 snRNP to the U4/U6 di-snRNP (Makarov, Makarova et al. 2000, Bottner, Schmidt et al. 2005, Liu, Rauhut et al. 2006). The genetic interaction indicates the NTE may be required for a function of Prp1 either in tri-snRNP formation, tri-snRNP incorporation in Complex B, or conversion of Complex B to Complex B^{act}. Finally, *cdc28-P8*, an allele of helicase Prp2, displayed no synthetic phenotype with *cwf10-ΔNTE*. Prp2 acts late in activation, facilitating the Complex B^{act} to Complex B' transition (Warkocki, Odenwälder et al. 2009, Lardelli, Thompson et al. 2010, Ohrt, Prior et al. 2012). These data further support a role for the NTE in an earlier step of spliceosome activation. The *S. cerevisiae Snu114ΔN* mutant did not interact with any U4 mutations that would stabilize U4/U6 basepairing, suggesting that the NTE does not facilitate the early activation step of U4/U6 unwinding (Frazer, Lovell et al. 2009). These data together may implicate the NTE in supporting the ejection of the tri-snRNP-specific proteins (including Prp1) from the spliceosome during the Complex B to Complex B^{act} transition. Further genetic experiments and high resolution structural data of the NTE in the tri-snRNP will greatly enhance our understanding of the NTE's role in spliceosome activation.

Crystallization of the U5 snRNP and the U5.U4/U6 tri-snRNP would provide insight into the NTE's orientation in these complexes. Preparing crystals suitable for structural analysis may be a more approachable goal than crystallization of the catalytic core, for several reasons. First, two other snRNPs have been crystallized, albeit with the removal of some associated protein and RNA regions that likely impaired crystallization. The human U1 snRNP has been crystallized using two separate methods (Pomeranz Krummel, Oubridge et al. 2009, Weber, Trowitzsch et al. 2010), and the U4 snRNP core, without U4-specific proteins, was crystallized from recombinant components (Leung, Nagai et al. 2011), providing a proof of concept for U5 snRNP crystallization Also, the U5 snRNP has two forms – a cytoplasmic form with Aar2, and a larger nuclear U5 snRNP in which Brr2 has replaced Aar2 (Weber, Cristao et al. 2011). Aar2 bound to ~1000 residues of Prp8 has already been crystallized (Galej, Oubridge et al. 2013). U5 snRNPs and U5.U4/U6 tri-snRNPs are present in HeLa and *S. cerevisiae* splicing extracts, without the further manipulation required to form and isolate catalytically active spliceosomes, and these complexes have already been successfully isolated and analyzed by 3D EM (Sander, Golas et al. 2006, Hacker, Sander et al. 2008).

Discovering the NTE's structure and origin

Our studies are the first to experimentally characterize the NTE's structure. Our biophysical methods uncovered regions of both structure and disorder in the recombinant NTE and provided insight regarding the secondary structure content of the domain. From several standpoints, it will be useful to pursue the generation of a high-resolution model of the NTE's structure. It is worth noting that, because the NTE contains some regions of

disorder, a binding partner may be necessary to provide a disorder-to-order transition that makes crystallization possible. First, the primary sequence of the NTE provides no clues as to what tertiary structure may be formed by the NTE. Structural data may thus unveil a new relationship between the primary sequence and either previously characterized or unique folds. Second, the tertiary structure may provide clues as to what type of molecules the NTE binds. *In vitro* RNA binding assays may be useful in determining if the NTE is a nucleic acid binding domain. Finally, having high-resolution structural data may lead to a hypothesis about the NTE's origins. Prp8, Cwf10's binding partner, has recently been shown to contain domains adapted from diverse sources. A set of domains in Prp8 resembles those of a fungal group II intron reverse transcriptase (Galej, Oubridge et al. 2013), while another domain was identified in deubiquitylating enzymes (Pena, Liu et al. 2007). Cwf10 is also a protein adapted from another system, as the bulk of Cwf10 is highly homologous to the eukaryotic ribosomal translocase EF2, at least at a primary sequence level (Fabrizio, Laggerbauer et al. 1997). It will be interesting to find out if the NTE is a known, but unrelated, domain that was fused to EF2 for the coupling of certain functions in the splicing process.

It may also be worth exploring the role of phosphorylation on the structure of the NTE. Several large-scale screens have identified phosphorylation sites on the *S. cerevisiae* and human NTEs (Li, Gerber et al. 2007, Albuquerque, Smolka et al. 2008, Olsen, Vermeulen et al. 2010, Rigbolt, Prokhorova et al. 2011), although no phosphorylated residues have been identified on the *S. pombe* NTE. It would be interesting to perform biophysical experiments, as in Chapter II, on phosphorylated and non-phosphorylated NTE to determine if phosphorylation acts as a switch for different

conformational arrangements, as it does with F-BAR protein Cdc15 (Roberts-Galbraith, Ohi et al. 2010).

Conclusion

In sum, our analyses of *S. pombe* Cwf10 NTE have helped to define its structural characteristics and its contributions to multiple steps of splicing. We expect this work will encourage further study of the NTE to determine the specific mechanisms it employs as a domain in the U5 snRNP and near the heart of the spliceosome.

APPENDIX A

EFFECTS OF VARIOUS STAINS AND TECHNIQUES ON ELECTRON MICROSCOPY IMAGE QUALITY OF A NEGATIVE STAIN PROTEIN SAMPLE

Purpose and Methodology

There are multiple heavy-metal stains used to coat biological specimens for imaging by transmission electron microscopy. Among those used to prepare stained grids of protein solutions, some of the more common are uranyl acetate, uranyl formate, phosphotungstic acid (PTA), ammonium molybdate, methylamine tungstate (NanoW), and methylamine vanadate (NanoVan). Additionally, staining qualities can be greatly influenced by the method of staining, including the steps of sample application, stain application, and stain removal/drying. While there are many other variables that influence staining qualities, I chose to investigate stain type and method of staining using a protein sample commonly used in EM characterization in our lab, *Helicobacter pylori* vacuolating toxin VacA, a 88-kDa monomer that oligomerizes, forming wheel-and-rod shape structures. All staining was performed on September 21, 2011.

Sample

VacA was obtained from the Tim Cover lab (Christian Gonzalez) in solution at 650 μg/ml. The protein was diluted to 108 μg/ml using 10 mM HEPES pH 7.9 and applied to grids.

Stains

PTA

2% PTA in water was made and refrigerated a few days before use. Stain solution has a shelf life of roughly three weeks. Purchased from Ted Pella (#19402).

Uranyl formate

0.7% weight/volume uranyl formate was prepared as per normal Melanie Ohi lab protocol. Stain was used the day of preparation and stored in a foil-wrapped tube to protect from light. Purchased from Polysciences, Inc. (#24762).

NanoW

NanoW was used as supplied in solution. Purchased from Nanoprobes (#2018-5ML).

NanoVan

NanoVan was used as supplied in solution. Purchased from Nanoprobes (#2011-5ML).

Methods of staining

Note that grids were not washed with buffer after sample application except in the case of the Melanie Ohi method and the Jim Stoops method with NanoW.

Melanie Ohi (modified)

Sample was applied to a glow-discharged grid and allowed to sit for two minutes. Sample was wicked from grid, and the grid was whisked in a small drop of buffer on parafilm. Buffer was wicked and grid was whisked again in buffer. Buffer was wicked and grid was whisked in a small drop of stain. Stain was wicked and grid was placed on top of a small drop of stain for 90 seconds. Stain was wicked and grid was air-waved dry.

Jim Stoops (also known as "deep staining")

A glow-discharged grid was held in forceps for the entire procedure. Sample (6 µL) was applied to the grid and allowed to sit 30-60 seconds. Sample was wicked from grid, with a small amount of sample allowed to remain on the grid. Immediately, 6 uL of stain was placed on the grid and wicked away. This step was repeated once more. Immediately 6µL of stain was placed on the grid and allowed to stand for 30-60 seconds and then wicked away. Grid was immediately air-waved dry. The grid was never allowed to dry completely between sample application and the final drying step.

Roger Craig (also known as the "Brandeis method")

Sample was applied to a glow-discharged grid held in forceps and allowed to sit for two minutes. Stain was loaded in a disposable bulb pipette. Forceps were held vertically with the grid end pointing towards floor. The pipette was held at 15-20 degrees from the vertical forceps with the pipette tip very near the grid. The grid was washed with stain by applying 4-6 continuous drops. The last drop was allowed to remain on the grid 30-60 seconds with the forceps returned to a flat surface. Stain was wicked away and the grid was waved in the air to dry.

Reactive buffers (named because of its successful use with sample in buffers that, with other methods, stains poorly)

Three drops of stain (150 µL each) were applied to parafilm. Sample (5 µL) was applied to a glow-discharged grid and allowed to sit 30-60 seconds. Grid was placed on top of the first stain drop, then moved to the second, and third drops. Grid was allowed to sit on top of the third stain drop for 30 seconds. Stain was wicked from the grid, and the grid was immediately air-waved dry.

Results and discussion

Representative images are shown for uranyl formate (Figure A-1), PTA (Figure A-2), NanoW (Figure A-3), and NanoVan (Figure A-4). These images have high background, likely due to the presence of VacA monomers (Yoshimasa Takizawa, personal communication). VacA oligomers were very clear under most combinations of stain/method except for the uranyl formate/reactive buffers combination, which produced fuzzy images with unusually high background (data not shown). Under some conditions, a variable depth of stain could be found on a single grid (for example, NanoW using the Jim Stoops method, Figure A-3A, B, and C). Side views of the oligomers were seen under several conditions (e.g. Figure A-1A and Figure A-2C).

Although uranyl formate is a good general purpose stain (Ohi, Li et al. 2004), it has been suggested that each protein sample will react differently to both stain and staining conditions (Radermacher, Ruiz et al. 2006 and Teresa Ruiz, personal communication). Therefore, it may be advisable to test a protein sample with different stains and staining conditions to obtain optimum particle images, which should positively influence all downstream processing steps in single-particle analysis.

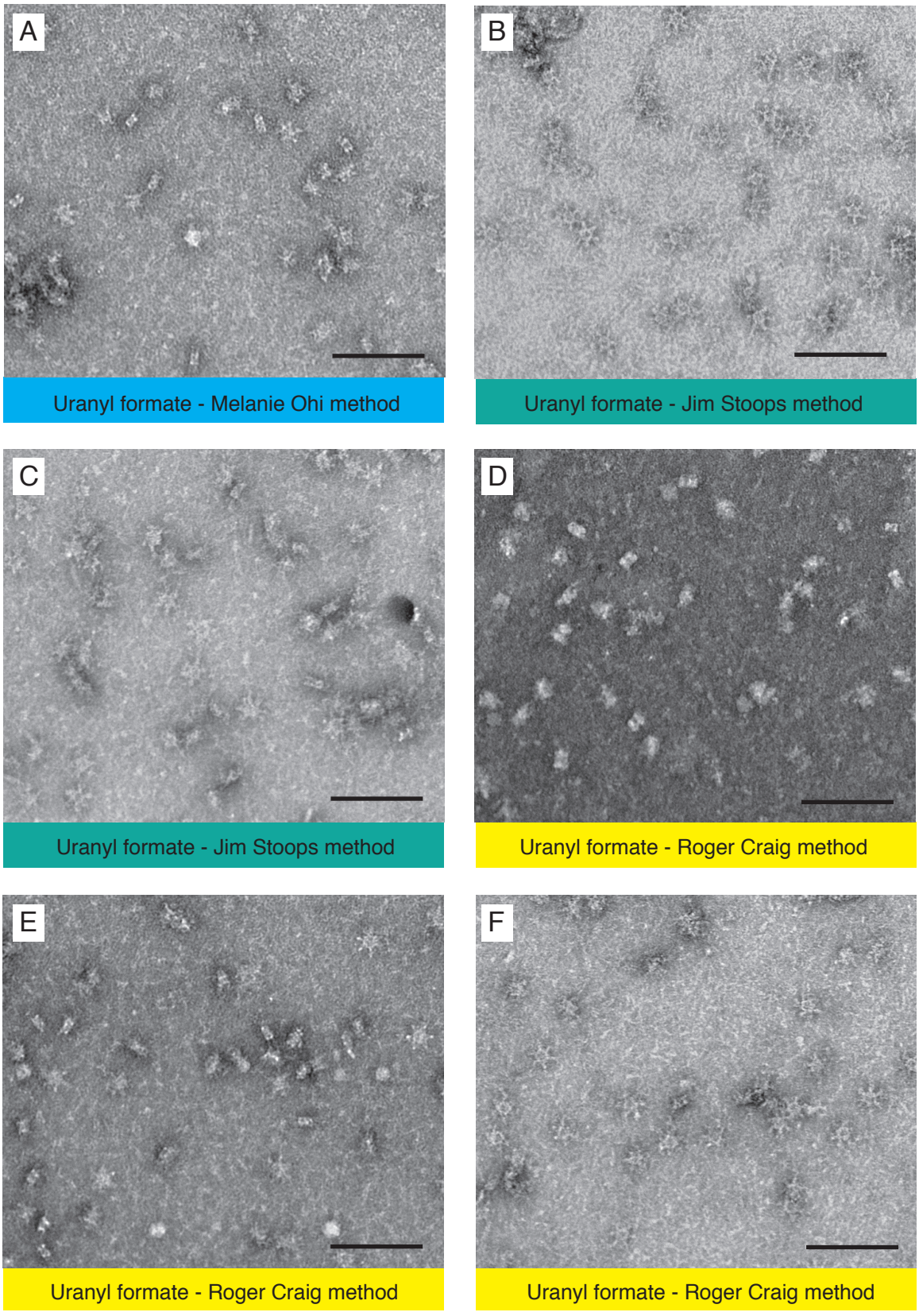

A
Uranyl formate - Melanie Ohi method
B
Uranyl formate - Jim Stoops method
C
Uranyl formate - Jim Stoops method
D
Uranyl formate - Roger Craig method
E
Uranyl formate - Roger Craig method
F
Uranyl formate - Roger Craig method

FIGURE A-1

FIGURE A-1 (continued)

Images of VacA stained with uranyl formate using various methods

VacA protein solution was stained with uranyl formate using the indicated method and imaged on an FEI Morgagni microscope at 28,000X. Images of the same stain/method are grouped together and color-coded. Scale bars represent 100 nm.

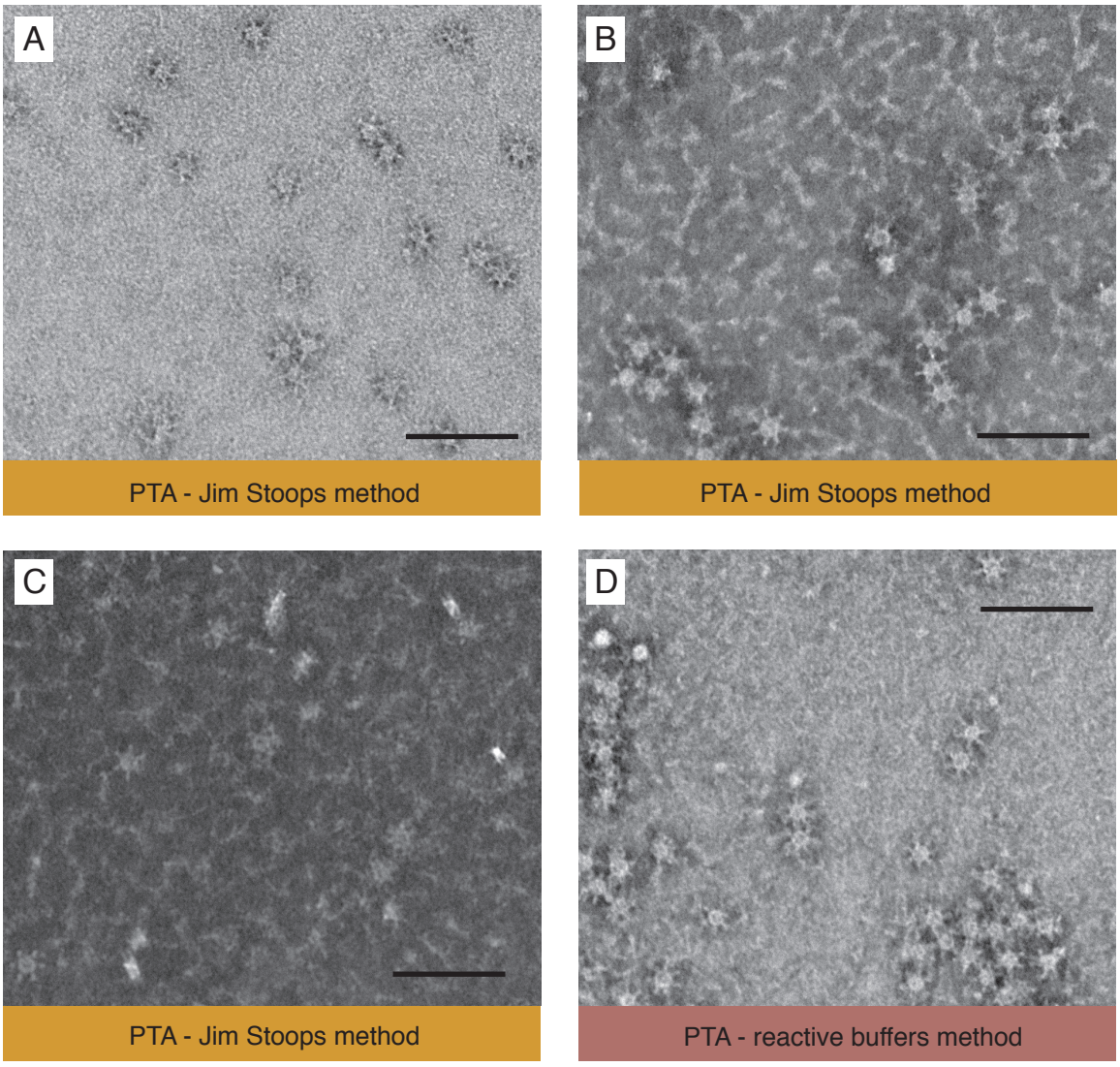
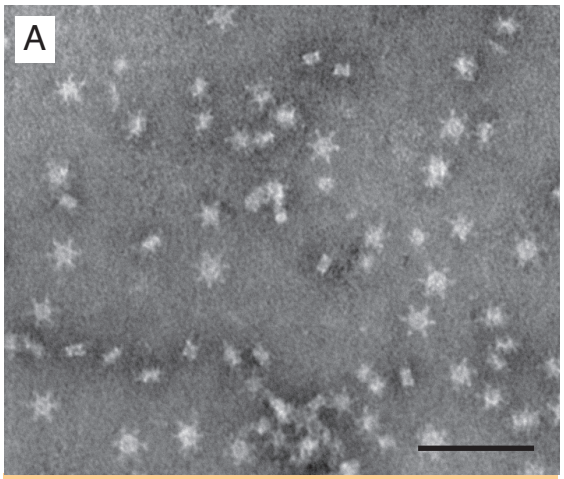


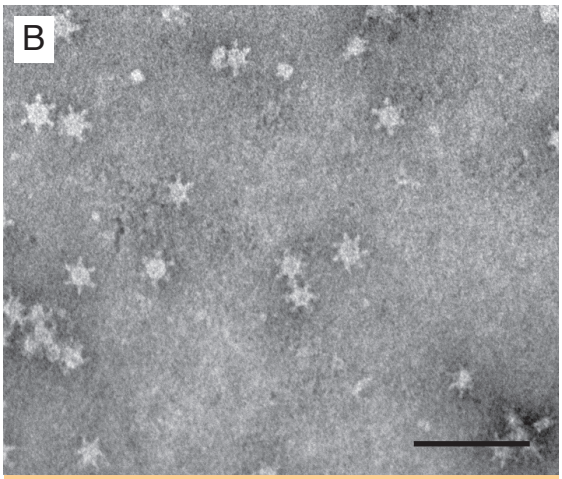
FIGURE A-2

Images of VacA stained with PTA using various methods

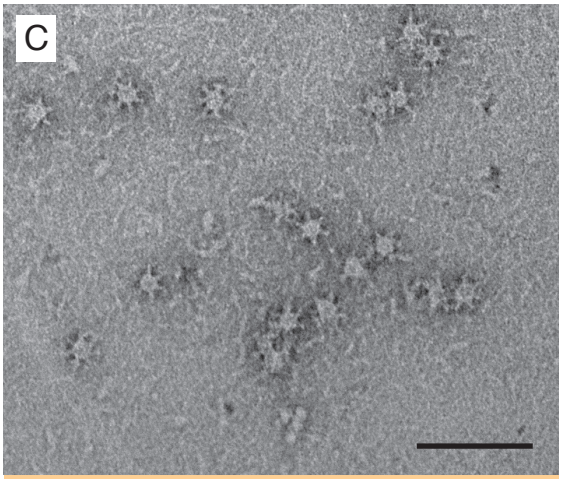
VacA protein solution was stained with phosphotungstic acid (PTA) using the indicated method and imaged on an FEI Morgagni microscope at 28,000X. Images of the same stain/method are grouped together and color-coded. Scale bars represent 100 nm.



NanoW - Jim Stoops method



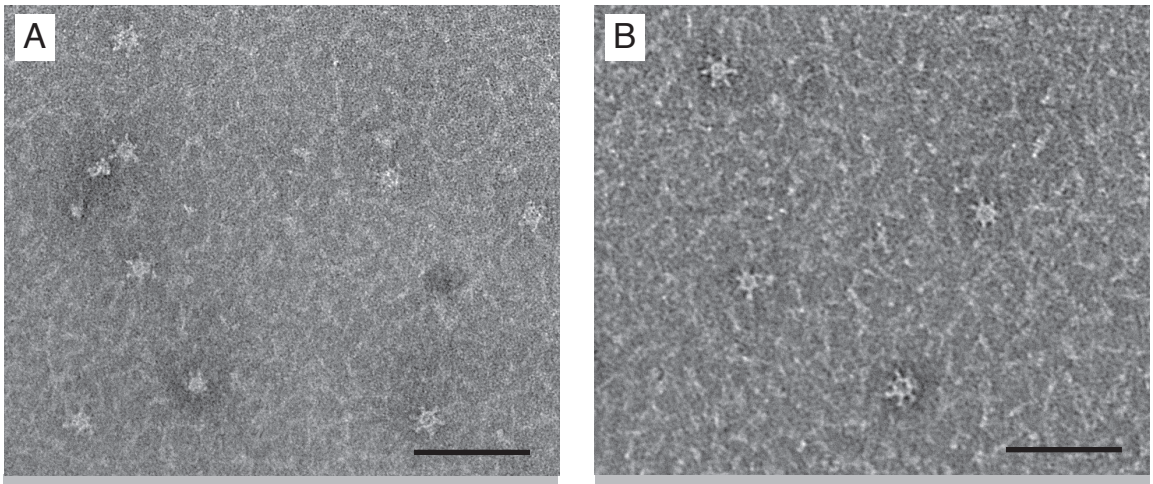
NanoW - Jim Stoops method



NanoW - Jim Stoops method

FIGURE A-3

Images of VacA stained with NanoW
VacA protein solution was stained with methylamine tungstate (NanoW) using the Jim Stoops method and imaged on an FEI Morgagni microscope at 28,000X.
Scale bars represent 100 nm.



NanoVan - Jim Stoops method

NanoVan - Jim Stoops method

FIGURE A-4

Images of VacA stained with NanoVan
VacA protein solution was stained with methylamine vanadate (NanoVan) using the Jim Stoops method and imaged on an FEI Morgagni microscope at 28,000X.
Scale bars represent 100 nm.

APPENDIX B

CONTRIBUTIONS OF NTC COMPONENTS TO DNA DAMAGE PHENOTYPES

Introduction

The NTC, as described in Chapter I, is a protein sub-complex of the spliceosome. While the NTC has a well-documented role in splicing, early evidence pointed to a role in DNA damage repair (Grey, Dusterhoft et al. 1996). While a DNA damage phenotype could arise secondarily from a splicing defect (discussed below), there has been an accumulation of evidence that NTC components function directly in DNA damage repair (reviewed in Legerski 2009, Chanarat and Strasser 2013). Such evidence comes primarily from yeast two-hybrid studies and other physical association data of NTC members with proteins involved in DNA damage repair. Kuraoka, et al., showed that human NTC member SYF1 interacts with xeroderma pigmentosum group A protein (XPA) and other transcription-coupled repair factors (Kuraoka, Ito et al. 2008). Budding yeast Prp19 associates with an endonuclease required for nucleotide excision repair (Revers, Cardone et al. 2002), and human Cdc5L interacts physically with DNA damage checkpoint protein ataxia-telangiectasia and Rad3-related (ATR) (Zhang, Kaur et al. 2009).

Despite these direct links, it is reasonable to suggest that inoperative spliceosome proteins will, through a splicing mechanism, impair the proper expression of some or all DNA damage repair factors, especially in higher eukaryotes where most genes contain introns. Additionally, the close proximity of splicing factors to genomic DNA during co-transcriptional splicing may present an opportunity to damage DNA: Li and Manley

demonstrated that knockdown of SR splicing factor ASF/SF2 led to R loop formation, leading to DNA damage (Li and Manley 2005). In their model, the presence of ASF/SF2 normally keeps the nascent RNA away from the vulnerable single strands of the DNA. When ASF/SF2 is not present, the transcript forms a hybrid with its single-strand DNA template (an R loop). This DNA/RNA hybrid can cause double-stranded breaks wherever transcription is taking place, and it can activate DNA damage repair checkpoints. Because ASF/SF2 is a non-NTC splicing factor whose loss can cause the appearance of DNA damage repair defects, we wondered whether the NTC truly is a DNA damage repair component, or whether NTC-associated DNA damage phenotypes could be due to interference with NTC's splicing function.

Results and discussion

To address this issue, we chose to investigate DNA damage and mitotic defect phenotypes in HeLa cells. By siRNA-depleting three NTC components, PRP19 (homolog of *S. cerevisiae* Prp19), PLRG1 (*S. cerevisiae* Prp46), and CDC5L (*S. cerevisiae* Cef1), and two non-NTC spliceosome RNA helicases, U5-200 (*S. cerevisiae* Brr2) and DHX8 (*S. cerevisiae* Prp22), we hoped to determine if the DNA damage phenotypes were different between NTC components and other spliceosome proteins.

Cells were transfected by siRNAs or a no-siRNA control for 72 hours. Cells were fixed and probed with an antibody against γ-H2AX, a genome-bound indicator of DNA damage. Twenty-five cells per siRNA were quantified for the amount of fluorescence signal. All siRNA-treated cells displayed increased levels of γ-H2AX, compared with empty-transfected control (Figure B-1A), indicating that the NTC and non-NTC splicing

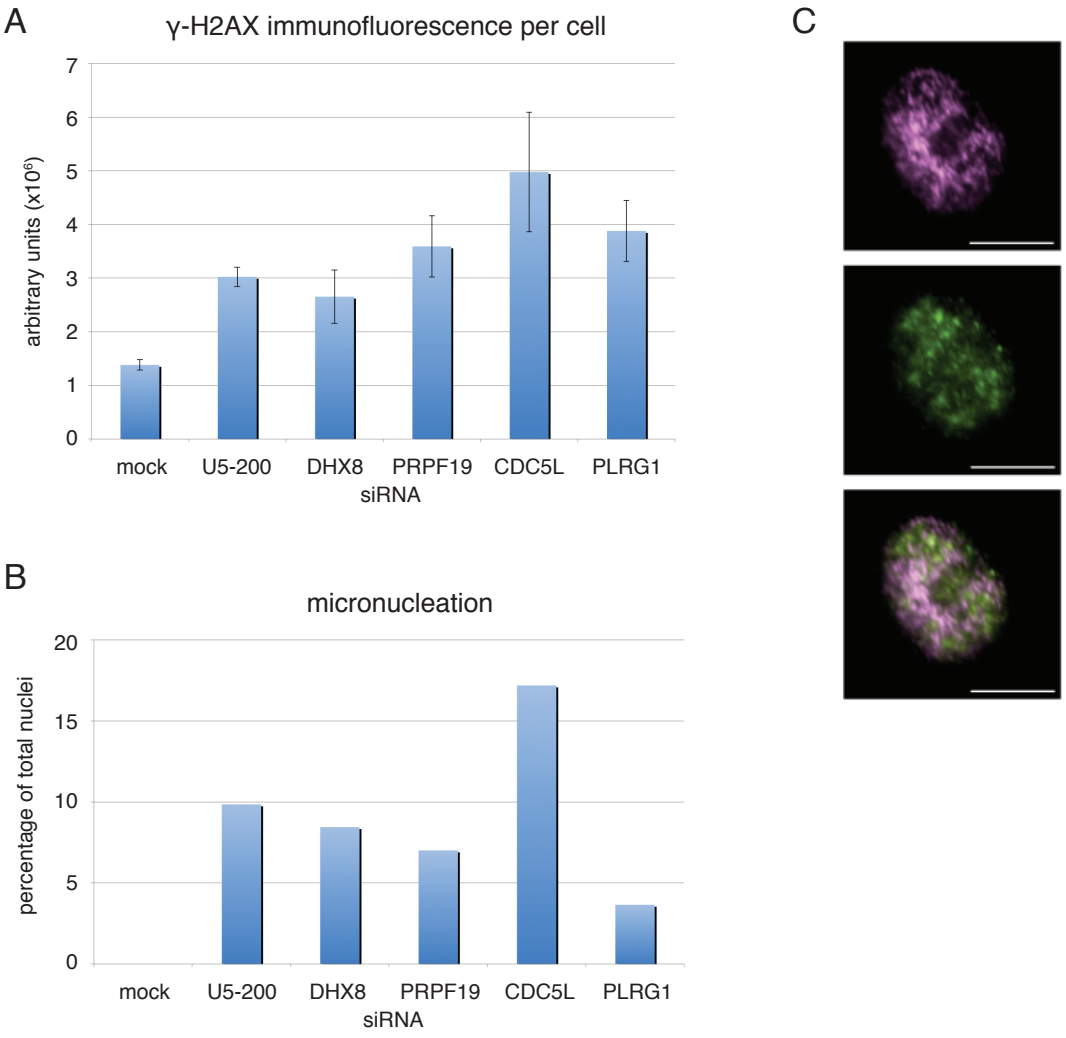


FIGURE B-1

siRNAs targeting NTC and other spliceosome components cause DNA damage
A) HeLa cells transfected with the indicated siRNAs for 72 hours were stained by anti-γ-H2AX, and the fluorescence intensity per cell was quantified. The value represents the average per-cell intensity. The error bar represents the standard error of the mean. At least 25 cells per condition were quantified. B) As in A, except that cells were stained for DNA by Hoechst, and >120 cells per genotype were counted. Values represent the percentage of cells with a micronucleation phenotype. C) HeLa cells were treated with 2mM hydroxyurea for four hours, fixed, and stained for γ-H2AX (magenta) and PRPF19 (green; Abcam ab27692). The bottom panel is an overlay in which co-localization is indicated by white. Scale bar represents 10 μm.

factors' depletion caused DNA damage. Although every NTC component had a higher per cell value than the two non-NTC components, statistical significance of the differences among the siRNA-treated cells was not determined.

Next cells were transfected as before for 72 hours and stained with DAPI to reveal DNA morphology. Micronucleation was a common phenotype observed. Micronucleation results from errors in chromosome segregation during mitosis and is a result of either chromosome breakage or dysfunction in the mitotic machinery (Norppa and Falck 2003). We quantified the percentage of nuclei that were micronucleated (Figure B-1B) and saw that all of the siRNAs caused micronucleation above baseline. Interestingly, siRNA against CDC5L led to the largest percentage of micronucleated cells, while PRP19 and PLRG1 siRNAs caused the smallest. As knockdown efficiency was not quantified, it is not clear whether this result indicates CDC5L loss causes more accumulation of micronucleated cells than loss of the other tested splicing factors.

Finally, we examined the behavior of NTC component PRPF19 under genotoxic stress conditions. PRPF19 distribution in nuclear speckles is not disrupted by hydroxyurea (HU) application (data not shown). Based on previous reports of NTC component physical interactions with DNA repair proteins, we examined co-localization of PRPF19 and γ-H2AX foci by immunofluorescence. Upon HU application, PRPF19 does not appear to co-localize with the DNA damage marker (Figure B-1C). These results together suggest that most or all of PRPF19 is not recruited to DNA damage sites under our conditions.

This brief set of experiments was designed to address whether NTC components have direct roles as DNA repair agents. We found that NTC-depleted HeLa cells had an

increased level of genotoxic stress; however, depleting cells of other splicing factors also increased genotoxic stress. Similarly, NTC and non-NTC-depleted HeLa cells underwent micronucleation. These results suggest that spliceosome components in general are important for DNA integrity; however, the results do not rule out an additional function specific to NTC components. The lack of co-localization of γ-H2AX foci with PRPF19 simply indicates that PRPF19 may not be present at the time of γ-H2AX signaling, and does not rule out a function for the protein at a different point in the DNA repair process. We recommend that additional investigation focus on understanding the potential role of improper splicing progression in causing genomic damage. With this foundation, the non-splicing roles of NTC members in DNA damage repair can be more thoroughly explored.

Additional Methods

Cells were transfected with HiPerFect (Qiagen) and siRNAs (Dharmacon ON-TARGETplus SMARTpool), 20 uL of 20 uM siRNA per 10 cm dish.

All images were taken using an Eclipse 90i (Nikon Instruments, Melville, NY) microscope equipped with Nikon objectives and a Cool SnapHQ2 CCD camera (Roper Scientific, Tucson, AZ). Immunofluorescence intensities were measured using NIS-Elements software (Nikon Instruments).

APPENDIX C

ISOLATING THE NTC

The NTC, as described in Chapter I, is a protein sub-complex of the spliceosome. There have been several attempts to isolate this sub-complex from both yeast and human cells. Previous characterization of the NTC in *S. pombe* suggests it is stably integrated into a larger U5.U2/U6 complex under a variety of purification conditions (McDonald, Ohi et al. 1999, Ohi, Link et al. 2002). However, Tarn, *et al*., were able to isolate the NTC from *S. cerevisiae* splicing extract using HA-tagged Prp19 (Tarn, Hsu et al. 1994). This purification, as detected by silver stain, spanned most fractions of a density gradient, suggesting a homogenous NTC purification had not been achieved. Several purifications of the NTC from human cells have culminated in the characterization of a fairly homogenous seven-protein complex that was amenable to 2D EM analysis (Grote, Wolf et al. 2010). Isolation of the NTC would be valuable in at least two applications: structural characterization and functional *in vitro* add-back experiments. Because of the NTC's integration into the catalytic and post-catalytic spliceosomes, high-resolution structural characterization of the NTC would enhance our understanding of spliceosome organization. Additionally, the isolation of a functional, homogenous NTC for use in *in vitro* splicing assays would enhance our understanding of spliceosome activation.

In our first attempt to characterize the *S. cerevisiae* NTC, we modified a method originally described in Tarn, et al. (Tarn, Hsu et al. 1994). This involved adjusting yeast cell lysate to 40% ammonium sulfate and removing the precipitant prior to purifying the

NTC with Prp19-TAP under high salt conditions (500 mM NaCl). We called this purification scheme "40P" Prp19-TAP. Our normal two-step TAP purification (Gould, Ren et al. 2004) uses 150 mM NaCl buffers throughout the purification, and under these conditions Prp19-TAP associates with the U5, U2, and U6 snRNAs (Ohi, Link et al. 2002). Interestingly, the complex appears to be less stable than the S. pombe U5.U2/U6 complex – by negative stain EM, *S. cerevisiae* Prp19-TAP particles appear more heterogeneous and fall apart more easily than *S. pombe* Prp19-TAP particles (Ohi, unpublished observation). Additionally, in *S. cerevisiae* lysates, multiple NTC components sediment at the top of the gradient, presumably as single proteins, in addition to sedimenting in a large 40S complex (Ohi and Gould 2002). We hoped to capitalize on the less-stable nature of *S. cerevisiae* U5.U2/U6 and isolate only NTC proteins by the addition of the ammonium sulfate fractionation step. As an initial determination of whether we had isolated the NTC instead of the larger 40S U5.U2/U6 complex, we analyzed the sedimentation pattern of Cwc2-myc, a protein that directly interacts with Prp19 and sediments in the U5.U2/U6 complex (Ohi and Gould 2002, Vander Kooi, Ren et al. 2010), in sucrose gradients by Western blotting. From this analysis, we saw that under the 40P conditions, Cwc2-myc still sedimented in a similar manner as Cwc2 purified under normal conditions (Figure C-1A). Analysis by mass spectrometry confirmed that U2 and U5 proteins were still associating with the 40P Prp19-TAP; however, we observed a few differences between the proteins associated with Prp19-TAP depending on whether the purification was done under 40P or normal conditions. Specifically, the 40P purification lacked Brr2 (U5 snRNP component), Lea1 (U2 snRNP component), and Ntc20 (NTC member) (data not shown). Finally, negative stain EM

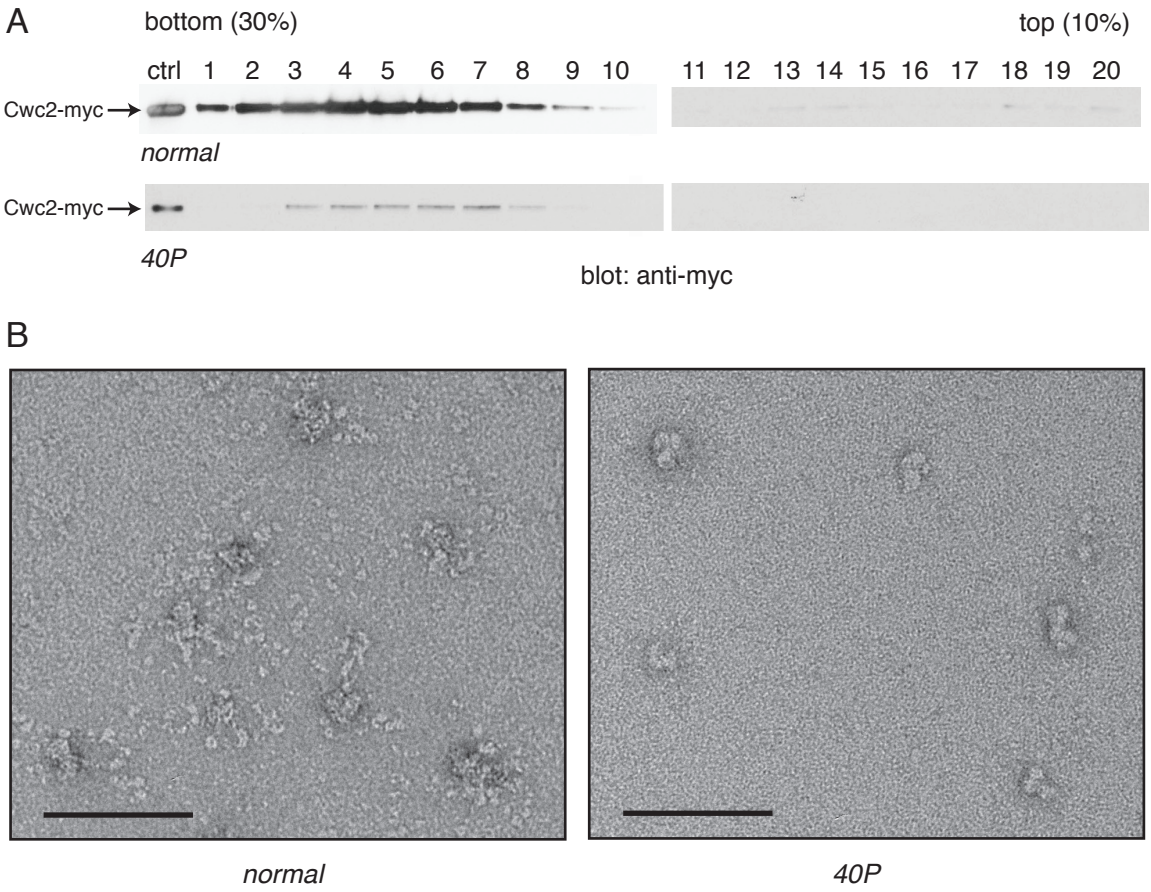


FIGURE C-1

Effects of ammonium sulfate extraction on spliceosome complexes
A) *prp19-TAP cwc2-myc* cells were lysed in NP-40 buffer (*normal*) or lysed in NP-40 buffer and then extracted with ammonium sulfate (*40P*). TAP-purified proteins were eluted and loaded on to a 10-30% sucrose gradient. Fractions collected were blotted with anti-myc, revealing the sedimentation pattern of Cwc2-myc. The "ctrl" lane was loaded with the TAP eluate. B) Electron micrographs of purified Prp19-TAP complexes, with *normal* and *40P* defined as in A. Scale bar represents 100 nm.

revealed that 40P Prp19-TAP particles were similar in size and shape to Prp19-TAP particles purified under normal TAP conditions (Ohi, Link et al. 2002), although with more rounded, less detailed densities (Figure C-1B). From these analyses, we concluded that the ammonium sulfate cut is unsuitable for destabilizing the NTC from larger spliceosome particles in *S. cerevisiae* lysates.

In a second attempt to characterize the *S. cerevisiae* NTC, we made whole cell splicing extract (WCSE) from *prp19-TAP cwc2-myc* cells using methods described in both Ansari and Schwer 1995 and Umen and Guthrie 1995. WCSE is prepared by a lysis and dialysis protocol different from our lab's normal native cell lysate protocol (Gould, Ren et al. 2004). As a first analysis, we characterized the sedimentation pattern of Cwc2-myc under these new conditions, without immunopurification (Figure C-2A). Promisingly, cells lysed using the WCSE approach had much more Cwc2-myc in low molecular weight fractions than did lysate made using our lab's normal protocol. Further adjustment of the WCSE to splicing conditions (final concentrations: 100 mM potassium phosphate, 5 mM magnesium chloride, 1.67 mM spermidine, and 5% PEG) with or without ATP caused an additional shift of Cwc2-myc away from higher molecular weight fractions (top of gradient) to lower molecular weight fractions (bottom of gradient). SDS-PAGE separation and silver stain analysis of purified Prp19-TAP from WCSE and WCSE under splicing conditions revealed markedly different band patterns than what was seen from Prp19-TAP lysed under conditions described in Ohi, et al. (Ohi, Link et al. 2002) (Figure C-2B). Finally, in an attempt to examine the structural homogeneity of these smaller NTC-containing complexes, we purified Prp19-TAP from WCSE and ran the eluted complex over a sucrose gradient under identical conditions to Figure C-1A,

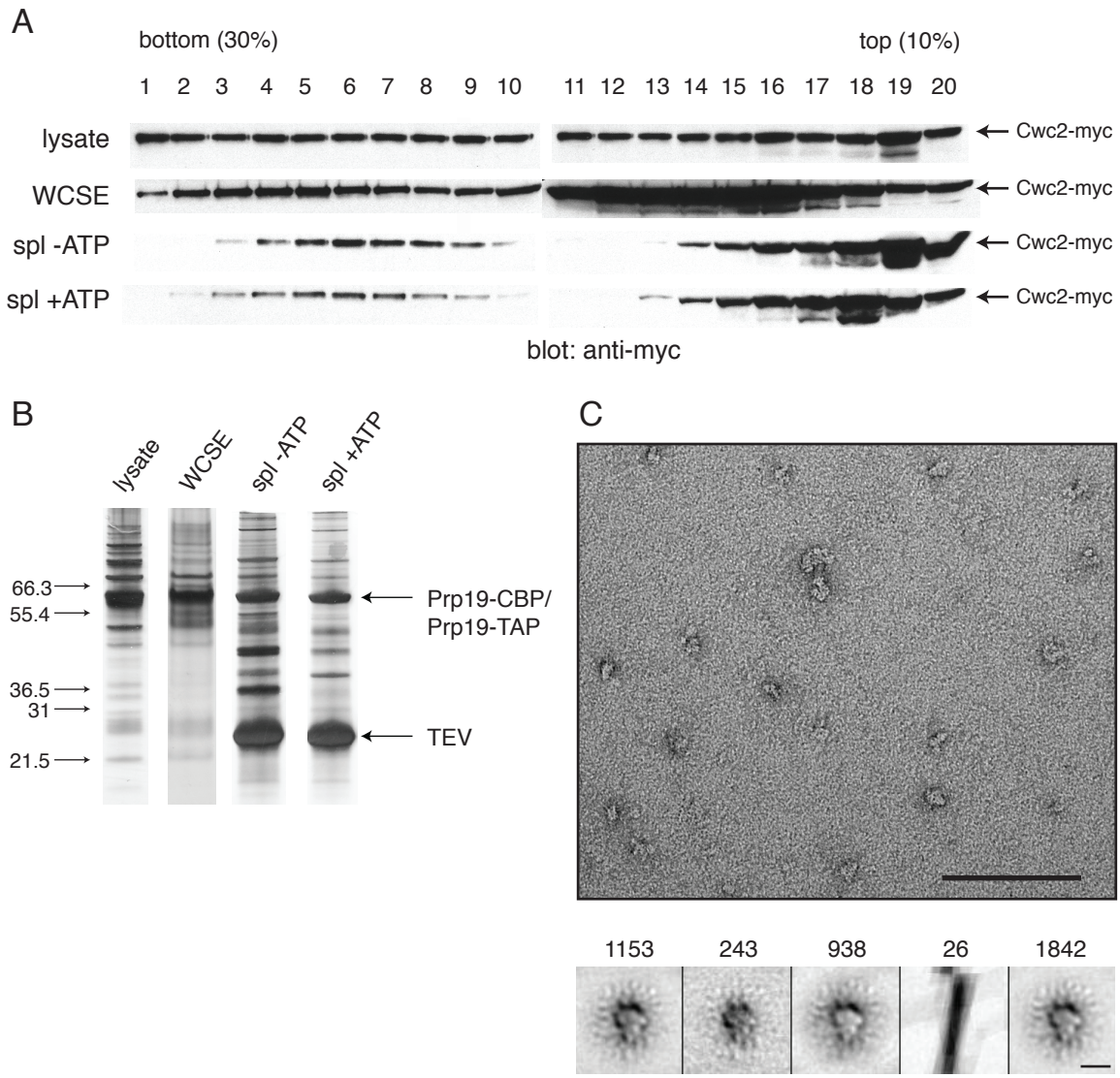


FIGURE C-2

NTC characterization in *S. cerevisiae* whole cell splicing extracts

A) *prp19-TAP cwc2-myc* cells were lysed in NP-40 buffer (lysate) or prepared as whole cell splicing extract (WCSE). WCSE was used in a splicing reaction without an added transcript and with (spl +ATP) or without (spl -ATP) ATP. These preparations were run on a 10-30% sucrose gradient and probed with anti-myc. B) Silver stain of Prp19-TAP purifications. Column descriptions are as in A. Purifications from lysate and WCSE were two step TAP procedures, while purifications from spl -ATP or +ATP were one step Dynabead-IgG procedures with subsequent elution with TEV protease (TEV). Molecular weight markers (left) are in kDa. C) top - Prp19-TAP was purified from WCSE and run on a glycerol gradient as in A, but with glutaraldehyde. Electron micrograph is of fraction 15. Scale bar represents 100 nm. bottom - Class averages of WCSE fraction 15 particles. Numbers indicate the number of particles in each class. Scale bar represents 10 nm.

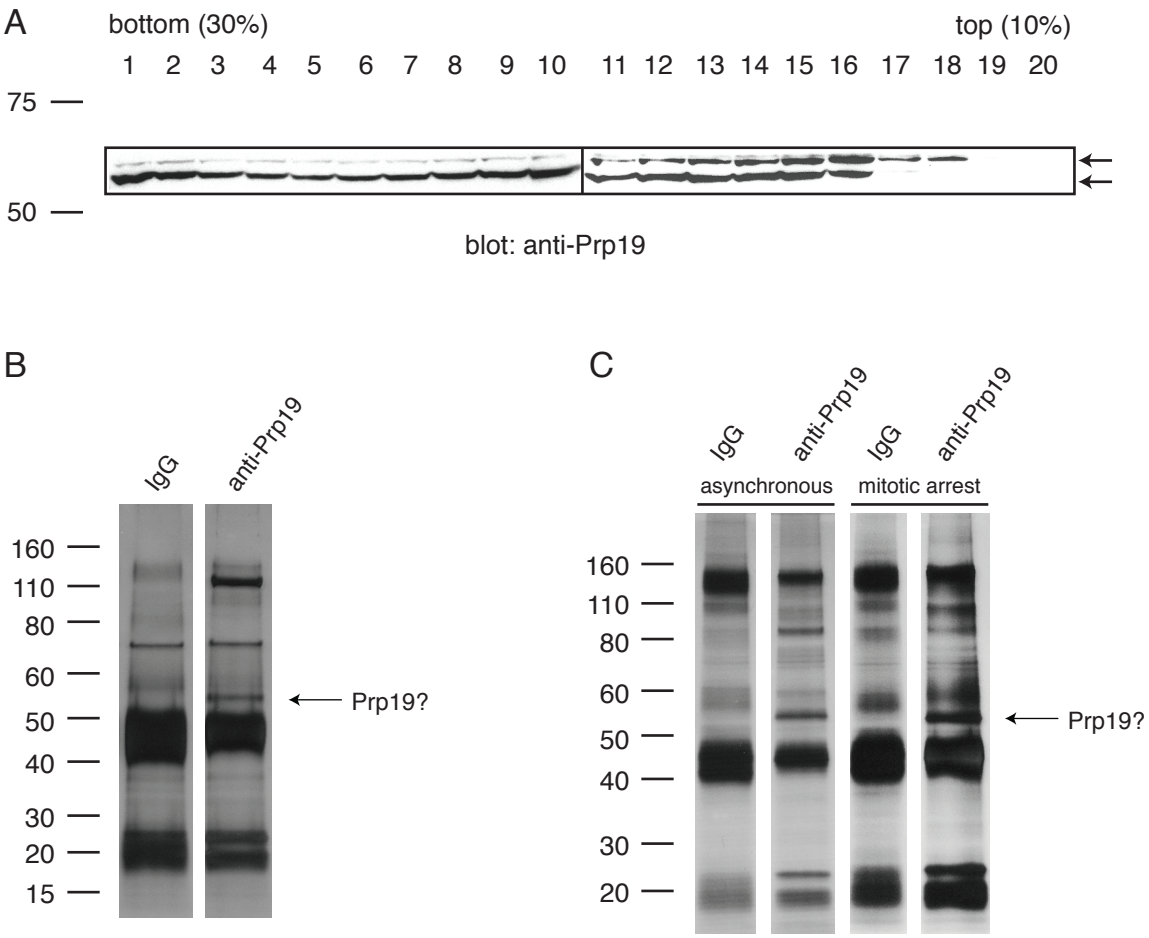


FIGURE C-3

Prp19 characterization in vertebrate model systems

A) *Xenopus laevis* egg extracts were run over a 10-30% sucrose gradient. Fractions were collected and blotted with anti-Prp19. Molecular weight markers (left) are in kDa. The observed top and bottom bands of Prp19 are indicated by arrows (right). B) Egg extracts were incubated with Affi-Prep protein A beads (Bio-Rad) conjugated to either rabbit IgG (Jackson Labs) or anti-Prp19. Washed beads were eluted by SDS and run on a gel. Silver stain and the possible location of the Prp19 band are shown. Molecular weight markers (left) are in kDa. C) HeLa cells were harvested after growth in serum-containing media (asynchronous) or after 15.5 hours in media with 10µM STLC (mitotic arrest). IPs, gels, staining, and markers are as in B.

except that the gradient contained glutaraldehyde as in the Grafix protocol (Kastner, Fischer et al. 2008). WCSE fraction 15 was imaged by negative-stain EM and showed detailed, yet heterogeneous particles (Figure C-1C, top). 4,202 particles were aligned to produce class averages using the software SPIDER. Although multiple processing parameters were tried, no good class averages could be obtained (Figure C-1C, bottom).

Next, we examined Prp19-associated proteins in *Xenopus laevis* egg extracts and HeLa cell lysates. *Xenopus* egg extracts appeared to contain two forms of Prp19 by Western blot, with one form predominant in high molecular weight fractions, and the other form in low molecular weight fractions (Figure C-3A, arrows). The identities of these two forms are unknown, but could indicate the presence of a modification, such as phosphorylation. Anti-human Prp19 antibody (Abcam ab27692) was used to immunoprecipitate Prp19 from the egg extracts (Figure C-3B), and both asynchronous and mitotic HeLa cells (Figure C-3C). Silver staining of all anti-Prp19 IPs clearly showed the presence of a band near the predicted molecular weight of Prp19 (55 kDa), and additional bands unique to the anti-Prp19 beads at 90-110 kDa could indicate a co-purifying NTC protein, such as CDC5L. The next step would be to analyze these purifications using mass spectrometry to determine which proteins are associating with Prp19. However, these initial results indicate that it may be possible to purify the NTC from either *X. laevis* or HeLa cell extracts.

Together, these results indicate that a homogenous NTC complex could possibly be isolated from the WCSE from yeast, even though our efforts have not yet been successful. It will be interesting to see if the human NTC identified in Grote, Wolf et al.

2010 can successfully be characterized by 3D EM, or if the proteolysis-resistant NTC core identified in the same work can be crystallized.

APPENDIX D

CYCLOPHILINS IN FISSION YEAST SPLICING

While there have been many functions attributed to enzymes such as kinases, RNA helicases, and GTPases in splicing, the functions of peptidyl-prolyl *cis-trans* isomerases (PPIases) are less clear. While most peptide bonds in biological systems are synthesized in the lower energy-state *trans* state, the bond preceding proline often occurs in *cis*, and can be interconverted to *trans*. The PPIases accelerate this proline isomerization, and targeted isomerization is an important regulatory mechanism for processes such as cell signaling and gene expression (see Introduction, Lu, Finn et al. 2007). There are at least eight PPIases that have been found by mass spectrometery to be associated with spliceosome components, almost all of them belonging to a sub-class of PPIases, cyclophilins, that generally have high affinity for the immunosuppressive drug cyclosporin A (Wang and Heitman 2005). While the spliceosome-associated cyclophilins show conservation of sequence amongst species, most of these cyclophilins have been lost from *Saccharomyces cerevisiae*, a common model organism used to study splicing. Interestingly, five of the eight spliceosome-associated cyclophilins (cyps) are present in the yeast *Schizosaccharomyces pombe*, and include Cyp1, Cyp8, Cyp9, Rct1, and Cwf27 (Figure D-1A). *S. pombe* Rct1 regulates phosphorylation of specific residues on RNA polymerase II's C-terminal domain (Gullerova, Barta et al. 2007). Cyp1 and Cwf27 have been found to associate with *S. pombe* U5.U2/U6 spliceosomes (Ohi, Link et al. 2002, Ren, McLean et al. 2011); however, no function has been associated with these proteins.

To further investigate the role of PPIases in the splicing reaction, we chose to examine *S. pombe* Cyp1, Cyp9, and Cwf27.

To confirm that these proteins are localized in the nucleus, we first attempted to C-terminally tag the endogenous loci of *cyp1*, *cyp9*, and *cwf27*. Live-cell imaging of GTP-tagged strains shows a strong, nuclear signal, indicating that the proteins are stable when tagged (Figure D-1B). We then performed TAP purifications using *cyp1-TAP*, *cyp9-TAP*, and *cwf27-TAP* strains. SDS-PAGE analysis followed by silver-staining of Cyp1-TAP bore a striking resemblance to the staining of Cwf2-TAP (Figure D-2A, lanes 3 and 1, respectively). Mass spectrometry analysis of Cwf2-TAP shows a similar protein composition to the U5.U2/U6 spliceosome (Ren, McLean et al. 2011). Additionally, the Cyp1-TAP and Cwf2-TAP negative-stain EM preparations showed particles of similar size and shape (Figure D-2B).

Similar characterization of the Cwf27-TAP and Cyp9-TAP purifications showed only a few bands by silver stain (Figure D-2A lanes 2 and 4), and no recognizable particles by EM (data not shown). Because the Cyp9-TAP purification yielded a strong band around the size expected for the bait protein (Figure D-2A "Cyp9-CBP"), it is likely that the Cyp9-TAP protein is stable but does not form strong associations with spliceosomes. Since Cwf27-TAP's expression was not confirmed by Western blot, and because there is no band near Cwf27-CBP's predicted molecular weight (Figure D-2A lane 2), these data are ambiguous as to whether Cwf27 was tagged improperly in this strain, or if the purification failed for technical reasons.

We next attempted to determine binding partners of Cwf27 using yeast two-hybrid analysis. Cwf27 contains a homolog in *S. cerevisiae*, Cwc27, which has been

A

human	*S. pombe*	*S. cerevisiae*
SDCCAG10	Cwf27/Cyp7	Cwc27
PPIL4	Rct1	n/a
PPIL1	Cyp1	n/a
PPWD1	Cyp9	n/a
PPIL2	Cyp8	n/a
PPIE	n/a	n/a
PPIG	n/a	n/a
PPIL3	n/a	n/a

B

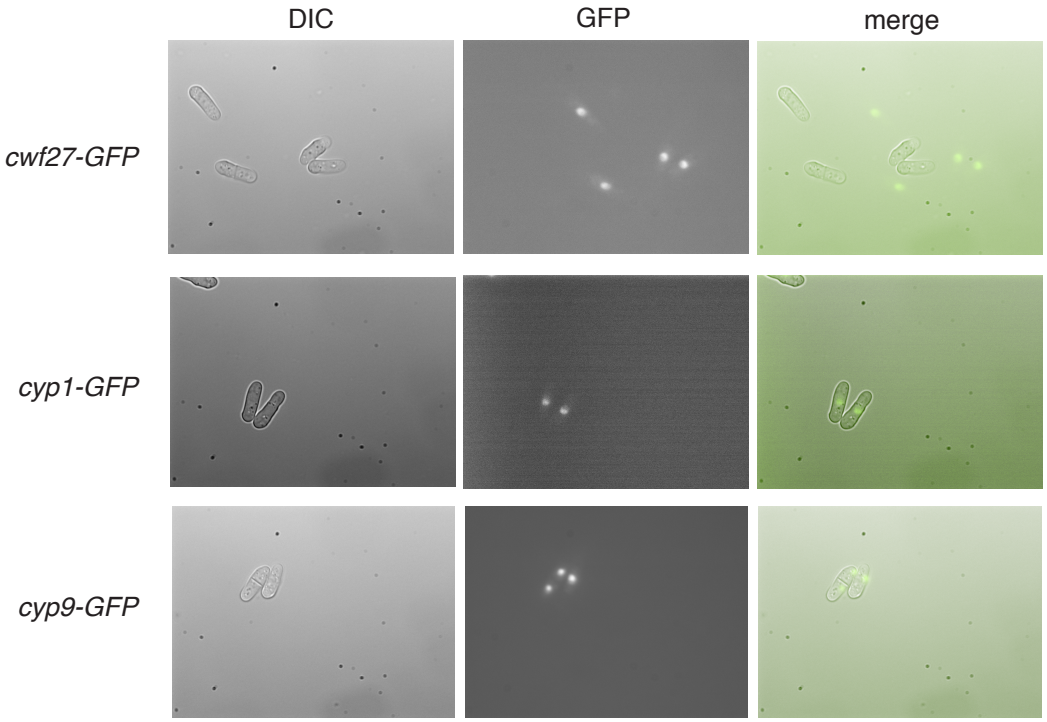


FIGURE D-1

Cyclophilin homologs and localization of GFP-tagged cyclophilins
A) Eight human spliceosome-associated cyclophilins are listed along with their homologs in fission and budding yeast. B) Cells were grown at room temperature and imaged live by DIC and a GFP channel. Positioning in the merge column is not accurate due to cell movement between the two imaging modes.

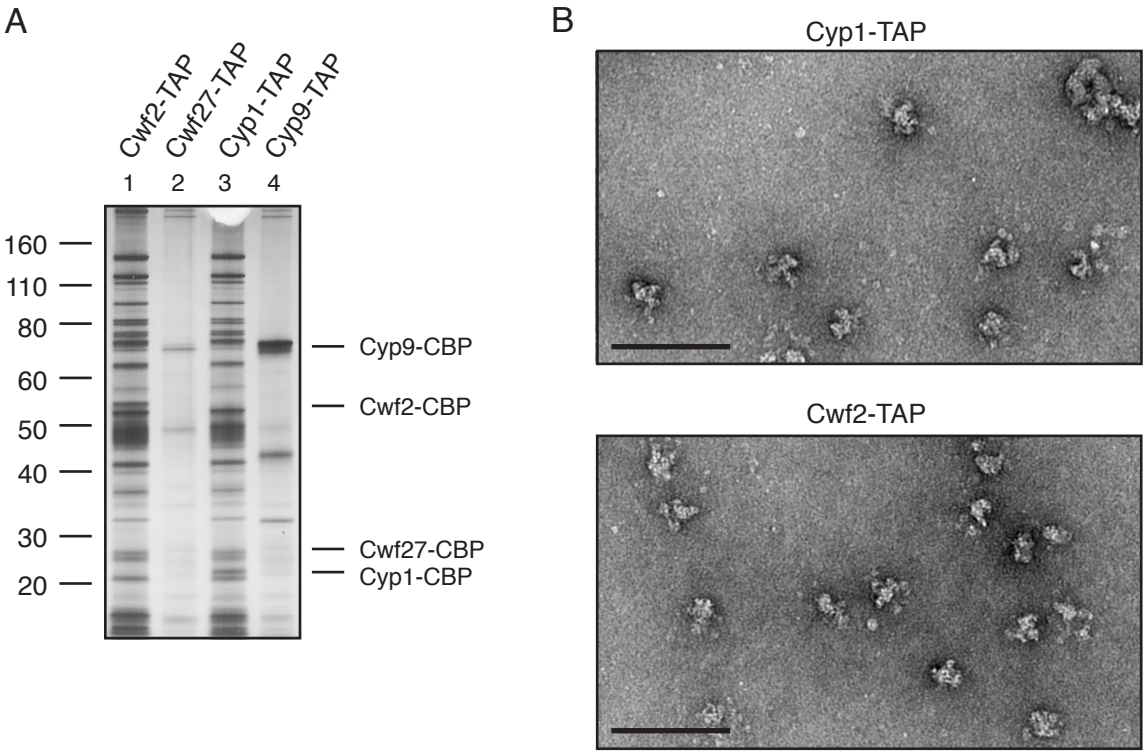


FIGURE D-2

Purification of cyclophilins from native lysates

A) Purifications of the indicated proteins were run on a 4-12% Bis-Tris SDS-PAGE gel and silver stained. Molecular weight markers are indicated on the left in kilodaltons. Protein names to the right of the gel indicate the predicted position of the bait fused to calmodulin binding protein (CBP). B) Purifications of the indicated proteins were negatively stained and imaged using an FEI Morgagni microscope at 100 kV. Scale bar represents 100 nm.

screened for genetic interactions in several large-scale studies. The interactions known with other spliceosome proteins were used to prepare targeted Y2H vectors for testing with Cwc27, and included the three subunits of the RES complex (Ist3, Bud13, and Pml1). Additionally, we tested two-hybrid interactions between Cwc27 and NTC components, for which vectors were readily available in our lab. Through this analysis, we found no positive interactions between Cwc27 and members of the RES and/or NTC complexes.

Next, we searched for potential dominant-negative growth effects by overexpression of the wild-type genes. We cloned either *cyp1* and *cwf27* cDNA into the pREP3X vector, as *cyp9* cDNA could not be isolated. Growth of strains carrying these vectors were not significantly reduced at 16, 25, or 37°C (Figure D-3A). However, spot assays, which can detect more subtle changes in growth, were not performed.

We investigated the effects of removing *cyp1*, *cyp9*, and *cwf27* from haploid *S. pombe* cells. Cells carrying the single deletions were viable and grew similar to wild-type at 25, 32, and 37°C (Figure D-3B). Additionally, the triple knockout strain *cyp1::ura4*$^+$ *cyp9::ura4*$^+$ *cwf27::ura4*$^+$ grew well at 25 and 37°C (Figure D-3C). From observation of live cells, it was unclear whether or not the triple mutant had abnormal cell morphology or length (data not shown), and cell length quantification was not performed.

Because of Cyp1's strong association with a possible spliceosome complex (Figure D-2A), we tested the *cyp1::ura4*$^+$ deletion strain for synthetic interactions with six known splicing alleles, *cdc5-120*, *cwf11Δ*, *dre4Δ*, *prp1-1*, *prp4-73*, and *prp5-1*. All double mutant strains were viable and showed no synthetic sick interactions when grown

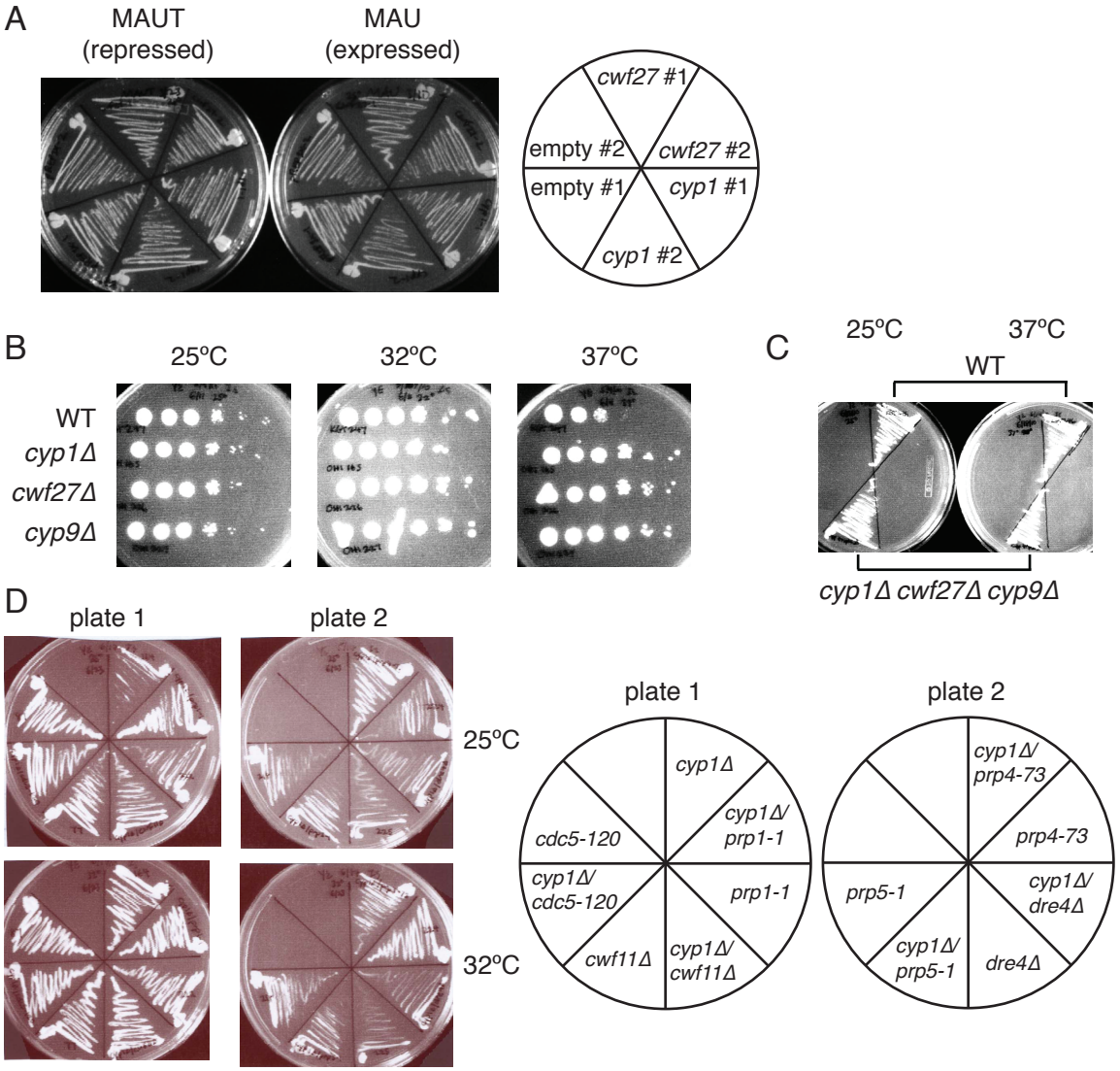


FIGURE D-3

Cyclophilin overexpression and deletion phenotypes
A) Wild-type cells were transformed with the pREP3X vector (empty) or with the indicated cyclophilin cDNA in pREP3X. Cells were grown at 25 on plates selecting for the plasmid and either repressing cyclophilin expression (MAU with 18.9µM thiamine) or expressing the cyclophilin (MAU). A pair of separate transformants was streaked for each genotype. B) Cells of the indicated genotypes were spotted on to YE plates at the indicated temperatures. C) Wild-type and *cyp1Δ cwf27Δ cyp9Δ* strains were streaked on to YE plates at the indicated temperatures. D) Single and double mutant strains were streaked on to YE at the indicated temperatures.

on YE at 25 and 32°C (Figure D-3D). Spot assays, which can detect more subtle changes in growth, were not performed.

From these experimental data, the only strong evidence for the involvement of any of these cyclophilins in splicing is the co-IP of a spliceosome-resembling particle by Cyp1-TAP (Figure D-2B). However, from my analysis, the enzymatic activity of these proteins does not seem to be important the splicing reaction, since even the triple knockout does not cause loss of cell viability. Although not essential for splicing, it is possible that the cyclophilins may be important for efficient splicing. One simple experiment to address this possibility would be to isolate cellular RNA from a wild-type control strain, each single deletion strain, and the triple-deletion strain and use RT-PCR analysis to determine if there is an accumulation of unspliced RNA in these strains.

APPENDIX E

MATERIALS AND METHODS

Strains, yeast methods, and molecular biology

Strains used in this study are listed in Table 1 of this appendix. Yeast strains were grown in yeast extract (YE) media or Edinburgh minimal media with appropriate supplements. The *spp41-1* (Schmidt, Richert et al. 1999) and *spp42-1* (Schmidt, Richert et al. 1999) ORFs were tagged endogenously at the 3' end with kanMX6 for genetic analyses as previously described (Bähler, Wu et al. 1998). Transformations were done as described (Keeney and Boeke 1994) for all tag insertions, gene replacements, and introduction of plasmids. Integration of tags was verified using whole-cell PCR and immunoblot analysis as appropriate. Crossing of tagged and mutated loci into other strains was accomplished using standard *S. pombe* mating, sporulation, and tetrad dissection techniques. For spot assays, cells were grown to midlog phase at 25°C and resuspended in water to achieve an equal OD_{595} value. 10-fold serial dilutions were made, and 2.5 µl of each dilution was plated on YE. Plates were incubated at the indicated temperatures for 2-9 days before imaging.

For induction of the nmt3X promoter (Maundrell 1993), the cells were first grown overnight in media containing 5 µg/ml of thiamine, then washed three times with media lacking thiamine and allowed to grow for at least 16 hours in thiamine-free media. For harvesting smaller cultures, the growth in 5 µg/ml of thiamine was omitted. Overexpression plasmids used include pREP3X *cwf10 1-135* (pOHI1038), pREP3X

cwf10 1-107 (pOHI1037), pREP41 NTAP (Tasto, Carnahan et al. 2001) (pOHI756), pREP41 NTAP-*cwf10 2-135* (pOHI1039), and pREP41 NTAP-*cwf10 2-26* (pOHI1076).

All plasmids were generated by standard molecular biology techniques. For gene replacements at the endogenous *cwf10*$^{+}$ locus, mutant open reading frames were generated using QuikChange II or QuikChange Lightning Multi technologies (Agilent Technologies, Santa Clara, CA). The mutant ORF and at least 500 bp of 5'- and 3'-flanking nucleotides were subcloned into the pIRT2 plasmid containing the *LEU2* marker. A diploid *cwf10*$^{+}$*/cwf10::ura4*$^{+}$ strain was transformed with pIRT2–*cwf10* mutant constructs and grown on minimal media lacking leucine, adenine, and uracil. Transformants were allowed to sporulate and stable haploid integrants were selected based on resistance to 5-FOA. Mutants were validated by whole-cell PCR with primers outside of the 5'- and 3'-flanking regions. pIRT2-*cwf10* mutant plasmids used include pOHI1035 (*2-135Δ*), pOHI978 (*2-127Δ*), pOHI1071 (*2-23Δ*), and pOHI1070 (*E/D-A*).

Liquid culture cell numbers were counted with a Z1 Coulter Counter (Beckman Coulter, Brea, CA) as previously described (Ohi, McCollum et al. 1994).

Random mutagenesis of *cwf10*$^{+}$:

A *kanMX6* cassette was inserted 3' of the *cwf10*$^{+}$ gene using OHI plasmid 725 and a previously described method (Bähler, Wu et al. 1998). The genomic region encompassing 500 bp upstream and downstream of *cwf10+::cwf10-kanMX6* was amplified by PCR. Error-prone PCR mutagenesis was performed on this template, essentially as described (Wilson and Keefe 2001). Briefly, the PCR reactions (300uL) contained 1X Taq buffer without magnesium chloride (NEB), 7 mM magnesium chloride,

0.5 mM manganese chloride, 0.2 mM dATP and dGTP, 1 mM dTTP and dCTP, 2 μM each primer, variable amounts of PCR template, and 15 units Taq polymerase (NEB). Full length *cwf10*: Reactions were split among several tubes and incubated under PCR cycling parameters: 30 sec at 95°C; 8 or 15 cycles of 30 sec at 95°C, and 7 min at 68°C; and 8 min at 68°C. C-terminal *cwf10*: Reactions were split among several tubes and incubated under PCR cycling parameters: 30 sec at 95°C; 15 cycles of 15 sec at 95°C, and 4 min at 68°C; and 8 min at 68°C. Reactions were recombined, the amplicon gel extracted or simply run through a DNA spin column, and transformed into strain OHI246. Transformed cells were plated on several YE plates, placed at 25°C, and the next day replica plated to YE + G418 at 25°C to select for transformants. Strains were then tested for conditional lethality by replica plating to YE and incubating at 36°C for several days. Several hundred clones were screened and nine clones (*cwf10-2* through *cwf10-10*) were chosen for further characterization. To determine the mutation locations, genomic DNA was prepared from each strain, the *cwf10* gene was PCR amplified, and the PCR product was directly sequenced.

Antibody preparation

His_6-Cwf10 (673-983) was expressed in *E. coli* BL21(DE3) cells (EMD Millipore, Battymarch Park, MA), purified with HisPur Cobalt agarose (Pierce/Thermo Scientific, Rockford, IL), and used to immunize rabbits (Cocalico Biologicals, Reamstown, PA). Cwf10-specific antibodies were affinity purified over NHS-activated sepharose fast flow 4 (GE Healthcare Life Sciences, Piscataway, NJ) covalently linked to His_6-Cwf10 (673-983).

Immunoprecipitations, immunoblotting, and sucrose gradients

"Native" and "denatured" whole-cell lysates were prepared as previously described, with leupeptin omitted from lysis buffers (McDonald, Ohi et al. 1999). For immunoblots, proteins were resolved by 10% SDS-PAGE (lysates except Brr2-HA), 4-12% Bis-Tris PAGE (Brr2-HA lysates), or 8% SDS-PAGE (lysate gradients) and transferred by electroblotting to a Protran nitrocellulose membrane (Whatman, GE Healthcare, Piscataway, NJ). Primary antibodies/antisera used include: anti-Cdc5 (1/5,000) (McDonald, Ohi et al. 1999) and anti-Asp1 (1/5,000) rabbit polyclonal antisera, anti-Cwf10 affinity purified rabbit polyclonal antibody (0.84 μg/ml), and anti-PSTAIR (detects *S. pombe* Cdc2) monoclonal antibody (1/10,000) (Sigma, St. Louis, MO). For some of the Cwf10 quantification, the antiserum was used instead of the antibody. Anti-myc (9E10, 0.3 μg/ml) and anti-HA (12CA5, 1 μg/ml) mouse monoclonal antibodies were used to detect Myc- and HA-tagged proteins, respectively. Primary antibodies were detected by secondary antibodies AlexaFluor 700 (Invitrogen, Life Technologies, Grand Island, NY) or IRDye 800CW (LI-COR Biosciences, Lincoln, NE) (1/10,00 dilution), and visualized and quantified using an Odyssey scanner and software (LI-COR Biosciences).

For gradients, a 15OD pellet was lysed and a 200 μl volume corresponding to 25% of the lysate was layered onto a 10 to 30% sucrose gradient and centrifuged at 25,000 RPM at 4°C for 16 h in a SW55Ti rotor (Beckman). Fractions from the gradients were collected manually and (a) TCA precipitated and resuspended in SDS sample buffer to detect protein or (b) extracted with 5:1 acid phenol:chloroform and resuspended in TBE-urea sample buffer (Life Technologies, Grand Island, NY) to detect RNA. Parallel

standard gradients contained thyroglobulin (19 S) and catalase (11.35 S) (HMW calibration kit; GE Healthcare, Piscataway, NJ) or 20% of lysate from a 20 OD pellet of *FAS2-V5* tagged *S. cerevisiae* (strain OHI375) (40 S marker).

Co-IPs for Western blot analysis were performed as in (Ren, McLean et al. 2011). TAPs (Tandem Affinity Purifications) were performed as described (Gould, Ren et al. 2004), with the following modifications: native lysis buffer was as described earlier in this section, calmodulin binding volume was reduced by 3 ml, CBB washes were 10 ml and 2 ml, respectively, and the final CBB buffer and the CEB buffer contained no detergent. For TAPs analyzed in Table 2-1 purifications were done using 75 mM NaCl.

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 either found in fractions collected from sucrose gradients or from NTAP-NTE and anti-snRNA cap (antitrimethylguanosine [m_3G], Millipore), RNAs were resolved using 6% Tris-Borate-EDTA–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 crosslinker; GE Healthcare, Piscataway, NJ), and detected by using γ-^{32}P ATP (PerkinElmer, Waltham, MA) labeled oligonucleotides complementary to *S. pombe* U1, U2, U4, U5, and U6 (for sequences see Table 2 of this appendix). Blots were exposed to phosphorimager screens for 14-18 hours and visualized using Typhoon 9200 or FLA-7000IP instruments (GE Healthcare, Piscataway, NJ). Quantification for snRNAs found in fractions collected from sucrose gradients was performed using ImageQuant TL 8.1

(GE Healthcare). For 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). 800 ng of RNA was used for each reaction. Three technical samples per genotype were processed. The resulting cDNA was PCR-amplified with *tbp1_a* or *mrps16_b* primers (Table 2, this appendix) for 27 cycles, and ethidium bromide-stained gels were imaged within a linear range and quantified with ImageQuant TL 8.1 (GE Healthcare).

RNA sequencing and analyses

Replicates were grown together and processed separately for all following steps. The Vanderbilt Technologies for Advanced Genomics Core Facility (VANTAGE, Nashville, TN) used the TruSeq Stranded mRNA Sample Preparation Kit (Illumina, San Diego, CA) to convert the mRNA in 100 ng of total RNA into a library of template molecules suitable for subsequent cluster generation and sequencing on the Illumina HiSeq 2500. The input total RNA was quality-checked by running an aliquot on the Agilent Bioanalyzer to confirm integrity. The Qubit RNA fluorometry assay was used to measure concentration. The input to library prep was 50 µl of 2 ng/µl DNase-treated total RNA. The total RNA underwent enrichment of the poly-A containing mRNA molecules using poly-T oligo-attached magnetic beads. Following purification, the eluted poly(A) RNA was cleaved into small fragments of 120-210 bp using divalent cations under elevated temperature. The cleaved RNA fragments were copied into first strand cDNA using SuperScript II reverse transcriptase and random primers. This was followed by second strand cDNA synthesis using DNA Polymerase I and RNase H. The cDNA

fragments were then put through an end-repair process, the addition of a single 'A' base, and ligation to the Illumina multiplexing adapters. The products were then purified and enriched with PCR to create the final cDNA sequencing library. The cDNA library underwent quality control by running on the Agilent Bioanalyzer HS DNA assay to confirm the final library size and on the Agilent Mx3005P qPCR machine using the KAPA Illumina library quantification kit to determine concentration. A 2 nM stock was created and samples pooled by molarity for multiplexing. From the pool, 12 pM was loaded into each well for the flow cell on the Illumina cBot for cluster generation. The flow cell was then loaded onto the Illumina HiSeq 2500 utilizing v3 chemistry and HTA 1.8. The raw sequencing reads in BCL format were processed through CASAVA-1.8.2 for FASTQ conversion and demultiplexing. The RTA chastity filter was used and only the PF (passfilter) reads were retained for further analysis. Raw expression data files are available from Gene Expression Omnibus (GEO accession: GSE47573; http://www.ncbi.nlm.nih.gov/geo/).

Genome level alignments and annotation. Paired-end reads of 76 base length (each end) originating from each sample were aligned using Bowtie 0.12.7 (Langmead, Trapnell et al. 2009) to the *S. pombe* genome sequence (Ensembl *S. pombe*, Build EF1, version 13) (Kersey, Lawson et al. 2010) as well as to the corresponding exon-exon junctions database (only the first part of the paired-end reads was considered). Up to 3 base-pair mismatches were allowed. Reads that matched multiple loci were removed from further analysis, and the resultant alignment files were processed to generate 'pile-ups' against each chromosome.

Exon-exon junctions. Searches were performed against the genome sequence combined with a dataset of known exon-exon junctions as defined by Ensembl *S. pombe* release-13. To ensure that a 76 base read mapped to a splice junction, only the last 70 bases of the first exon and the first 70 bases of the second exon were considered (if the exon exceeded length 70). In this way, reads that overlapped a junction by less than 6 nucleotides were excluded. Reads that matched to more than one junction or elsewhere in the genome were also discarded.

Defining Intronic, Antisense, and Intergenic Regions. The known annotated set of *S. pombe* loci (7,022; Ensembl version 13, as before) was used to define unambiguous antisense loci (*i.e.,* those that exactly mirror known annotated transcripts without overlapping nearby genes), and unique accessions were assigned ("anti_xxx"; 3,097). Using this augmented annotation, intergenic regions were defined (*i.e.* regions between known annotated and antisense loci on each strand), and unique accessions were assigned ("inter_xxx"; 8,810). Unique accessions were also assigned to all known introns (5,361). Thus, in total 21,193 loci were interrogated across the 4 samples.

Normalization, Fold Changes, and Differential Expression. Differential expression between samples was determined using the DESeq Bioconductor package (Anders and Huber 2010). A cutoff of ±2 fold-change and corrected p-value <0.05 were applied to derive a list of differentially expressed loci.

RNA-seq Expression level. Normalized expression levels (E) for individual exons and introns were calculated using the formula [1] as described in (Mortazavi, Williams et al. 2008, Bradford, Hey et al. 2010) (RPKM measure). Briefly, the number of reads (R) detected across a given region at a given sample (i) was multiplied by a constant (C =

1×10^9) and divided by the total number of reads at that sample (Ti) multiplied by the region's length (L).

[1]

$$E = \log_2\left(C\frac{R_i}{T_i L}\right)$$

A small constant was added (10^{-5}) to all expression values to avoid taking logarithms of zero. Gene level expression values were summarized using exon data.

Splicing Efficiency and Differential Splicing Significance. Splicing efficiency (SE) reflects the proportion of spliced mRNA signal relative to pre-mRNA signal. Splicing efficiency is computed by dividing junction reads (JR; aka trans-reads) by reads that straddle an exon-intron boundary (only the upstream 5' exon relative to the intron was considered; EI; formula 2).

[2]

$$SE = \log_2\left(\frac{JR}{EI}\right)$$

A Cochran-Mantel-Haenszel (CMH) chi-squared test for repeated test of independence, which accounts for biological replicates (Agresti 1990), was applied to identify statistically significant introns (*i.e.*, those that display differences in their splicing efficiency between samples; mantelhaen.test command in R). The False Discovery Rate (q-value) was computed using the bioconductor qvalue package (Storey 2002) and a cutoff of $q < 0.05$ applied.

Expression and Purification of Cwf10(1-135)His_6

The N-terminal 1-135 amino acid sequence of Cwf10 with 6 C-terminal histidine residues was cloned into pET15b (NcoI/BamHI; plasmid pOHI1020) (EMD Millipore) and transformed into *E. coli* Rosetta 2 (DE3) pLysS cells (EMD Millipore). Cells were grown in Terrific broth (Invitrogen, Grand Island, NY) to an OD_{595} of ~0.8 and cold shocked for 20 minutes on ice. Upon addition of 1 mM IPTG, the plasmid was overexpressed for 20h at 15°C. Cells were lysed in PBS (pH 7.0), 350 mM NaCl, 0.1% Triton X-100 and one SIGMA*FAST* protease tablet (Sigma-Aldrich, St. Louis, MO). Cwf10(1-135)His_6 was purified using two 5 ml Histrap HP columns (GE Healthcare, Waukesha, WI) and a 2.5-500 mM imidazole linear gradient. Cwf10(1-135)His_6 was further purified using anion-exchange (Uno Q1, Bio-Rad) chromatography in buffer A (50 mM Tris-HCl, pH 8.0) using a 0-1 M NaCl linear gradient, followed by gel filtration (Superdex 200, GE Healthcare) in 25 mM Tris-HCl (pH 7.3), 100 mM NaCl, and 1 mM EDTA.

Circular Dichroism

Purified Cwf10(1-135)His_6 was analyzed using a Jasco J-810 spectropolarimeter (Jasco Analytical Instruments, Easton, MD). Far-UV data were collected at a protein concentration of 0.18 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 20 nm/min from 198 to 260 nm. Far-UV spectra were collected in duplicate and background-corrected against a buffer blank. Spectra were analyzed using the program K2D2 (Perez-Iratxeta and Andrade-Navarro 2008) to estimate secondary structure. Near-UV data were collected at a protein

concentration of 2.01 mg/ml in a 1 cm quartz cuvette. Spectra were collected with an average time of 4 s for each point and a step size of 10 nm/min from 250 to 330 nm. For both Cwf10(1-135)His_6 and Cwf10(1-135)His_6 in 6 M guanidine-HCl five spectra were collected for near-UV data and background-corrected against a buffer blank. Data were converted to mean residue ellipticity $[\theta]_m$ (degrees cm^2dmol^{-1}) using $[\theta]_m = \theta/(10lcn)$, where θ is the measured ellipticity, l is the cell path length in cm, c is the molar concentration of protein in mol/liter, and n is the number of residues/chain.

Analytical ultracentrifugation

Purified Cwf10(1-135)His_6 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 at 0.45 mg/ml. Samples were loaded into double-sector cells (path length of 1.2 cm) with charcoal-filled Epon centerpieces and saphire windows. Sedfit (version 12.0) (Schuck 2000) was used to analyze velocity scans using every five scans from a total of 167 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 5 S.

NMR analysis

Cwf10(1-135)His_6 was purified as above except cells were grown and expressed in M9 media supplemented with ^{15}N-labeled ammonium chloride (Cambridge Isotopes, Andover, MA) as the only nitrogen source, purified as described above, and 10% D_2O was added to the final sample. During purification half of the sample was left in gel filtration buffer (25 mM Tris-HCl [pH 7.3], 100 mM NaCl, and 1 mM EDTA) and the

other half was buffer exchanged into gel filtration buffer plus 6 M guanidine-HCl. Standard sensitivity enhanced echo/anti-echo ^{15}N-^{1}H HSQC NMR data were collected at 25°C for both samples and at 50°C for the gel filtration buffer sample (no Guanidine-HCl) using a 600 MHz Bruker AVIII spectrometer (Bruker, Billerica, MA) with a CPQCI probe and z-axis gradient. The spectra were processed using Topspin 3.2 (Bruker, Billerica, MA). The indirect dimension was four times zero filled to a final matrix of 2048 x 1024 data points and 72^0 and 90^0 squared sine bell apodization was applied in the F2 and F1 dimensions respectively. Spectra were further analyzed with Sparky (T. D. Goddard and D. G. Kneller, University of California, San Francisco).

Mass spectrometry

TAP elutions were trichloroacetic acid (TCA)-precipitated, resolubilized in 8 M urea, 100 mM Tris pH 8.5, reduced and alkylated, and then diluted back to 2 M urea and digested overnight with trypsin as described (MacCoss, McDonald et al. 2002). Resulting peptides (corresponding to about 5% of the TAP eluate) were analyzed by a 70 minute data-dependent 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 (MS/MS) was collected throughout the run and dynamic exclusion was enabled to minimize acquisition of redundant spectra. MS/MS spectra were searched via SEQUEST against a *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).

Negative-stain EM

Uranyl formate-stained samples were prepared for EM as described (Ohi, Li et al. 2004). In brief, 2.5 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 Morgagni electron microscope (FEI, Hillsboro, OR) operated at an acceleration voltage of 100 kV. Images were recorded at a magnification of 28,000 and collected using a 1K x 1K CCD camera (AMT, Wobum, MA).

Appendix E Table 1: *Schizosaccharomyces pombe* strains used in this study

Strain	Genotype	Source
OHI246	*h- leu1-32 ura4-D18 ade6-M210*	K. Gould[a]
OHI408	*h- cwf10 2-127Δ ade6-M21X leu1-32 ura4-D18*	This work
OHI479	*h+ brr2-HA::kanR ade6-M210 leu1-32 ura4-D18*	Adapted from K. Gould[a]
OHI480	*h+ prp3-linker3XV5 cwf10 2-127Δ ade6-M21X leu1-32 ura4-D18*	This work
OHI371	*h+ prp3-linker3XV5*	This work
OHI482	*h- brr2-HA::kanR cwf10 2-127Δ ade6-M21X leu1-32 ura4-D18*	This work
OHI323	*h- prp1-myc13X::kanR ade6-M210 leu1-32 ura4-D18*	Carnahan, et al., 2005
OHI439	*h- prp1-myc13X::kanR cwf10 2-127Δ ade6-M21X leu1-32 ura4-D18*	This work
OHI448	*h- cwf10 2-127Δ::kanR ade6-M21X leu1-32 ura4-D18*	This work
OHI463	*h+ cdc28-P8*	Adapted from Nasmyth and Nurse, 1981
OHI467	*h+ cdc28-P8 cwf10 2-127Δ::kanR ade6-M21X ura4-D18*	This work
OHI464	*h+ prp10-1*	Adapted from Urushiyama, et al., 1996
OHI468	*h+ prp10-1 cwf10 2-127Δ::kanR*	This work
OHI466	*h+ spp42-1::kanR leu1-32*	Adapted from Schmidt, et al, 1999
OHI465	*h+ spp41-1::kanR leu1-32 ura4-D18*	Adapted from Schmidt, et al, 1999
OHI471	*h+ spp41-1::kanR cwf10 2-127Δ ade6-M21X leu1-32 ura4-D18*	This work
OHI322	*h+ prp1-4 ura4-D18*	Adapted from Urushiyama, et al., 1996
OHI321	*h+ prp4-73 leu1-32 ura4-294*	Adapted from Rosenberg, et al., 1991
OHI462	*h- prp4-73 cwf10 2-127Δ::kanR leu1-32 ura4-*	This work

	D18 or -294	
OHI478	*h+ aar2::kanR ade6-M216 leu1-32 ura4-D18*	Kim, et al., 2010
OHI481	*h- aar2::kanR cwf10 2-127Δ ade6-M21X leu1-32 ura4-D18*	This work
OHI487	*h+ prp17::ura4+ ade6-M21X leu1-32 ura4-D18*	Sapra, et al., 2004
OHI488	*h+ prp17::ura4+ cwf10 2-127Δ::kanR ade6-M21X leu1-32 ura4-D18*	This work
OHI483	*h+ cwf10 2-23Δ ade6-M21X leu1-32 ura4-D18*	This work
OHI383	*h-/h+ cwf10+/cwf10::ura4+ ade6-M210/ade6-M216 leu1-32/leu1-32 ura4-D18/ura4-D18*	This work
OHI260	*h- cdc5-TAP::kanR*	Ohi, et al., 2002

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

Appendix E, Table 2: Primers used in this study

Analysis	Name	Purpose	Primer sequence(s) (5'-3')
RT-PCR	tbp1_a F BL365	amplify first *tbp1* intron	CGCTTTACCCACCACGGCCTCGCAAG
RT-PCR	tbp1_a R BL366	amplify first *tbp1* intron	TTCTGCATTACGTGCATGTAGCGC
RT-PCR	mrps16_b F BL541	amplify second *mrps16* intron	AAAGCACCTCAAGCCAAACCTATCGA
RT-PCR	mrps16_b R BL542	amplify second *mrps16* intron	TCAAGAAACGGGTTTTTTAGGAAGCAACT
Northern	BL432	hybridize U1	ATAACTGAGTCTCCACAATTCTTGTG
Northern	BL429	hybridize U2	AGAACAGATACTACACTTGATCTAAG
Northern	BL433	hybridize U4	AAAAGTTCCTCACTGATAAGCGTAAT
Northern	BL434	hybridize U5	ACAGTCAAATTAGCACACCTTACAAA
Northern	BL435	hybridize U6	TCTTCTCTGTATCGTTTCAATTTGAC

BIBLIOGRAPHY

Achsel, T., K. Ahrens, H. Brahms, S. Teigelkamp and R. Lührmann (1998). "The human U5-220kD protein (hPrp8) forms a stable RNA-free complex with several U5-specific proteins, including an RNA unwindase, a homologue of ribosomal elongation factor EF-2, and a novel WD-40 protein." Mol Cell Biol **18**(11): 6756-6766.

Agafonov, D. E., J. Deckert, E. Wolf, P. Odenwälder, S. Bessonov, C. L. Will, H. Urlaub and R. Lührmann (2011). "Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method." Mol Cell Biol **31**(13): 2667-2682.

Agresti, A. (1990). Categorical data analysis. New York, Wiley.

Albuquerque, C. P., M. B. Smolka, S. H. Payne, V. Bafna, J. Eng and H. Zhou (2008). "A multidimensional chromatography technology for in-depth phosphoproteome analysis." Mol Cell Proteomics **7**(7): 1389-1396.

Anders, S. and W. Huber (2010). "Differential expression analysis for sequence count data." Genome Biol **11**(10): R106.

Ansari, A. and B. Schwer (1995). "SLU7 and a novel activity, SSF1, act during the PRP16-dependent step of yeast pre-mRNA splicing." EMBO J **14**(16): 4001-4009.

Bach, M., G. Winkelmann and R. Lührmann (1989). "20S small nuclear ribonucleoprotein U5 shows a surprisingly complex protein composition." Proc Natl Acad Sci U S A **86**(16): 6038-6042.

Bähler, J., J. Q. Wu, M. S. Longtine, N. G. Shah, A. McKenzie, 3rd, A. B. Steever, A. Wach, P. Philippsen and J. R. Pringle (1998). "Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*." Yeast **14**(10): 943-951.

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.

Bayne, E. H., M. Portoso, A. Kagansky, I. C. Kos-Braun, T. Urano, K. Ekwall, F. Alves, J. Rappsilber and R. C. Allshire (2008). "Splicing factors facilitate RNAi-directed silencing in fission yeast." Science **322**(5901): 602-606.

Behrens, S. E., F. Galisson, P. Legrain and R. Lührmann (1993). "Evidence that the 60-kDa protein of 17S U2 small nuclear ribonucleoprotein is immunologically and functionally related to the yeast PRP9 splicing factor and is required for the efficient formation of prespliceosomes." Proc Natl Acad Sci U S A **90**(17): 8229-8233.

Behrens, S. E., K. Tyc, B. Kastner, J. Reichelt and R. Lührmann (1993). "Small nuclear ribonucleoprotein (RNP) U2 contains numerous additional proteins and has a bipartite RNP structure under splicing conditions." Mol Cell Biol **13**(1): 307-319.

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**(6): 1737-1748.

Bellare, P., E. C. Small, X. Huang, J. A. Wohlschlegel, J. P. Staley and E. J. Sontheimer (2008). "A role for ubiquitin in the spliceosome assembly pathway." Nat Struct Mol Biol **15**(5): 444-451.

Ben-Shem, A., L. Jenner, G. Yusupova and M. Yusupov (2010). "Crystal structure of the eukaryotic ribosome." Science **330**(6008): 1203-1209.

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.

Bindereif, A. and M. R. Green (1987). "An ordered pathway of snRNP binding during mammalian pre-mRNA splicing complex assembly." EMBO J **6**(8): 2415-2424.

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.

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.

Boon, K. L., C. M. Norman, R. J. Grainger, A. J. Newman and J. D. Beggs (2006). "Prp8p dissection reveals domain structure and protein interaction sites." RNA **12**(2): 198-205.

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**(3): 151-161.

Bradford, J. R., Y. Hey, T. Yates, Y. Li, S. D. Pepper and C. J. Miller (2010). "A comparison of massively parallel nucleotide sequencing with oligonucleotide microarrays for global transcription profiling." BMC Genomics **11**: 282.

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.

Bringmann, P., B. Appel, J. Rinke, R. Reuter, H. Theissen and R. Lührmann (1984). "Evidence for the existence of snRNAs U4 and U6 in a single ribonucleoprotein complex and for their association by intermolecular base pairing." EMBO J **3**(6): 1357-1363.

Brosi, R., H. P. Hauri and A. Kramer (1993). "Separation of splicing factor SF3 into two components and purification of SF3a activity." J Biol Chem **268**(23): 17640-17646.

Bryson, K., L. J. McGuffin, R. L. Marsden, J. J. Ward, J. S. Sodhi and D. T. Jones (2005). "Protein structure prediction servers at University College London." Nucleic Acids Res **33**(Web Server issue): W36-38.

Buhler, M. and D. Moazed (2007). "Transcription and RNAi in heterochromatic gene silencing." Nat Struct Mol Biol **14**(11): 1041-1048.

Carnahan, R. H., A. Feoktistova, L. Ren, S. Niessen, J. R. Yates, 3rd and K. L. Gould (2005). "Dim1p is required for efficient splicing and export of mRNA encoding lid1p, a component of the fission yeast anaphase-promoting complex." Eukaryot Cell **4**(3): 577-587.

Chan, S. P., D. I. Kao, W. Y. Tsai and S. C. Cheng (2003). "The Prp19p-associated complex in spliceosome activation." Science **302**(5643): 279-282.

Chanarat, S. and K. Strasser (2013). "Splicing and beyond: The many faces of the Prp19 complex." Biochim Biophys Acta **1833**(10): 2126-2134.

Cheng, S. C., A. N. Newman, R. J. Lin, G. D. McFarland and J. N. Abelson (1990). "Preparation and fractionation of yeast splicing extract." Methods Enzymol **181**: 89-96.

Chiang, T. W. and S. C. Cheng (2013). "A weak spliceosome-binding domain of Yju2 functions in the first step and bypasses Prp16 in the second step of splicing." Mol Cell Biol **33**(9): 1746-1755.

Chiara, M. D., L. Palandjian, R. Feld Kramer and R. Reed (1997). "Evidence that U5 snRNP recognizes the 3' splice site for catalytic step II in mammals." EMBO J **16**(15): 4746-4759.

Coelho Ribeiro, M. L., J. Espinosa, S. Islam, O. Martinez, J. J. Thanki, S. Mazariegos, T. Nguyen, M. Larina, B. Xue and V. N. Uversky (2013). "Malleable ribonucleoprotein machine: protein intrinsic disorder in the *Saccharomyces cerevisiae* spliceosome." PeerJ **1**: e2.

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

Cordin, O., D. Hahn and J. D. Beggs (2012). "Structure, function and regulation of spliceosomal RNA helicases." Curr Opin Cell Biol **24**(3): 431-438.

Deckert, J., K. Hartmuth, D. Boehringer, N. Behzadnia, C. L. Will, B. Kastner, H. Stark, H. Urlaub and R. Lührmann (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.

Dix, I., C. S. Russell, R. T. O'Keefe, A. J. Newman and J. D. Beggs (1998). "Protein-RNA interactions in the U5 snRNP of *Saccharomyces cerevisiae*." RNA **4**(12): 1675-1686.

Dumesic, P. A., P. Natarajan, C. Chen, I. A. Drinnenberg, B. J. Schiller, J. Thompson, J. J. Moresco, J. R. Yates, 3rd, D. P. Bartel and H. D. Madhani (2013). "Stalled spliceosomes are a signal for RNAi-mediated genome defense." Cell **152**(5): 957-968.

Ekwall, K., G. Cranston and R. C. Allshire (1999). "Fission yeast mutants that alleviate transcriptional silencing in centromeric flanking repeats and disrupt chromosome segregation." Genetics **153**(3): 1153-1169.

Fabrizio, P. and J. Abelson (1990). "Two domains of yeast U6 small nuclear RNA required for both steps of nuclear precursor messenger RNA splicing." Science **250**(4979): 404-409.

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.

Fantes, P. A. (1977). "Control of cell size and cycle time in *Schizosaccharomyces pombe*." J Cell Sci **24**: 51-67.

Fourmann, J. B., J. Schmitzova, 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.

Frazer, L. N., S. C. Lovell and R. T. O'Keefe (2009). "Analysis of synthetic lethality reveals genetic interactions between the GTPase Snu114p and snRNAs in the catalytic core of the *Saccharomyces cerevisiae* spliceosome." Genetics **183**(2): 497-515-491SI-494SI.

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**(6): 927-938.

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.

Grainger, R. J. and J. D. Beggs (2005). "Prp8 protein: at the heart of the spliceosome." RNA **11**(5): 533-557.

Grewal, S. I. and S. Jia (2007). "Heterochromatin revisited." Nat Rev Genet **8**(1): 35-46.

Grey, M., A. Dusterhoft, J. A. Henriques and M. Brendel (1996). "Allelism of PSO4 and PRP19 links pre-mRNA processing with recombination and error-prone DNA repair in *Saccharomyces cerevisiae*." Nucleic Acids Res **24**(20): 4009-4014.

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

Gullerova, M., A. Barta and Z. J. Lorkovic (2007). "Rct1, a nuclear RNA recognition motif-containing cyclophilin, regulates phosphorylation of the RNA polymerase II C-terminal domain." Mol Cell Biol **27**(10): 3601-3611.

Habara, Y., S. Urushiyama, T. Shibuya, Y. Ohshima and T. Tani (2001). "Mutation in the prp12+ gene encoding a homolog of SAP130/SF3b130 causes differential inhibition of pre-mRNA splicing and arrest of cell-cycle progression in *Schizosaccharomyces pombe*." RNA **7**(5): 671-681.

Hacker, I., B. Sander, M. M. Golas, E. Wolf, E. Karagoz, 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.

Hartmuth, K., H. Urlaub, H. P. Vornlocher, C. L. Will, M. Gentzel, M. Wilm and R. Lührmann (2002). "Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method." Proc Natl Acad Sci U S A **99**(26): 16719-16724.

Hartwell, L. H. (1967). "Macromolecule synthesis in temperature-sensitive mutants of yeast." J Bacteriol **93**(5): 1662-1670.

He, Y., G. R. Andersen and K. H. Nielsen (2010). "Structural basis for the function of DEAH helicases." EMBO Rep **11**(3): 180-186.

Hegele, A., A. Kamburov, A. Grossmann, C. Sourlis, S. Wowro, M. Weimann, C. L. Will, V. Pena, R. Lührmann and U. Stelzl (2012). "Dynamic protein-protein interaction wiring of the human spliceosome." Mol Cell **45**(4): 567-580.

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.

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.

Jankowsky, E., C. H. Gross, S. Shuman and A. M. Pyle (2000). "The DExH protein NPH-II is a processive and directional motor for unwinding RNA." Nature **403**(6768): 447-451.

Jankowsky, E., C. H. Gross, S. Shuman and A. M. Pyle (2001). "Active disruption of an RNA-protein interaction by a DExH/D RNA helicase." Science **291**(5501): 121-125.

Jones, D. T. (1999). "Protein secondary structure prediction based on position-specific scoring matrices." J Mol Biol **292**(2): 195-202.

Jones, M. H., D. N. Frank and C. Guthrie (1995). "Characterization and functional ordering of Slu7p and Prp17p during the second step of pre-mRNA splicing in yeast." Proc Natl Acad Sci U S A **92**(21): 9687-9691.

Jørgensen, R., P. A. Ortiz, A. Carr-Schmid, P. Nissen, T. G. Kinzy and G. R. Andersen (2003). "Two crystal structures demonstrate large conformational changes in the eukaryotic ribosomal translocase." Nat Struct Biol **10**(5): 379-385.

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**(3): 265-269.

Karaduman, R., P. Dube, H. Stark, P. Fabrizio, B. Kastner and R. Lührmann (2008). "Structure of yeast U6 snRNPs: arrangement of Prp24p and the LSm complex as revealed by electron microscopy." RNA **14**(12): 2528-2537.

Kastner, B., N. Fischer, M. M. Golas, B. Sander, P. Dube, D. Boehringer, K. Hartmuth, J. Deckert, F. Hauer, E. Wolf, H. Uchtenhagen, H. Urlaub, F. Herzog, J. M. Peters, D. Poerschke, R. Lührmann and H. Stark (2008). "GraFix: sample preparation for single-particle electron cryomicroscopy." Nat Methods **5**(1): 53-55.

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.

Kersey, P. J., D. Lawson, E. Birney, P. S. Derwent, M. Haimel, J. Herrero, S. Keenan, A. Kerhornou, G. Koscielny, A. Kähäri, R. J. Kinsella, E. Kulesha, U. Maheswari, K. Megy, M. Nuhn, G. Proctor, D. Staines, F. Valentin, A. J. Vilella and A. Yates (2010). "Ensembl Genomes: extending Ensembl across the taxonomic space." Nucleic Acids Res **38**(Database issue): D563-569.

Kim, D. U., J. Hayles, D. Kim, V. Wood, H. O. Park, M. Won, H. S. Yoo, T. Duhig, M. Nam, G. Palmer, S. Han, L. Jeffery, S. T. Baek, H. Lee, Y. S. Shim, M. Lee, L. Kim, K. S. Heo, E. J. Noh, A. R. Lee, Y. J. Jang, K. S. Chung, S. J. Choi, J. Y. Park, Y. Park, H. M. Kim, S. K. Park, H. J. Park, E. J. Kang, H. B. Kim, H. S. Kang, H. M. Park, K. Kim, K. Song, K. B. Song, P. Nurse and K. L. Hoe (2010). "Analysis of a genome-wide set of gene deletions in the fission yeast *Schizosaccharomyces pombe*." Nat Biotechnol **28**(6): 617-623.

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.

Korneta, I., M. Magnus and J. M. Bujnicki (2012). "Structural bioinformatics of the human spliceosomal proteome." Nucleic Acids Res **40**(15): 7046-7065.

Kuhn, A. N., Z. Li and D. A. Brow (1999). "Splicing factor Prp8 governs U4/U6 RNA unwinding during activation of the spliceosome." Mol Cell **3**(1): 65-75.

Kuraoka, I., S. Ito, T. Wada, M. Hayashida, L. Lee, M. Saijo, Y. Nakatsu, M. Matsumoto, T. Matsunaga, H. Handa, J. Qin, Y. Nakatani and K. Tanaka (2008). "Isolation of XAB2 complex involved in pre-mRNA splicing, transcription, and transcription-coupled repair." J Biol Chem **283**(2): 940-950.

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. Blocker, 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 and Y. J. Chen (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.

Langmead, B., C. Trapnell, M. Pop and S. L. Salzberg (2009). "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome." Genome Biol **10**(3): R25.

Lardelli, R. M., J. X. Thompson, J. R. Yates, 3rd 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.

Legerski, R. J. (2009). "The Pso4 complex splices into the DNA damage response." Cell Cycle **8**(21): 3448-3449.

Lerner, M. R. and J. A. Steitz (1979). "Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus." Proc Natl Acad Sci U S A **76**(11): 5495-5499.

Leung, A. K., K. Nagai and J. Li (2011). "Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis." Nature **473**(7348): 536-539.

Li, X., S. A. Gerber, A. D. Rudner, S. A. Beausoleil, W. Haas, J. Villen, J. E. Elias and S. P. Gygi (2007). "Large-scale phosphorylation analysis of alpha-factor-arrested *Saccharomyces cerevisiae*." J Proteome Res **6**(3): 1190-1197.

Li, X. and J. L. Manley (2005). "Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability." Cell **122**(3): 365-378.

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.

Liu, Z. R., B. Laggerbauer, R. Lührmann and C. W. Smith (1997). "Crosslinking of the U5 snRNP-specific 116-kDa protein to RNA hairpins that block step 2 of splicing." RNA **3**(11): 1207-1219.

Lu, K. P., G. Finn, T. H. Lee and L. K. Nicholson (2007). "Prolyl cis-trans isomerization as a molecular timer." Nat Chem Biol **3**(10): 619-629.

Lützelberger, M., C. A. Bottner, W. Schwelnus, S. Zock-Emmenthal, A. Razanau and N. F. Käufer (2009). "The N-terminus of Prp1 (Prp6/U5-102 K) is essential for spliceosome activation *in vivo*." Nucleic Acids Res **38**(5): 1610-1622.

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, 3rd (2002). "Shotgun identification of protein modifications from protein complexes and lens tissue." Proc Natl Acad Sci U S A **99**(12): 7900-7905.

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, T. Achsel and R. Lührmann (2000). "The human homologue of the yeast splicing factor prp6p contains multiple TPR elements and is stably associated with the U5 snRNP via protein-protein interactions." J Mol Biol **298**(4): 567-575.

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.

Makarova, O. V., E. M. Makarov, H. Urlaub, C. L. Will, M. Gentzel, M. Wilm and R. Lührmann (2004). "A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing." EMBO J **23**(12): 2381-2391.

Maniatis, T. and B. Tasic (2002). "Alternative pre-mRNA splicing and proteome expansion in metazoans." Nature **418**(6894): 236-243.

Marone, M., S. Mozzetti, D. De Ritis, L. Pierelli and G. Scambia (2001). "Semiquantitative RT-PCR analysis to assess the expression levels of multiple transcripts from the same sample." Biol Proced Online **3**: 19-25.

Mathew, R., K. Hartmuth, S. Mohlmann, H. Urlaub, R. Ficner and R. Lührmann (2008). "Phosphorylation of human PRP28 by SRPK2 is required for integration of the U4/U6-U5 tri-snRNP into the spliceosome." Nat Struct Mol Biol **15**(5): 435-443.

Maundrell, K. (1993). "Thiamine-repressible expression vectors pREP and pRIP for fission yeast." Gene **123**(1): 127-130.

Mayas, R. M., H. Maita, D. R. Semlow and J. P. Staley (2010). "Spliceosome discards intermediates via the DEAH box ATPase Prp43p." Proc Natl Acad Sci U S A **107**(22): 10020-10025.

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**(8): 5352-5362.

Montes, M., S. Becerra, M. Sanchez-Alvarez and C. Sune (2012). "Functional coupling of transcription and splicing." Gene **501**(2): 104-117.

Mortazavi, A., B. A. Williams, K. McCue, L. Schaeffer and B. Wold (2008). "Mapping and quantifying mammalian transcriptomes by RNA-Seq." Nat Methods **5**(7): 621-628.

Mozaffari-Jovin, S., T. Wandersleben, K. F. Santos, C. L. Will, R. Lührmann and M. C. Wahl (2013). "Inhibition of RNA helicase Brr2 by the C-terminal tail of the spliceosomal protein Prp8." Science **341**(6141): 80-84.

Nancollis, V., J. P. Ruckshanthi, L. N. Frazer and R. T. O'Keefe (2013). "The U5 snRNA internal loop 1 is a platform for Brr2, Snu114 and Prp8 protein binding during U5 snRNP assembly." J Cell Biochem.

Nasmyth, K. and P. Nurse (1981). "Cell division cycle mutants altered in DNA replication and mitosis in the fission yeast *Schizosaccharomyces pombe*." Mol Gen Genet **182**(1): 119-124.

Neubauer, G., A. Gottschalk, P. Fabrizio, B. Séraphin, R. Lührmann and M. Mann (1997). "Identification of the proteins of the yeast U1 small nuclear ribonucleoprotein complex by mass spectrometry." Proc Natl Acad Sci U S A **94**(2): 385-390.

Newman, A. J. and C. Norman (1992). "U5 snRNA interacts with exon sequences at 5' and 3' splice sites." Cell **68**(4): 743-754.

Norppa, H. and G. C. Falck (2003). "What do human micronuclei contain?" Mutagenesis **18**(3): 221-233.

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.

O'Keefe, R. T. and A. J. Newman (1998). "Functional analysis of the U5 snRNA loop 1 in the second catalytic step of yeast pre-mRNA splicing." EMBO J **17**(2): 565-574.

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 U S A **104**(9): 3195-3200.

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.

Ohrt, T., M. Prior, J. Dannenberg, P. Odenwalder, O. Dybkov, N. Rasche, J. Schmitzova, I. Gregor, P. Fabrizio, J. Enderlein and R. Lührmann (2012). "Prp2-mediated protein rearrangements at the catalytic core of the spliceosome as revealed by dcFCCS." RNA **18**(6): 1244-1256.

Olsen, J. V., M. Vermeulen, A. Santamaria, C. Kumar, M. L. Miller, L. J. Jensen, F. Gnad, J. Cox, T. S. Jensen, E. A. Nigg, S. Brunak and M. Mann (2010). "Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis." Sci Signal **3**(104): ra3.

Pan, Q., O. Shai, L. J. Lee, B. J. Frey and B. J. Blencowe (2008). "Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing." Nat Genet **40**(12): 1413-1415.

Pauling, M. H., D. S. McPheeters and M. Ares, Jr. (2000). "Functional Cus1p is found with Hsh155p in a multiprotein splicing factor associated with U2 snRNA." Mol Cell Biol **20**(6): 2176-2185.

Pawlicki, J. M. and J. A. Steitz (2010). "Nuclear networking fashions pre-messenger RNA and primary microRNA transcripts for function." Trends Cell Biol **20**(1): 52-61.

Pena, V., S. Liu, J. M. Bujnicki, R. Lührmann and M. C. Wahl (2007). "Structure of a multipartite protein-protein interaction domain in splicing factor prp8 and its link to retinitis pigmentosa." Mol Cell **25**(4): 615-624.

Pena, V., A. Rozov, P. Fabrizio, R. Lührmann and M. C. Wahl (2008). "Structure and function of an RNase H domain at the heart of the spliceosome." EMBO J **27**(21): 2929-2940.

Perez-Iratxeta, C. and M. A. Andrade-Navarro (2008). "K2D2: estimation of protein secondary structure from circular dichroism spectra." BMC Struct Biol **8**: 25.

Peske, F., N. B. Matassova, A. Savelsbergh, M. V. Rodnina and W. Wintermeyer (2000). "Conformationally restricted elongation factor G retains GTPase activity but is inactive in translocation on the ribosome." Mol Cell **6**(2): 501-505.

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.

Pomeranz Krummel, D. A., C. Oubridge, A. K. Leung, J. Li and K. Nagai (2009). "Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution." Nature **458**(7237): 475-480.

Potashkin, J., D. Kim, M. Fons, T. Humphrey and D. Frendewey (1998). "Cell-division-cycle defects associated with fission yeast pre-mRNA splicing mutants." Curr Genet **34**(3): 153-163.

Potashkin, J., R. Li and D. Frendewey (1989). "Pre-mRNA splicing mutants of *Schizosaccharomyces pombe*." EMBO J **8**(2): 551-559.

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

Radermacher, M., T. Ruiz, T. Clason, S. Benjamin, U. Brandt and V. Zickermann (2006). "The three-dimensional structure of complex I from Yarrowia lipolytica: a highly dynamic enzyme." J Struct Biol **154**(3): 269-279.

Reidt, U., M. C. Wahl, D. Fasshauer, D. S. Horowitz, R. Lührmann and R. Ficner (2003). "Crystal structure of a complex between human spliceosomal cyclophilin H and a U4/U6 snRNP-60K peptide." J Mol Biol **331**(1): 45-56.

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**(2): e16719.

Revers, L. F., J. M. Cardone, D. Bonatto, J. Saffi, M. Grey, H. Feldmann, M. Brendel and J. A. Henriques (2002). "Thermoconditional modulation of the pleiotropic sensitivity phenotype by the *Saccharomyces cerevisiae* PRP19 mutant allele *pso4-1*." Nucleic Acids Res **30**(22): 4993-5003.

Rigbolt, K. T., T. A. Prokhorova, V. Akimov, J. Henningsen, P. T. Johansen, I. Kratchmarova, M. Kassem, M. Mann, J. V. Olsen and B. Blagoev (2011). "System-

wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation." Sci Signal **4**(164): rs3.

Ritchie, D. B., M. J. Schellenberg, E. M. Gesner, S. A. Raithatha, D. T. Stuart and A. M. Macmillan (2008). "Structural elucidation of a PRP8 core domain from the heart of the spliceosome." Nat Struct Mol Biol **15**(11): 1199-1205.

Roberts-Galbraith, R. H., M. D. Ohi, B. A. Ballif, J. S. Chen, I. McLeod, W. H. McDonald, S. P. Gygi, J. R. Yates, 3rd and K. L. Gould (2010). "Dephosphorylation of F-BAR protein Cdc15 modulates its conformation and stimulates its scaffolding activity at the cell division site." Mol Cell **39**(1): 86-99.

Rodnina, M. V., A. Savelsbergh, V. I. Katunin and W. Wintermeyer (1997). "Hydrolysis of GTP by elongation factor G drives tRNA movement on the ribosome." Nature **385**(6611): 37-41.

Rodnina, M. V. and W. Wintermeyer (2011). "The ribosome as a molecular machine: the mechanism of tRNA-mRNA movement in translocation." Biochem Soc Trans **39**(2): 658-662.

Romfo, C. M., S. Lakhe-Reddy and J. A. Wise (1999). "Molecular genetic analysis of U2AF59 in *Schizosaccharomyces pombe*: differential sensitivity of introns to mutational inactivation." RNA **5**(1): 49-65.

Rosenberg, G. H., S. K. Alahari and N. F. Kaufer (1991). "*prp4* from *Schizosaccharomyces pombe*, a mutant deficient in pre-mRNA splicing isolated using genes containing artificial introns." Mol Gen Genet **226**(1-2): 305-309.

Ruby, S. W. and J. Abelson (1991). "Pre-mRNA splicing in yeast." Trends Genet **7**(3): 79-85.

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**(2): 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.

Sapra, A. K., Y. Arava, P. Khandelia and U. Vijayraghavan (2004). "Genome-wide analysis of pre-mRNA splicing: intron features govern the requirement for the second-step factor, Prp17 in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*." J Biol Chem **279**(50): 52437-52446.

Schellenberg, M. J., T. Wu, D. B. Ritchie, S. Fica, J. P. Staley, K. A. Atta, P. LaPointe and A. M. MacMillan (2013). "A conformational switch in PRP8 mediates metal ion

coordination that promotes pre-mRNA exon ligation." Nat Struct Mol Biol **20**(6): 728-734.

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.

Schneider, M., H. H. Hsiao, C. L. Will, R. Giet, H. Urlaub and R. Lührmann (2010). "Human PRP4 kinase is required for stable tri-snRNP association during spliceosomal B complex formation." Nat Struct Mol Biol **17**(2): 216-221.

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

Schwelnus, W., K. Richert, F. Opitz, T. Gross, Y. Habara, T. Tani and N. F. Kaufer (2001). "Fission yeast Prp4p kinase regulates pre-mRNA splicing by phosphorylating a non-SR-splicing factor." EMBO Rep **2**(1): 35-41.

Sengupta, J., J. Nilsson, R. Gursky, M. Kjeldgaard, P. Nissen and J. Frank (2008). "Visualization of the eEF2-80S ribosome transition-state complex by cryo-electron microscopy." J Mol Biol **382**(1): 179-187.

Shi, Y., B. Reddy and J. L. Manley (2006). "PP1/PP2A phosphatases are required for the second step of Pre-mRNA splicing and target specific snRNP proteins." Mol Cell **23**(6): 819-829.

Singh, R. K. and T. A. Cooper (2012). "Pre-mRNA splicing in disease and therapeutics." Trends Mol Med **18**(8): 472-482.

Small, E. C., S. R. Leggett, A. A. Winans and J. P. Staley (2006). "The EF-G-like GTPase Snu114p regulates spliceosome dynamics mediated by Brr2p, a DExD/H box ATPase." Mol Cell **23**(3): 389-399.

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**(13): 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.

Spahn, C. M., M. G. Gomez-Lorenzo, R. A. Grassucci, R. Jorgensen, G. R. Andersen, R. Beckmann, P. A. Penczek, J. P. Ballesta and J. Frank (2004). "Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation." EMBO J **23**(5): 1008-1019.

Sridharan, V., J. Heimiller and R. Singh (2011). "Genomic mRNA profiling reveals compensatory mechanisms for the requirement of the essential splicing factor U2AF." Mol Cell Biol **31**(4): 652-661.

Staley, J. P. and C. Guthrie (1998). "Mechanical devices of the spliceosome: motors, clocks, springs, and things." Cell **92**(3): 315-326.

Stark, H., M. V. Rodnina, H. J. Wieden, M. van Heel and W. Wintermeyer (2000). "Large-scale movement of elongation factor G and extensive conformational change of the ribosome during translocation." Cell **100**(3): 301-309.

Stevens, S. W. and J. Abelson (1999). "Purification of the yeast U4/U6.U5 small nuclear ribonucleoprotein particle and identification of its proteins." Proc Natl Acad Sci U S A **96**(13): 7226-7231.

Storey, J. D. (2002). "A direct approach to false discovery rates." Journal of the Royal Statistical Society **Series B**(64): 479-498.

Tabach, Y., A. C. Billi, G. D. Hayes, M. A. Newman, O. Zuk, H. Gabel, R. Kamath, K. Yacoby, B. Chapman, S. M. Garcia, M. Borowsky, J. K. Kim and G. Ruvkun (2013). "Identification of small RNA pathway genes using patterns of phylogenetic conservation and divergence." Nature **493**(7434): 694-698.

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.

Tasto, J. J., R. H. Carnahan, W. H. McDonald and K. L. Gould (2001). "Vectors and gene targeting modules for tandem affinity purification in *Schizosaccharomyces pombe*." Yeast **18**(7): 657-662.

Taylor, D. J., J. Nilsson, A. R. Merrill, G. R. Andersen, P. Nissen and J. Frank (2007). "Structures of modified eEF2 80S ribosome complexes reveal the role of GTP hydrolysis in translocation." EMBO J **26**(9): 2421-2431.

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.

Turner, I. A., C. M. Norman, M. J. Churcher and A. J. Newman (2006). "Dissection of Prp8 protein defines multiple interactions with crucial RNA sequences in the catalytic core of the spliceosome." RNA **12**(3): 375-386.

Umen, J. G. and C. Guthrie (1995). "A novel role for a U5 snRNP protein in 3' splice site selection." Genes Dev **9**(7): 855-868.

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.

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

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.

Walbott, H., S. Mouffok, R. Capeyrou, S. Lebaron, O. Humbert, H. van Tilbeurgh, Y. Henry and N. Leulliot (2010). "Prp43p contains a processive helicase structural architecture with a specific regulatory domain." EMBO J **29**(13): 2194-2204.

Wang, E. T., R. Sandberg, S. Luo, I. Khrebtukova, L. Zhang, C. Mayr, S. F. Kingsmore, G. P. Schroth and C. B. Burge (2008). "Alternative isoform regulation in human tissue transcriptomes." Nature **456**(7221): 470-476.

Wang, P. and J. Heitman (2005). "The cyclophilins." Genome Biol **6**(7): 226.

Wang, X., S. Zhang, J. Zhang, X. Huang, C. Xu, W. Wang, Z. Liu, J. Wu and Y. Shi (2010). "A large intrinsically disordered region in SKIP and its disorder-order transition induced by PPIL1 binding revealed by NMR." J Biol Chem **285**(7): 4951-4963.

Ward, J. J., L. J. McGuffin, K. Bryson, B. F. Buxton and D. T. Jones (2004). "The DISOPRED server for the prediction of protein disorder." Bioinformatics **20**(13): 2138-2139.

Warkocki, Z., P. Odenwälder, J. Schmitzova, F. Platzmann, H. Stark, H. Urlaub, R. Ficner, P. Fabrizio and R. Lührmann (2009). "Reconstitution of both steps of *Saccharomyces cerevisiae* splicing with purified spliceosomal components." Nat Struct Mol Biol **16**(12): 1237-1243.

Weber, G., V. F. Cristao, L. A. F. de, K. F. Santos, N. Holton, J. Rappsilber, J. D. Beggs and M. C. Wahl (2011). "Mechanism for Aar2p function as a U5 snRNP assembly factor." Genes Dev **25**(15): 1601-1612.

Weber, G., S. Trowitzsch, B. Kastner, R. Lührmann and M. C. Wahl (2010). "Functional organization of the Sm core in the crystal structure of human U1 snRNP." EMBO J **29**(24): 4172-4184.

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

Wilson, D. S. and A. D. Keefe (2001). "Random mutagenesis by PCR." Curr Protoc Mol Biol **Chapter 8**: Unit8 3.

Yang, K., L. Zhang, T. Xu, A. Heroux and R. Zhao (2008). "Crystal structure of the beta-finger domain of Prp8 reveals analogy to ribosomal proteins." Proc Natl Acad Sci U S A **105**(37): 13817-13822.

Yang, Q., M. Del Campo, A. M. Lambowitz and E. Jankowsky (2007). "DEAD-box proteins unwind duplexes by local strand separation." Mol Cell **28**(2): 253-263.

Yang, V. W., M. R. Lerner, J. A. Steitz and S. J. Flint (1981). "A small nuclear ribonucleoprotein is required for splicing of adenoviral early RNA sequences." Proc Natl Acad Sci U S A **78**(3): 1371-1375.

Yates, J. R., 3rd, 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.

Zhang, L., X. Li and R. Zhao (2013). "Structural analyses of the pre-mRNA splicing machinery." Protein Sci **22**(6): 677-692.

Zhang, L., J. Shen, M. T. Guarnieri, A. Heroux, K. Yang and R. Zhao (2007). "Crystal structure of the C-terminal domain of splicing factor Prp8 carrying retinitis pigmentosa mutants." Protein Sci **16**(6): 1024-1031.

Zhang, L., T. Xu, C. Maeder, L. O. Bud, J. Shanks, J. Nix, C. Guthrie, J. A. Pleiss and R. Zhao (2009). "Structural evidence for consecutive Hel308-like modules in the spliceosomal ATPase Brr2." Nat Struct Mol Biol **16**(7): 731-739.

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.