Structural and Biochemical Insights into U6 snRNP Assembly and Recycling

By

Allison Louise Didychuk

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Biophysics)

at the

UNIVERSITY OF WISCONSIN-MADISON

2017

Date of final oral examination: December 7th, 2017

The dissertation is approved by the following members of the Final Oral Committee:

Samuel E. Butcher, Professor, Biochemistry

David A. Brow, Professor, Biomolecular Chemistry

Aaron A. Hoskins, Assistant Professor, Biochemistry

John L. Markley, Professor, Biochemistry

David A. Wassarman, Professor, Medical Genetics

ABSTRACT

Splicing is essential for gene expression in all eukaryotes and is catalyzed by the spliceosome, a macromolecular machine that must assemble and undergo elaborate conformational and compositional changes on the pre-mRNA substrate during each round of splicing activation and catalysis. Sub-complexes of the spliceosome consisting of a small nuclear RNA (snRNA) and associated proteins (called snRNPs) must undergo their own assembly processes prior to assembly of the spliceosome on pre-mRNA. This thesis describes work on several early steps in the assembly pathway of U6 snRNA on its journey into snRNPs and functional spliceosomes.

I studied an early processing step of U6 snRNA in which a 3'-5' exonuclease, Usb1, trims the 3' tail of U6 and leaves a phosphate group. I characterized Usb1 from *S. cerevisiae* and *H. sapiens* and demonstrated that the two homologs have divergent activities that result in different 3' end modifications. I also demonstrated that Usb1 activity is essential for U6 snRNP formation, as Usb1 activity prevents binding of Lhp1 (which binds after transcription termination) and promotes binding of the U6 snRNP protein complex Lsm2-8.

After formation of the U6 snRNP, an internal stem loop in U6 must be unwound and base paired to U4 snRNA in order to form the U4/U6 di-snRNP and U4/U6.U5 tri-snRNP. This large conformational rearrangement is catalyzed by the U6 snRNP protein Prp24. I developed an *in vitro* annealing assay to monitor Prp24-mediated U4/U6 annealing. Using this assay, I characterized structural elements in Prp24 and U6 and measured their contribution to annealing.

I then discuss work towards determining the structure of the U4/U6 di-snRNA via a combination of NMR and SAXS/WAXS. When compared to the structure of the U4/U6.U5 tri-snRNP, it becomes clear that associated proteins markedly change the conformation of U4/U6, which may have implications for the subsequent steps of spliceosome activation. My work has significantly contributed to our understanding of how U6 is modified and chaperoned into the splicing cycle and how its protein binding partners influence its assembly.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Sam Butcher. Sam has been an amazing and patient mentor to me over the past five years. Sam's unwavering support and ceaseless positivity have kept me afloat throughout numerous failed experiments, setbacks, and crises of confidence. He has been an unfailing champion and has lifted me up academically and professionally, giving me every opportunity possible to grow and develop as a scientist and community member. I will always appreciate his support and how he consistently treated me and my ideas with the utmost respect.

I have also been incredibly fortunate to have had the privilege of working closely with and benefiting from the suggestions, conversations, and guidance of Dave Brow and Aaron Hoskins. Dave's extensive knowledge of the literature and history of U6 have been an invaluable resource. His willingness to troubleshoot ideas for new experiments, suggestions, and insightful comments on manuscripts have been instrumental throughout my graduate career. I have also learned many lessons and directly benefited from Aaron's enthusiasm, creativity, and vision. I am extremely lucky to have benefited from Dave and Aaron's mentoring and how they have consistently gone above and beyond to support me. I am also lucky to have had John Markley and Dave Wassarman on my committee, who have supported my work with their thoughtful feedback and questions.

Both scientifically and personally, I had the good fortune to befriend Tucker Carrocci at RNA 2014. Since that time, he has taught me, encouraged me, and pushed me to become a better scientist. His knowledge of both techniques and the literature have been priceless. He has also been the best conference buddy I could ever ask for. Eric Montemayor was also helpful in getting many projects off the ground and in training me. His knowledge and commitment to rigorous science have been invaluable.

I count myself extremely lucky to have had the opportunity to mentor three stellar undergraduates: Andy DeLaitsch, Matt Larson, Stefani Lucarelli. They have made my time in lab more enjoyable through their positivity, their unique perspective, and their excitement about

science. Working alongside, teaching, and learning from Andy, Matt, and Stefani was one of the highlights of graduate school for me.

I would also like to thank current and former Butcher, Brow, and Hoskins labs members, especially Jordan Burke, Katie Mouzakis, Josh Larson, Josh Paulson, Maggie Rodgers, Sarah Hansen, Clarisse van der Feltz, Allyson Yake, and Yuichiro Nomura, for making the tri- (now tetra!) lab community such an amazing place to work on splicing. Additionally, the larger scientific community in the Biochemistry department, particularly Ronnie Frederick, Kai Cai, Lai Bergeman, Mark Anderson, Milo Westler, Gabriel Cornilescu, Marco Tonelli, Katie Henzler-Wildman, past and present members of the Markley Lab, Fox Lab, and NMRFAM, along with students and faculty of the Biophysics Graduate Program, Bob Smith and Craig Bingman, have helped me along the way, broadened my horizons, and made working here easier and more fun.

I was incredibly fortunate to rotate in the lab with Lauren Michael, who taught me how to pour my first gel. She and her husband Nick have since become our Madison family over countless nights of board games and bad movies. I have had an immense amount of support and encouragement from my family and friends. My parents, sister, and brother have encouraged me to pursue my goals for as long as I can remember, and the opportunities afforded to me wouldn’t have been possible without their support. I am forever grateful to them. Finally, I would like to give the biggest acknowledgement to my husband Sam. He has supported me since the day we met, and I can’t imagine being half as successful or a tenth as happy without him, his endless encouragement, and love.

TABLE OF CONTENTS

List of Figures

Appendix 3: Screening for mutations that result in a yeast-like human Usb1

Appendix 4: Overexpression of U6 pathway factors complement deletion of Usb1

Appendix 5: NMR analysis of yUsb1

Appendix 6: Conformational changes of Prp24 during U6 binding and U4/U6 annealing

Appendix 7: Additional insights into Prp24-mediated annealing

List of Tables

Chapter 1: Thesis Overview

The assembly process of the U6 snRNP is incredibly complex. U6 snRNA undergoes a many processing steps in its initial biogenesis and undergoes large conformation changes during every round of splicing. The work described in this dissertation focuses on early steps in the assembly pathway of U6 on its pilgrimage into functional spliceosomes. In chapter 2, I review the life cycle of U6 snRNA. Despite its discovery nearly forty years ago, we lack molecular level detail for how U6 undergoes its conformational changes and how it is regulated through transcription, modification, and degradation, among other regulatory steps. I highlight and summarize what is known about U6 snRNA, which spans multiple areas of RNA biology.

In chapter 3, I discuss an early processing step of U6 wherein a 3'-5' exonuclease, Usb1, trims the 3' tail of U6 and leaves a phosphate group. Using a combination of structural biology, *in vitro* biochemical assays, and *in vivo* experiments, I characterized Usb1 from *S. cerevisiae* and *H. sapiens* and demonstrated that the two homologs have divergent activity that results in different 3' end modifications. I demonstrated that Usb1 activity is essential to U6 snRNP formation, as Usb1 activity prevents binding of Lhp1 (which binds after transcription termination) and promotes binding of the U6 snRNP protein complex Lsm2-8. Thus, modification by Usb1 is important to shuttle U6 snRNA into the U6 snRNP. This work was published in *Nature Communications*.

After formation of the U6 snRNP, an internal stem loop in U6 must be unwound and base paired to U4 snRNA in order to form the U4/U6 di-snRNP and U4/U6.U5 tri-snRNP. This large conformational rearrangement is catalyzed by the U6 snRNP protein Prp24. In chapter 4, I discuss work towards understanding the mechanism of Prp24-mediated U4/U6 annealing. I developed an *in vitro* annealing assay to monitor U4/U6 annealing. Using this assay, I characterized structural elements in Prp24 and U6 and measured their contribution to annealing. A large, positively charged groove in Prp24 formed by three of its four RRMs is essential for annealing *in vitro*. Additionally, a metastable intramolecular stem within U6 called the telestem promotes annealing. Finally, inclusion of the Lsm2-8 ring stimulated annealing. Together, these data allowed us to propose a model where Lsm2-8 binding and stabilization of the telestem both promote telestem

formation, which in turn positions the RRMs of Prp24 in a conformation able to form the electropositive groove. This work was published in *Nucleic Acids Research*.

In Chapter 5, I discuss work towards determining the structure of the U4/U6 di-snRNA via a combination of NMR and SAXS/WAXS. This methods-driven approach revealed that the core of the U4/U6 di-snRNA lacks structural dynamics in solution and that the preferred conformation is dramatically different from the structure observed in the structure of the U4/U6.U5 tri-snRNP, demonstrating that di- and tri-snRNP proteins promote structural rearrangements that may be essential for its function. This work was published in the *Journal of Molecular Biology*.

Chapter 6 contains a summary of my work and provides future directions and a discussion of the important remaining unanswered questions in the field. This includes proposed experiments to further our understanding of the biological role of Usb1, why it possesses divergent phosphodiesterase activity, and to examine how Prp24 mediates U4/U6 annealing.

Chapter 2: The life of U6 small nuclear RNA, from cradle to grave

The work presented in this chapter is being prepared for submission as a review article in the following form:

Didychuk, A.L., Butcher, S.E., Brow, D.A. The life of U6 small nuclear RNA, from cradle to grave.

Author contributions:

A.L.D. wrote the article and made figures with revisions and guidance from S.E.B. and D.A.B.

2.1 Abstract

Removal of introns from precursor messenger RNA and some non-coding transcripts is an essential step in eukaryotic gene expression. In the nucleus, this process of RNA splicing is carried out by the spliceosome, a multi-megaDalton macromolecular machine whose core components are conserved from yeast to humans. In addition to many proteins, the spliceosome contains five uridine-rich small nuclear RNAs (snRNAs) that undergo an elaborate series of conformational changes to correctly recognize the splice sites and catalyze intron removal. Decades of biochemical and genetic data, along with recent cryo-EM structures, unequivocally demonstrate that U6 snRNA forms much of the catalytic core of the spliceosome and is highly dynamic, interacting with three snRNAs, the pre-mRNA substrate, and >25 protein partners throughout the splicing cycle. This review summarizes the current state of knowledge on how U6 snRNA is synthesized, modified, incorporated into snRNPs and spliceosomes, recycled, and degraded.

2.2 Introduction

In the nucleus of eukaryotic cells, removal of intervening sequences (introns) from precursor messenger RNA (pre-mRNA) and ligation of the remaining mRNA segments (exons) occurs prior to export of mRNA to the cytoplasm and translation by the ribosome. This essential step in gene expression, known as pre-mRNA splicing, must occur with single-nucleotide precision to ensure proper gene expression. Some noncoding RNAs also contain introns that must be removed by the same process to allow proper RNA function.

The pre-mRNA splicing reaction consists of two transesterification steps and closely resembles the reaction carried out by group II self-splicing introns. First, the 2'-oxygen of a nucleotide near the 3' end of the intron (the branch site, most often an adenosine) is activated for nucleophilic attack of the 5' exon-intron junction (the 5' splice site) to form a branched “lariat” intron-3' exon intermediate. Subsequently, nucleophilic attack of the released 5' exon on the intron-3' exon junction (the 3' splice site) produces a ligated 5' exon-3' exon product and free lariat intron (WAHL *et al.* 2009; WILL AND LUHRMANN 2011). U6 snRNA sits at the heart of the spliceosome, where it coordinates the magnesium ions required for splicing chemistry and, with help from U2 and U5 snRNAs, positions the substrate for the splicing reaction.

Pre-mRNA splicing is catalyzed by the spliceosome, a complex macromolecular machine that consists of five snRNAs (U1, U2, U4, U5, and U6) and their associated proteins (small nuclear ribonucleoprotein complexes; snRNPs), as well as a protein-only NineTeen Complex (NTC) and a number of accessory proteins (WAHL *et al.* 2009; WILL AND LUHRMANN 2011). A small percentage of human introns, as well as those of certain other metazoan and protozoan taxa, have non-canonical splice sites and are recognized and removed by the minor “U12-type” spliceosome. The minor spliceosome shares U5 snRNA with the major spliceosome, but contains the U11, U12, U4atac, and U6atac snRNAs, which replace U1, U2, U4 and U6 snRNAs, respectively (TARN AND STEITZ 1996a; TARN AND STEITZ 1996b). Both the major and minor spliceosomes assemble *de*

novo on each intron and, after splicing, dissociate into their component snRNPs and accessory proteins. The dynamic nature of the spliceosome makes its study challenging.

Early steps in recognition of the intron include base pairing of U1 snRNA in the U1 snRNP to the 5' splice site and U2 snRNA in the U2 snRNP to the branch site, forming the A complex (also known as the prespliceosome). The A complex is joined by a preassembled U4/U6.U5 triple (tri)-snRNP to form the B complex (**Figure 2-1**). In tri-snRNP nomenclature, the slash indicates base pairing between U4 and U6 snRNAs and the period conveys that U5 snRNP associates with the U4/U6 di-snRNP through protein-protein and RNA-protein interactions. During activation of the assembled spliceosome for the first catalytic step, large-scale conformational rearrangements cause U1 and U4 snRNPs to be ejected, leaving the U2/U6.U5 complex together with the NTC (B^{act}) and accessory factors to catalyze the chemical steps (WILL AND LUHRMANN 2011). U6 must undergo several RNA base pairing exchanges: disruption of U4/U6 in favor of a mutually exclusive interaction with U2, and base pairing to the 5'SS. The timing of these events and the interactions controlling activation are just beginning to be elucidated (SHCHERBAKOVA *et al.* 2013; HOSKINS *et al.* 2016).

Recent cryo-EM structures of spliceosomal complexes including the U4/U6.U5 tri-snRNP (AGAFONOV *et al.* 2016; NGUYEN *et al.* 2016; WAN *et al.* 2016b), the pre-catalytic B complex (PLASCHKA *et al.* 2017), and the catalytic U2/U6.U5 spliceosome at different stages of splicing (YAN *et al.* 2015; GALEJ *et al.* 2016; RAUHUT *et al.* 2016; WAN *et al.* 2016a; YAN *et al.* 2016; BERTRAM *et al.* 2017; FICA *et al.* 2017; WAN *et al.* 2017; YAN *et al.* 2017; ZHANG *et al.* 2017) have allowed for near-atomic level views of the extensive structural rearrangements that must take place during splicing. After each round of splicing, the snRNPs and NTC dissociate from the intron and are used again. U6, however, is released as a free RNA and so must bind its own snRNP proteins again before assembling with the U4 and U5 snRNPs to reconstitute the tri-snRNP (FOURMANN *et al.* 2013). The splicing cycle is described in greater detail in recent reviews (WILL AND LUHRMANN 2011; HOSKINS AND MOORE 2012).

A role for U6 in the catalytic center of the spliceosome had long been proposed based on a number of lines of evidence. The discovery that group II self-splicing introns use the same two-step chemical mechanism as the spliceosome (PEEBLES *et al.* 1986; VAN DER VEEN *et al.* 1986) suggested that the two enzymes share catalytic RNA elements (SHARP 1985; CECH 1986; MADHANI AND GUTHRIE 1992; MEFFORD AND STALEY 2009). The high degree of conservation in U6 RNA sequence (BROW AND GUTHRIE 1988) and the presence of introns in the most conserved region of U6 in some organisms (TANI AND OHSHIMA 1989; TANI AND OHSHIMA 1991), potentially due to reverse splicing, hinted at the proximity of U6 to the active site of the spliceosome (BROW AND GUTHRIE 1989). Subsequent crosslinking and genetic studies indicated that the strictly conserved "ACAGA-box" sequence of U6 base pairs to the intron 5' splice site in the catalytically active spliceosome (SAWA AND ABELSON 1992; SAWA AND SHIMURA 1992; WASSARMAN AND STEITZ 1992; KANDELS-LEWIS AND SERAPHIN 1993; LESSER AND GUTHRIE 1993; SONTHEIMER AND STEITZ 1993), and mutational studies identified functional similarities between Domain V of group II introns and the U2/U6 base-paired region (MADHANI AND GUTHRIE 1992; CHANFREAU AND JACQUIER 1994; PEEBLES *et al.* 1995; SHUKLA AND PADGETT 2002; MEFFORD AND STALEY 2009). Elegant biochemical experiments confirmed that U6 snRNA is responsible for coordinating catalytic metal ions essential for both transesterification steps (YEAN *et al.* 2000; FICA *et al.* 2013).

At just over 100 nucleotides in length, U6 snRNA is highly economical, as nearly every part of the RNA possesses some essential function. In addition to its catalytic role at the heart of the spliceosome, U6 snRNA is notable for undergoing extensive structural rearrangements, including unwinding and reformation of stable internal secondary structure, and for directly interacting with >25 proteins during a single round of splicing. Here, we summarize the life cycle of the U6 snRNA (**Figure 2-2**), and highlight differences found among the eukaryotes. Commonalities in the U6 snRNA life cycle illustrate conservation across more than a billion years of eukaryotic evolution. We focus primarily on U6 snRNA from the best-studied organisms: yeast (S. *cerevisiae*, and to a lesser extent, S. *pombe*) and human.

Figure 2-1. A snRNP-centric view of splicing.

The spliceosome is composed of small nuclear ribonucleoprotein particles (snRNPs) and a protein-only complex called the NTC, which are represented as colored circles. Additional transiently-bound proteins (not shown) are also necessary for progression through the splicing cycle. The U1, U2, U4, U5, and U6 snRNPs consist of the small nuclear RNA (snRNA) for which they are named and associated proteins. The snRNPs and NTC undergo ordered assembly on the pre-mRNA and experience both conformational and compositional changes throughout the cycle. After splicing is complete, the snRNPs and NTC are released and reused (dotted lines).

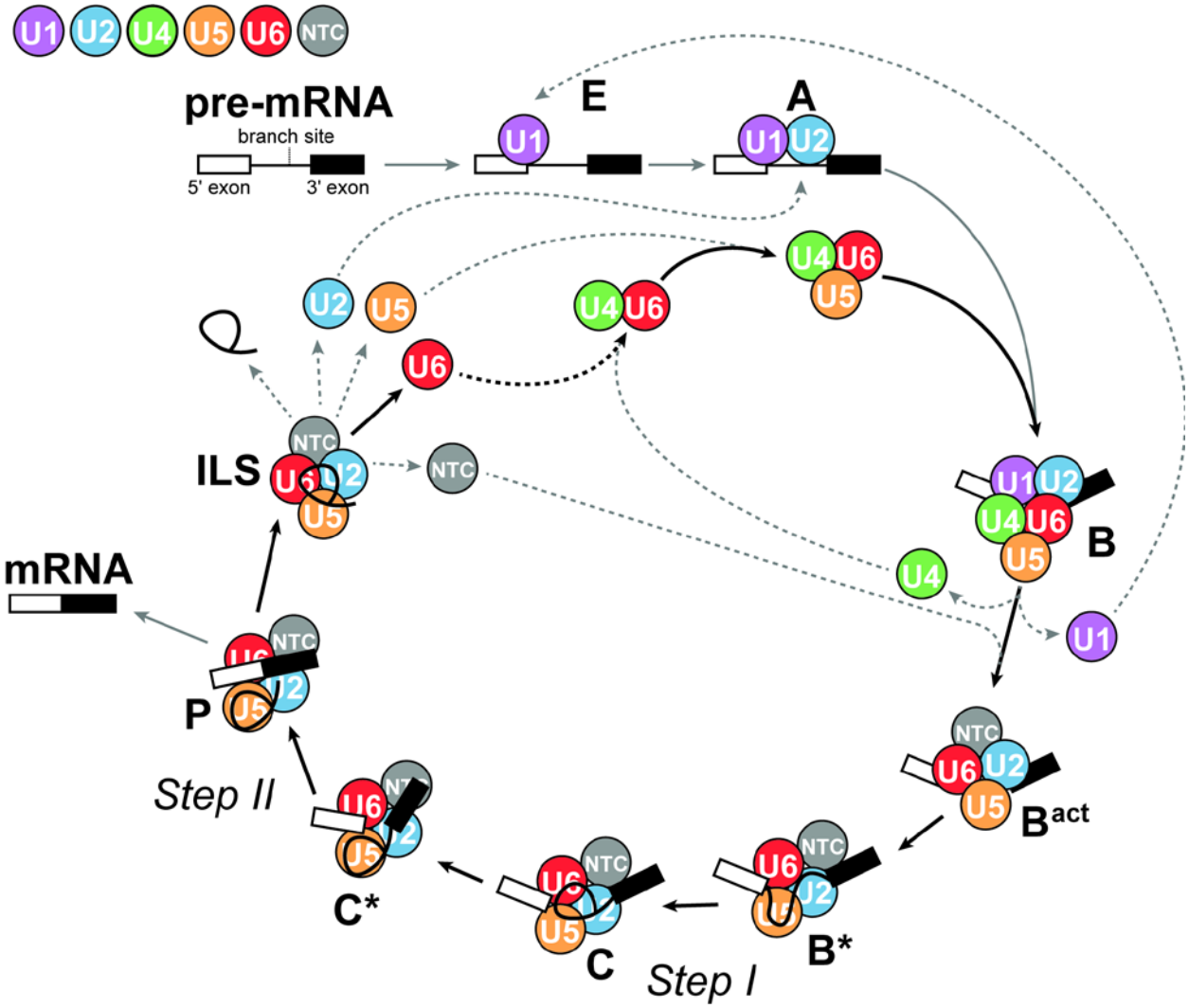
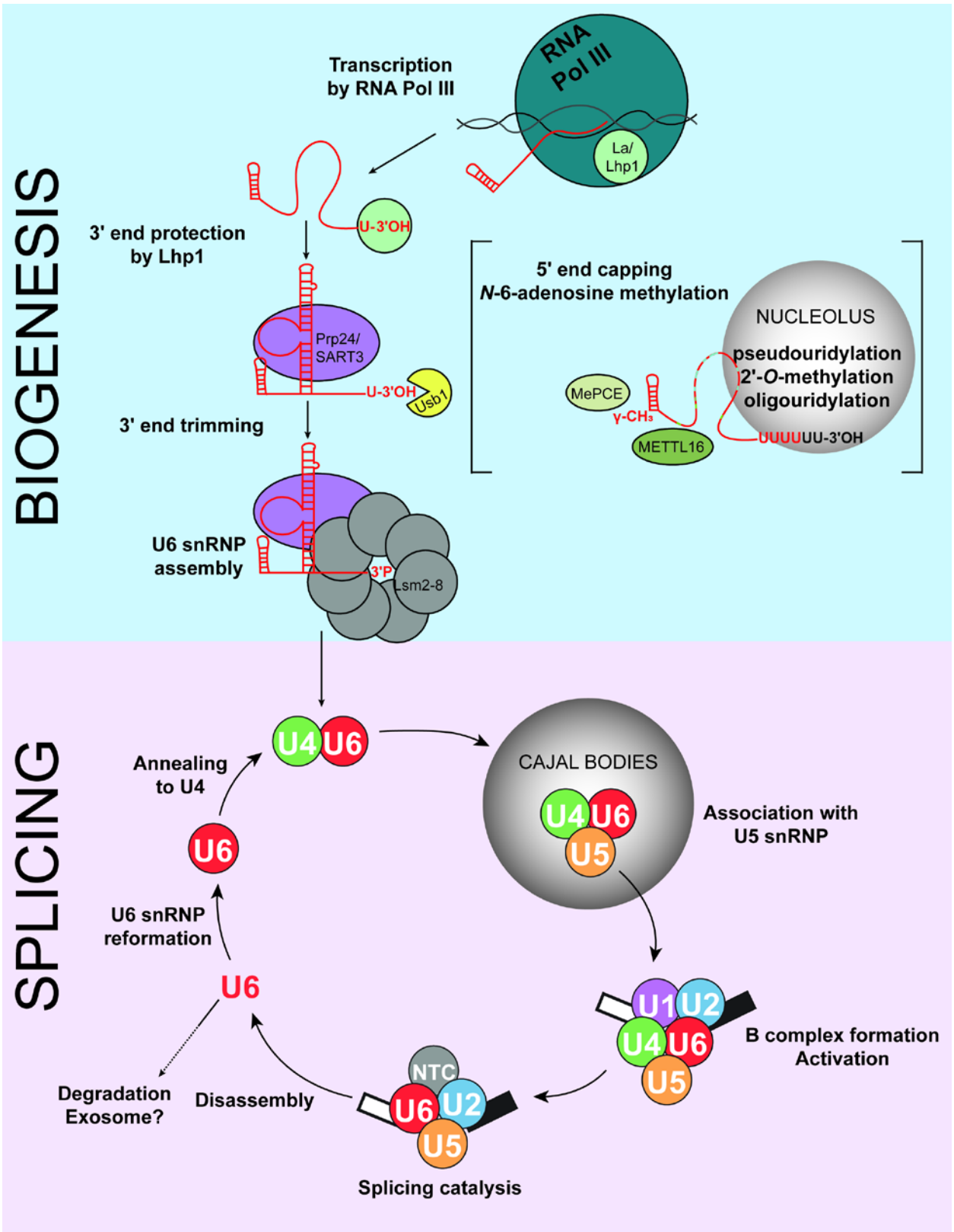


Figure 2-2. Summary of the U6 lifecycle.

Key steps in U6 snRNA biogenesis and assembly (top) and incorporation into the splicing cycle (bottom) are conserved in eukaryotes. Many additional modification steps, represented in brackets, occur in *S. pombe* and humans.


BIOGENESIS
Transcription by RNA Pol III
RNA Pol III
La/ Lhp1
U-3'OH
3' end protection by Lhp1
Prp24/ SART3
U-3'OH
Usb1
3' end trimming
U6 snRNP assembly
3'P
Lsm2-8
5' end capping
N-6-adenosine methylation
NUCLEOLUS
pseudouridylation
2'-O-methylation
oligouridylation
MePCE
γ-CH3
METTL16
UUUUUU-3'OH
SPLICING
U4 U6
Annealing to U4
U6
U6 snRNP reformation
U6
CAJAL BODIES
U4 U6
U5
Association with U5 snRNP
U1 U2
U4 U6
U5
B complex formation
Activation
NTC
U6 U2
U5
Splicing catalysis
Disassembly
Degradation
Exosome?

2.3 U6 snRNA gene transcription

2.3.1 Number of U6 genes

S. cerevisiae has a single genomic copy of U6 snRNA, encoded by the *SNR6* gene on chromosome XII (BROW AND GUTHRIE 1988). However, a survey of 145 fungal genomes identified species with up to 20 U6 gene copies and an average of 2.3 per genome (CANZLER *et al.* 2016). In contrast, there are >900 copies of U6 distributed throughout the human genome, although the majority of these are likely pseudogenes and not transcriptionally active (DOUCET *et al.* 2015). At least four human U6 genes encoding identical RNAs are transcriptionally active to various degrees (DOMITROVICH AND KUNKEL 2003). Additionally, a variant of human U6 snRNA with nine substitutions and one nucleotide deletion is expressed under the control of an internal promoter, unlike other transcriptionally active human U6 genes (TICHELAAR *et al.* 1994; TICHELAAR *et al.* 1998). The presence of multiple U6 genes of varying transcriptional activity has complicated their individual study, and whether paralogous but divergent U6 snRNAs exhibit differences in modification, localization or function is poorly understood. The U6atac RNA is a paralog of U6 that functions in the minor spliceosome and is even further diverged in sequence from the other transcribed U6 snRNAs (TARN AND STEITZ 1996a).

2.3.2 Transcription of U6 genes by RNA polymerase III

Unlike the other spliceosomal snRNAs, which are synthesized by RNA polymerase II (Pol II), U6 is synthesized by RNA polymerase III (Pol III) (REDDY *et al.* 1987; MOENNE *et al.* 1990). While the sequence of U6 snRNA is highly conserved between yeast and humans, its Pol III promoter structure is divergent. In yeast, the U6 promoter region is similar to tRNA gene promoters (ESCHENLAUER *et al.* 1993) in that it contains A and B block elements (BROW AND GUTHRIE 1990), as well as a TATA box that is bound by TATA-binding protein (TBP) (MARGOTTIN *et al.* 1991) (**Figure 2-3A**). In *S. cerevisiae*, the B block is located downstream of the U6 coding region. Consequently, the distance between the A and B blocks is atypical (~200 nts vs. the

normal 30-90 nts in tRNA genes), and this extended spacing is important for transcription *in vivo* (KAISER *et al.* 2004).

The B block is essential for transcription of the U6 gene *in vivo* and is the binding site for transcription initiation factor TFIIIC, which also contacts the A block (BURNOL *et al.* 1993; ESCHENLAUER *et al.* 1993; KAISER AND BROW 1995). TFIIIC recruits TFIIIB (composed of TBP, Bdp1, and Brf1), which binds over the TATA box and in turn recruits Pol III (GERLACH *et al.* 1995). Transcription is promoted by but not dependent on the presence of the TATA box and a stretch of thymidines just downstream of it (BURNOL *et al.* 1993; ESCHENLAUER *et al.* 1993; MARTIN *et al.* 2001).

The non-histone chromatin protein Nhp6 is also important for transcription of SNR6 (KRUPPA *et al.* 2001; MARTIN *et al.* 2001), In a heterologous, *in vitro* chromatin assembly system, a nucleosome positioned between the A and B blocks brings the regions close together for optimal binding by TFIIIC (SHIVASWAMY *et al.* 2004), but the micrococcal nuclease footprint of native chromatin assembled between the A and B blocks *in vivo* is shorter than expected for an intact nucleosome (GERLACH *et al.* 1995). It is possible that Nhp6 modifies the structure of a nucleosome bound to *SNR6* (STILLMAN 2010) (**Figure 2-3A**). Alternatively, Nhp6 may favor a bent conformation of DNA that promotes TFIIIB binding (BRAGLIA *et al.* 2007).

U6 gene promoter structure in fungi is flexible and can include or exclude identifiable TATA boxes, intragenic A blocks, and downstream B blocks (CANZLER *et al.* 2016). *S. pombe* has a similar promoter structure to *S. cerevisiae*, with the exception that the B block element is located intragenically in an intron (FRENDEWEY *et al.* 1990).

In humans, U6 RNA is transcribed from a Pol III Type III promoter, a promoter architecture that also drives 7SK RNA and RNase P RNA transcription. Most characterized human U6 genes do not contain intragenic promoter elements, but rather exclusively contain upstream promoter elements (DAS *et al.* 1988). These U6 genes have a TATA box 30 base pairs upstream of the transcription start site, a proximal sequence element (PSE) 50 base pairs upstream, and a distal

sequence element (DSE; also called an OCT site) 250 base pairs upstream, which closely resembles the promoter structure for Pol II-transcribed genes (**Figure 2-3B**). Interestingly, the PSE and DSE are similar to, and can be switched with, corresponding elements from the Pol II-synthesized snRNAs (KUNKEL AND PEDERSON 1988), and it is inclusion of the TATA box that specifies Pol III rather than Pol II transcription for snRNAs (LOBO AND HERNANDEZ 1989). Other organisms have divergent methods of specifying Pol II vs III transcription, such as the sequence of the PSE or the distance between the PSE and the TATA box (reviewed in (HERNANDEZ 2001)).

The PSE is required for basal transcription while the DSE is required for efficient transcription. The PSE is bound by the small nuclear RNA activating protein complex (SNAPc), which also binds the promoters of Pol II-transcribed snRNAs. The DSE is bound by the Oct-1 and Staf proteins (**Figure 2-3B**). Cooperative binding of Oct-1 and SNAPc is enhanced by a nucleosome positioned between the DSE and PSE (ZHAO *et al.* 2001). Oct-1 binding is negatively regulated by p38 kinase (LIN AND NATARAJAN 2012), while Staf binding is influenced by chromatin-modifying enzymes (YUAN *et al.* 2007). The TATA box is bound by TFIIIB2, composed of TBP, Bdp1, and Brf2. TFIIIC is not required as there is no A or B block. Interestingly, yeast also contain an upstream PSE-like sequence that is not required for transcription (ESCHENLAUER *et al.* 1993) and an upstream positioned nucleosome (ARIMBASSERI AND BHARGAVA 2008), suggesting that while some similarities remain, the promoter structure in yeast and humans has diverged significantly.

Chromatin immunoprecipitation detects very low levels of Pol II over the *S. cerevisiae* U6 gene, similar to that detected over the Pol II-silenced rDNA and telomeres (STEINMETZ *et al.* 2006). Furthermore, a hypomorphic mutation in the Sen1 helicase increases Pol II levels at all these loci (STEINMETZ *et al.* 2006), and an anti-sense transcript of the U6 gene contains a high-affinity Nrd1 binding site that promotes Sen1-dependent Pol II termination (STEINMETZ AND BROW 1998). Thus, like the rDNA and telomeres (VASILJEVA *et al.* 2008), the *S. cerevisiae* U6 gene may be silenced

for Pol II by an unknown mechanism coupled to Sen1-dependent termination of an anti-sense transcript.

Interestingly, in humans, transcription of U6 is also dependent upon the interaction of Pol II at a site ~300 bp upstream of the gene, a phenomenon shown to be generally true for Pol III-transcribed genes (LISTERMAN *et al.* 2007; BARSKI *et al.* 2010; OLER *et al.* 2010). This may result from the influence of chromatin remodeling through recruitment of Pol II transcription factors that are also used in Pol III transcription (RAHA *et al.* 2010). Transcription of U6atac is also dependent on both Pol II and Pol III (YOUNIS *et al.* 2013). Pol II influence on U6 transcription does not appear to be evolutionarily conserved, as Pol II does not associate near the yeast U6 gene *SNR6* (STEINMETZ et al. 2006).

Transcription termination of U6 is caused by a stretch of dA's in the template strand at the 3' end of the gene, although the number of dA's for efficient termination vary in eukaryotes (ARIMBASSERI *et al.* 2013). The La protein (Lhp1 in yeast) facilitates transcription termination (GOTTLIEB AND STEITZ 1989a; GOTTLIEB AND STEITZ 1989b), promotes RNA polymerase recycling and transcription reinitiation (MARAIA *et al.* 1994; MARAIA 1996; FRENCH *et al.* 2008), and binds the 3' end of newly transcribed U6 RNAs (RINKE AND STEITZ 1985). Pol III transcription is terminated without the help of additional termination factors. Weak base pairing between the nascent oligo (rU) and the template strand oligo (dA) stretch, as well as interactions between the nontemplate strand oligo (dT) stretch and a subunit of Pol III, terminate transcription (BOGENHAGEN AND BROWN 1981; HAMADA *et al.* 2000; ARIMBASSERI AND MARAIA 2015). Yeast U6 is terminated by a stretch of ten dAs, leaving an RNA product with a heterogeneous U-tail length consisting of 4-7 uridines (BROW AND GUTHRIE 1990).

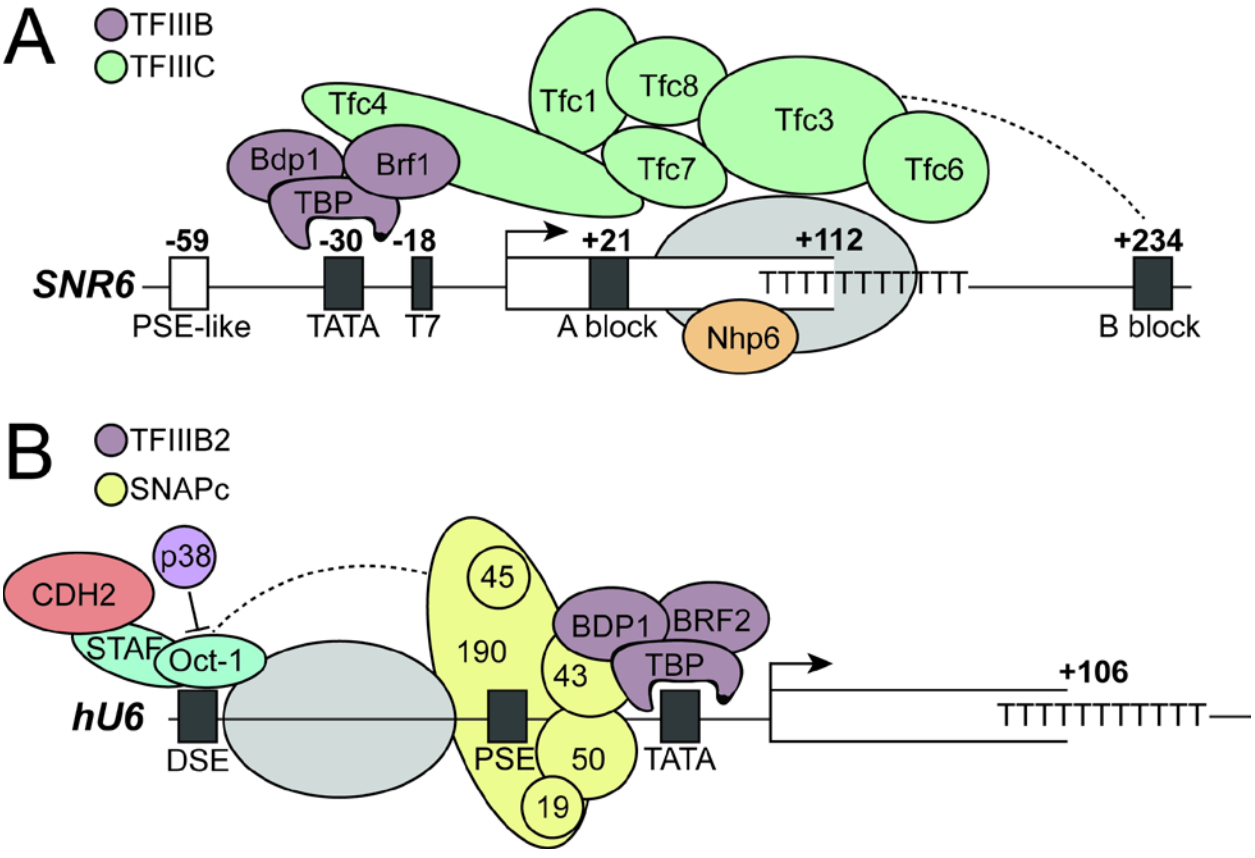
How U6 is transcriptionally regulated during the cell cycle, development, in different tissues, and in disease states has been poorly studied. It has been noted that tissue-specific differences of U6 expression level exist (SPANIEL *et al.* 2013). Whether this difference is due to tissue-specific stability or transcriptional activity is not clear (SPANIEL *et al.* 2013). Tissue-specific

differences in snRNP distribution have also been reported (HAMM AND MATTAJ 1989). Transcriptional activity represents the first step of regulation in the lifecycle of U6, yet how these differences occur or what their effect is on splicing efficiency or fidelity has not been investigated.

Figure 2-3. U6 snRNA promoter structure is divergent in eukaryotes.

(A) U6 promoter structure in *S. cerevisiae*. U6 is under the control of a Pol III Type II promoter, with an upstream TATA box, an internal A block, and downstream B block. The TFIIIC complex recognizes the A and B blocks and directs binding of the TFIIIB complex to the TATA box. Nhp6 also promotes transcription, but its binding site is uncertain. A possible nucleosome is indicated by a grey oval.

(B) U6 promoter structure in *Homo sapiens*. U6 is under the control of a Pol III Type III promoter, where promoter elements (in black) are exclusively upstream of the transcription start site. The TATA box is recognized by the TFIIIB2 complex, while the PSE is recognized by the SNAPc complex. Factors including OCT1, STAF, and CDH2 interact with the DSE. p38 inhibits Oct-1 binding to the DSE. A nucleosome between the DSE and PSE enhances Oct-1 and SNAPc binding.



2.4 Localization during biogenesis

It is generally thought that U6 remains in the nucleus after transcription and for all biogenesis steps (VANKAN *et al.* 1990; PESSA *et al.* 2008), while the PolI II-synthesized RNAs have a cytoplasmic assembly step (reviewed in (MATERA AND WANG 2014)). At some point after transcription termination, the La/Lhp1 protein is displaced from the 3' end of U6 snRNA. Binding by La/Lhp1 is not sufficient for complete nuclear retention in *Xenopus* oocytes (BOELENS *et al.* 1995). Instead, binding of the Lsm2-8 heteroheptameric protein ring acts as the primary nuclear retention signal (SPILLER *et al.* 2007a; SPILLER *et al.* 2007b).

U6 is modified (pseudouridylated and 2'-*O*-methylated) in the nucleolus by snoRNPs (TYCOWSKI *et al.* 1998; GANOT *et al.* 1999; LANGE AND GERBI 2000), while the PolI II-synthesized snRNAs are modified in Cajal bodies by scaRNPs (DARZACQ *et al.* 2002). U6 is not retained in the nucleolus, but rather transiently passes through to obtain modifications (LANGE AND GERBI 2000). Nucleolar localization of *Xenopus* U6 does not depend on base pairing interactions with U4 or U2, or formation of the U6 5'SL (GERBI AND LANGE 2002) but both nucleolar localization and Cajal body localization do depend on the 3' oligo-U tail (GERBI AND LANGE 2002). Furthermore, localization does not depend on post-transcriptional modifications such as pseudouridylation or 2'-*O*-methylation. In *S. cerevisiae*, U6 is not constitutively pseudouridylated and is not 2'-*O*-methylated, and therefore is thought to be entirely nucleoplasmic (BERTRAND *et al.* 1998).

The presence of snRNPs in subnuclear organelles was first identified in the early 1990s (CARMO-FONSECA *et al.* 1991a; CARMO-FONSECA *et al.* 1991b; MATERA AND WARD 1993). Work over the past decade has revealed that U6 can be assembled into the U4/U6 di-snRNP and U4/U6.U5 tri-snRNP in subnuclear compartments called Cajal bodies (CBs) (STANEK *et al.* 2003; STANEK AND NEUGEBAUER 2004; KLINGAUF *et al.* 2006; NOVOTNY *et al.* 2011; NOVOTNY *et al.* 2015). CBs are also the compartment in which the Sm class snRNAs are post-transcriptionally modified by scaRNPs (JADY *et al.* 2003). In humans, U6 snRNA is localized to CBs through its interaction with the U6 snRNP protein SART3 (aka hPrp24, p110, or Tip110) (STANEK *et al.* 2003).

The N-terminal half-a-tetratricopeptide (HAT) domains of SART3 are necessary for U6 localization to CBs (STANEK *et al.* 2003). Localization of SART3 depends upon its interaction with coilin, the major structural CB protein (XU *et al.* 2005). Interestingly, U4/U6•SART3, rather than U6•SART3, accumulates in CBs, suggesting that CBs may be a site for snRNP assembly (STANEK AND NEUGEBAUER 2004). While formation of U4/U6 is enhanced ~10-fold in CBs, in part due to the increased local concentration of snRNAs (KLINGAUF *et al.* 2006; NOVOTNY *et al.* 2011), snRNP formation can also occur elsewhere in the nucleus as CBs are non-essential (LEMM *et al.* 2006). SART3 localizes within CBs transiently (on the order of a few seconds) (DUNDR *et al.* 2004) and re-enters new CBs after a period in the nucleoplasm (STANEK *et al.* 2008), suggesting that SART3 (and U6 along with it) localizes to CBs during both initial biogenesis and recycling after a round of splicing. Depletion of factors needed for tri-snRNP formation (hPrp6 and hPrp8) induces formation of CBs and results in accumulation of U4/U6 in CBs (NOVOTNY *et al.* 2011; NOVOTNY *et al.* 2015). Because SART3 interacts with U6 and U4/U6 but is displaced upon formation of U4/U6.U5 (BELL *et al.* 2002), it can anchor immature U4/U6 di-snRNPs to CBs through its interaction with coilin (NOVOTNY *et al.* 2015). Localization of unassembled U6 snRNP in sub-nuclear CBs may prevent incorporation of immature splicing components into active spliceosomes. Retention of U6 snRNP in CBs is a convenient checkpoint for the cell to ensure that mature U6 (in the tri-snRNP) is used in splicing.

In yeast, the U6 gene is localized near the nucleolar periphery (BELAGAL *et al.* 2016). Although yeast also have nucleolar bodies akin to CBs (VERHEGGEN *et al.* 2002), little work has been done to determine if and how sub-nuclear localization plays a role in U6 snRNP biogenesis in yeast. By analogy, work on the biogenesis of the U3 snoRNP has demonstrated that there are differences in localization during biogenesis in yeast and humans, but that overall the processes are similar (VERHEGGEN *et al.* 2002). It remains to be determined if and why yeast U6 localizes to any sub-nuclear compartments, and if the U4/U6.U5 tri-snRNP is assembled in a specific region in yeast nuclei.

2.5 Posttranscriptional modification

U6 snRNA is extensively modified in metazoans, and many of these post-transcriptional modifications are evolutionarily conserved even in *S. pombe*. In contrast, *S. cerevisiae* U6 snRNA contains very few modifications, and only one known obligate modification (3' end processing). The post-transcriptional modifications of U6 in *S. cerevisiae* and humans are summarized in **Table 2-1** and discussed below. Post-transcriptional modifications of U6 are likely to be involved in modulation of U6 interactions with other RNAs and proteins throughout the splicing cycle. However, the precise effect that modifications have on U6 RNA structure or interactions with protein and other snRNAs is poorly understood, as is the timing and cellular location of where these modifications occur.

2.5.1 5' Capping

Instead of the 2,2,7-trimethylguanosine (TMG) cap that the other snRNAs possess or the 7-methylguanosine (m^7G/cap-0) cap of mRNAs, human U6 possesses a γ-monomethyl phosphate 5' modification (SINGH AND REDDY 1989). The γ-monomethyl cap is shared by other Pol III transcripts, including the non-coding 7SK RNA (GUPTA *et al.* 1990a). However, capping of U6 small nuclear RNA *in vitro* is not dependent upon transcription, suggesting that capping of U6 is not obligatorily co-transcriptional *in vivo* (GUPTA *et al.* 1990b). Capping is dependent upon the 5' stem loop and is also sequence-dependent (SINGH *et al.* 1990), unlike co-transcriptional capping of Pol II transcripts. It is not known if *S. cerevisiae* U6 has a γ-monomethyl cap, although deletion of its first 11 nucleotides results in installation of a TMG cap despite still being transcribed by RNA Pol III (KWAN *et al.* 2000). This finding implies that yeast U6 normally receives a 5' modification that prevents TMG capping, but that disruption of conserved sequences at the base of the 5'-stem, which are the determinant for γ-monomethyl capping in humans (SINGH *et al.* 1990), blocks this modification.

The function of the U6 cap is not known, although it may be important for stability (SHUMYATSKY *et al.* 1993). TMG capping of other snRNAs is important for transport into the

nucleus, but the γ-monomethyl cap of U6 likely does not impact transport (SHUMYATSKY *et al.* 1993) or retention in the nucleus (SPILLER *et al.* 2007a). The γ-monomethyl cap could also play a role in displacing La/Lhp1 from U6 RNA, as La interacts with the 5' triphosphate prior to capping (BHATTACHARYA *et al.* 2002).

A 130 kDa protein responsible for cap formation in U6 and 7SK was isolated in 1994 (SHIMBA AND REDDY 1994), but the gene was not identified until 2007 (JERONIMO *et al.* 2007). Bin3 (renamed methyl phosphate capping enzyme, MePCE) is responsible for installation of the γ-monomethyl cap on 7SK and U6 RNA. While U6 RNA co-purified with MePCE, knockdown of MePCE reduced levels of 7SK RNA but not U6 RNA (JERONIMO *et al.* 2007). It is currently not clear how or to what degree the 5' cap of U6 contributes to stability or function. Interestingly, *S. cerevisiae* does not have a clear homolog of MePCE, and the methylation status of the 5' triphosphate is unknown. For a comprehensive review on the enzyme Bin3/MePCE, see reference (COSGROVE *et al.* 2012).

2.5.2 3' end modifications

After synthesis by Pol III, U6 is left with a oligo-U tail of variable length with a terminal 2',3' cis diol. The oligo-U tail and terminal 2' and 3' hydroxyl groups are bound by the La protein (Lhp1 in yeast) (STEFANO 1984; RINKE AND STEITZ 1985; TERNS *et al.* 1992). While the length of the oligo-U tail is initially heterogeneous in part due to Pol III termination, U6 is both post-transcriptionally shortened and extended (oligouridylated) (REDDY *et al.* 1987).

U6 snRNA is 3' oligouridylated in many eukaryotes, including *S. pombe* and humans. A U6 terminal uridyl transferase (TUTase) activity was identified in 1998 (TRIPPE *et al.* 1998) and the gene, TUT1, was identified in 2006 (TRIPPE *et al.* 2006). Characterization of this enzyme revealed that it requires the presence of uridine at the 3' end of U6 and that it can add three additional uridines *in vitro* (TRIPPE *et al.* 2003). Interestingly, TUT1 may recognize additional secondary structure in U6 and therefore may preferentially oligouridylate U6 while the RNA is in a certain conformation (YAMASHITA *et al.* 2017). TUT1 localizes to the nucleolus, suggesting that

oligouridylation may occur alongside other modifications such as pseudouridylation and 2'-*O*-methylation (discussed in section 2.5.3 and 2.5.4) (TRIPPE *et al.* 2006).

The enzyme responsible for 3' tail shortening was characterized in the late 1990's (BOOTH AND PUGH 1997), but the gene was not identified for an additional fifteen years (MROCZEK *et al.* 2012; SHCHEPACHEV *et al.* 2012; HILCENKO *et al.* 2013). The enzyme was named U six biogenesis protein 1 (Usb1), and is a 3'-5' exonuclease (MROCZEK *et al.* 2012; SHCHEPACHEV *et al.* 2012; HILCENKO *et al.* 2013). Usb1 trims back the oligo-U tail and leaves a phosphate group on the terminal nucleotide. In humans, the terminal nucleotide has a 2',3'-cyclic phosphate, while in yeast the terminal phosphate is a 3' non-cyclic phosphate (LUND AND DAHLBERG 1992; DIDYCHUK *et al.* 2017). The majority of human U6 snRNAs end with five terminal uridines and a 2',3'-cyclic phosphate (SHCHEPACHEV *et al.* 2015).

Some metazoans do not have a homolog of Usb1. These organisms, including *C. elegans*, do not possess the phosphoryl or cyclic phosphate modifications, but instead have either a 2',3'-cis diol in their mature form or an intriguing "blocked" modification (LUND AND DAHLBERG 1992). Interestingly, *C. elegans* does not have a Usb1 homolog, but rather has a TUTase called USIP-1 (RUEGGER *et al.* 2015). The USIP-1 enzyme interacts with the U6 snRNP protein SART3 (discussed in section 2.6.2). The chemical identity of the 3' end produced by USIP-1 is not known.

The 3' tail of U6 can also be adenylated (CHEN *et al.* 2000). Adenylation inhibits uridylation, and may target U6 for degradation. In cells deficient for Usb1 activity, U6 with adenylated tails accumulate (HILCENKO *et al.* 2013; SHCHEPACHEV *et al.* 2015). Polyadenylation may result in targeting U6 to the nuclear exosome. Human Usb1 is active on polyadenylate tails, and therefore can counteract polyadenylation (HILCENKO *et al.* 2013). It is unclear what the polyadenylating enzyme is, or if it can function on the product of human Usb1, a 2',3'-cyclic phosphate. An additional phosphatase enzyme may be necessary to remove the cyclic phosphate and allow the polyadenylating enzyme to function.

The timing and location of 3' end modification by Usb1 is not known. An early study suggested that the length and modification of U6 changed during the splicing cycle (TAZI *et al.* 1993). Tazi *et al*. made the intriguing observation that 3' end modification of U6 required pre-mRNA, raising the possibility that U6 may not be processed until after a round of splicing is complete. In this way, after a successful first round of splicing, the cell could mark functional U6 to be protected and kept for additional rounds of splicing. However, a later study showed that depletion of U1 and U2 snRNAs had no effect on U6 processing (GU *et al.* 1997). It is possible that the 3' end of U6 is altered throughout the splicing cycle, as the 3' end binding Lsm2-8 complex (discussed in section 2.6.3) dissociates during spliceosome activation, which may render the U6 3' end accessible for modification. Interestingly, Usb1 localizes to the nucleus but not the nucleolus (MROCZEK *et al.* 2012). This implies that in humans, Usb1 recognizes a TUT1-extended U6 in the nucleoplasm after modification in the nucleolus but before it has traveled to Cajal bodies for di- and tri-snRNP assembly. It is unclear if TUT1 can function on the terminal 2',3'-cyclic phosphate of mature, Usb1-processed U6, and it is also unclear if uridylation occurs only once after transcription termination, or if U6 is trimmed and extended repeatedly throughout its life cycle.

Modification of the U6 3' end is important for recognition by U6 3' end binding proteins. The La protein binds U6 immediately after transcription termination and specifically recognizes the 2',3'-cis diol. Indeed, while a subset of U6 molecules found in the cell have cis diols and are bound by La (RINKE AND STEITZ 1985), maturation of U6 and incorporation into the U6 snRNP requires the 3' modification by Usb1. Modification of the 3' end reduces the ability of La to bind U6 (TERNS *et al.* 1992) and is important for binding of the Lsm2-8 ring (LICHT *et al.* 2008). Interestingly, human Lsm2-8 preferentially binds U6 containing a 2',3'-cyclic phosphate modification (LICHT *et al.* 2008), while *S. cerevisiae* Lsm2-8 prefers to bind a non-cyclic phosphate, suggesting that Usb1 and Lsm2-8 co-evolved to allow for tight binding by Lsm2-8 on the product of Usb1 activity (DIDYCHUK *et al.* 2017).

2.5.3 Pseudouridylation

Pseudouridylation of RNA is thought to be structurally stabilizing because of increased base stacking (DAVIS 1995) as well as the added hydrogen bonding potential and conformational flexibility due to the C-C glycosidic bond (CHARETTE AND GRAY 2000). The effects of pseudouridylation on structure and RNA-protein interactions in U2 snRNA have been well-studied (YU *et al.* 1998; NEWBY AND GREENBAUM 2001; WU *et al.* 2016; VAN DER FELTZ *et al.* 2017), but the effects of pseudouridylation on U6 RNA have not. In *S. cerevisiae*, U6 is not constitutively pseudouridylated (MASSENET *et al.* 1999). Recently, position U28 was found to be inducibly pseudouridylated during filamentous growth by the pseudouridine synthase Pus1 (BASAK AND QUERY 2014). Interestingly, mutations within U6 far from U28 (U36C and G50U) increased pseudouridylation of U28. These mutations likely disrupt interactions of U6 with U6 snRNP protein Prp24 (MONTEMAYOR *et al.* 2014) and would also likely alter interactions within the activated spliceosome. It is not clear at which step in the U6 lifecycle U6-U28 is pseudouridylated, or how changing RNA-protein interactions elsewhere in the RNA increases pseudouridylation of U6-U28. Furthermore, it is not clear how pseudouridylation of U6-U28 would affect RNA-RNA or RNA-proteins interactions in different splicing complexes. In the U4/U6.U5 tri-snRNP, U6-U28 interacts with U5-U99 (NGUYEN *et al.* 2016), whereas in the C complex, U6-U28 is within 5 Å of Bud31 and Ecm2 (GALEJ *et al.* 2016). Understanding how pseudouridylation of U6-U28 affects RNP structure, and understanding how pseudouridylation of U6 induces filamentous growth in yeast, could reveal important new insights into spliceosome biology.

In contrast, human U6 and U6atac are constitutively pseudouridylated at several positions (MASSENET AND BRANLANT 1999; MASSENET *et al.* 1999). These positions (U31, U40, and U86 in U6; U83 in U6atac) (**Figure 2-4**) (MASSENET AND BRANLANT 1999) are in functionally important parts of the RNA, although it is unclear if or how modifications in U6 can stabilize its structure. Recently solved cryo-EM structures have allowed for structural analysis of splicing complexes;

however, they lack the resolution to unambiguously identify posttranscriptional modifications. Improved resolution may lead to a better understanding of how modifications in U6 function.

Localization in the nucleolus is sufficient for formation of both the three pseudouridines in human U6 and the eight 2'-*O*-methyl groups (discussed in section 2.5.4), suggesting that all of the machinery required for these modifications is present in the nucleolus (GANOT *et al.* 1999). Localization of U6 during its biogenesis is discussed in section 2.4.

2.5.4 Ribose 2'-*O*-methylation

Ribose 2'-*O*-methylation is a ubiquitous modification that can significantly stabilize RNA, either structurally (SASHITAL *et al.* 2007) or metabolically (by reducing hydrolytic or nucleolytic cleavage). Like pseudouridylation, ribose methylation occurs in the nucleolus (GANOT *et al.* 1999). Human U6 contains eight ribose methylations, at positions A47, A53, G54, C60, C62, C63, A70, and C77, which could stabilize secondary structure either in the U6 snRNP (the internal stem loop; ISL) or in U4/U6 (Stem I and Stem II) (**Figure 2-5**). Ribose methylation requires a methyltransferase enzyme and "methylation guide" (mg) snoRNAs, including mgU6-47 and mgU6-77 (TYCOWSKI *et al.* 1998), mgU6-53 (GANOT *et al.* 1999), and MBII-166 (which modifies U6-C60) (HUTTENHOFER *et al.* 2001). Modification of many of the U6 ribose groups are conserved in plants (KISS *et al.* 1987) and in *S. pombe* (GU *et al.* 1996). The mgU6-47 snoRNA is conserved in *S. pombe* and is required for methylation. However, while disruption of the mgU6-47 gene resulted in complete loss of methylation, only a small cold-sensitive splicing defect was observed (ZHOU *et al.* 2002). Interestingly, human U6atac has no ribose 2'-*O*-methyl groups, despite forming a homologous ISL in which ribose methylations in U6 are concentrated (MASSENET AND BRANLANT 1999).

Recent RiboMeth-seq results show that the level of ribose modification varies in different cell types (KROGH *et al.* 2017). While ribose methylation may not be essential, its conservation across species suggests it (along with pseudouridylation and other modifications) may play a role

in fine-tuning the stability of U6 and its complexes to modulate splicing, perhaps in response to variation in cellular temperatures that occur even in homeotherms.

2.5.5 *N*-6-adenosine and *N*-2-guanosine methylation

Human U6 contains a single *N*-6-methyladenosine modification (m^6A) at position A43 (SHIMBA *et al.* 1995). Study of m^6A modification has recently undergone a renaissance due to transcriptome-wide mapping studies that show it is a ubiquitous and regulated modification in mRNA (reviewed in (MEYER AND JAFFREY 2014)). Most mRNA m^6A residues are found within a consensus sequence (GAC or AAC) and are modified by the METTL3 complex. U6-A43 does not lie within a consensus sequence and is not modified by METTL3, but rather by a recently identified methyltransferase, METTL16 (PENDLETON *et al.* 2017; WARDA *et al.* 2017). Interestingly, METTL16 activity is dependent on both sequence and secondary structure. This is consistent with earlier reports that modification of U6-A43 depends on formation of the adjacent ISL secondary structure within U6 (SHIMBA *et al.* 1995). It is unclear when A43 is modified to m^6A, but *in vitro* studies suggest that the conversion is rapid (SHIMBA *et al.* 1995). A recent report showed that METTL16 associates on U6 alongside the capping enzyme MePCE (see section 2.5.1) and the 3' end binding protein La (WARDA *et al.* 2017). Determining the order of U6 modifications and the effect of multiple modifying enzymes competing for the same substrate may yield interesting insights into U6 biogenesis.

Modified U6 is incorporated into U4/U6 snRNPs, suggesting it is functional (BRINGMANN AND LUHRMANN 1987). The modification at position A43 lies in the center of the ACAGAGA motif, a perfectly conserved and essential motif that base pairs to the intron adjacent to the 5' splice site and helps organize the spliceosome active site through formation of base triple interactions (FICA *et al.* 2014). Biophysical studies of m^6A modification within double-stranded RNA indicate that it destabilizes base pairing, but at the ends of helices is stabilizing when stacked and unpaired (ROOST *et al.* 2015). In a recent structure of a human C* complex, U6-A43 is unpaired across

from an adenosine in the intron (BERTRAM *et al.* 2017). The methylation state of the nucleotide could not be determined in the structure.

Human U6 also contains a single *N*-2-guanosine methylation (m^2G) at position G72 (EPSTEIN *et al.* 1980). The role of this modification is unknown, as G72 is expected to fall within a G-C base pair in both the U6 ISL and within U4/U6 Stem II (**Figure 2-4, 5**) and because m^2G is expected to be isoenergetic with guanosine in G-C base pairs (RIFE *et al.* 1998). Like base methylation of A43, methylation of G72 may also depend on secondary structure (GANOT *et al.* 1999).

2.5.6 Splicing of U6 snRNA

The sole copy of the U6 gene in *S. pombe* has an intron that is removed by the spliceosome (TANI AND OHSHIMA 1989). The intron is located immediately adjacent to the highly conserved catalytic AGC triad that is essential for splicing catalysis, leading to the hypothesis that the intron in *S. pombe* U6 RNA arose from reverse-splicing of an excised intron into a U6 molecule in the active spliceosome, followed by reverse transcription and incorporation into the genome (BROW AND GUTHRIE 1989; TANI AND OHSHIMA 1991). This hypothesis suggested that U6 (and specifically this region of U6) are near the active site of the spliceosome. Indeed, it is now well-established that this region of U6 forms the catalytic core of the spliceosome. Recent bioinformatics analysis shows that many yeast species have evolutionarily unrelated introns and that these introns are distributed throughout the U6 sequence (CANZLER *et al.* 2016). These introns (up to four introns in a single fungal U6 gene) possess canonical 5' splice sites, branch sites, and 3' splice sites, but are otherwise not conserved in sequence. Phylogenetic analysis suggests that while the presence of some introns in related species likely arose from a common ancestor, many of the U6 introns may be species-specific insertions (CANZLER *et al.* 2016). Experimental validation that U6 genes containing introns are expressed and spliced is necessary. It is possible that species with multiple U6 paralogs of which some contain introns may use splicing as a method to regulate U6 levels.

Spliceosomal introns can also be found in U2 (TAKAHASHI *et al.* 1993) and U5 snRNAs (TAKAHASHI *et al.* 1996), supporting the hypothesis that RNAs close to the active site of the spliceosome can be targets of reverse splicing. However, spliceosomal introns have also been identified in U1 snRNA (TAKAHASHI *et al.* 1996), which departs before the active site of the spliceosome is formed, and in rRNA genes (BHATTACHARYA *et al.* 2000). It is not clear what the determinants of novel spliceosomal intron insertion in ncRNAs are, but possibilities include RNA abundance, structure, or proximity to the spliceosomal active site.

Figure 2-4. Sequence and putative secondary structure of *S. cerevisiae* U6, human U6, and U6atac.

The secondary structure of U6 from S. cerevisiae (left) is based on the structure of the U6 snRNP core (Montemayor et al. 2014), and includes the 5'SL, telestem, asymmetric bulge, ISL, and 3' tail. The secondary structure of human U6 and U6atac within the U6 snRNP have not been experimentally determined, and are shown with secondary structure to mimic that of yeast U6. Human U6 contains a 5'SL, the ISL, and 3' tail, and may contain an asymmetric bulge and telestem region. U6atac lacks a 5'SL but contains an additional 3'SL. Constitutively modified nucleotides are highlighted in red. Catalytically important residues are boxed in red.

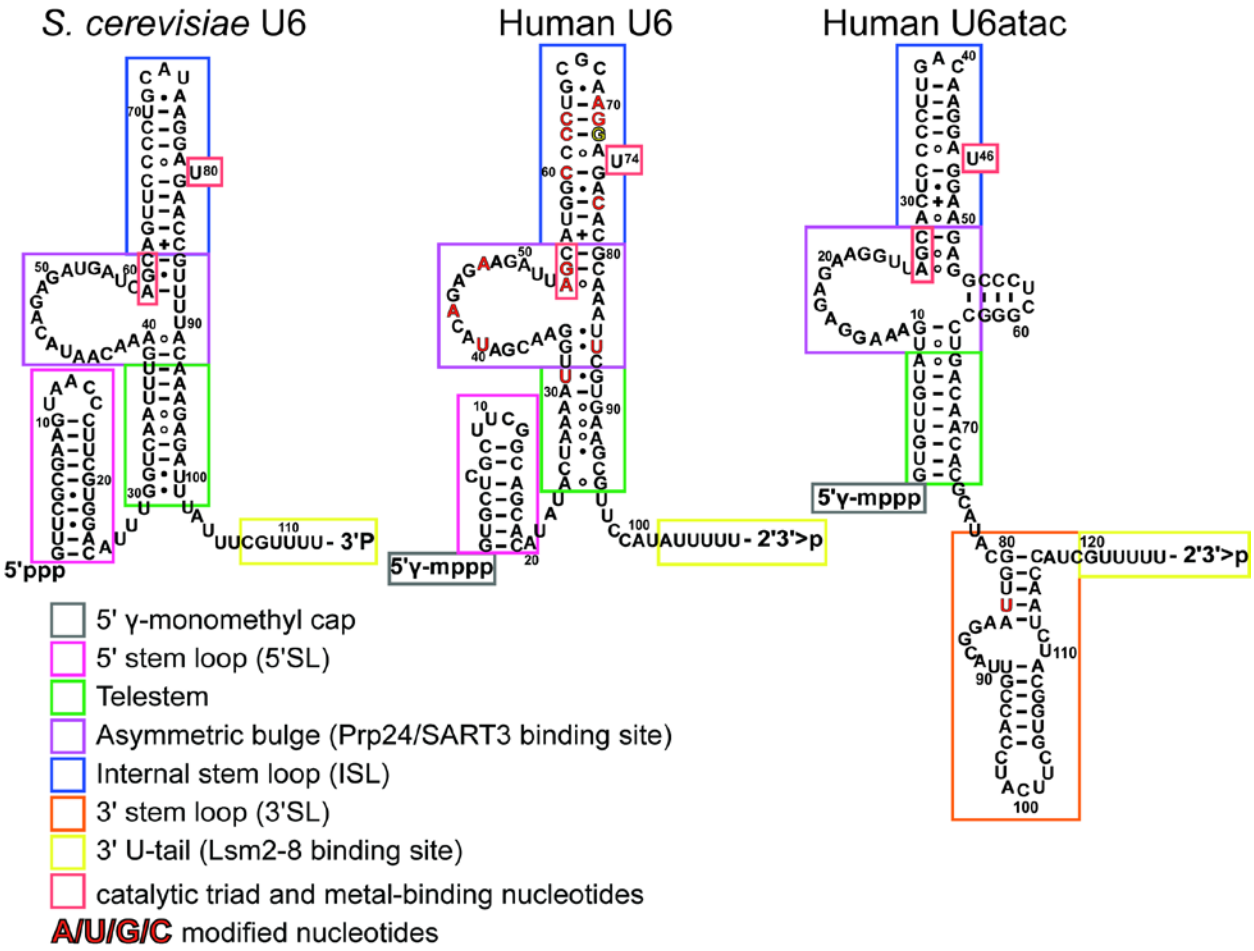


Figure 2-5. U6 undergoes large conformational changes during the splicing cycle.

(A) Cartoon of base pairing throughout the splicing cycle. For simplicity, minimal interacting sequences of U2 and U5 snRNAs are shown.

(B) Structure of U6 and its partners in different splicing complexes.

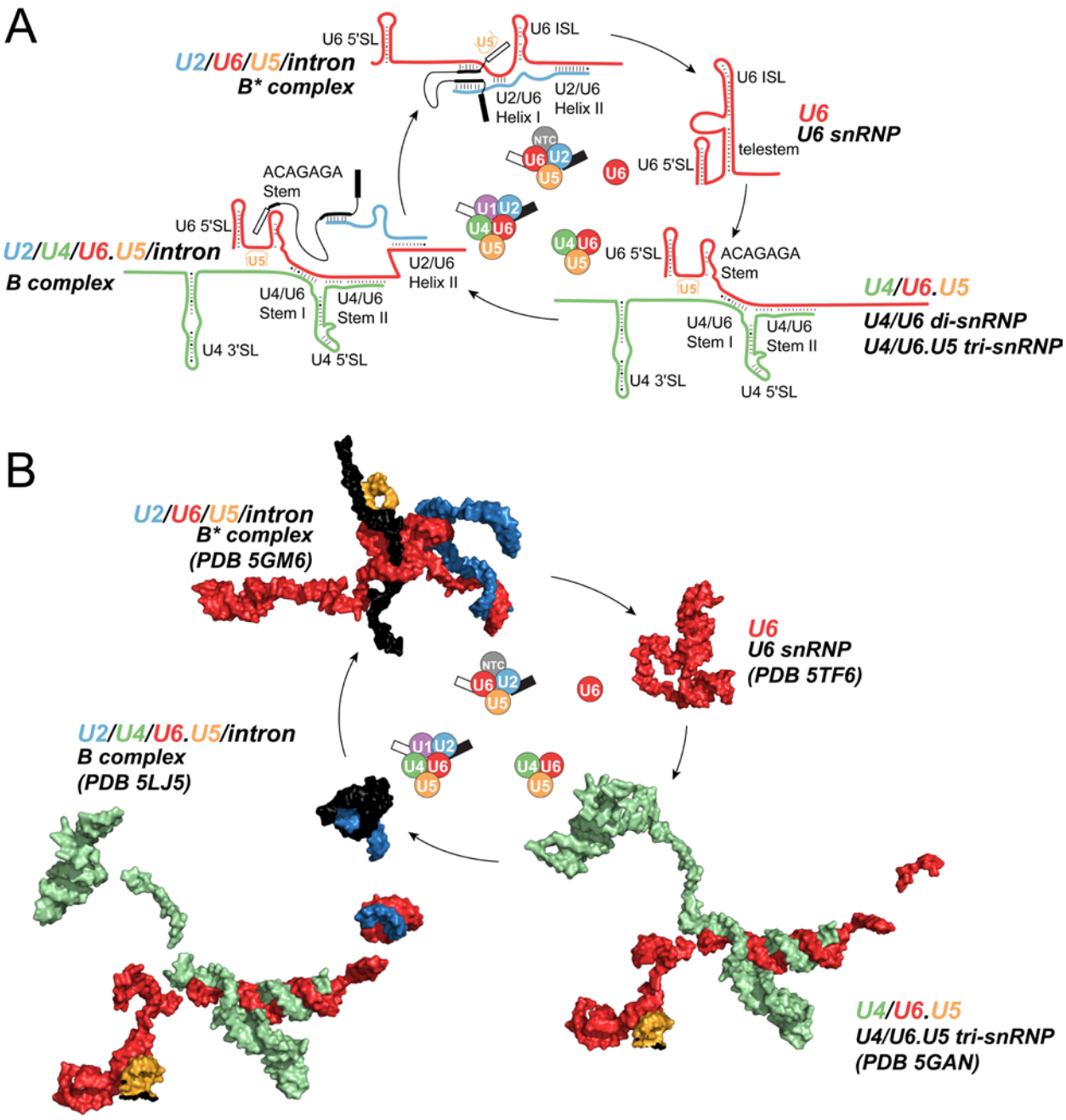


Table 2-1. Post-transcriptional modifications of U6 snRNA in humans (*Hs*) and *S. cerevisiae* (*Sc*).

Modification	Organism	Position	Enzyme
Terminal modifications			
5' capping	*Hs*	5' γ-monomethyl	Bin3/MePCE
3' end trimming	*Sc*	3' terminal 3' phosphate	Usb1
	Hs	3' terminal 2',3'-cyclic phosphate	Usb1/Mpn1
3' end oligouridylation	*Hs*	3' end	U6 TUTase
3' end oligoadenylation	*Hs*	3' end	TRAMP?
Internal modifications			
Pseudouridylation (ψ)	*Sc*	U28	Pus1
	Hs	U31	?
	Hs	U40	?
	Hs	U86	?
Ribose 2'-*O*-methylation (2'-*O*-me)	*Hs*	A47	mgU6-47 (SNORD7)
		A53	mgU6-53 (SNORD8), mgU6-53B (SNORD9)
		G54	?
		C60	MBII-166
		C62	?
		C63	?
		A70	?
		C77	mgU6-77 (SNORD10)
N-6-adenosine methylation (m^6A)	*Hs*	A43	METTL16
N-2-guanosine methylation (m^2G)	*Hs*	G72	?

2.6 The U6 snRNP

2.6.1 U6 snRNA secondary structure

U6 is the most conserved of the snRNAs in sequence and length (BROW AND GUTHRIE 1988). The least conserved region is the 5' stem-loop (5'SL), which varies in both sequence and length across eukarya despite maintaining a stem-loop structure (ROIHA *et al.* 1989). *C. merolae*, a moderate thermophile, contains a minimal spliceosome that lacks U1 snRNA, yet has an unusually stable 5'SL with 22 base pairs (versus 10 base pairs in *S. cerevisiae*) (STARK *et al.* 2015). The 5'SL persists throughout the various U6-containing splicing complexes. It does not have any known catalytic function, but does correspond to part of the A-block promoter element in *S. cerevisiae* and is necessary for efficient 5' capping in humans (discussed in section 2.5.1). The 5' capping determinants are conserved in yeast phyla (ROIHA *et al.* 1989), further evidence that 5' capping is also conserved.

The remaining portion of U6 forms several mutually exclusive structures. While in the U6 snRNP, U6 consists of the 5'SL, the telestem, and the internal stem-loop (ISL, previously called the 3' stem-loop or intramolecular stem-loop) (**Figure 2-4, 5**). While the 5'SL likely remains stably base paired throughout the splicing cycle, the less stable telestem and ISL helices make dynamic transitions into other secondary structures during spliceosome assembly and activation. Incorporation of U6 into the spliceosome requires unwinding of the ISL and subsequent base pairing to U4, a mutually exclusive interaction. As observed in the U4/U6.U5 tri-snRNP structure (AGAFONOV *et al.* 2016; NGUYEN *et al.* 2016; WAN *et al.* 2016b), the telestem is unwound when U6 is paired to U4 snRNA. However, *in vitro* data suggests the telestem can form transiently in free U4/U6 RNA (BROW AND VIDAVER 1995; RODGERS *et al.* 2016). During spliceosome activation, U4 is unwound from U4/U6 by the helicase Brr2 (RAGHUNATHAN AND GUTHRIE 1998a), U6 base pairs with U2 (MADHANI AND GUTHRIE 1992), and the U6 ISL reforms (**Figure 2-5**). The telestem is mutually exclusive with formation of U2/U6 helix II (**Figure 2-5**). The telestem enhances Prp24 affinity and Prp24-mediated annealing (DIDYCHUK *et al.* 2016) and also destabilizes U4/U6 (BROW

AND VIDAVER 1995; RODGERS *et al.* 2016). The dynamic telestem and U6 ISL structures may help U6 molecules avoid kinetic or thermodynamic folding traps during spliceosome assembly and activation.

Essential nucleotides in the U6 ACAGAGA box and the AGC triad were identified early on (FABRIZIO AND ABELSON 1990; MADHANI *et al.* 1990). During activation, the ACAGAGA box interacts with the 5' splice site (KANDELS-LEWIS AND SERAPHIN 1993; LESSER AND GUTHRIE 1993; JOHNSON AND ABELSON 2001; RAUHUT *et al.* 2016; YAN *et al.* 2016). The AGC triad is important for intramolecular base pairing within U6 as well as pairing to U2 (in U2/U6 helix I) and U4 (in U4/U6 Stem I) (**Figure 2-5**). The AGC triad makes intramolecular base-triple tertiary contacts with the last two nucleotides of the ACAGAGA sequence (HILLIKER AND STALEY 2004; FICA *et al.* 2014; GALEJ *et al.* 2016; RAUHUT *et al.* 2016; WAN *et al.* 2016a; YAN *et al.* 2016; BERTRAM *et al.* 2017; FICA *et al.* 2017; YAN *et al.* 2017; ZHANG *et al.* 2017). The AGC triad and base triple interactions are conserved in the self-splicing Group II introns, which share a common evolutionary ancestor with the spliceosome (PYLE 2016). A conserved bulged residue in the ISL (U80 in *S. cerevisiae*) is involved in catalysis by coordinating a catalytic metal ion and by forming a base triple with the C of the AGC triad (FICA *et al.* 2013; FICA *et al.* 2014). Mutation of U80 to C or A has no obvious effect on yeast growth, while mutation to a G is lethal unless U6-C67 is mutated to prevent pairing across the ISL and U4-G14 is mutated to maintain pairing with U6 residue 67 (MCMANUS *et al.* 2007). Thus, the tertiary structure of this region of the ISL is more important than the primary structure.

U6atac lacks a 5'SL, but contains a similar central secondary structure (**Figure 2-4**). Indeed, the yeast or human U6 ISL can functionally replace the U6atac ISL (SHUKLA AND PADGETT 2001). U6atac contains an additional stem loop 3' of the ISL, which is important for targeting to the minor spliceosome (DIETRICH *et al.* 2009).

2.6.2 Prp24/SART3

The yeast Prp24 protein was discovered through a genetic selection for suppressors of mutations in U4 RNA that disrupt U4/U6 pairing (SHANNON AND GUTHRIE 1991) as well as a screen for heat-sensitive mutations in splicing (VIJAYRAGHAVAN *et al.* 1989). Mutations in Prp24 were also isolated in a selection for suppressors of a mutation that stabilizes the U6 ISL (VIDAVER *et al.* 1999; MONTEMAYOR *et al.* 2014), underscoring its importance for the structural transitions of U6 snRNA. In *S. cerevisiae*, Prp24 contains three canonical RNA recognition motifs (RRMs) and a fourth "occluded" RRM in which the typical RNA binding surface, a four-stranded β-sheet, is masked by additional helices (RADER AND GUTHRIE 2002; KWAN AND BROW 2005; BAE *et al.* 2007; MARTIN-TUMASZ *et al.* 2011; MONTEMAYOR *et al.* 2014). The human homolog of Prp24, SART3, has only two RRMs (which may correspond to yeast RRMs 1 and 2 or 2 and 3), and instead has multiple N-terminal half-a-tetratricopeptide (HAT) repeats (BELL *et al.* 2002; RADER AND GUTHRIE 2002; KWAN AND BROW 2005). SART3 has recently been shown to dimerize through these HAT repeats (PARK *et al.* 2016; ZHANG *et al.* 2016), suggesting that Prp24 and SART3 may both require four RRMs to function efficiently. Prp24 from *S. pombe* may be an evolutionary intermediate in that it has four RRMs plus the HAT domains. All Prp24 homologs contain a highly conserved short peptide at the C-terminus called the SNFFL box, which is important for interactions between Prp24 and Lsm2-8 (FROMONT-RACINE *et al.* 2000; RADER AND GUTHRIE 2002). Prp24 is essential for unwinding the U6 ISL and annealing U4 to U6 (discussed in section 2.7.1) (GHETTI *et al.* 1995; RADER AND GUTHRIE 2002; DIDYCHUK *et al.* 2016).

2.6.3 Lsm2-8

The Lsm proteins are paralogs of the Sm proteins that bind to U1, U2, U4 and U5 snRNAs, and are related to bacterial Hfq and Sm-like archaeal proteins (MURA *et al.* 2013). Lsm proteins are small (9-21 kDa) and assemble into at least 2 different heteroheptameric, ring-shaped complexes, Lsm1-7 and Lsm2-8. The Lsm1-7 ring localizes to the cytosol and is involved in the 5'-3' mRNA decay pathway (THARUN *et al.* 2000). The Lsm2-8 ring incorporates the Lsm8 protein

in place of Lsm1 and localizes to the nucleus where it binds the 3' U-tail of U6 (ACHSEL *et al.* 1999; MAYES *et al.* 1999). The *S. cerevisiae* Lsm2-8 complex preferentially binds U6 3' ends modified with a terminal phosphate (DIDYCHUK *et al.* 2017), while the metazoan Lsm2-8 complex preferentially binds a terminal 2',3'-cyclic phosphate (LICHT *et al.* 2008), as discussed above in section 2.5.2. Structures of the Lsm1-7 and Lsm2-8 complexes revealed that the two rings are exceedingly similar, with <1 Å rmsd between subunits Lsm2-7 in each structure and between Lsm1 and Lsm8 (SHARIF AND CONTI 2013; ZHOU *et al.* 2014a; ZHOU *et al.* 2014b). The crystal structure of the *S. cerevisiae* Lsm2-8 bound to a short RNA oligo was determined and revealed that the 3' end of U6 is recognized in the center of the ring (ZHOU *et al.* 2014a). The Lsm2-8 complex can also be observed in cryo-EM structures of the U4/U6.U5 tri-snRNP, albeit at lower resolutions (AGAFONOV *et al.* 2016; NGUYEN *et al.* 2016; WAN *et al.* 2016b).

Unlike the Sm ring, the Lsm2-8 ring is stable in the absence of RNA (ACHSEL *et al.* 1999) and assembles without assistance from other proteins. The Sm ring, in contrast, requires the SMN complex and Gemin proteins to assemble three Sm subcomplexes in order to assemble and bind RNA (reviewed in (BATTLE *et al.* 2006; MATERA AND WANG 2014). The Lsm2-8 ring is likely imported into the nucleus via importin β/KAP95 (SPILLER *et al.* 2007a), where the N- and C-terminal domains of Lsm8 act as nuclear retention signals for U6 (SPILLER *et al.* 2007a; SPILLER *et al.* 2007b; REIJNS *et al.* 2009).

2.6.4 Structure of the U6 snRNP

The structure of the core of the U6 snRNP, with the majority of both U6 snRNA (nucleotides 30-101) and Prp24 (all four RRMs), has been determined (MONTEMAYOR *et al.* 2014). To date, this is the only structure of a protein containing four RRMs bound to RNA. It reveals how three of the RRMs cooperate to specifically and tightly bind U6 RNA. RRMs typically bind 3-4 single-stranded nucleotides across a beta sheet surface using consensus RNP1 and RNP2 motifs (CLERY *et al.* 2008). In contrast, the Prp24 protein directly contacts 20 nucleotides of U6 RNA via interactions with RRM domains 2-4. The first RRM does not contact U6 snRNA, but forms a large

protein-protein interface with RRM2. RRMs 2 and 3 bind the asymmetric bulge in a relatively canonical RRM fashion, targeting the ACAGA box and sequences immediately upstream, respectively. The fourth RRM binds primarily to the ISL and telestem via the two amphipathic helices that occlude the beta-sheet face of the RRM (MARTIN-TUMASZ *et al.* 2011). Surprisingly, the RRM3/RRM4 linker, a single aspartate residue, passes through the asymmetric bulge and RRM4 makes tertiary interactions with RRM2, forming an entwined topology consisting of interlocked rings of protein and RNA (MONTEMAYOR *et al.* 2014). The entwined topology is further closed off by the ISL and telestem and is exceptionally stable, as the complex is resistant to 2 M monovalent salt (SHANNON AND GUTHRIE 1991). RRMs 1, 2, and 4 together form a large (~20 Å wide), positively charged groove. In the crystal lattice, the U6 ISL of a neighboring complex is packed into the groove, leading to the hypothesis that this groove might stabilize U4/U6 pairing during Prp24-mediated stimulation of U4/U6 di-snRNP formation (MONTEMAYOR *et al.* 2014).

The U6•Prp24 structure contained three substitutions to stabilize the RNA (MONTEMAYOR *et al.* 2014) including A62G, which stabilizes the base of the ISL and causes a cold sensitive phenotype *in vivo* (FORTNER *et al.* 1994). Interestingly, suppressors of U6-A62G cold sensitivity localize to the RNA-protein interface, suggesting that detrimental overstabilization of the U6 ISL structure can be neutralized by destabilization of the RNA-protein interface *in vivo* (MONTEMAYOR *et al.* 2014). A second structure with the wild-type ISL crystallized in a different space group, yet exhibits the same interlocked topology (MONTEMAYOR *et al.* 2017). Stabilization of the telestem with U100C/U101C mutations, which improve both Prp24 affinity for U6 as well as Prp24-mediated U4/U6 annealing activity (discussed in section 2.7.1), likely also stabilizes the interlocked topology (DIDYCHUK *et al.* 2016). Together, these data strongly suggest that the interlocked topology is a biologically functional structure.

The structure of the entire snRNP, with full-length RNA and including both Prp24 and Lsm2-8, remains unsolved. It is known that Prp24 and Lsm2-8 have at least one protein-protein interaction (via the Prp24 SNFFL box), but where this interaction occurs on the Lsm2-8 ring and

how the individual components are oriented with respect to one another has not been determined. Uncovering how Prp24 and Lsm2-8 interact will undoubtedly be vital to understanding the mechanism of Prp24-mediated U4/U6 annealing.

2.7. U4/ U6 di-snRNP and U4/U6.U5 tri-snRNP

2.7.1 U4/U6 annealing

In order to be incorporated into the assembling spliceosome, the ISL of U6 must be completely unwound and base paired to U4 snRNA via the RNA chaperone activity of Prp24. This assembly step may serve both to proofread U6 structure by ensuring that only U6 that can bind Prp24 and base pair to U4 is incorporated into the spliceosome. In addition, U4 may act as an "antisense negative regulator" of the catalytic residues of U6 by sequestering them in the double stranded helixes of U4/U6 prior to spliceosome activation (GUTHRIE 1991).

It has long been known that the U6 snRNP protein Prp24 can anneal U4 and U6 RNAs in an ATP-independent manner (GHETTI *et al.* 1995; RAGHUNATHAN AND GUTHRIE 1998b). The 20 Å-wide electropositive groove of Prp24 is critical for U4/U6 annealing but not binding of U6, suggesting that it is the "active site" for U4/U6-annealing (DIDYCHUK *et al.* 2016). Stabilization of the telestem in U6 also contributes to efficient annealing, possibly by stabilizing the formation of the electropositive groove-containing topology (DIDYCHUK *et al.* 2016). The Lsm2-8 ring enhances Prp24-mediated annealing *in vitro*, but does not facilitate U4/U6 annealing by itself (DIDYCHUK *et al.* 2016). Interestingly, Prp24 binds the U4/U6 product complex *in vitro* (GHETTI et al. 1995; DIDYCHUK et al. 2016) yet does not stay associated with the U4/U6 di-snRNP in yeast *in vivo* (SHANNON AND GUTHRIE 1991). How the interlocked topology of U6•Prp24 is resolved during annealing, and how Prp24 is displaced from the U4/U6 di-snRNP, is unclear.

The ISL is not contacted by protein in the either of the determined U6 snRNP core structures. In the ISL-stabilized (A62G) structure, the top of the ISL is disordered and not visible (MONTEMAYOR *et al.* 2014). In the wild-type U6 snRNP core structure, there are two complexes in the asymmetric unit (MONTEMAYOR *et al.* 2017). In one, the ISL is largely disordered and in the other, the ISL displays a 20° bend relative to the A62G ISL (MONTEMAYOR *et al.* 2017). These observations, along with NMR measurements and MD simulations (BLAD *et al.* 2005; VENDITTI *et al.* 2009), suggest that the U6 ISL is inherently dynamic. Since the ISL must be completely

unwound during U4/U6 annealing, dynamics likely play an important role in the annealing pathway. The stabilized A62G ISL in isolation has a melting temperature of ~63°C (SASHITAL *et al.* 2003) while the wild-type ISL is less stable. Binding of Prp24 may facilitate U4/U6 formation by allowing the ISL to remain in a dynamic, meta-stable conformation while adjacent to the electropositive groove in Prp24. It is unlikely that the ISL needs to fully melt in order for annealing to be initiated, as local unwinding of a few base pairs could allow formation of a nascent U4/U6 duplex to invade the ISL and facilitate the subsequent exchange of intra- for intermolecular base pairs via branch migration (MESELSON 1972; SIGAL AND ALBERTS 1972).

In humans, SART3 (hPrp24) functions in a similar role in both U4/U6 annealing (BELL *et al.* 2002; MEDENBACH *et al.* 2004) and U4atac/U6atac annealing (DAMIANOV *et al.* 2004). Although the HAT domains are dispensable for binding U6, both the HAT domain and the two RRMs are required for efficient U4/U6 recycling (BELL *et al.* 2002; MEDENBACH *et al.* 2004). Surprisingly, the C-terminal tail (the SNFFL box), which is known to interact with the Lsm2-8 ring (RADER AND GUTHRIE 2002) is not necessary for efficient recycling *in vitro* (MEDENBACH *et al.* 2004). Instead, an interaction with the U4 snRNP protein 90K (hPrp3) mediated by the HAT domain of SART3 is necessary for efficient recycling (MEDENBACH *et al.* 2004). It has also been reported that human Lsm2-8 can recycle U4/U6 in the absence of other factors (ACHSEL *et al.* 1999). As monomeric SART3 has only two RRMs, it is unlikely to form an electropositive groove as observed in the yeast U6•Prp24 structure, although may potentially do so via the aforementioned dimerization of the HAT domains (PARK *et al.* 2016). Alternatively, the role of the electropositive groove of Prp24 may have been functionally replaced by protein-protein interactions between SART3 and hPrp3 (MEDENBACH *et al.* 2004).

The U4 snRNP proteins Snu13 and Prp31 and U4/U6 di-snRNP proteins Prp3/Prp4 (plus CypH in humans) may also contribute to the annealing process. Snu13 and Prp31 influence the conformation of the U4 5'SL, which may hold U4 snRNA in a conformation amenable to annealing with U6 (HARDIN *et al.* 2015). Indeed, the structure of protein-free U4/U6 (CORNILESCU *et al.* 2016)

is significantly different from its structure while in U4/U6.U5 tri-snRNP (NGUYEN *et al.* 2016). Prp3 binds U4/U6 Stem II and a region on U6 3' of the stem, suggesting that its binding site does not exist until U4/U6 is formed (LIU *et al.* 2015; NGUYEN *et al.* 2016). However, it may transiently capture this single-stranded region of U6 from U6•Prp24 to accelerate U4/U6 annealing (LIU *et al.* 2015). The contribution of additional proteins to Prp24-mediated annealing has not been measured, but may reveal additional RNA-protein or protein-protein interactions that accelerate the assembly process. Understanding how the mechanism of Prp24-mediated annealing in both yeast and humans could give new insight into conserved versus divergent mechanisms of RNA chaperones.

2.7.2 Formation and structure of the U4/U6.U5 tri-snRNP

The tri-snRNP is formed through protein-protein and protein-RNA interactions between the U4/U6 di-snRNP and the U5 snRNP. Prp31 is necessary for assembly of both major and minor U4/U6.U5 tri-snRNPs in humans (SCHNEIDER *et al.* 2002). Knockdown of both hPrp31 and the tri-snRNP-specific protein hPrp6 leads to an accumulation of U4/U6 di-snRNPs in Cajal bodies (NOVOTNY *et al.* 2015) (discussed in section 2.4). The structure of the tri-snRNP from yeast and humans has revealed myriad protein-protein contacts that hold the U4/U6 di-snRNP and U5 snRNP together (AGAFONOV *et al.* 2016; NGUYEN *et al.* 2016; WAN *et al.* 2016b). Prp6 is the “glue” holding the tri-snRNP together, as its N-terminal region contacts U5 snRNP component Prp8 (and Brr2 in the human tri-snRNP) while its C-terminal region contacts U4 snRNP components Snu13, Prp31, Prp3, and Prp4 (GALISSON AND LEGRAIN 1993; AGAFONOV *et al.* 2016; NGUYEN *et al.* 2016; WAN *et al.* 2016b).

The structure of the U4/U6.U5 tri-snRNP revealed an unexpected short internal stem in U6 5' of U4/U6 Stem I consisting of nucleotides 33-50 (NGUYEN *et al.* 2016). Interestingly, this stem loop contains the ACAGAGA sequence which must pair to the 5' SS. This ACAGAGA stem loop is stabilized by Dib1 and Prp8 and is also present in the B complex structure (PLASCHKA *et al.* 2017). As Dib1 is not observed in B^{act} or subsequent catalytic structures, it likely departs during

activation and the ACAGAGA stem unwinds, allowing recognition of the intron and catalysis to occur. The ACAGAGA stem was not resolved in the structure of the human tri-snRNP, despite Dim1 (the homolog of Dib1) binding near this region (AGAFONOV *et al.* 2016). The RNA-RNA and protein-RNA contacts in the tri-snRNP effectively sequester all of the catalytic elements of U6: the ACAGAGA sequence is partially occluded by the ACAGAGA stem loop, the AGC triad is base paired in U4/U6 Stem I, and U80 is base paired in U4/U6 Stem II.

The most significant difference in the structures of U4/U6.U5 tri-snRNP from human and yeast is in the placement of the helicase Brr2. In the yeast tri-snRNP, Brr2 binds U4 in a single stranded region between the U4 3'SL and U4/U6 Stem I, suggesting that it is poised for unwinding U4/U6 (NGUYEN *et al.* 2016). Indeed, yeast tri-snRNP samples are disassembled in the presence of ATP (RAGHUNATHAN AND GUTHRIE 1998a; NGUYEN *et al.* 2015). In contrast, Brr2 in the human tri-snRNP is on the opposite end of the complex ~200 Å away, far from its U4/U6 Stem I substrate, and the tri-snRNP is insensitive to ATP-dependent disassembly (AGAFONOV *et al.* 2016). This difference is likely due to the presence of Sad1 in the human tri-snRNP, where Sad1 both stabilizes the interaction between U4/U6 and U5 snRNPs and tethers Brr2 in a “preactivation” position away from its substrate (AGAFONOV *et al.* 2016). Yeast Sad1, in contrast, does not co-purify with the tri-snRNP (STEVENS *et al.* 2001; NGUYEN *et al.* 2016). Despite interacting weakly with the tri-snRNP in yeast, Sad1 still prevents Brr2-mediated unwinding (HUANG *et al.* 2014).

The minor spliceosome tri-snRNP (U4atac/U6atac.U5) contains many of the same components as the major spliceosome U4/U6.U5 (SCHNEIDER *et al.* 2002). The additional 3' stem loop in U6atac (**Figure 2-4**) may differentiate the major and minor spliceosomes, as this stem loop is both important for targeting U6atac to the minor spliceosome (DIETRICH *et al.* 2009) and for interacting with protein 65K, a U11/U12-specific protein (SINGH *et al.* 2016). It remains to be seen if the major and minor spliceosomes have additional conformational or compositional differences, and how the structures of the major and minor tri-snRNP differ.

Post-translational modifications within tri-snRNP proteins affect both tri-snRNP assembly and stability as well as incorporation into the spliceosome. Human Prp4 kinase, named after its *S. pombe* ortholog (ALAHARI *et al.* 1993; ECKERT *et al.* 2016) and not the *S. cerevisiae* U4/U6 di-snRNP protein, phosphorylates Prp6 and Prp31, which promotes tri-snRNP formation (SCHNEIDER *et al.* 2010a). *S. pombe* and *Fusarium graminearum* Prp4 kinase exhibit genetic and physical interactions with Prp8, Brr2, Prp6, and Prp31 (SCHMIDT *et al.* 1999; BOTTNER *et al.* 2005; GAO *et al.* 2016). Ubiquitination of Prp8 in yeast (BELLARE *et al.* 2008) and of Prp3 in humans (SONG *et al.* 2010) also may influence tri-snRNP formation and stability. In addition to this, SUMOylation of Prp3 plays a role in tri-snRNP assembly (POZZI *et al.* 2017). Characterizing how post-translational modifications affect tri-snRNP structure, function, and interactions, as well as how these post-translational modifications are regulated in normal and disease states is an important future direction towards understanding how tri-snRNP assembly is regulated *in vivo*.

2.8 U6 in the catalytic spliceosome

2.8.1 Activation

U6 base pairs with the 5'SS after remodeling of the U1-5'SS interaction (SAWA AND ABELSON 1992) by the DEAD-box helicase Prp28 (STALEY AND GUTHRIE 1999). This is an important proof-reading step to ensure that stable and correct U6-5'SS duplexes are formed (YANG *et al.* 2013). U2 base pairs to U6 via formation of U2/U6 helix II (WASSARMAN AND STEITZ 1992; SCHNEIDER *et al.* 2010b) (**Figure 2-5**). Then, U4/U6 Stems I and II are completely unwound in order for the U6 ISL to reform and for U6 to base pair with U2 (**Figure 2-5**). Ejection of U4 depends on the U5-snRNP component helicase Brr2, which unwinds U4/U6 (RAGHUNATHAN AND GUTHRIE 1998a); reviewed in (ABSMEIER *et al.* 2016). Brr2 activity is regulated by protein-protein contacts with itself and Prp8, which can both up- and downregulate its unwinding activity (MOZAFFARI-JOVIN *et al.* 2012; MOZAFFARI-JOVIN *et al.* 2013; ABSMEIER *et al.* 2015). Intramolecular folding of U6 contributes to efficient unwinding (RODGERS *et al.* 2016; THEUSER *et al.* 2016). Once unwound, U4 snRNA departs with Snu13 and Prp31 (and likely the U4 Sm ring), while Prp3/Prp4 are released separately (THEUSER *et al.* 2016).

Brr2-mediated unwinding and U4 snRNP departure typically occur before arrival of the NTC (HOSKINS *et al.* 2016). Recent data on human spliceosomes chemically stalled between the transition from B to B^{act} reveal that the Lsm2-8 ring is present on U6 after U4 departure (SIDAROVICH *et al.* 2017). The NTC is required for dissociation of the Lsm ring in yeast (CHAN *et al.* 2003). Recent structures of the spliceosome after activation show that the 3' end of U6 (the binding site of the Lsm ring) is partially occluded by Syf1 of the NTC, suggesting that the NTC may directly displace the Lsm ring. It is not clear if 3' end modifying enzymes such as Usb1 and U6-TUTase can access the 3' end of U6 in these complexes, but high throughput studies suggest that Usb1 may interact with components of the NTC (HAZBUN *et al.* 2003; KROGAN *et al.* 2006; YU *et al.* 2008).

2.8.2 Conformational changes during catalysis

In the past two years, a multitude of spliceosomal structures have been determined at different catalytic steps. Structures of spliceosomes paused before activation (B complex) (PLASCHKA *et al.* 2017) the first catalytic step (B^{act} complex)(RAUHUT *et al.* 2016; YAN *et al.* 2016), after first step chemistry (C complex)(GALEJ *et al.* 2016; WAN *et al.* 2016a), remodeled before the second catalytic step (C* complex)(BERTRAM *et al.* 2017; FICA *et al.* 2017; YAN *et al.* 2017; ZHANG *et al.* 2017), and after ligated exon departure (ILS)(YAN *et al.* 2015; WAN *et al.* 2017) reveal that, despite large-scale conformational and compositional changes elsewhere in the spliceosome (**Table 2-2**), the structure of U6 snRNA changes very little during these transitions. Indeed, the structure of U6 snRNA at each of these steps is nearly indistinguishable (**Figure 2-6A**), even when comparing structures isolated from *S. cerevisiae* and *S. pombe*. U2/U6 Helix Ia and Ib (composed of nucleotides 55-61 in U6) and Helix II (nucleotides 93-102 in U6) remain base paired and unchanged throughout the structures (**Figure 2-6B**). The ISL (nucleotides 64 to 84) is also structurally identical. Yeast U6 residue C66 is bulged from the ISL and sits in a pocket composed of the NTC and NTC-associated proteins Clf1, Cef1, Prp45, and Ecm2. The structure of the ISL within the spliceosomal complexes is significantly different from the free RNA or when within the U6 snRNP due to the many protein contacts surrounding it. The 3' tail of U6 (nucleotides 105-112) is unstructured in all available cryo-EM structures and is not modeled in any of the structures. Small local base pairing changes do occur between the ACAGAGA box and the intron, especially U6 nucleotides G50 and A51. However, local changes in accessibility, flexibility, or structure may occur in solution, with important implications for activation and catalysis (BAO *et al.* 2017).

In contrast, U2 snRNA and the branch site helix, along with many protein partners, undergo >50 Å motions and large rotations. Analogy to group II introns, along with detailed genetic and biochemical studies, suggested that the active site for first and second step chemistry are the same (CHANFREAU AND JACQUIER 1994; MARCIA AND PYLE 2012; FICA *et al.* 2013; FICA *et al.* 2014), and that large conformational changes are required in order to move the first step product (the 2'-

5' branch site) out of the single active site, followed by positioning of the 3' exon in the active site for second step chemistry (exon ligation). The first hints of these large scale motions came from single molecule experiments (ABELSON *et al.* 2010; CRAWFORD *et al.* 2013). Recent cryo-EM structures of splicing complexes before and after the first and second steps of splicing corroborate the hypothesis that U6 forms a rigid active site in the spliceosome and that its substrates must be positioned within it for catalysis to occur (YAN *et al.* 2015; GALEJ *et al.* 2016; RAUHUT *et al.* 2016; WAN *et al.* 2016a; YAN *et al.* 2016; BERTRAM *et al.* 2017; FICA *et al.* 2017; WAN *et al.* 2017; YAN *et al.* 2017; ZHANG *et al.* 2017).

2.8.3 Metal ion binding in the spliceosome active site

A two-metal mechanism for RNA-mediated splicing catalysis was proposed in 1993 (STEITZ AND STEITZ 1993). This model stated, by analogy to protein phosphoryl transfer mechanisms, that two divalent metal ions situated 3.9 Å apart stabilize the nucleophile and the leaving group during transesterification. Elegant thiophilic metal ion rescue experiments with phosphorothioate substitutions revealed that the U6 ISL coordinates the two Mg^{+2} ions required for catalysis via five non-bridging phosphate oxygens (YEAN *et al.* 2000; FICA *et al.* 2013). Electron density in the active site of recent spliceosome cryo-EM structures has been modeled as metal ions (WAN *et al.* 2016a; YAN *et al.* 2016; FICA *et al.* 2017; YAN *et al.* 2017; ZHANG *et al.* 2017). However, the current resolution of these cryo-EM structures precludes the unambiguous placement of metal ions into the electron density, such that the exact location of magnesium ions in the active site, and positioning and identity of other structural metal ions, remains to be determined.

Figure 2-6. *S. cerevisiae* U6 does not undergo conformational changes during the transition from B^{act} to ILS complexes.

(A) Superimposition of U6 snRNA structures from the following *S. cerevisiae* spliceosomal complexes: B^{act} (PDB 5GM6, red and PDB 5LQW, salmon), C (PDB 5GMK, orange and PDB 5LJ5, light orange), C* (PDB 5WSG, lemon and PDB 5MQ0, lime), and the *S. pombe* ILS (PDB 3JB9, blue).

(B) Secondary structure of U2/U6 in C* complex (PDB 5MQ0). Protein contacts to U6 are shown in green. Base triple interactions are shown in dashed lines.

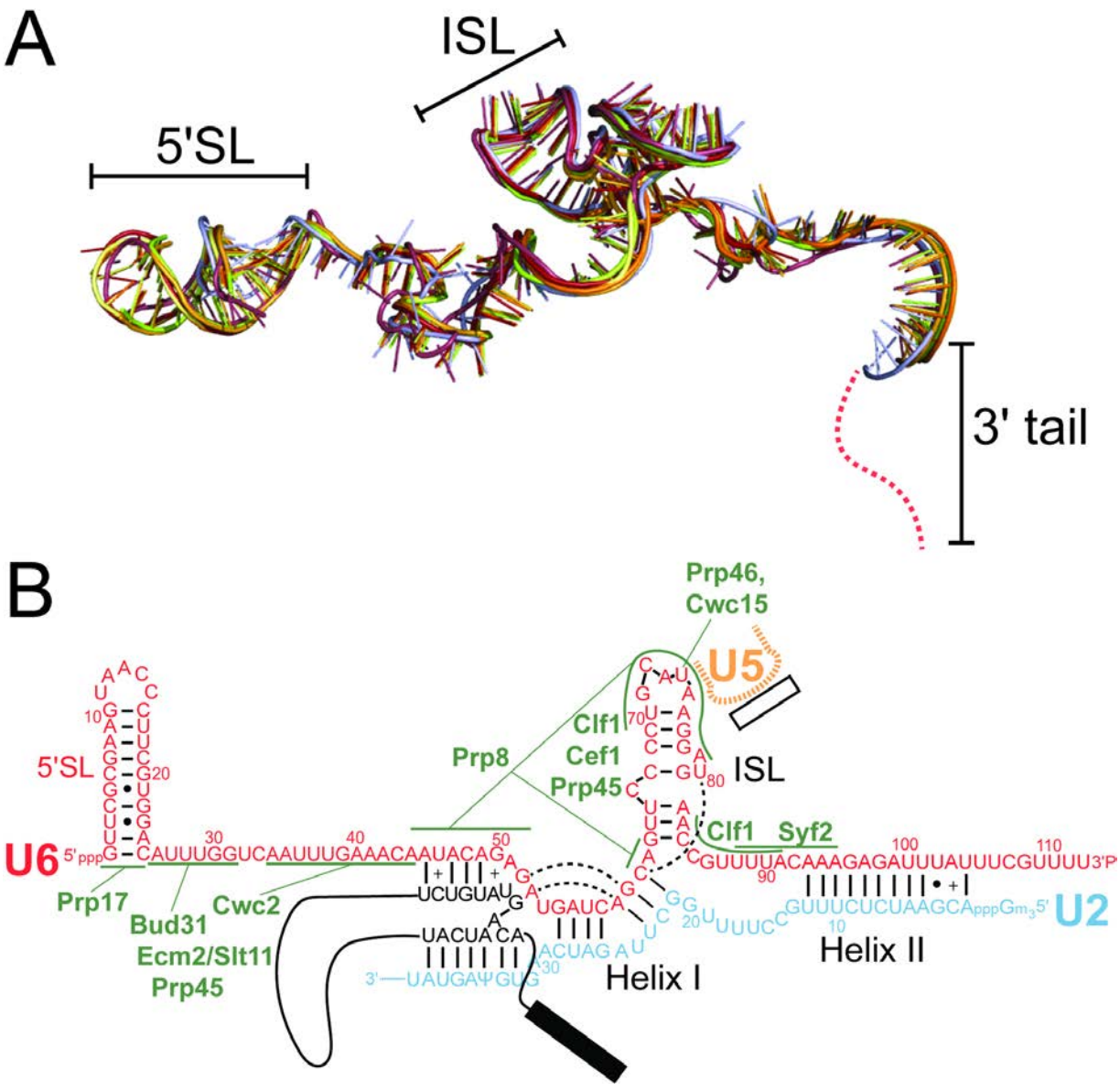


Table 2-2. RNA-RNA and protein-RNA contacts in U6-containing complexes.
RNAs and proteins within 5 Å of U6 snRNA in available structures. RNAs are shown in italics. Protein components of different snRNPs are highlighted in color, including U2 (blue), U4 (green), U5 (orange), and U6 (red). Protein-only complexes NTC (grey) and NTR (purple) are also included; additional factors are not highlighted. All structures are from *S. cerevisiae*.

	U6 snRNP (PDB 4N0T)	U4/U6.U5 tri-snRNP (PDB 5GAN)	B complex (PDB 5NRL)	B^{act} complex (PDB 5LQW)	C complex (PDB 5GMK)	C* complex (PDB 5MQ0)
Directly contacting U6 snRNA		*U4 snRNA*	*U2 snRNA*	*U2 snRNA*	*U2 snRNA*	*U2 snRNA*
		U5 snRNA	*U4 snRNA*	*U5 snRNA*	*U5 snRNA*	*U5 snRNA*
			U5 snRNA	*Intron RNA*	*Intron RNA*	*Intron RNA*
	Prp24		*Intron RNA*			
	Lsm2	Lsm2	Lsm2	Prp8	Prp8	Prp8
	Lsm3	Lsm3	Lsm3	Bud31	Bud31	Bud31
	Lsm4	Lsm4	Lsm4	Cwc2	Cwc2	Cwc2
	Lsm5	Lsm5	Lsm5	Ecm2/Slt11	Ecm2/Slt11	Ecm2/Slt11
	Lsm6	Lsm6	Lsm6	Prp45	Prp45	Prp45
	Lsm7	Lsm7	Lsm7	Prp46	Prp46	Prp46
	Lsm8	Lsm8	Lsm8	Clf1	Cwc15	Cwc15
		Prp3	Prp3	Syf1	Clf1	Clf1
		Snu13	Snu13	Cef1	Syf1	Cef1
		Prp31	Prp4		Cef1	Syf2
		Prp8	Prp8		Isy1	Prp17
		Brr2	Dib1		Cwc25	
		Dib1	Snu23		Yju2	
			Prp38			
In complex, but not directly contacting U6 snRNA		U5 Sm ring	U5 Sm Ring	U5 Sm ring	U5 Sm ring	U5 Sm ring
		Snu114	Snu114	Snu114	Snu114	Snu114
		Prp4	Brr2	Brr2	Brr2	Msl1
		U4 Sm ring	U4 Sm ring	Rse1	Msl1	Lea1
		Prp6	Prp31	Rds3	Lea1	U2 Sm ring
			Rse1	Hsh155	U2 Sm ring	Prp19
			Rds3	U2 Sm ring	Prp19	Snt309
			Hsh155	Prp2	Snt309	Syf1
			U2 Sm ring	Cwc26	Prp16	Slu7
			Prp9/Prp11/Prp21	Cwc22	Cwc22	Prp18
			Msl1	Pml1	Cwc21	Cwc22
			Lea1	Ist3		Cwc21
			Hsh49			Prp22
			Cus1			
			Ysf3			
			Snu66			
			Prp6			
			Spp381			

2.9 Recycling and degradation

2.9.1 Recycling

After a round of intron removal and splicing, the spliceosome undergoes an active disassembly process in which the component snRNPs are recycled for a new round of splicing. The U2/U6.U5-intron lariat complex is disassembled by the ATP-dependent DEAH-box helicase Prp43 (MARTIN *et al.* 2002). Prp43 interacts with cofactors Ntr1 and Ntr2 to disassemble the intron-lariat spliceosome (ILS) and release the component snRNPs and the lariat-intron (TSAI *et al.* 2005). In *in vitro* assays, U6 is released as the free snRNA (TSAI *et al.* 2005; FOURMANN *et al.* 2013; FOURMANN *et al.* 2016). In the cell, U6 is presumably recognized during or after release from the ILS by Prp24 and/or Lsm2-8. Prp24 is required in yeast even when mutations in U4 and U6 allow for bypass of stable U4/U6 base pairing, indicating that Prp24 has an essential function other than U4/U6 annealing, perhaps in driving ILS dissociation by capturing U6 (BURKE *et al.* 2015). It is not known if recycling of U6 after a round of splicing follows the same pathway as initial assembly of the U6 snRNP, i.e., U6 is bound by Prp24 and Lsm2-8, then assembled into the U4/U6 di-snRNP.

2.9.2 Degradation

The degradation pathway of U6 is poorly understood. The balance between uridylation, polyadenylation, and trimming by Usb1 (see section 2.5.2) affects the half-life of U6 in metazoans (SHCHEPACHEV *et al.* 2012; HILCENKO *et al.* 2013; SHCHEPACHEV *et al.* 2015). End processing by Usb1 has a protective effect against degradation in *S. cerevisiae*, *S. pombe*, and human cells, suggesting that the modification may protect against other cellular nucleases (MROCZEK *et al.* 2012; SHCHEPACHEV *et al.* 2012; HILCENKO *et al.* 2013; SHCHEPACHEV *et al.* 2015). However, it is not known which nucleases are responsible for degrading U6 in Usb1-depleted cells or during the normal U6 life cycle. Additionally, other cellular deadenylases can interact with and influence the 3' sequence of U6 snRNA (SHUKLA AND PARKER 2017). How the activities of different 3' modifying enzymes affect the stability and functionality of U6 is poorly understood.

How is U6 snRNA normally degraded? Recent data suggests that the exosome may play a role in normal turnover of spliceosomal RNAs, as knockdown of the exosome exonuclease Rrp44 resulted in higher levels of U6 (ZHANG *et al.* 2015). Presence of a terminal phosphate (in yeast) or 2',3'-cyclic phosphate (in humans) on U6 may inhibit degradation by the exosome, as the nuclear exosome exonuclease Rrp6 is inactive on a 3' phosphate-terminated RNA (BURKARD AND BUTLER 2000). Additional enzymes may be necessary to dephosphorylate the 3' end of U6 prior to adenylation and decay. It has been demonstrated that Mtr4 and Trf4 (components of the TRAMP complex) as well as Rrp6 (a component of the nuclear exosome) interact with components of the U4/U6.U5 tri-snRNP (NAG AND STEITZ 2012). While the primary reason for recruitment of the exosome machinery may be to degrade improperly spliced mRNAs or excised introns, it may also have a role in degrading “expired” snRNAs. The TRAMP complex is recruited to the intron even before splicing is complete and may in fact help stimulate splicing (KONG *et al.* 2014). Thus, the components of the nuclear decay machinery are temporally and spatially near the spliceosomal RNAs. Supporting this idea, in *S. cerevisiae*, deletion of the poly(A) polymerase from the TRAMP complex, Trf4, results in a small increase in U6 levels (COPELA *et al.* 2008). Moreover, Rrp44, Rrp6, and Trf4 can be crosslinked to U6 (SCHNEIDER *et al.* 2012). However, deletion of Trf4 in *S. pombe* does not affect polyadenylation of U6, suggesting that other poly(A) polymerases are involved or can complement loss of Trf4 (SHCHEPACHEV *et al.* 2015). Further study on the determinants of U6 decay, including the adenylation state of the 3' tail, will help to define how this metabolically stable RNA is monitored, which may provide another mechanism for spliceosome fidelity.

2.10 Conclusions and perspectives

X-ray crystallography and cryo-EM approaches have led to all-atom models of multiple spliceosomal complexes that contain U6 snRNA. These structures allow us to directly visualize the vast conformational changes that U6 must undergo during spliceosome assembly and activation, and highlight the large number of dynamic steps that must occur to form such complex structures. Each assembly and processing step represents an opportunity for quality control, ensuring that functional spliceosomes are prepared to carry out a step of gene expression that is essential in all eukaryotes. U6, U2 and U5 snRNA as well as many proteins must come together to form the catalytic core of the spliceosome, where U6 snRNA coordinates the magnesium ions required for splicing catalysis. Regulation and monitoring of U6 throughout its lifetime is crucial for retaining the spliceosome's catalytic activity and to ensure that only full-length, properly modified U6 molecules are incorporated into the spliceosome. Research in the coming years will further illuminate how U6 is regulated (via transcriptional activity, post-transcriptional modification, RNA-protein interactions, and sub-nuclear localization) and assembled into the spliceosome.

2.11 References

Abelson, J., M. Blanco, M. A. Ditzler, F. Fuller, P. Aravamudhan *et al.*, 2010 Conformational dynamics of single pre-mRNA molecules during in vitro splicing. Nat Struct Mol Biol 17**:** 504-512.

Absmeier, E., K. F. Santos and M. C. Wahl, 2016 Functions and regulation of the Brr2 RNA helicase during splicing. Cell Cycle 15**:** 3362-3377.

Absmeier, E., J. Wollenhaupt, S. Mozaffari-Jovin, C. Becke, C. T. Lee *et al.*, 2015 The large N-terminal region of the Brr2 RNA helicase guides productive spliceosome activation. Genes Dev 29**:** 2576-2587.

Achsel, T., H. Brahms, B. Kastner, A. Bachi, M. Wilm *et al.*, 1999 A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. EMBO J 18**:** 5789-5802.

Agafonov, D. E., B. Kastner, O. Dybkov, R. V. Hofele, W. T. Liu *et al.*, 2016 Molecular architecture of the human U4/U6.U5 tri-snRNP. Science 351**:** 1416-1420.

Alahari, S. K., H. Schmidt and N. F. Kaufer, 1993 The fission yeast prp4+ gene involved in pre-mRNA splicing codes for a predicted serine/threonine kinase and is essential for growth. Nucleic Acids Res 21**:** 4079-4083.

Arimbasseri, A. G., and P. Bhargava, 2008 Chromatin structure and expression of a gene transcribed by RNA polymerase III are independent of H2A.Z deposition. Mol Cell Biol 28**:** 2598-2607.

Arimbasseri, A. G., and R. J. Maraia, 2015 Mechanism of Transcription Termination by RNA Polymerase III Utilizes a Non-template Strand Sequence-Specific Signal Element. Mol Cell 58**:** 1124-1132.

Arimbasseri, A. G., K. Rijal and R. J. Maraia, 2013 Transcription termination by the eukaryotic RNA polymerase III. Biochim Biophys Acta 1829**:** 318-330.

Bae, E., N. J. Reiter, C. A. Bingman, S. S. Kwan, D. Lee *et al.*, 2007 Structure and interactions of the first three RNA recognition motifs of splicing factor prp24. J Mol Biol 367**:** 1447-1458.

Bao, P., C. Hobartner, K. Hartmuth and R. Luhrmann, 2017 Yeast Prp2 liberates the 5' splice site and the branch site adenosine for catalysis of pre-mRNA splicing. RNA.

Barski, A., I. Chepelev, D. Liko, S. Cuddapah, A. B. Fleming *et al.*, 2010 Pol II and its associated epigenetic marks are present at Pol III-transcribed noncoding RNA genes. Nat Struct Mol Biol 17**:** 629-634.

Basak, A., and C. C. Query, 2014 A pseudouridine residue in the spliceosome core is part of the filamentous growth program in yeast. Cell Rep 8**:** 966-973.

Battle, D. J., M. Kasim, J. Yong, F. Lotti, C. K. Lau *et al.*, 2006 The SMN complex: an assembly machine for RNPs. Cold Spring Harb Symp Quant Biol 71**:** 313-320.

Belagal, P., C. Normand, A. Shukla, R. Wang, I. Leger-Silvestre *et al.*, 2016 Decoding the principles underlying the frequency of association with nucleoli for RNA polymerase III-transcribed genes in budding yeast. Mol Biol Cell 27**:** 3164-3177.

Bell, M., S. Schreiner, A. Damianov, R. Reddy and A. Bindereif, 2002 p110, a novel human U6 snRNP protein and U4/U6 snRNP recycling factor. EMBO J 21**:** 2724-2735.

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

Bertram, K., D. E. Agafonov, W. T. Liu, O. Dybkov, C. L. Will *et al.*, 2017 Cryo-EM structure of a human spliceosome activated for step 2 of splicing. Nature 542**:** 318-323.

Bertrand, E., F. Houser-Scott, A. Kendall, R. H. Singer and D. R. Engelke, 1998 Nucleolar localization of early tRNA processing. Genes Dev 12**:** 2463-2468.

Bhattacharya, D., F. Lutzoni, V. Reeb, D. Simon, J. Nason *et al.*, 2000 Widespread occurrence of spliceosomal introns in the rDNA genes of ascomycetes. Mol Biol Evol 17**:** 1971-1984.

Bhattacharya, R., K. Perumal, K. Sinha, R. Maraia and R. Reddy, 2002 Methylphosphate cap structure in small RNAs reduces the affinity of RNAs to La protein. Gene Expr 10**:** 243-253.

Blad, H., N. J. Reiter, F. Abildgaard, J. L. Markley and S. E. Butcher, 2005 Dynamics and metal ion binding in the U6 RNA intramolecular stem-loop as analyzed by NMR. J Mol Biol 353**:** 540-555.

Boelens, W. C., I. Palacios and I. W. Mattaj, 1995 Nuclear retention of RNA as a mechanism for localization. RNA 1**:** 273-283.

Bogenhagen, D. F., and D. D. Brown, 1981 Nucleotide sequences in Xenopus 5S DNA required for transcription termination. Cell 24**:** 261-270.

Booth, B. L., Jr., and B. F. Pugh, 1997 Identification and characterization of a nuclease specific for the 3' end of the U6 small nuclear RNA. J Biol Chem 272**:** 984-991.

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

Braglia, P., S. L. Dugas, D. Donze and G. Dieci, 2007 Requirement of Nhp6 proteins for transcription of a subset of tRNA genes and heterochromatin barrier function in Saccharomyces cerevisiae. Mol Cell Biol 27**:** 1545-1557.

Bringmann, P., and R. Luhrmann, 1987 Antibodies specific for N6-methyladenosine react with intact snRNPs U2 and U4/U6. FEBS Lett 213**:** 309-315.

Brow, D. A., and C. Guthrie, 1988 Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. Nature 334**:** 213-218.

Brow, D. A., and C. Guthrie, 1989 Splicing a spliceosomal RNA. Nature 337**:** 14-15.

Brow, D. A., and C. Guthrie, 1990 Transcription of a yeast U6 snRNA gene requires a polymerase III promoter element in a novel position. Genes Dev 4**:** 1345-1356.

Brow, D. A., and R. M. Vidaver, 1995 An element in human U6 RNA destabilizes the U4/U6 spliceosomal RNA complex. RNA 1**:** 122-131.

Burkard, K. T., and J. S. Butler, 2000 A nuclear 3'-5' exonuclease involved in mRNA degradation interacts with Poly(A) polymerase and the hnRNA protein Npl3p. Mol Cell Biol 20**:** 604-616.

Burke, J. E., S. E. Butcher and D. A. Brow, 2015 Spliceosome assembly in the absence of stable U4/U6 RNA pairing. RNA 21**:** 923-934.

Burnol, A. F., F. Margottin, P. Schultz, M. C. Marsolier, P. Oudet *et al.*, 1993 Basal promoter and enhancer element of yeast U6 snRNA gene. J Mol Biol 233**:** 644-658.

Canzler, S., P. F. Stadler and J. Hertel, 2016 U6 snRNA intron insertion occurred multiple times during fungi evolution. RNA Biol 13**:** 119-127.

Carmo-Fonseca, M., R. Pepperkok, B. S. Sproat, W. Ansorge, M. S. Swanson *et al.*, 1991a In vivo detection of snRNP-rich organelles in the nuclei of mammalian cells. EMBO J 10**:** 1863-1873.

Carmo-Fonseca, M., D. Tollervey, R. Pepperkok, S. M. Barabino, A. Merdes *et al.*, 1991b Mammalian nuclei contain foci which are highly enriched in components of the pre-mRNA splicing machinery. EMBO J 10**:** 195-206.

Cech, T. R., 1986 The generality of self-splicing RNA: relationship to nuclear mRNA splicing. Cell 44**:** 207-210.

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

Chanfreau, G., and A. Jacquier, 1994 Catalytic site components common to both splicing steps of a group II intron. Science 266**:** 1383-1387.

Charette, M., and M. W. Gray, 2000 Pseudouridine in RNA: what, where, how, and why. IUBMB Life 49**:** 341-351.

Chen, Y., K. Sinha, K. Perumal and R. Reddy, 2000 Effect of 3' terminal adenylic acid residue on the uridylation of human small RNAs in vitro and in frog oocytes. RNA 6**:** 1277-1288.

Clery, A., M. Blatter and F. H. Allain, 2008 RNA recognition motifs: boring? Not quite. Curr Opin Struct Biol 18**:** 290-298.

Copela, L. A., C. F. Fernandez, R. L. Sherrer and S. L. Wolin, 2008 Competition between the Rex1 exonuclease and the La protein affects both Trf4p-mediated RNA quality control and pre-tRNA maturation. RNA 14**:** 1214-1227.

Cornilescu, G., A. L. Didychuk, M. L. Rodgers, L. A. Michael, J. E. Burke *et al.*, 2016 Structural Analysis of Multi-Helical RNAs by NMR-SAXS/WAXS: Application to the U4/U6 di-snRNA. J Mol Biol 428**:** 777-789.

Cosgrove, M. S., Y. Ding, W. A. Rennie, M. J. Lane and S. D. Hanes, 2012 The Bin3 RNA methyltransferase targets 7SK RNA to control transcription and translation. Wiley Interdiscip Rev RNA 3**:** 633-647.

Crawford, D. J., A. A. Hoskins, L. J. Friedman, J. Gelles and M. J. Moore, 2013 Single-molecule colocalization FRET evidence that spliceosome activation precedes stable approach of 5' splice site and branch site. Proc Natl Acad Sci U S A 110**:** 6783-6788.

Damianov, A., S. Schreiner and A. Bindereif, 2004 Recycling of the U12-type spliceosome requires p110, a component of the U6atac snRNP. Mol Cell Biol 24**:** 1700-1708.

Darzacq, X., B. E. Jady, C. Verheggen, A. M. Kiss, E. Bertrand *et al.*, 2002 Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. EMBO J 21**:** 2746-2756.

Das, G., D. Henning, D. Wright and R. Reddy, 1988 Upstream regulatory elements are necessary and sufficient for transcription of a U6 RNA gene by RNA polymerase III. EMBO J 7**:** 503-512.

Davis, D. R., 1995 Stabilization of RNA stacking by pseudouridine. Nucleic Acids Res 23**:** 5020-5026.

Didychuk, A. L., E. J. Montemayor, D. A. Brow and S. E. Butcher, 2016 Structural requirements for protein-catalyzed annealing of U4 and U6 RNAs during di-snRNP assembly. Nucleic Acids Res 44**:** 1398-1410.

Didychuk, A. L., E. J. Montemayor, T. J. Carrocci, A. T. DeLaitsch, S. E. Lucarelli *et al.*, 2017 Usb1 controls U6 snRNP assembly through evolutionarily divergent cyclic phosphodiesterase activities. Nat Commun 8**:** 497.

Dietrich, R. C., R. A. Padgett and G. C. Shukla, 2009 The conserved 3' end domain of U6atac snRNA can direct U6 snRNA to the minor spliceosome. RNA 15**:** 1198-1207.

Domitrovich, A. M., and G. R. Kunkel, 2003 Multiple, dispersed human U6 small nuclear RNA genes with varied transcriptional efficiencies. Nucleic Acids Res 31**:** 2344-2352.

Doucet, A. J., G. Droc, O. Siol, J. Audoux and N. Gilbert, 2015 U6 snRNA Pseudogenes: Markers of Retrotransposition Dynamics in Mammals. Mol Biol Evol 32**:** 1815-1832.

Dundr, M., M. D. Hebert, T. S. Karpova, D. Stanek, H. Xu *et al.*, 2004 In vivo kinetics of Cajal body components. J Cell Biol 164**:** 831-842.

Eckert, D., N. Andree, A. Razanau, S. Zock-Emmenthal, M. Lutzelberger *et al.*, 2016 Prp4 Kinase Grants the License to Splice: Control of Weak Splice Sites during Spliceosome Activation. PLoS Genet 12**:** e1005768.

Epstein, P., R. Reddy, D. Henning and H. Busch, 1980 The nucleotide sequence of nuclear U6 (4.7 S) RNA. J Biol Chem 255**:** 8901-8906.

Eschenlauer, J. B., M. W. Kaiser, V. L. Gerlach and D. A. Brow, 1993 Architecture of a yeast U6 RNA gene promoter. Mol Cell Biol 13**:** 3015-3026.

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**:** 404-409.

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

Fica, S. M., C. Oubridge, W. P. Galej, M. E. Wilkinson, X. C. Bai *et al.*, 2017 Structure of a spliceosome remodelled for exon ligation. Nature 542**:** 377-380.

Fica, S. M., N. Tuttle, T. Novak, N. S. Li, J. Lu *et al.*, 2013 RNA catalyses nuclear pre-mRNA splicing. Nature 503**:** 229-234.

Fortner, D. M., R. G. Troy and D. A. Brow, 1994 A stem/loop in U6 RNA defines a conformational switch required for pre-mRNA splicing. Genes Dev 8**:** 221-233.

Fourmann, J. B., O. Dybkov, D. E. Agafonov, M. J. Tauchert, H. Urlaub *et al.*, 2016 The target of the DEAH-box NTP triphosphatase Prp43 in Saccharomyces cerevisiae spliceosomes is the U2 snRNP-intron interaction. Elife 5.

Fourmann, J. B., J. Schmitzova, H. Christian, H. Urlaub, R. Ficner *et al.*, 2013 Dissection of the factor requirements for spliceosome disassembly and the elucidation of its dissociation products using a purified splicing system. Genes Dev 27**:** 413-428.

French, S. L., Y. N. Osheim, D. A. Schneider, M. L. Sikes, C. F. Fernandez *et al.*, 2008 Visual analysis of the yeast 5S rRNA gene transcriptome: regulation and role of La protein. Mol Cell Biol 28**:** 4576-4587.

Frendewey, D., I. Barta, M. Gillespie and J. Potashkin, 1990 Schizosaccharomyces U6 genes have a sequence within their introns that matches the B box consensus of tRNA internal promoters. Nucleic Acids Res 18**:** 2025-2032.

Fromont-Racine, M., A. E. Mayes, A. Brunet-Simon, J. C. Rain, A. Colley *et al.*, 2000 Genome-wide protein interaction screens reveal functional networks involving Sm-like proteins. Yeast 17**:** 95-110.

Galej, W. P., M. E. Wilkinson, S. M. Fica, C. Oubridge, A. J. Newman *et al.*, 2016 Cryo-EM structure of the spliceosome immediately after branching. Nature 537**:** 197-201.

Galisson, F., and P. Legrain, 1993 The biochemical defects of prp4-1 and prp6-1 yeast splicing mutants reveal that the PRP6 protein is required for the accumulation of the [U4/U6.U5] tri-snRNP. Nucleic Acids Res 21**:** 1555-1562.

Ganot, P., B. E. Jady, M. L. Bortolin, X. Darzacq and T. Kiss, 1999 Nucleolar factors direct the 2'-O-ribose methylation and pseudouridylation of U6 spliceosomal RNA. Mol Cell Biol 19**:** 6906-6917.

Gao, X., Q. Jin, C. Jiang, Y. Li, C. Li *et al.*, 2016 FgPrp4 Kinase Is Important for Spliceosome B-Complex Activation and Splicing Efficiency in Fusarium graminearum. PLoS Genet 12**:** e1005973.

Gerbi, S. A., and T. S. Lange, 2002 All small nuclear RNAs (snRNAs) of the [U4/U6.U5] Tri-snRNP localize to nucleoli; Identification of the nucleolar localization element of U6 snRNA. Mol Biol Cell 13**:** 3123-3137.

Gerlach, V. L., S. K. Whitehall, E. P. Geiduschek and D. A. Brow, 1995 TFIIIB placement on a yeast U6 RNA gene in vivo is directed primarily by TFIIIC rather than by sequence-specific DNA contacts. Mol Cell Biol 15**:** 1455-1466.

Ghetti, A., M. Company and J. Abelson, 1995 Specificity of Prp24 binding to RNA: a role for Prp24 in the dynamic interaction of U4 and U6 snRNAs. RNA 1**:** 132-145.

Gottlieb, E., and J. A. Steitz, 1989a Function of the mammalian La protein: evidence for its action in transcription termination by RNA polymerase III. EMBO J 8**:** 851-861.

Gottlieb, E., and J. A. Steitz, 1989b The RNA binding protein La influences both the accuracy and the efficiency of RNA polymerase III transcription in vitro. EMBO J 8**:** 841-850.

Gu, J., J. R. Patton, S. Shimba and R. Reddy, 1996 Localization of modified nucleotides in Schizosaccharomyces pombe spliceosomal small nuclear RNAs: modified nucleotides are clustered in functionally important regions. RNA 2**:** 909-918.

Gu, J., G. Shumyatsky, N. Makan and R. Reddy, 1997 Formation of 2',3'-cyclic phosphates at the 3' end of human U6 small nuclear RNA in vitro. Identification of 2',3'-cyclic

phosphates at the 3' ends of human signal recognition particle and mitochondrial RNA processing RNAs. J Biol Chem 272**:** 21989-21993.

Gupta, S., R. K. Busch, R. Singh and R. Reddy, 1990a Characterization of U6 small nuclear RNA cap-specific antibodies. Identification of gamma-monomethyl-GTP cap structure in 7SK and several other human small RNAs. J Biol Chem 265**:** 19137-19142.

Gupta, S., R. Singh and R. Reddy, 1990b Capping of U6 small nuclear RNA in vitro can be uncoupled from transcription. J Biol Chem 265**:** 9491-9495.

Guthrie, C., 1991 Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein. Science 253**:** 157-163.

Hamada, M., A. L. Sakulich, S. B. Koduru and R. J. Maraia, 2000 Transcription termination by RNA polymerase III in fission yeast. A genetic and biochemically tractable model system. J Biol Chem 275**:** 29076-29081.

Hamm, J., and I. W. Mattaj, 1989 An abundant U6 snRNP found in germ cells and embryos of Xenopus laevis. EMBO J 8**:** 4179-4187.

Hardin, J. W., C. Warnasooriya, Y. Kondo, K. Nagai and D. Rueda, 2015 Assembly and dynamics of the U4/U6 di-snRNP by single-molecule FRET. Nucleic Acids Res 43**:** 10963-10974.

Hazbun, T. R., L. Malmstrom, S. Anderson, B. J. Graczyk, B. Fox *et al.*, 2003 Assigning function to yeast proteins by integration of technologies. Mol Cell 12**:** 1353-1365.

Hernandez, N., 2001 Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. J Biol Chem 276**:** 26733-26736.

Hilcenko, C., P. J. Simpson, A. J. Finch, F. R. Bowler, M. J. Churcher *et al.*, 2013 Aberrant 3' oligoadenylation of spliceosomal U6 small nuclear RNA in poikiloderma with neutropenia. Blood 121**:** 1028-1038.

Hilliker, A. K., and J. P. Staley, 2004 Multiple functions for the invariant AGC triad of U6 snRNA. RNA 10**:** 921-928.

Hoskins, A. A., and M. J. Moore, 2012 The spliceosome: a flexible, reversible macromolecular machine. Trends Biochem Sci 37**:** 179-188.

Hoskins, A. A., M. L. Rodgers, L. J. Friedman, J. Gelles and M. J. Moore, 2016 Single molecule analysis reveals reversible and irreversible steps during spliceosome activation. Elife 5.

Huang, Y. H., C. S. Chung, D. I. Kao, T. C. Kao and S. C. Cheng, 2014 Sad1 counteracts Brr2-mediated dissociation of U4/U6.U5 in tri-snRNP homeostasis. Mol Cell Biol 34**:** 210-220.

Huttenhofer, A., M. Kiefmann, S. Meier-Ewert, J. O'Brien, H. Lehrach *et al.*, 2001 RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. EMBO J 20**:** 2943-2953.

Jady, B. E., X. Darzacq, K. E. Tucker, A. G. Matera, E. Bertrand *et al.*, 2003 Modification of Sm small nuclear RNAs occurs in the nucleoplasmic Cajal body following import from the cytoplasm. EMBO J 22**:** 1878-1888.

Jeronimo, C., D. Forget, A. Bouchard, Q. Li, G. Chua *et al.*, 2007 Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. Mol Cell 27**:** 262-274.

Johnson, T. L., and J. Abelson, 2001 Characterization of U4 and U6 interactions with the 5' splice site using a S. cerevisiae in vitro trans-splicing system. Genes Dev 15**:** 1957-1970.

Kaiser, M. W., and D. A. Brow, 1995 Lethal mutations in a yeast U6 RNA gene B block promoter element identify essential contacts with transcription factor-IIIC. J Biol Chem 270**:** 11398-11405.

Kaiser, M. W., J. Chi and D. A. Brow, 2004 Position-dependent function of a B block promoter element implies a specialized chromatin structure on the S.cerevisiae U6 RNA gene, SNR6. Nucleic Acids Res 32**:** 4297-4305.

Kandels-Lewis, S., and B. Seraphin, 1993 Involvement of U6 snRNA in 5' splice site selection. Science 262**:** 2035-2039.

Kiss, T., M. Antal and F. Solymosy, 1987 Plant small nuclear RNAs. II. U6 RNA and a 4.5SI-like RNA are present in plant nuclei. Nucleic Acids Res 15**:** 543-560.

Klingauf, M., D. Stanek and K. M. Neugebauer, 2006 Enhancement of U4/U6 small nuclear ribonucleoprotein particle association in Cajal bodies predicted by mathematical modeling. Mol Biol Cell 17**:** 4972-4981.

Kong, K. Y., H. M. Tang, K. Pan, Z. Huang, T. H. Lee *et al.*, 2014 Cotranscriptional recruitment of yeast TRAMP complex to intronic sequences promotes optimal pre-mRNA splicing. Nucleic Acids Res 42**:** 643-660.

Krogan, N. J., G. Cagney, H. Yu, G. Zhong, X. Guo *et al.*, 2006 Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature 440**:** 637-643.

Krogh, N., M. Kongsbak-Wismann, C. Geisler and H. Nielsen, 2017 Substoichiometric ribose methylations in spliceosomal snRNAs. Org Biomol Chem.

Kruppa, M., R. D. Moir, D. Kolodrubetz and I. M. Willis, 2001 Nhp6, an HMG1 protein, functions in SNR6 transcription by RNA polymerase III in S. cerevisiae. Mol Cell 7**:** 309-318.

Kunkel, G. R., and T. Pederson, 1988 Upstream elements required for efficient transcription of a human U6 RNA gene resemble those of U1 and U2 genes even though a different polymerase is used. Genes Dev 2**:** 196-204.

Kwan, S., V. L. Gerlach and D. A. Brow, 2000 Disruption of the 5' stem-loop of yeast U6 RNA induces trimethylguanosine capping of this RNA polymerase III transcript in vivo. RNA 6**:** 1859-1869.

Kwan, S. S., and D. A. Brow, 2005 The N- and C-terminal RNA recognition motifs of splicing factor Prp24 have distinct functions in U6 RNA binding. RNA 11**:** 808-820.

Lange, T. S., and S. A. Gerbi, 2000 Transient nucleolar localization Of U6 small nuclear RNA in Xenopus Laevis oocytes. Mol Biol Cell 11**:** 2419-2428.

Lemm, I., C. Girard, A. N. Kuhn, N. J. Watkins, M. Schneider *et al.*, 2006 Ongoing U snRNP biogenesis is required for the integrity of Cajal bodies. Mol Biol Cell 17**:** 3221-3231.

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

Licht, K., J. Medenbach, R. Luhrmann, C. Kambach and A. Bindereif, 2008 3'-cyclic phosphorylation of U6 snRNA leads to recruitment of recycling factor p110 through LSm proteins. RNA 14**:** 1532-1538.

Lin, B. R., and V. Natarajan, 2012 Negative regulation of human U6 snRNA promoter by p38 kinase through Oct-1. Gene 497**:** 200-207.

Listerman, I., A. S. Bledau, I. Grishina and K. M. Neugebauer, 2007 Extragenic accumulation of RNA polymerase II enhances transcription by RNA polymerase III. PLoS Genet 3**:** e212.

Liu, S., S. Mozaffari-Jovin, J. Wollenhaupt, K. F. Santos, M. Theuser *et al.*, 2015 A composite double-/single-stranded RNA-binding region in protein Prp3 supports tri-snRNP stability and splicing. Elife 4**:** e07320.

Lobo, S. M., and N. Hernandez, 1989 A 7 bp mutation converts a human RNA polymerase II snRNA promoter into an RNA polymerase III promoter. Cell 58**:** 55-67.

Lund, E., and J. E. Dahlberg, 1992 Cyclic 2',3'-phosphates and nontemplated nucleotides at the 3' end of spliceosomal U6 small nuclear RNA's. Science 255**:** 327-330.

Madhani, H. D., R. Bordonne and C. Guthrie, 1990 Multiple roles for U6 snRNA in the splicing pathway. Genes Dev 4**:** 2264-2277.

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

Maraia, R. J., 1996 Transcription termination factor La is also an initiation factor for RNA polymerase III. Proc Natl Acad Sci U S A 93**:** 3383-3387.

Maraia, R. J., D. J. Kenan and J. D. Keene, 1994 Eukaryotic transcription termination factor La mediates transcript release and facilitates reinitiation by RNA polymerase III. Mol Cell Biol 14**:** 2147-2158.

Marcia, M., and A. M. Pyle, 2012 Visualizing group II intron catalysis through the stages of splicing. Cell 151**:** 497-507.

Margottin, F., G. Dujardin, M. Gerard, J. M. Egly, J. Huet *et al.*, 1991 Participation of the TATA factor in transcription of the yeast U6 gene by RNA polymerase C. Science 251**:** 424-426.

Martin-Tumasz, S., A. C. Richie, L. J. Clos, 2nd, D. A. Brow and S. E. Butcher, 2011 A novel occluded RNA recognition motif in Prp24 unwinds the U6 RNA internal stem loop. Nucleic Acids Res 39**:** 7837-7847.

Martin, A., S. Schneider and B. Schwer, 2002 Prp43 is an essential RNA-dependent ATPase required for release of lariat-intron from the spliceosome. J Biol Chem 277**:** 17743-17750.

Martin, M. P., V. L. Gerlach and D. A. Brow, 2001 A novel upstream RNA polymerase III promoter element becomes essential when the chromatin structure of the yeast U6 RNA gene is altered. Mol Cell Biol 21**:** 6429-6439.

Massenet, S., and C. Branlant, 1999 A limited number of pseudouridine residues in the human atac spliceosomal UsnRNAs as compared to human major spliceosomal UsnRNAs. RNA 5**:** 1495-1503.

Massenet, S., Y. Motorin, D. L. Lafontaine, E. C. Hurt, H. Grosjean *et al.*, 1999 Pseudouridine mapping in the Saccharomyces cerevisiae spliceosomal U small nuclear RNAs (snRNAs) reveals that pseudouridine synthase pus1p exhibits a dual substrate specificity for U2 snRNA and tRNA. Mol Cell Biol 19**:** 2142-2154.

Matera, A. G., and Z. Wang, 2014 A day in the life of the spliceosome. Nat Rev Mol Cell Biol 15**:** 108-121.

Matera, A. G., and D. C. Ward, 1993 Nucleoplasmic organization of small nuclear ribonucleoproteins in cultured human cells. J Cell Biol 121**:** 715-727.

Mayes, A. E., L. Verdone, P. Legrain and J. D. Beggs, 1999 Characterization of Sm-like proteins in yeast and their association with U6 snRNA. EMBO J 18**:** 4321-4331.

McManus, C. J., M. L. Schwartz, S. E. Butcher and D. A. Brow, 2007 A dynamic bulge in the U6 RNA internal stem-loop functions in spliceosome assembly and activation. RNA 13**:** 2252-2265.

Medenbach, J., S. Schreiner, S. Liu, R. Luhrmann and A. Bindereif, 2004 Human U4/U6 snRNP recycling factor p110: mutational analysis reveals the function of the tetratricopeptide repeat domain in recycling. Mol Cell Biol 24**:** 7392-7401.

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

Meselson, M., 1972 Formation of hybrid DNA by rotary diffusion during genetic recombination. J Mol Biol 71**:** 795-798.

Meyer, K. D., and S. R. Jaffrey, 2014 The dynamic epitranscriptome: N6-methyladenosine and gene expression control. Nat Rev Mol Cell Biol 15**:** 313-326.

Moenne, A., S. Camier, G. Anderson, F. Margottin, J. Beggs *et al.*, 1990 The U6 gene of Saccharomyces cerevisiae is transcribed by RNA polymerase C (III) in vivo and in vitro. EMBO J 9**:** 271-277.

Montemayor, E. J., E. C. Curran, H. H. Liao, K. L. Andrews, C. N. Treba *et al.*, 2014 Core structure of the U6 small nuclear ribonucleoprotein at 1.7-A resolution. Nat Struct Mol Biol 21**:** 544-551.

Montemayor, E. J., A. L. Didychuk, H. Liao, P. Hu, D. A. Brow *et al.*, 2017 Structure and conformational plasticity of the U6 small nuclear ribonucleoprotein core. Acta Crystallogr D Struct Biol 73**:** 1-8.

Mozaffari-Jovin, S., K. F. Santos, H. H. Hsiao, C. L. Will, H. Urlaub *et al.*, 2012 The Prp8 RNase H-like domain inhibits Brr2-mediated U4/U6 snRNA unwinding by blocking Brr2 loading onto the U4 snRNA. Genes Dev 26**:** 2422-2434.

Mozaffari-Jovin, S., T. Wandersleben, K. F. Santos, C. L. Will, R. Luhrmann *et al.*, 2013 Inhibition of RNA helicase Brr2 by the C-terminal tail of the spliceosomal protein Prp8. Science 341**:** 80-84.

Mroczek, S., J. Krwawicz, J. Kutner, M. Lazniewski, I. Kucinski *et al.*, 2012 C16orf57, a gene mutated in poikiloderma with neutropenia, encodes a putative phosphodiesterase responsible for the U6 snRNA 3' end modification. Genes Dev 26**:** 1911-1925.

Mura, C., P. S. Randolph, J. Patterson and A. E. Cozen, 2013 Archaeal and eukaryotic homologs of Hfq: A structural and evolutionary perspective on Sm function. RNA Biol 10**:** 636-651.

Nag, A., and J. A. Steitz, 2012 Tri-snRNP-associated proteins interact with subunits of the TRAMP and nuclear exosome complexes, linking RNA decay and pre-mRNA splicing. RNA Biol 9**:** 334-342.

Newby, M. I., and N. L. Greenbaum, 2001 A conserved pseudouridine modification in eukaryotic U2 snRNA induces a change in branch-site architecture. RNA 7**:** 833-845.

Nguyen, T. H., W. P. Galej, X. C. Bai, C. Oubridge, A. J. Newman *et al.*, 2016 Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 A resolution. Nature 530**:** 298-302.

Nguyen, T. H., W. P. Galej, X. C. Bai, C. G. Savva, A. J. Newman *et al.*, 2015 The architecture of the spliceosomal U4/U6.U5 tri-snRNP. Nature 523**:** 47-52.

Novotny, I., M. Blazikova, D. Stanek, P. Herman and J. Malinsky, 2011 In vivo kinetics of U4/U6.U5 tri-snRNP formation in Cajal bodies. Mol Biol Cell 22**:** 513-523.

Novotny, I., A. Malinova, E. Stejskalova, D. Mateju, K. Klimesova *et al.*, 2015 SART3-Dependent Accumulation of Incomplete Spliceosomal snRNPs in Cajal Bodies. Cell Rep 10**:** 429-440.

Oler, A. J., R. K. Alla, D. N. Roberts, A. Wong, P. C. Hollenhorst *et al.*, 2010 Human RNA polymerase III transcriptomes and relationships to Pol II promoter chromatin and enhancer-binding factors. Nat Struct Mol Biol 17**:** 620-628.

Park, J. K., T. Das, E. J. Song and E. E. Kim, 2016 Structural basis for recruiting and shuttling of the spliceosomal deubiquitinase USP4 by SART3. Nucleic Acids Res 44**:** 5424-5437.

Peebles, C. L., P. S. Perlman, K. L. Mecklenburg, M. L. Petrillo, J. H. Tabor *et al.*, 1986 A self-splicing RNA excises an intron lariat. Cell 44**:** 213-223.

Peebles, C. L., M. Zhang, P. S. Perlman and J. S. Franzen, 1995 Catalytically critical nucleotide in domain 5 of a group II intron. Proc Natl Acad Sci U S A 92**:** 4422-4426.

Pendleton, K. E., B. Chen, K. Liu, O. V. Hunter, Y. Xie *et al.*, 2017 The U6 snRNA m6A Methyltransferase METTL16 Regulates SAM Synthetase Intron Retention. Cell 169**:** 824-835 e814.

Pessa, H. K., C. L. Will, X. Meng, C. Schneider, N. J. Watkins *et al.*, 2008 Minor spliceosome components are predominantly localized in the nucleus. Proc Natl Acad Sci U S A 105**:** 8655-8660.

Plaschka, C., P. C. Lin and K. Nagai, 2017 Structure of a pre-catalytic spliceosome. Nature 546**:** 617-621.

Pozzi, B., L. Bragado, C. L. Will, P. Mammi, G. Risso *et al.*, 2017 SUMO conjugation to spliceosomal proteins is required for efficient pre-mRNA splicing. Nucleic Acids Res 45**:** 6729-6745.

Pyle, A. M., 2016 Group II Intron Self-Splicing. Annu Rev Biophys 45**:** 183-205.

Rader, S. D., and C. Guthrie, 2002 A conserved Lsm-interaction motif in Prp24 required for efficient U4/U6 di-snRNP formation. RNA 8**:** 1378-1392.

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

Raghunathan, P. L., and C. Guthrie, 1998b A spliceosomal recycling factor that reanneals U4 and U6 small nuclear ribonucleoprotein particles. Science 279**:** 857-860.

Raha, D., Z. Wang, Z. Moqtaderi, L. Wu, G. Zhong *et al.*, 2010 Close association of RNA polymerase II and many transcription factors with Pol III genes. Proc Natl Acad Sci U S A 107**:** 3639-3644.

Rauhut, R., P. Fabrizio, O. Dybkov, K. Hartmuth, V. Pena *et al.*, 2016 Molecular architecture of the Saccharomyces cerevisiae activated spliceosome. Science 353**:** 1399-1405.

Reddy, R., D. Henning, G. Das, M. Harless and D. Wright, 1987 The capped U6 small nuclear RNA is transcribed by RNA polymerase III. J Biol Chem 262**:** 75-81.

Reijns, M. A., T. Auchynnikava and J. D. Beggs, 2009 Analysis of Lsm1p and Lsm8p domains in the cellular localization of Lsm complexes in budding yeast. FEBS J 276**:** 3602-3617.

Rife, J. P., C. S. Cheng, P. B. Moore and S. A. Strobel, 1998 N 2-methylguanosine is iso-energetic with guanosine in RNA duplexes and GNRA tetraloops. Nucleic Acids Res 26**:** 3640-3644.

Rinke, J., and J. A. Steitz, 1985 Association of the lupus antigen La with a subset of U6 snRNA molecules. Nucleic Acids Res 13**:** 2617-2629.

Rodgers, M. L., A. L. Didychuk, S. E. Butcher, D. A. Brow and A. A. Hoskins, 2016 A multi-step model for facilitated unwinding of the yeast U4/U6 RNA duplex. Nucleic Acids Res 44**:** 10912-10928.

Roiha, H., E. O. Shuster, D. A. Brow and C. Guthrie, 1989 Small nuclear RNAs from budding yeasts: phylogenetic comparisons reveal extensive size variation. Gene 82**:** 137-144.

Roost, C., S. R. Lynch, P. J. Batista, K. Qu, H. Y. Chang *et al.*, 2015 Structure and thermodynamics of N6-methyladenosine in RNA: a spring-loaded base modification. J Am Chem Soc 137**:** 2107-2115.

Ruegger, S., T. S. Miki, D. Hess and H. Grosshans, 2015 The ribonucleotidyl transferase USIP-1 acts with SART3 to promote U6 snRNA recycling. Nucleic Acids Res 43**:** 3344-3357.

Sashital, D. G., A. M. Allmann, S. R. Van Doren and S. E. Butcher, 2003 Structural basis for a lethal mutation in U6 RNA. Biochemistry 42**:** 1470-1477.

Sashital, D. G., V. Venditti, C. G. Angers, G. Cornilescu and S. E. Butcher, 2007 Structure and thermodynamics of a conserved U2 snRNA domain from yeast and human. RNA 13**:** 328-338.

Sawa, H., and J. Abelson, 1992 Evidence for a base-pairing interaction between U6 small nuclear RNA and 5' splice site during the splicing reaction in yeast. Proc Natl Acad Sci U S A 89**:** 11269-11273.

Sawa, H., and Y. Shimura, 1992 Association of U6 snRNA with the 5'-splice site region of pre-mRNA in the spliceosome. Genes Dev 6**:** 244-254.

Schmidt, H., K. Richert, R. A. Drakas and N. F. Kaufer, 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**:** 1183-1191.

Schneider, C., G. Kudla, W. Wlotzka, A. Tuck and D. Tollervey, 2012 Transcriptome-wide analysis of exosome targets. Mol Cell 48**:** 422-433.

Schneider, C., C. L. Will, O. V. Makarova, E. M. Makarov and R. Luhrmann, 2002 Human U4/U6.U5 and U4atac/U6atac.U5 tri-snRNPs exhibit similar protein compositions. Mol Cell Biol 22**:** 3219-3229.

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

Schneider, M., C. L. Will, M. Anokhina, J. Tazi, H. Urlaub *et al.*, 2010b Exon definition complexes contain the tri-snRNP and can be directly converted into B-like precatalytic splicing complexes. Mol Cell 38**:** 223-235.

Shannon, K. W., and C. Guthrie, 1991 Suppressors of a U4 snRNA mutation define a novel U6 snRNP protein with RNA-binding motifs. Genes Dev 5**:** 773-785.

Sharif, H., and E. Conti, 2013 Architecture of the Lsm1-7-Pat1 complex: a conserved assembly in eukaryotic mRNA turnover. Cell Rep 5**:** 283-291.

Sharp, P. A., 1985 On the origin of RNA splicing and introns. Cell 42**:** 397-400.

Shchepachev, V., H. Wischnewski, E. Missiaglia, C. Soneson and C. M. Azzalin, 2012 Mpn1, mutated in poikiloderma with neutropenia protein 1, is a conserved 3'-to-5' RNA exonuclease processing U6 small nuclear RNA. Cell Rep 2**:** 855-865.

Shchepachev, V., H. Wischnewski, C. Soneson, A. W. Arnold and C. M. Azzalin, 2015 Human Mpn1 promotes post-transcriptional processing and stability of U6atac. FEBS Lett 589**:** 2417-2423.

Shcherbakova, I., A. A. Hoskins, L. J. Friedman, V. Serebrov, I. R. Correa, Jr. *et al.*, 2013 Alternative spliceosome assembly pathways revealed by single-molecule fluorescence microscopy. Cell Rep 5**:** 151-165.

Shimba, S., J. A. Bokar, F. Rottman and R. Reddy, 1995 Accurate and efficient N-6-adenosine methylation in spliceosomal U6 small nuclear RNA by HeLa cell extract in vitro. Nucleic Acids Res 23**:** 2421-2426.

Shimba, S., and R. Reddy, 1994 Purification of human U6 small nuclear RNA capping enzyme. Evidence for a common capping enzyme for gamma-monomethyl-capped small RNAs. J Biol Chem 269**:** 12419-12423.

Shivaswamy, S., G. A. Kassavetis and P. Bhargava, 2004 High-level activation of transcription of the yeast U6 snRNA gene in chromatin by the basal RNA polymerase III transcription factor TFIIIC. Mol Cell Biol 24**:** 3596-3606.

Shukla, G. C., and R. A. Padgett, 2001 The intramolecular stem-loop structure of U6 snRNA can functionally replace the U6atac snRNA stem-loop. RNA 7**:** 94-105.

Shukla, G. C., and R. A. Padgett, 2002 A catalytically active group II intron domain 5 can function in the U12-dependent spliceosome. Mol Cell 9**:** 1145-1150.

Shukla, S., and R. Parker, 2017 PARN Modulates Y RNA Stability and Its 3'-End Formation. Mol Cell Biol 37.

Shumyatsky, G., D. Wright and R. Reddy, 1993 Methylphosphate cap structure increases the stability of 7SK, B2 and U6 small RNAs in Xenopus oocytes. Nucleic Acids Res 21**:** 4756-4761.

Sidarovich, A., C. L. Will, M. M. Anokhina, J. Ceballos, S. Sievers *et al.*, 2017 Identification of a small molecule inhibitor that stalls splicing at an early step of spliceosome activation. Elife 6.

Sigal, N., and B. Alberts, 1972 Genetic recombination: the nature of a crossed strand-exchange between two homologous DNA molecules. J Mol Biol 71**:** 789-793.

Singh, J., K. Sikand, H. Conrad, C. L. Will, A. A. Komar *et al.*, 2016 U6atac snRNA stem-loop interacts with U12 p65 RNA binding protein and is functionally interchangeable with the U12 apical stem-loop III. Sci Rep 6**:** 31393.

Singh, R., S. Gupta and R. Reddy, 1990 Capping of mammalian U6 small nuclear RNA in vitro is directed by a conserved stem-loop and AUAUAC sequence: conversion of a noncapped RNA into a capped RNA. Mol Cell Biol 10**:** 939-946.

Singh, R., and R. Reddy, 1989 Gamma-monomethyl phosphate: a cap structure in spliceosomal U6 small nuclear RNA. Proc Natl Acad Sci U S A 86**:** 8280-8283.

Song, E. J., S. L. Werner, J. Neubauer, F. Stegmeier, J. Aspden *et al.*, 2010 The Prp19 complex and the Usp4Sart3 deubiquitinating enzyme control reversible ubiquitination at the spliceosome. Genes Dev 24**:** 1434-1447.

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

Spaniel, C., M. Honda, S. R. Selitsky, D. Yamane, T. Shimakami *et al.*, 2013 microRNA-122 abundance in hepatocellular carcinoma and non-tumor liver tissue from Japanese patients with persistent HCV versus HBV infection. PLoS One 8**:** e76867.

Spiller, M. P., K. L. Boon, M. A. Reijns and J. D. Beggs, 2007a The Lsm2-8 complex determines nuclear localization of the spliceosomal U6 snRNA. Nucleic Acids Res 35**:** 923-929.

Spiller, M. P., M. A. Reijns and J. D. Beggs, 2007b Requirements for nuclear localization of the Lsm2-8p complex and competition between nuclear and cytoplasmic Lsm complexes. J Cell Sci 120**:** 4310-4320.

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

Stanek, D., and K. M. Neugebauer, 2004 Detection of snRNP assembly intermediates in Cajal bodies by fluorescence resonance energy transfer. J Cell Biol 166**:** 1015-1025.

Stanek, D., J. Pridalova-Hnilicova, I. Novotny, M. Huranova, M. Blazikova *et al.*, 2008 Spliceosomal small nuclear ribonucleoprotein particles repeatedly cycle through Cajal bodies. Mol Biol Cell 19**:** 2534-2543.

Stanek, D., S. D. Rader, M. Klingauf and K. M. Neugebauer, 2003 Targeting of U4/U6 small nuclear RNP assembly factor SART3/p110 to Cajal bodies. J Cell Biol 160**:** 505-516.

Stark, M. R., E. A. Dunn, W. S. Dunn, C. J. Grisdale, A. R. Daniele *et al.*, 2015 Dramatically reduced spliceosome in Cyanidioschyzon merolae. Proc Natl Acad Sci U S A 112**:** E1191-1200.

Stefano, J. E., 1984 Purified lupus antigen La recognizes an oligouridylate stretch common to the 3' termini of RNA polymerase III transcripts. Cell 36**:** 145-154.

Steinmetz, E. J., and D. A. Brow, 1998 Control of pre-mRNA accumulation by the essential yeast protein Nrd1 requires high-affinity transcript binding and a domain implicated in RNA polymerase II association. Proc Natl Acad Sci U S A 95**:** 6699-6704.

Steinmetz, E. J., C. L. Warren, J. N. Kuehner, B. Panbehi, A. Z. Ansari *et al.*, 2006 Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. Mol Cell 24**:** 735-746.

Steitz, T. A., and J. A. Steitz, 1993 A general two-metal-ion mechanism for catalytic RNA. Proc Natl Acad Sci U S A 90**:** 6498-6502.

Stevens, S. W., I. Barta, H. Y. Ge, R. E. Moore, M. K. Young *et al.*, 2001 Biochemical and genetic analyses of the U5, U6, and U4/U6 x U5 small nuclear ribonucleoproteins from Saccharomyces cerevisiae. RNA 7**:** 1543-1553.

Stillman, D. J., 2010 Nhp6: a small but powerful effector of chromatin structure in Saccharomyces cerevisiae. Biochim Biophys Acta 1799**:** 175-180.

Takahashi, Y., T. Tani and Y. Ohshima, 1996 Spliceosomal introns in conserved sequences of U1 and U5 small nuclear RNA genes in yeast Rhodotorula hasegawae. J Biochem 120**:** 677-683.

Takahashi, Y., S. Urushiyama, T. Tani and Y. Ohshima, 1993 An mRNA-type intron is present in the Rhodotorula hasegawae U2 small nuclear RNA gene. Mol Cell Biol 13**:** 5613-5619.

Tani, T., and Y. Ohshima, 1989 The gene for the U6 small nuclear RNA in fission yeast has an intron. Nature 337**:** 87-90.

Tani, T., and Y. Ohshima, 1991 mRNA-type introns in U6 small nuclear RNA genes: implications for the catalysis in pre-mRNA splicing. Genes Dev 5**:** 1022-1031.

Tarn, W. Y., and J. A. Steitz, 1996a Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. Science 273**:** 1824-1832.

Tarn, W. Y., and J. A. Steitz, 1996b A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro. Cell 84**:** 801-811.

Tazi, J., T. Forne, P. Jeanteur, G. Cathala and C. Brunel, 1993 Mammalian U6 small nuclear RNA undergoes 3' end modifications within the spliceosome. Mol Cell Biol 13**:** 1641-1650.

Terns, M. P., E. Lund and J. E. Dahlberg, 1992 3'-end-dependent formation of U6 small nuclear ribonucleoprotein particles in Xenopus laevis oocyte nuclei. Mol Cell Biol 12**:** 3032-3040.

Tharun, S., W. He, A. E. Mayes, P. Lennertz, J. D. Beggs *et al.*, 2000 Yeast Sm-like proteins function in mRNA decapping and decay. Nature 404**:** 515-518.

Theuser, M., C. Hobartner, M. C. Wahl and K. F. Santos, 2016 Substrate-assisted mechanism of RNP disruption by the spliceosomal Brr2 RNA helicase. Proc Natl Acad Sci U S A 113**:** 7798-7803.

Tichelaar, J. W., B. Knerer, A. Vrabel and E. D. Wieben, 1994 Transcription of a variant human U6 small nuclear RNA gene is controlled by a novel, internal RNA polymerase III promoter. Mol Cell Biol 14**:** 5450-5457.

Tichelaar, J. W., E. D. Wieben, R. Reddy, A. Vrabel and P. Camacho, 1998 In vivo expression of a variant human U6 RNA from a unique, internal promoter. Biochemistry 37**:** 12943-12951.

Trippe, R., E. Guschina, M. Hossbach, H. Urlaub, R. Luhrmann *et al.*, 2006 Identification, cloning, and functional analysis of the human U6 snRNA-specific terminal uridylyl transferase. RNA 12**:** 1494-1504.

Trippe, R., H. Richly and B. J. Benecke, 2003 Biochemical characterization of a U6 small nuclear RNA-specific terminal uridylyltransferase. Eur J Biochem 270**:** 971-980.

Trippe, R., B. Sandrock and B. J. Benecke, 1998 A highly specific terminal uridylyl transferase modifies the 3'-end of U6 small nuclear RNA. Nucleic Acids Res 26**:** 3119-3126.

Tsai, R. T., R. H. Fu, F. L. Yeh, C. K. Tseng, Y. C. Lin *et al.*, 2005 Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2. Genes Dev 19**:** 2991-3003.

Tycowski, K. T., Z. H. You, P. J. Graham and J. A. Steitz, 1998 Modification of U6 spliceosomal RNA is guided by other small RNAs. Mol Cell 2**:** 629-638.

van der Feltz, C., A. C. DeHaven and A. A. Hoskins, 2017 Stress-induced Pseudouridylation Alters the Structural Equilibrium of Yeast U2 snRNA Stem II. J Mol Biol.

van der Veen, R., A. C. Arnberg, G. van der Horst, L. Bonen, H. F. Tabak *et al.*, 1986 Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing in vitro. Cell 44**:** 225-234.

Vankan, P., C. McGuigan and I. W. Mattaj, 1990 Domains of U4 and U6 snRNAs required for snRNP assembly and splicing complementation in Xenopus oocytes. EMBO J 9**:** 3397-3404.

Vasiljeva, L., M. Kim, N. Terzi, L. M. Soares and S. Buratowski, 2008 Transcription termination and RNA degradation contribute to silencing of RNA polymerase II transcription within heterochromatin. Mol Cell 29**:** 313-323.

Venditti, V., L. Clos, 2nd, N. Niccolai and S. E. Butcher, 2009 Minimum-energy path for a u6 RNA conformational change involving protonation, base-pair rearrangement and base flipping. J Mol Biol 391**:** 894-905.

Verheggen, C., D. L. Lafontaine, D. Samarsky, J. Mouaikel, J. M. Blanchard *et al.*, 2002 Mammalian and yeast U3 snoRNPs are matured in specific and related nuclear compartments. EMBO J 21**:** 2736-2745.

Vidaver, R. M., D. M. Fortner, L. S. Loos-Austin and D. A. Brow, 1999 Multiple functions of Saccharomyces cerevisiae splicing protein Prp24 in U6 RNA structural rearrangements. Genetics 153**:** 1205-1218.

Vijayraghavan, U., M. Company and J. Abelson, 1989 Isolation and characterization of pre-mRNA splicing mutants of Saccharomyces cerevisiae. Genes Dev 3**:** 1206-1216.

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

Wan, R., C. Yan, R. Bai, G. Huang and Y. Shi, 2016a Structure of a yeast catalytic step I spliceosome at 3.4 A resolution. Science 353**:** 895-904.

Wan, R., C. Yan, R. Bai, J. Lei and Y. Shi, 2017 Structure of an Intron Lariat Spliceosome from Saccharomyces cerevisiae. Cell 171**:** 120-132 e112.

Wan, R., C. Yan, R. Bai, L. Wang, M. Huang *et al.*, 2016b The 3.8 A structure of the U4/U6.U5 tri-snRNP: Insights into spliceosome assembly and catalysis. Science 351**:** 466-475.

Warda, A. S., J. Kretschmer, P. Hackert, C. Lenz, H. Urlaub *et al.*, 2017 Human METTL16 is a N6-methyladenosine (m6A) methyltransferase that targets pre-mRNAs and various non-coding RNAs. EMBO Rep 18**:** 2004-2014.

Wassarman, D. A., and J. A. Steitz, 1992 Interactions of small nuclear RNA's with precursor messenger RNA during in vitro splicing. Science 257**:** 1918-1925.

Will, C. L., and R. Luhrmann, 2011 Spliceosome structure and function. Cold Spring Harb Perspect Biol 3.

Wu, G., H. Adachi, J. Ge, D. Stephenson, C. C. Query *et al.*, 2016 Pseudouridines in U2 snRNA stimulate the ATPase activity of Prp5 during spliceosome assembly. EMBO J 35**:** 654-667.

Xu, H., R. S. Pillai, T. N. Azzouz, K. B. Shpargel, C. Kambach *et al.*, 2005 The C-terminal domain of coilin interacts with Sm proteins and U snRNPs. Chromosoma 114**:** 155-166.

Yamashita, S., Y. Takagi, T. Nagaike and K. Tomita, 2017 Crystal structures of U6 snRNA-specific terminal uridylyltransferase. Nat Commun 8**:** 15788.

Yan, C., J. Hang, R. Wan, M. Huang, C. C. Wong *et al.*, 2015 Structure of a yeast spliceosome at 3.6-angstrom resolution. Science 349**:** 1182-1191.

Yan, C., R. Wan, R. Bai, G. Huang and Y. Shi, 2016 Structure of a yeast activated spliceosome at 3.5 A resolution. Science 353**:** 904-911.

Yan, C., R. Wan, R. Bai, G. Huang and Y. Shi, 2017 Structure of a yeast step II catalytically activated spliceosome. Science 355**:** 149-155.

Yang, F., X. Y. Wang, Z. M. Zhang, J. Pu, Y. J. Fan *et al.*, 2013 Splicing proofreading at 5' splice sites by ATPase Prp28p. Nucleic Acids Res 41**:** 4660-4670.

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

Younis, I., K. Dittmar, W. Wang, S. W. Foley, M. G. Berg *et al.*, 2013 Minor introns are embedded molecular switches regulated by highly unstable U6atac snRNA. Elife 2**:** e00780.

Yu, H., P. Braun, M. A. Yildirim, I. Lemmens, K. Venkatesan *et al.*, 2008 High-quality binary protein interaction map of the yeast interactome network. Science 322**:** 104-110.

Yu, Y. T., M. D. Shu and J. A. Steitz, 1998 Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. EMBO J 17**:** 5783-5795.

Yuan, C. C., X. Zhao, L. Florens, S. K. Swanson, M. P. Washburn *et al.*, 2007 CHD8 associates with human Staf and contributes to efficient U6 RNA polymerase III transcription. Mol Cell Biol 27**:** 8729-8738.

Zhang, L., Y. Wan, G. Huang, D. Wang, X. Yu *et al.*, 2015 The exosome controls alternative splicing by mediating the gene expression and assembly of the spliceosome complex. Sci Rep 5**:** 13403.

Zhang, Q., R. Harding, F. Hou, A. Dong, J. R. Walker *et al.*, 2016 Structural Basis of the Recruitment of Ubiquitin-specific Protease USP15 by Spliceosome Recycling Factor SART3. J Biol Chem 291**:** 17283-17292.

Zhang, X., C. Yan, J. Hang, L. I. Finci, J. Lei *et al.*, 2017 An Atomic Structure of the Human Spliceosome. Cell 169**:** 918-929 e914.

Zhao, X., P. S. Pendergrast and N. Hernandez, 2001 A positioned nucleosome on the human U6 promoter allows recruitment of SNAPc by the Oct-1 POU domain. Mol Cell 7**:** 539-549.

Zhou, H., Y. Q. Chen, Y. P. Du and L. H. Qu, 2002 The Schizosaccharomyces pombe mgU6-47 gene is required for 2'-O-methylation of U6 snRNA at A41. Nucleic Acids Res 30**:** 894-902.

Zhou, L., J. Hang, Y. Zhou, R. Wan, G. Lu *et al.*, 2014a Crystal structures of the Lsm complex bound to the 3' end sequence of U6 small nuclear RNA. Nature 506**:** 116-120.

Zhou, L., Y. Zhou, J. Hang, R. Wan, G. Lu *et al.*, 2014b Crystal structure and biochemical analysis of the heptameric Lsm1-7 complex. Cell Res 24**:** 497-500.

Chapter 3: Usb1 controls U6 snRNP assembly through evolutionarily divergent cyclic phosphodiesterase activities

This chapter is published in the following form:

Didychuk, A.L., Montemayor, E.J., Carrocci, T.J., DeLaitsch, A.T., Lucarelli, S.E., Westler, W.M., Brow, D.A., Hoskins, A.A., Butcher, S.E. (2017) Usb1 controls U6 snRNP assembly through evolutionarily divergent cyclic phosphodiesterase activities. *Nature Communications* 8(1):497.

Author contributions:

A.L.D., E.J.M., T.J.C., and S.E.B. designed research; A.L.D. analyzed data. A.L.D., E.J.M., T.J.C, A.T.D., and S.E.L. performed research. A.L.D., E.J.M., and A.T.D. prepared crystallization samples. A.L.D. and E.J.M. performed structure determination. A.L.D. and S.E.L. performed biochemical assays. A.L.D. and T.J.C. generated the Usb1 deletion yeast strain and carried out experiments. T.J.C. performed Western blots. W.M.W. wrote NMR pulse programs and assisted with data collection and analysis. A.L.D. and S.E.B. wrote the paper with input from A.A.H., D.A.B., E.J.M., and T.J.C.

3.1 Abstract

U6 small nuclear ribonucleoprotein (snRNP) biogenesis is essential for spliceosome assembly, but not well understood. Here, we report structures of the U6 RNA processing enzyme Usb1 from yeast and a substrate analog bound complex from humans. Unlike the human ortholog, we show that yeast Usb1 has cyclic phosphodiesterase activity that leaves a terminal 3' phosphate which prevents over-processing. Usb1 processing of U6 RNA dramatically alters its affinity for cognate RNA-binding proteins. We reconstitute the post-transcriptional assembly of yeast U6 snRNP *in vitro*, which occurs through a complex series of hand-offs involving 10 proteins (Lhp1, Prp24, Usb1 and Lsm2-8) and anti-cooperative interactions between Prp24 and Lhp1. We propose a model for U6 snRNP assembly that explains how evolutionarily divergent and seemingly antagonistic proteins cooperate to protect and chaperone the nascent snRNA during its journey to the spliceosome.

3.2 Introduction

Splicing of precursor messenger RNA (pre-mRNA) is an essential process in all eukaryotes and is catalyzed by the spliceosome. The spliceosome is a dynamic macromolecular machine composed primarily of five ribonucleoprotein particles known as the U1, U2, U4, U5 and U6 snRNPs, each containing a small nuclear RNA (snRNA) and numerous proteins. The highly conserved U6 snRNA coordinates magnesium ions in the active site that are required for splicing catalysis[1]. Unlike the other snRNAs, U6 is synthesized by RNA polymerase III (Pol III) and, like other Pol III transcripts, its transcription is terminated stochastically when the polymerase encounters a stretch of adenines in the template strand (**Figure 3-1A**)[2-4] with *S. cerevisiae* requiring at least six sequential adenines in the template strand to terminate efficiently[5]. Nascent U6 terminates in a 3' polyuridine tract of heterogeneous length (4-8 uridines) with terminal 2' and 3' hydroxyl groups (a cis diol) (**Figure 3-1A**) and is bound by the La protein (Lhp1 in *S. cerevisiae*)[6,7]. However, the predominant form of mature U6 *in vivo* in most organisms characterized to date does not contain a terminal cis diol and is not bound by La protein[8].

Post-transcriptional exonucleolytic processing of U6 is directed by the U six biogenesis protein 1 (Usb1)[9,10], a 3'-5' exonuclease belonging to the 2H phosphodiesterase superfamily of enzymes[9,11]. Usb1 is an essential protein in *S. cerevisiae*[9,12], but not in *S. pombe*[10] and metazoans[10,11,13]. In yeast, it is likely that immature U6 RNA is the only essential target of Usb1, as overexpression of this snRNA rescues deletion of yUsb1[11]. Human Usb1 is involved in processing precursors of both U6 and the minor spliceosome snRNA U6atac[13]. Loss-of-function mutations in human Usb1 are associated with poikiloderma with neutropenia, a rare skin disease that is also associated with loss of white blood cells[14].

In humans, processing of U6 by Usb1 creates a terminal 2',3'-cyclic phosphate[11] which stimulates binding of Lsm2-8[15]. Binding of Lsm2-8 in turn promotes formation of the U6 and U4/U6 snRNPs[16-18]. Thus, the identity of the 3' end of U6 RNA is a crucial determinant in the assembly of U6-containing snRNPs. However, the end modification of the terminal nucleotide in yeast differs

from that in humans, with yeast primarily containing a non-cyclic phosphate group[8]. Prior to this work, it was not known if yUsb1 directly promotes formation of the terminal phosphate, or if an additional cyclic phosphodiesterase (CPDase) acts on U6 RNA after exonucleolytic processing. Furthermore, the position of the terminal phosphate on yeast U6 RNA (at either the 2' or 3' oxygen) was not known, and cannot be resolved in recent cryo-EM structures due to local resolutions of >7 Å for the U6 3' tail[19-24].

To investigate the mechanism of U6 RNA processing and snRNP assembly, we characterized the structure and activities of yeast and human Usb1. We determined the 1.8 Å crystal structure of the catalytic domain of yUsb1, and a 1.4 Å co-crystal structure of hUsb1 bound to the substrate analog uridine 5'-monophosphate (5'UMP). We demonstrate the importance of the identity of the 3' end of U6 to snRNP formation and show how U6 RNA is involved in a series of protein-mediated hand-offs prior to formation of the mature U6 snRNP.

3.3 Results

3.3.1 Yeast Usb1 exhibits cyclic phosphodiesterase activity

We sought to understand the activity of Usb1 from *S. cerevisiae* (yUsb1). To this end, we prepared full-length yUsb1 protein and tested it for exoribonuclease activity *in vitro* using oligonucleotide model substrates. When incubated with RNA terminating in multiple uridines and a cis diol, yUsb1 predominately removed only 1 nucleotide from the 3' end of an RNA, with 80% of the substrate converted into the *n-1* product (**Figure 3-1B**). Interestingly, 15% of the *n-1* product had a slightly slower mobility consistent with it containing a cyclic phosphate 3' end[25]. Removal of one or two additional nucleotides occurs infrequently (15% and 3% of the total product, respectively) (**Figure 3-1B**). Substrates with a deoxyuridine at the *n-1* position were not processed by yUsb1 (**Figure 3-2A**), indicating that yUsb1 acts exclusively as a 3'-5' exonuclease and that additional cleavage products are due to inefficient re-processing of *n-1* RNA. Additionally, we find that yUsb1 removes 1-3 nucleotides regardless of the length of the polyuridine tail (**Figure 3-2B**).

We investigated the product of yUsb1 processing by exploiting the disparate activities of Calf Intestinal Phosphatase (CIP) and T4 polynucleotide kinase (PNK) (**Figure 3-1B**). Both CIP and T4 PNK can remove terminal 3' phosphoryl groups from oligonucleotides to produce cis diols, but only T4 PNK can remove both cyclic and noncyclic terminal phosphates. Treatment of yUsb1-processed RNA with CIP or T4 PNK both resulted in reduced mobility of the products, consistent with the presence of a noncyclic phosphate group on yUsb1-processed RNA (**Figure 3-1B**). Thus, yUsb1 directly catalyzes the formation of a non-cyclic phosphate, and U6 3' end processing does not require a *trans* acting 2',3'-cyclic phosphodiesterase in yeast. Yeast Usb1 catalyzes two distinct chemical reactions: (1) exonucleolytic removal of a terminal uridine (“first step”) and (2) cyclic phosphodiesterase (CPDase) ring opening to leave a non-cyclic phosphate (“second step”).

To unambiguously identify whether yUsb1 leaves a 2' or 3' phosphate, we investigated the CPDase activity of yUsb1 in isolation from its exonuclease activity using NMR spectroscopy

(**Figure 3-1C**). The ^{31}P chemical shift for a 2',3'-cyclic phosphate is ~20 ppm, whereas non-cyclic 2' and 3' phosphates of uridine monophosphate (UMP) have unique and well-resolved chemical shifts between 3-3.5 ppm (**Figure 3-1C**). When 2',3'-cUMP is incubated with yUsb1, a new peak at 3.4 ppm is formed that corresponds precisely to the chemical shift of ^{31}P in a 3'UMP standard (**Figure 3-1C**)[26]. Additionally, the H6 resonance of uracil is well-documented to be highly sensitive to the position of the phosphate at the 2' vs. 3' position[26,27] and further confirms that the product is a 3' phosphate (**Figure 3-3A**). Finally, two-dimensional ^{1}H-^{31}P heteronuclear multiple bond correlation (HMBC) and ^{1}H-^{1}H correlation spectroscopy (COSY) spectra unambiguously show that the product is a 3' phosphate (**Figure 3-3B,C**). These data demonstrate that yUsb1 has CPDase activity that catalyzes the formation of a terminal 3' phosphate, unlike metazoan Usb1, which lacks CPDase activity altogether[10,11].

To determine how yUsb1 can occasionally remove more than one nucleotide, we tested its ability to process RNA substrates with different 3' ends. yUsb1 is less efficient on 2' phosphate-terminated substrates and is inactive on RNAs with a terminal 3' phosphate (**Figure 3-1D**). Substrates terminating in a 2',3'-cyclic phosphate are processed by Usb1 with slower kinetics and to a lesser extent (**Figure 3-1D**, lane 6-8). These data demonstrate that yUsb1 is incapable of further processing its dominant 3' phosphate product and that successive (*n-2* and *n-3*) products are likely formed from a cyclic phosphate intermediate that is reprocessed prior to second step CPDase chemistry. The amount of slower mobility *n-1* cyclic phosphate product is consistent with the relative kinetics of *n-2* product formation (**Figure 3-1D**). The presence of both exonuclease and CPDase activity, along with inactivity on 3' phosphate-terminated substrates, reveals an elegant mechanism for ensuring that yUsb1 does not over process and degrade U6 RNA.

Figure 3-1. yUsb1 acts as a 3'-5' exonuclease and CPDase *in vitro*.

A) U6 snRNA is synthesized by RNA Polymerase III. Transcription termination produces a heterogeneous U6 with a 5-8 nucleotide U-tail. Processing by yUsb1 shortens the U-tail and leaves a phosphoryl group.

B) Usb1 removes nucleotides from the 3' end of RNAs. The 5' labelled U6 95-112+3U oligonucleotide cis diol substrate (lane 2) is insensitive to CIP (lane 3) or T4 PNK (lane 4) treatment. Incubation with yUsb1 for one hour results in a shorter product (lane 5). Similar reactivity of the product to both CIP (lane 6) and T4 PNK (lane 7) indicates that the product is a non-cyclic phosphate. An alkaline hydrolysis ladder (lane 1) shows the mobility of oligonucleotide products of different lengths.

C) Top: One-dimensional ^{31}P NMR spectra of 2',3'-cUMP shows a single peak at 20 ppm. A 3'-UMP standard has a single peak at 3.4 ppm. When 2',3'-cUMP is incubated with AtRNL, which leaves a 2' phosphate[8], there is a single peak at 3.2 ppm. Incubation of 2',3'-cUMP with yUsb1 produces a new signal at 3.4 ppm. Bottom: zoom of dashed region in top panel.

D) Time course of Usb1 processing on RNAs with different 3' end modifications. yUsb1 is most active on RNA substrates with a cis diol (lanes 1-4), less active on those with a 2',3'-cyclic phosphate (>p; lanes 5-8) or 2' phosphates (2'P; lanes 9-12), and is inactive on 3' phosphate ends (3'P; lane 13-16).

E) Model describing the dual activities of yUsb1.

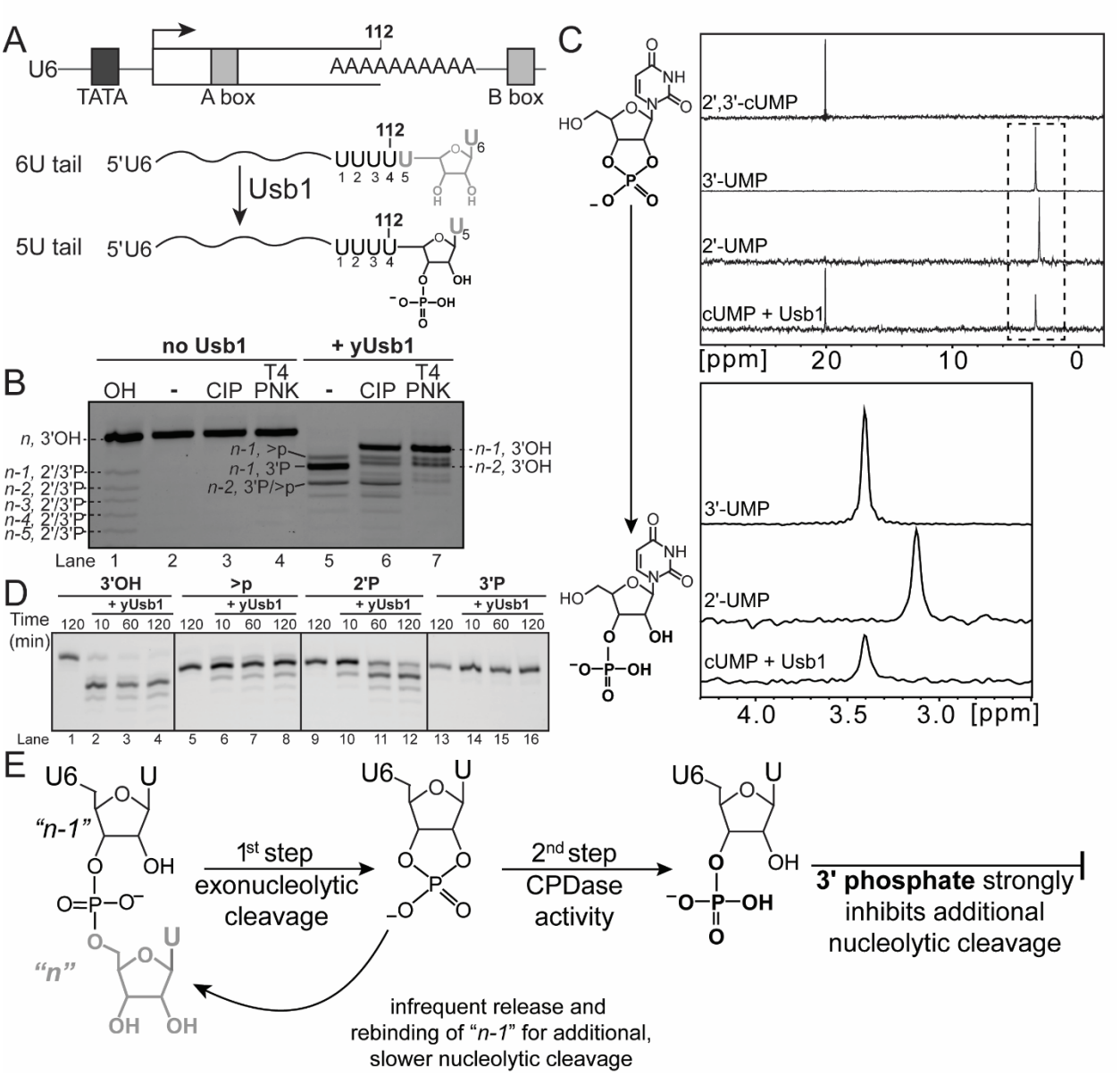

A
U6
TATA
A box
112
AAAAAAAAAA
B box
6U tail
5'U6
UUUUU
1 2 3 4 5
Usb1
5U tail
UUUU
1 2 3 4
B
no Usb1
+ yUsb1
OH
-
CIP
T4 PNK
n, 3'OH
n-1, 2'/3'P
n-2, 2'/3'P
n-3, 2'/3'P
n-4, 2'/3'P
n-5, 2'/3'P
n-1, >p
n-1, 3'P
n-2, 3'P/>p
n-1, 3'OH
n-2, 3'OH
Lane 1 2 3 4 5 6 7
C
2',3'-cUMP
3'-UMP
2'-UMP
cUMP + Usb1
[ppm] 20 10 0
4.0 3.5 3.0 [ppm]
D
3'OH
>p
2'P
3'P
+ yUsb1
Time (min)
120 10 60 120
Lane 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
E
"n-1"
"n"
1st step
exonucleolytic cleavage
2nd step
CPDase activity
3' phosphate strongly inhibits additional nucleolytic cleavage
infrequent release and rebinding of "n-1" for additional, slower nucleolytic cleavage

Figure 3-2. Usb1 is inactive on substrates with a deoxyuridine adjacent to the scissile phosphate.

A) Time course of Usb1 processing on substrates with three additional uridines (U6-7U, lanes 3-9) or where the *n-1* nucleotide is a deoxyuridine (U6-dU, lanes 10-16) indicate that yUsb1 processing is exonucleolytic and depends on a 2' hydroxyl. Alkaline hydrolysis ladders of the RNAs (lanes 1-2) show the mobility of oligonucleotide products of different length.

B) Time course of Usb1 processing on substrates with three additional uridines (U6-7U, lanes 3-9) or six additional uridines (U6-10U, lanes 10-16) indicate that yUsb1 processing is independent of U-tail length. Alkaline hydrolysis ladders of the RNAs (lanes 1-2) show the mobility of oligonucleotide products of different length.

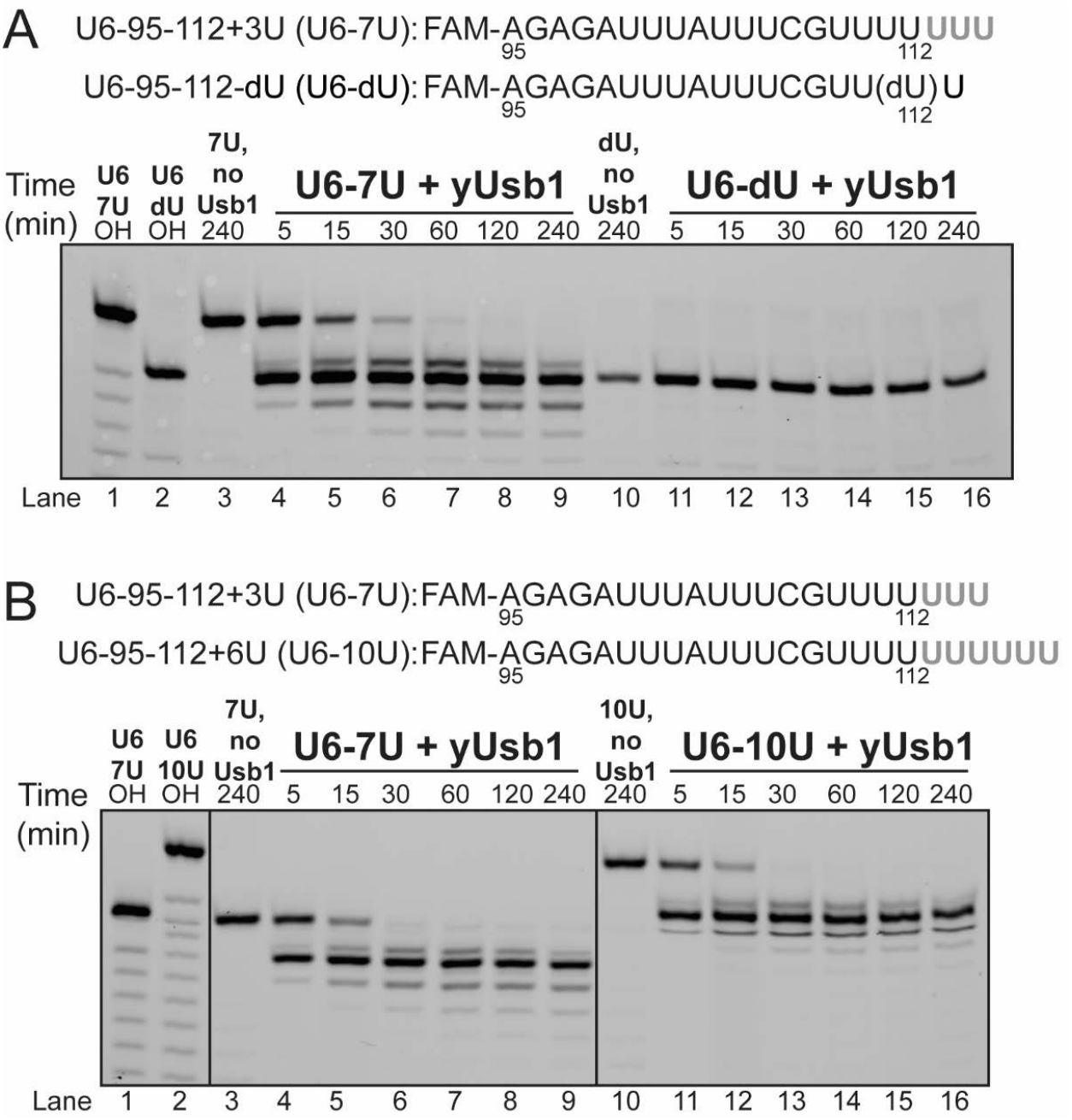
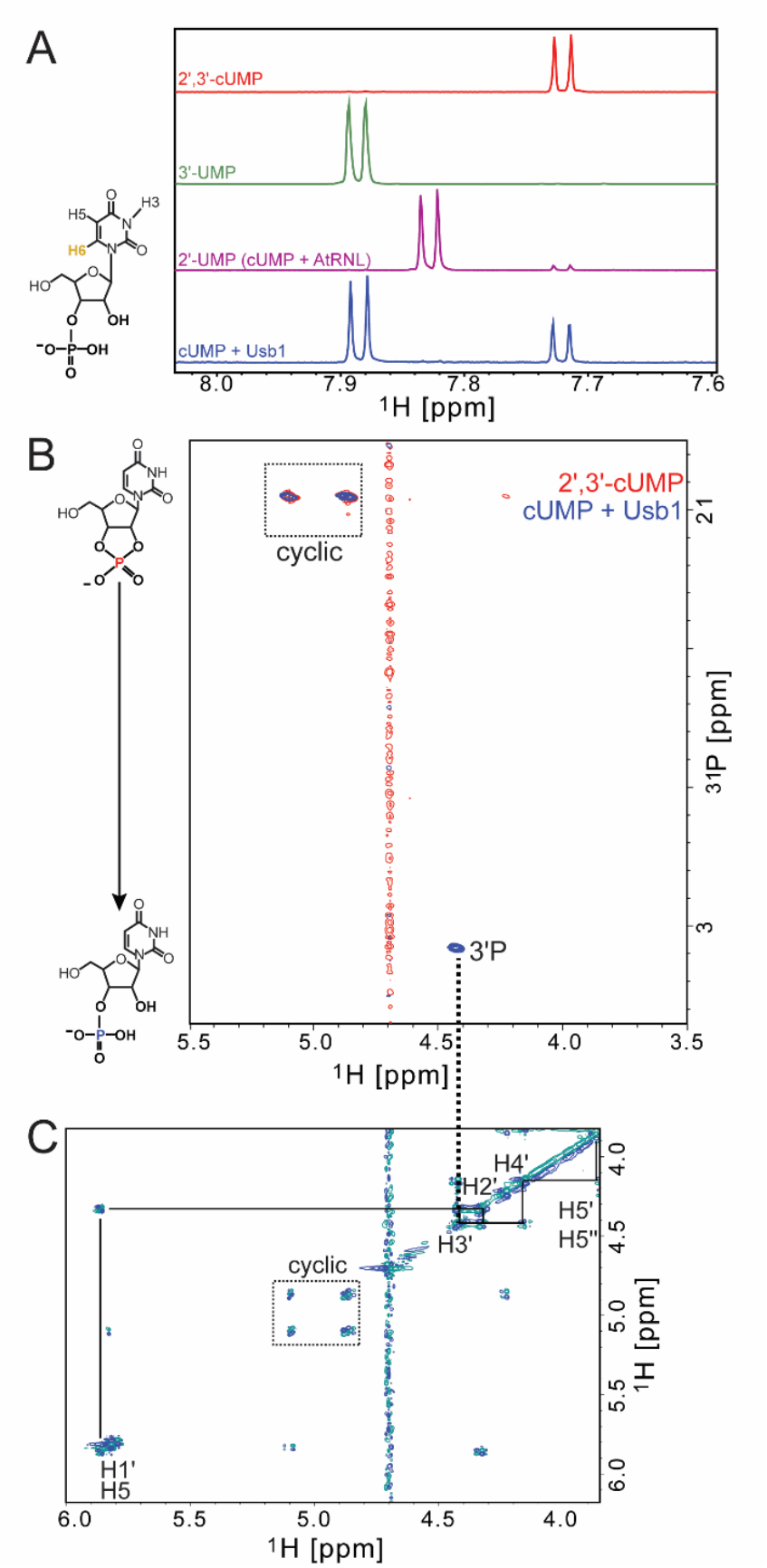

A
U6-95-112+3U (U6-7U): FAM-AGAGAUUUAUUUCGUUUUUUU
95
112
U6-95-112-dU (U6-dU): FAM-AGAGAUUUAUUUCGUU(dU)U
95
112
Time (min)
U6 7U OH
U6 dU OH
7U, no Usb1 240
U6-7U + yUsb1
5 15 30 60 120 240
dU, no Usb1 240
U6-dU + yUsb1
5 15 30 60 120 240
Lane 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
B
U6-95-112+3U (U6-7U): FAM-AGAGAUUUAUUUCGUUUUUUU
95
112
U6-95-112+6U (U6-10U): FAM-AGAGAUUUAUUUCGUUUUUUUUUU
95
112
Time (min)
U6 7U OH
U6 10U OH
7U, no Usb1 240
U6-7U + yUsb1
5 15 30 60 120 240
10U, no Usb1 240
U6-10U + yUsb1
5 15 30 60 120 240
Lane 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Figure 3-3. NMR of Usb1-treated 2',3'-cUMP reveals Usb1 has CPDase activity and leaves a 3' phosphate.

A) 1D-^{1}H spectra focusing on the ^{1}H chemical shift region for uracil H6 protons. 2',3'-cyclic UMP (red) was incubated with AtRNL (purple) or Usb1 (blue). The spectra of a 3'UMP standard is also shown. The H6 proton peaks are split by proton-proton coupling between H5 and H6.

B) 2D-^{1}H-^{31}P HMBC overlay of 2',3'-cUMP before (red) and after incubation with AtRNL (purple) or Usb1 (blue). (c) 2D-^{1}H-^{1}H COSY of 2',3'-cUMP after incubation with Usb1.


A
2',3'-cUMP
3'-UMP
2'-UMP (cUMP + AtRNL)
cUMP + Usb1
H5
H3
H6
HO
OH
8.0
7.9
7.8
7.7
7.6
1H [ppm]
B
2',3'-cUMP
cUMP + Usb1
cyclic
3'P
21
3
31P [ppm]
5.5
5.0
4.5
4.0
3.5
1H [ppm]
C
H4'
H2'
H5'
H5"
H3'
cyclic
H1'
H5
4.0
4.5
5.0
5.5
6.0
1H [ppm]

3.3.2 The conserved architecture and active site of Usb1

We determined the crystal structure of the catalytic domain of yUsb1 (a.a. 71-290) to 1.8 Å resolution (**Figure 3-4A** and **Table 3-1**). The enzyme exhibits a typical 2H phosphodiesterase fold[28] with an active site containing two H-X-S motifs (**Figure 3-4B**). yUsb1 crystallized in the presence of 2 M ammonium sulfate, and the resulting structure contained a sulfate ion coordinated within the active site which likely mimics the binding mode of the scissile phosphate (**Figure 3-4B** and **Figure 3-5**). NE2 of H231 is 2.8 Å from the sulfate ion, while H133 is farther away (4 Å). Residues 107-115 could not be modeled in our structure and are presumed to be disordered. Despite low sequence identity (<20%), the structure of yUsb1 is strikingly homologous to that of human Usb1[11], with nearly superimposable active sites (rmsd for H/S in active site = 0.3 Å).

Figure 3-4. Structure of yUsb1 and structure of hUsb1 with a substrate analog bound in the active site.

A) Structure of yUsb1 at 1.8 Å. Residues 1-70 were excluded from the crystallizable construct as they were predicted to be unstructured. The two active site H-x-S motifs are shown in ball-and-stick.

B) Active site of yUsb1. A sulfate ion is coordinated in the center of the H-x-S motifs. A water molecule is coordinated by histidine 133 and serine 135.

C) Structure of hUsb1 with 5'UMP in the active site (blue; 5'UMP in yellow) is similar to the structure of yUsb1 (green). Residues 1-78 were truncated to facilitate crystallization. The two active site H-x-S motifs are shown in sticks.

D) Active site of hUsb1. 5'UMP is coordinated in the center of the H-x-S motifs. The 5' phosphate is positioned similarly to the sulfate ion in the yUsb1 structure (b). Tyrosine 202 makes a stacking interaction with the uracil of 5'UMP. Hydrogen bonds between histidine 208 and the O5' oxygen and between serine 122, histidine 120, and serine 210 and the phosphate hold 5'UMP in the active site.

E) Time course comparing yUsb1 (lanes 1-5) and hUsb1 (lanes 6-10) activity. An alkaline hydrolysis ladder (lane 11) shows the mobility of oligonucleotide products of different length.

F) Single time-point (60 minute) comparison of yUsb1 (y) and hUsb1 (h) on RNA substrates with different 3' modifications. Both yUsb1 and hUsb1 are most active on 2',3'-cis diol RNAs (lanes 1-3), less active on cyclic phosphate (lanes 4-6) and 2' phosphate (lanes 7-9) RNAs, and inactive on 3' phosphate RNAs.

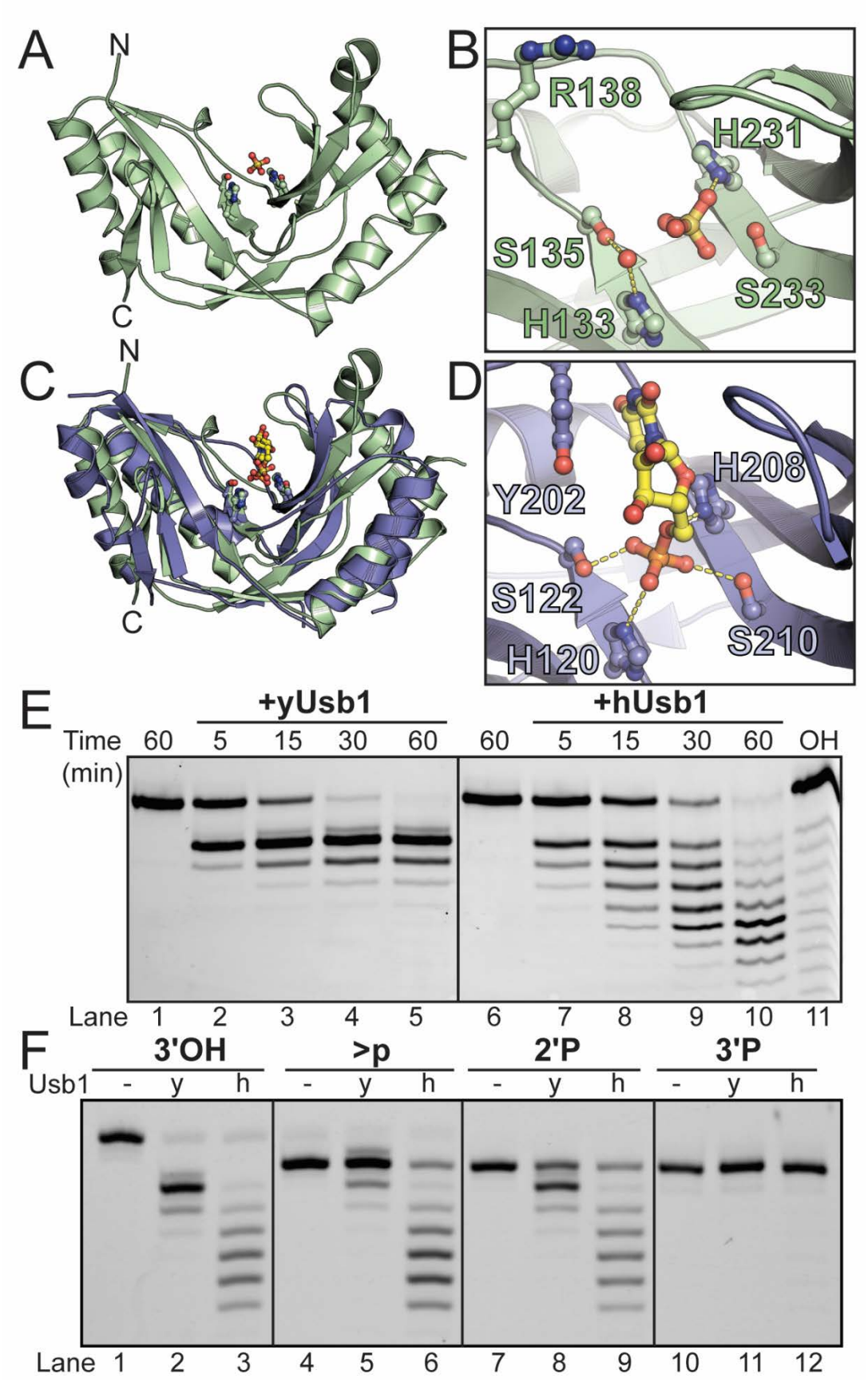

A
N
C
B
R138
H231
S135
S233
H133
C
N
C
D
H208
Y202
S122
S210
H120
E
+yUsb1
+hUsb1
Time (min)
60
5
15
30
60
60
5
15
30
60
OH
Lane
1
2
3
4
5
6
7
8
9
10
11
F
3'OH
>p
2'P
3'P
Usb1
-
y
h
-
y
h
-
y
h
-
y
h
Lane
1
2
3
4
5
6
7
8
9
10
11
12

Figure 3-5. Active site densities and annealed omit map.

A) Stereo view of the $2mF_o$-DF_c electron density map of the yUsb1 active site, contoured at 1σ.

B) Stereo view of the $2mF_o$-DF_c electron density map of the hUsb1/5'UMP active site, contoured at 1σ.

C) Annealed omit map of the hUsb1/5'UMP active site. The occupancy of the bound ligand was refined to a value of 0.72 in phenix.refine. The residual F_o-F_c density near to the C5 atom overlaps with a bound water molecule in the ligand-free structure reported previously[11] (PDB 4H7W).

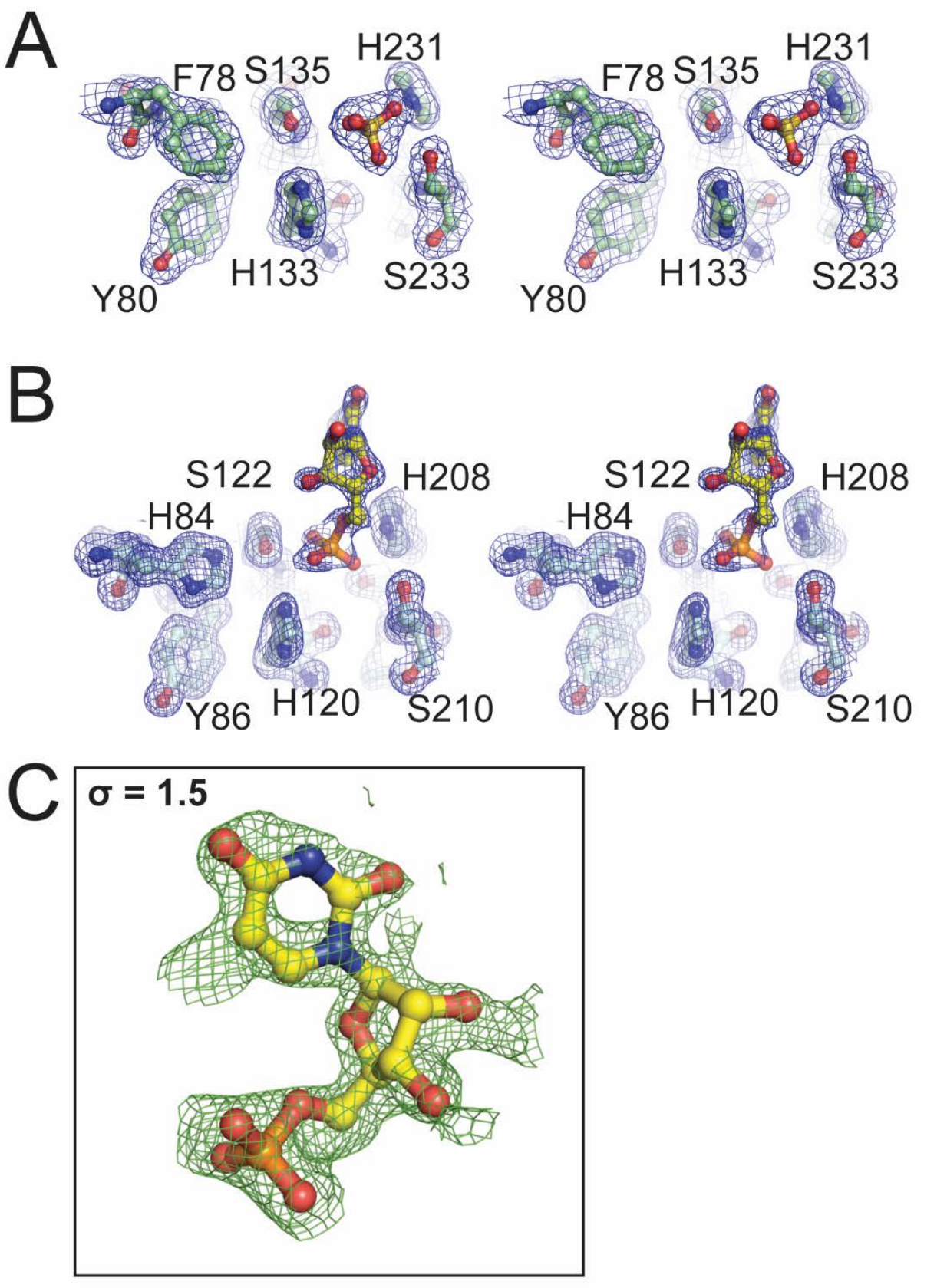

A
F78
S135
H231
H133
S233
Y80
F78
S135
H231
H133
S233
Y80
B
S122
H208
H84
Y86
H120
S210
S122
H208
H84
Y86
H120
S210
C
σ = 1.5

Table 3-1. Data collection and refinement statistics.

	hUsb1 + 5'UMP PDB 5V1M	yUsb1	yUsb1 phasing
Data Collection			
Wavelength (Å)	0.9795	0.9786	0.9786
Resolution range (Å)*	53.2-1.47 (1.50-1.47)	78.9-1.80 (1.84-1.80)	81.9-2.00 (2.05-2.00)
Space group	$P2_1$	$I4_122$	$I4_122$
Unit cell dimensions (Å)	42.4, 53.2, 46.6 β = 106.95°	157.9, 157.9, 44.4	163.6, 163.6, 43.8
Total reflections*	224,991 (10,848)	382,211 (21,365)	597,214 (42,693)
Unique reflections*	33,607 (1,621)	26,187 (1,517)	20,480 (1,464)
Multiplicity*	6.7 (6.7)	14.6 (14.1)	29.2 (29.2)
Completeness (%)*	99.3 (99.8)	99.4 (98.0)	100 (100)
Mean I/σ(I)*	12.8 (1.2)	23.3 (1.4)	13.8 (1.1)
Wilson B-factor	16.3	28.6	49.9
R-merge*	0.08 (1.51)	0.09 (2.23)	0.25 (4.84)
$CC_{1/2}$*	0.999 (0.581)	1.000 (0.446)	0.999 (0.457)
Refinement			FOM = 0.58
R_{work}/R_{free}*	0.16/0.19 (0.30/0.36)	0.19/0.23 (0.34/0.39)	
Total number of atoms	3,414	1,897	
macromolecules	3,125	1,746	
ligands	131	35	
water	158	116	
RMS(bonds)	0.023	0.018	
RMS(angles)	2.062	1.681	
Ramachandran favored	98.4 %	95.7 %	
Ramachandran outliers	0 %	0.96 %	
Average B factor (Å^2)	28.1	36.1	
protein	26.9	35.6	
ligands/ions	42.9	46.7	
solvent	38.7	41.2	

*Values shown in parentheses are for the highest resolution shell.

3.3.3 Structure of human Usb1 with a substrate analog

To understand how Usb1 recognizes RNA and catalyzes 3' end processing, we sought to determine the co-crystal structure of Usb1 with a substrate analog. We obtained a structure of truncated hUsb1 (a.a. 78-265) with uridine 5'-monophosphate (5'UMP) at 1.4 Å resolution (**Figure 3-4C**), using the same crystal form as Hilcenko *et al*.[11], and incorporation of ligand via soaking of 10 mM 5'UMP at pH ~ 6.5. In the structure of the apo enzyme at pH 5.6 (PDB ID 4H7W), the active site histidine H120 appears in two conformations, with one conformation tilted away from the active site and a proximal conformation stabilized by the coordination of a water molecule[11]. In the 5'UMP-containing structure, H120 adopts the proximal conformation, and is 3.1 Å from an oxygen of the phosphate group in 5'UMP (**Figure 3-4D**). 5'UMP is held in place by hydrogen bonding interactions with active site residues H120, S122, S210 and H208, and a stacking interaction with Y202. The 5' phosphate of 5'UMP is positioned near the center of the active site, close to the position of the sulfate ion in the yeast active site which would correspond to the scissile phosphate that is involved in the first exonucleolytic step. We note that the hydrogen bonding interaction between H208 and the O5' oxygen is positioned for general acid catalysis (**Figure 3-4D**), consistent with an enzymatic mechanism proposed elsewhere[11]. We hypothesize that the binding mode of 5'UMP in our structure of hUsb1 is analogous to how the last nucleotide of an RNA substrate would be coordinated within the active site.

3.3.4 Comparison of human and yeast Usb1 enzymatic activities

We further characterized the enzymatic activities of human and yeast Usb1 in order to better understand how they differ with respect to U6 processing. hUsb1 removes multiple nucleotides over the course of an hour (**Figure 3-4E**) as observed previously[11]. Interestingly, hUsb1 is also strongly inhibited by a 3' phosphate terminated RNA (**Figure 3-4F** and ref. 11). Thus, the mechanistic underpinning for the ability of hUsb1 to remove multiple nucleotides from U6 arises from its inability to catalyze “second step” cyclic phosphate ring opening. The structure of hUsb1 with 5'UMP shows that RNAs with 3' modifications would be sterically occluded by a

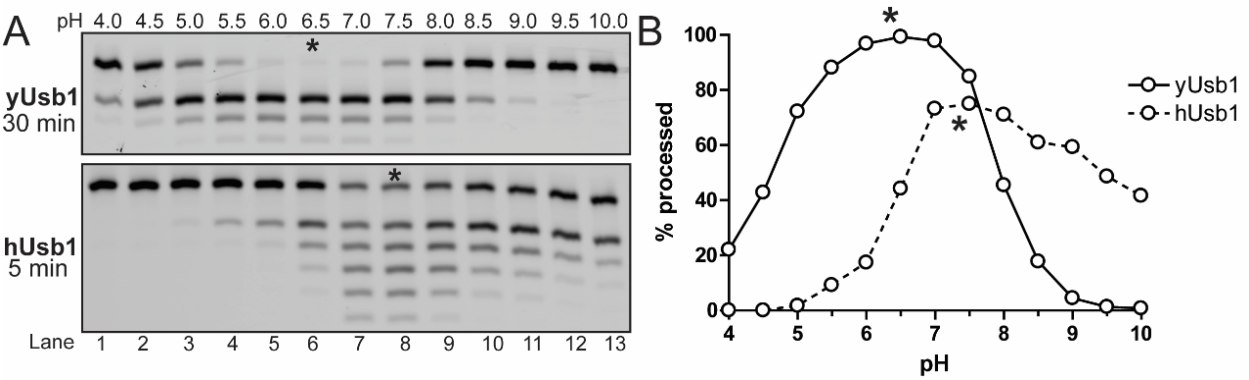
loop spanning residues 161-167, suggesting that this region of the protein may contribute to substrate specificity.

Human Usb1 was less active in our assay conditions (pH 6.5) than in previous assays conducted at pH 8.0[11]. To reconcile this difference, we monitored the pH dependence of human Usb1 and found that it had a pH optimum of 7.5, in contrast to yeast Usb1, which exhibited a significantly different activity range and a pH optimum of 6.5 (**Figure 3-6A,B**). yUsb1 remains partially active even at pH 4, whereas hUsb1 is completely inactive. In contrast, at pH 10, yUsb1 is completely inactive, whereas hUsb1 retains activity. This difference in pH optimum suggests that the active site histidine residues in yeast and hUsb1 have markedly different pK_a's, likely due to different active site microenvironments. For example, hUsb1 has a histidine (H84) adjacent to the active site that can hydrogen bond to the active site serine S122[11]. In yeast, this position is a phenylalanine (F78) which cannot form an analogous hydrogen bond.

Figure 3-6. yUsb1 and hUsb1 have different pH optimums.

A) yUsb1 and hUsb1 have different pH optimums (indicated by an asterisk) and different tolerances for low and high pH. Single time-point comparison of yUsb1 (top; 30 minute time point) and hUsb1 (bottom; 5 minute time point) in conditions with pH ranging from 4.0 - 10.0. yUsb1 is most active at pH ~6.5 and is active over a range from pH 4.0 - 8.5. hUsb1 is most active at pH ~7.5 and is active over a range from pH 5.5 - 10.0.

B) Quantification of (a) showing the amount of substrate depleted as a function of pH.



3.3.5 A residue adjacent to the active site influences activity

As yeast and human Usb1 exhibit remarkably similar H-X-S active sites, we inspected residues surrounding the active site for additional explanations as to how yeast and human Usb1 exhibit divergent enzymatic behaviors. We observed that Phe78 in yeast and His84 in human Usb1 are structurally homologous, suggesting a possible role for influencing water nucleophilicity during second step chemistry, or modulation of active site pK_a (see above section) (**Figure 3-7A**). With the exception of *S. pombe* Usb1*,* aromaticity (but not amino acid identity) is conserved at this position in other organisms (**Figure 3-7B**). We therefore asked if mutations in yUsb1 at position 78 would have a phenotype in *S. cerevisiae*. Indeed, mutating F78 to alanine causes a slow growth phenotype *in vivo* (**Figure 3-7C**) and a >300-fold reduction in the rate of processing *in vitro* (**Figure 3-7D**). Surprisingly, F78H has no observable growth phenotype (**Figure 3-7C**), yet has a 47-fold reduction in *in vitro* processing rate (**Figure 3-7D**). Thus, significant reductions in Usb1 activity but not complete loss of the protein (**Table 3-2**)[11] are tolerated by yeast.

Since mutation of F78 to histidine (as in hUsb1) supported yeast growth, we next asked if hUsb1 could complement deletion of *USB1* in yeast. When expressed from a high-copy plasmid using a GPD promoter, wild-type hUsb1 allows for an extreme slow growth phenotype, with small colonies visible only after >3 days (**Figure 3-7E**). We substituted H84 in hUsb1 for the phenylalanine found in that position in yUsb1 and determined if that improved yeast proliferation. Indeed, hUsb1-H84F rescues yeast growth, whereas H84A does not, suggesting that H84F is a true gain-of-function mutant (**Figure 3-7E**). Thus, mutations at this position greatly affect both *in vitro* processing rate and *in vivo* viability. Conversely, substitution of many perfectly conserved residues outside of the active site (i.e. yUsb1 Y80) has no effect on yeast viability (**Table 3-2**).

Figure 3-7. Residue F78 influences RNA processing by yUsb1.

A) Overlay of yUsb1 and hUsb1 active sites shows that yUsb1-F78 and hUsb1-H84 are positioned similarly adjacent to the active site.

B) Sequence alignment of the region surrounding yUsb1-F78/hUsb1-H84. Several residues (yUsb1 W75 and Y80) are perfectly or near-perfectly conserved, while conservation of position F78 is limited to aromaticity.

C) Usb1 mutants F78A and F78H can complement genomic deletion of *USB1* using plasmid shuffle and assaying for growth on media containing 5-FOA. Usb1-F78A results in a growth defect.

D) Quantification of the rate of Usb1 processing on a fluorescent substrate (as in Figure 3-1) with different substitutions at position 78. Plotted data points represent the average of three technical replicates ± s.d.

E) Overexpression of yUsb1 (under control of the GPD promoter) complements deletion of *USB1*, but overexpression of hUsb1 results in an extreme slow growth phenotype. Overexpression of hUsb1-H84F partially rescues growth, while hUsb1-H84A does not.

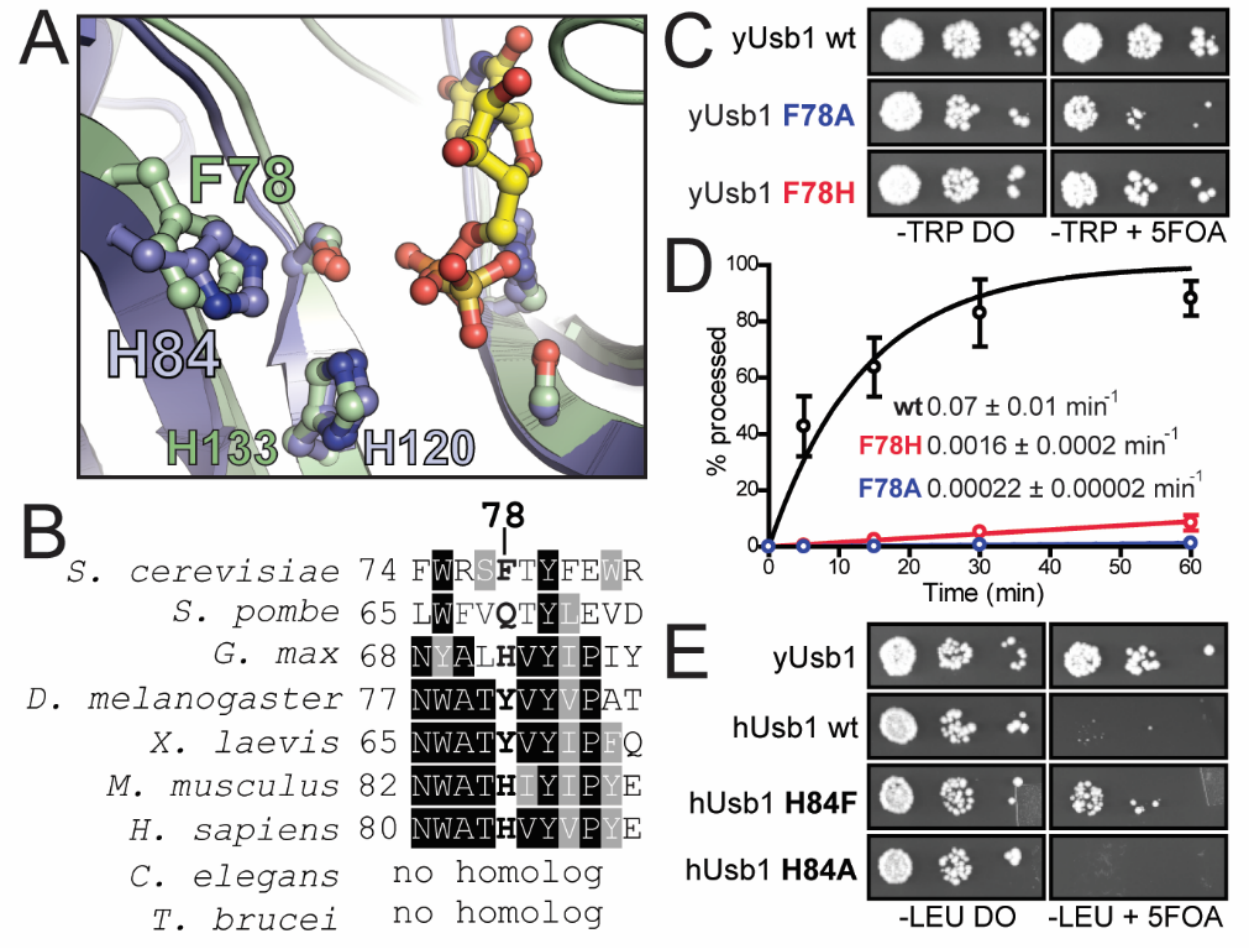

A
F78
H84
H133
H120
B
78
S. cerevisiae 74 FWRSFTYFEWR
S. pombe 65 LWFVQTYLEVD
G. max 68 NYALHVYIPIY
D. melanogaster 77 NWATYVYVPAT
X. laevis 65 NWATYVYIPFQ
M. musculus 82 NWATHIYIPYE
H. sapiens 80 NWATHVYVPYE
C. elegans no homolog
T. brucei no homolog
C
yUsb1 wt
yUsb1 F78A
yUsb1 F78H
-TRP DO
-TRP + 5FOA
D
% processed
Time (min)
wt 0.07 ± 0.01 min-1
F78H 0.0016 ± 0.0002 min-1
F78A 0.00022 ± 0.00002 min-1
E
yUsb1
hUsb1 wt
hUsb1 H84F
hUsb1 H84A
-LEU DO
-LEU + 5FOA

Table 3-2. Summary of the effect of mutations and truncations of yUsb1 *in vivo*.

Mutations or truncations were introduced into pRS414-Usb1 and plated on media lacking TRP and containing 5-FOA. Ability to complement loss of the *URA3* marked wild-type Usb1 plasmid is listed in the right hand column, where "-" indicates no growth and "+++" indicates growth similar to wild-type Usb1. * indicates strains that were not tested by serial dilution.

USB1 allele	*USB1Δ* complementation
pRS414 backbone	
Vector	-
wild type	+++
Point mutations	
W75A	+++
F78A	+
F78H	+++
Y80A	+++
H133A*	-
H133N*	-
H231A*	-
H231N*	-
N-terminal truncations	
usb1Δ1-20	+++
usb1Δ1-41	+++
usb1Δ1-42	+++
usb1Δ1-43	+++
usb1Δ1-44	-
usb1Δ1-45	-
usb1Δ1-46	-
usb1Δ1-47	-
usb1Δ1-48	-
usb1Δ1-49	-
usb1Δ1-50	-
usb1Δ1-50 + SV40 NLS	-
usb1Δ1-70 + SV40 NLS	-

3.3.6 The N-terminus of Usb1 is essential for yeast viability

Orthologs of Usb1 contain an N-terminal region that is predicted to be disordered[29,30]. Surprisingly, the N-terminal region is more conserved across species than the catalytic domain (**Figure 3-8A**). Although the length of the domain is highly variable, there are two regions that are conserved in sequence: a $Y(S)_N[D/E]$ motif in the first 20 amino acids, and a proline-rich region in the center of the domain. The first serine-rich motif could be a site for post-translational modification, while the second, proline-rich region could be important for a protein-protein interaction, stability, or expression of Usb1, as proline-rich regions are important for many protein-protein interactions[31]. In this region of ~20 amino acids, ~20% of the residues (and up to 35% in *G. max* Usb1) are prolines (**Figure 3-8A**). We asked what role, if any, the N-terminal domain plays in catalysis or function of yUsb1 *in vivo*.

We first compared the *in vitro* activity of the catalytic domain of yUsb1 (a.a. 71-290) to that of full-length yUsb1. The N-terminal region (residues 1-70) has no observable effect in our *in vitro* exonuclease assay as the rate and extent of processing for full-length and the catalytic domain of yUsb1 are indistinguishable (**Figure 3-8B**). Hilcenko *et al.*[11] previously reported that an N-terminal deletion of yUsb1 (Usb1 77-290) failed to complement *USB1Δ* in yeast. Our crystal structure reveals that residues 73-76 form the start of a β-strand and make several intramolecular contacts that are likely important for folding. However, we find that yUsb1 71-290 (Δ70) is also insufficient to support yeast growth (**Figure 3-8C**). This result is surprising, because the catalytic domain is fully active *in vitro* and because the N-terminal region (residues 1-70) is predicted to be largely disordered. By making successive truncations (**Table 3-2**), we found that residues 43-290 are essential for viability (**Figure 3-8C**). This essential fragment correlates with the start of the proline-rich region (**Figure 3-8A**). The inviability of Δ70 or Δ50 alleles of *USB1* could not be rescued by inclusion of an N-terminal SV40 nuclear localization signal (NLS) (**Table 3-2**). This suggests that the role of the N-terminus is more complex than controlling subcellular localization of Usb1.

When full-length and truncations of yUsb1 are expressed under control of a GPD promoter to achieve greater expression levels, we find that more extensive truncations of the N-terminus can complement *USB1Δ*; however, the Δ70 allele is still lethal (**Figure 3-9A**). Western blot analysis shows that truncation of the N-terminal region of yUsb1 results in reduced levels of Usb1 (**Figure 3-9B**). Deletion of the first 42 or 44 residues results in a significantly reduced signal from Usb1, and deletion of the first 70 amino acids results in an undetectable amount of Usb1, suggesting that the N-terminus of Usb1 plays a role in protein stability.

Taking advantage of the viable hUsb1-H84F mutation, we discovered that the N-terminal region of hUsb1 (residues 1-78) is also required for growth (**Figure 3-8D**). This result suggests that the N-terminal domain may possess a conserved function. How the N-terminal domain influences stability, whether through an as-yet unidentified interaction or through another mechanism, remains to be determined.

Figure 3-8. Usb1 catalytic activity alone is not sufficient for yeast viability.

A) Sequence alignment of N-terminal regions of Usb1. While the length of the N-terminus is poorly conserved (ranging from between 46-103 residues in length), two regions of sequence conservation exist near the N-terminus (with a conserved Y-poly S-[D/E] motif) and in the center of the N-terminal region (the proline-rich region). Proline composition of this region is indicated to the right of the alignment.

B) Time course comparing the activities of full-length Usb1 (lanes 2-9) and the catalytic domain (residues 71-290) (lanes 10-17) over the course of 18 hours shows similar rate and extent of processing for both proteins. Alkaline hydrolysis ladders (lanes 1 and 18) show the mobility of oligonucleotide products of different length.

C) N-terminal truncated forms of Usb1 show different capacity to support complementation of *USB1Δ*. Removing the first 42 residues supports growth, whereas removing 44 residues or more does not.

D) hUsb1-H84F does not complement *USB1Δ* if the homologous N-terminal region (Δ78) is deleted.

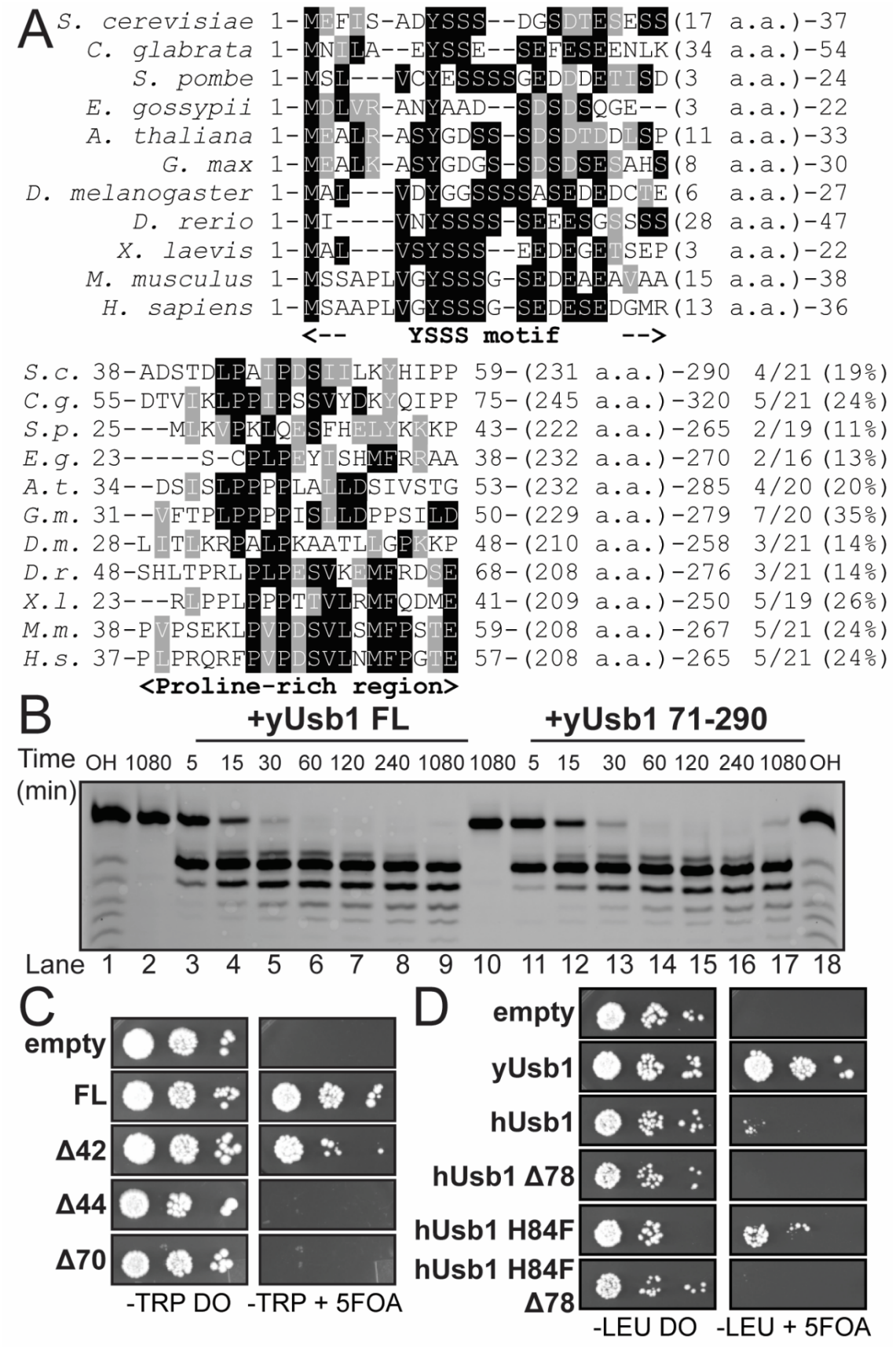
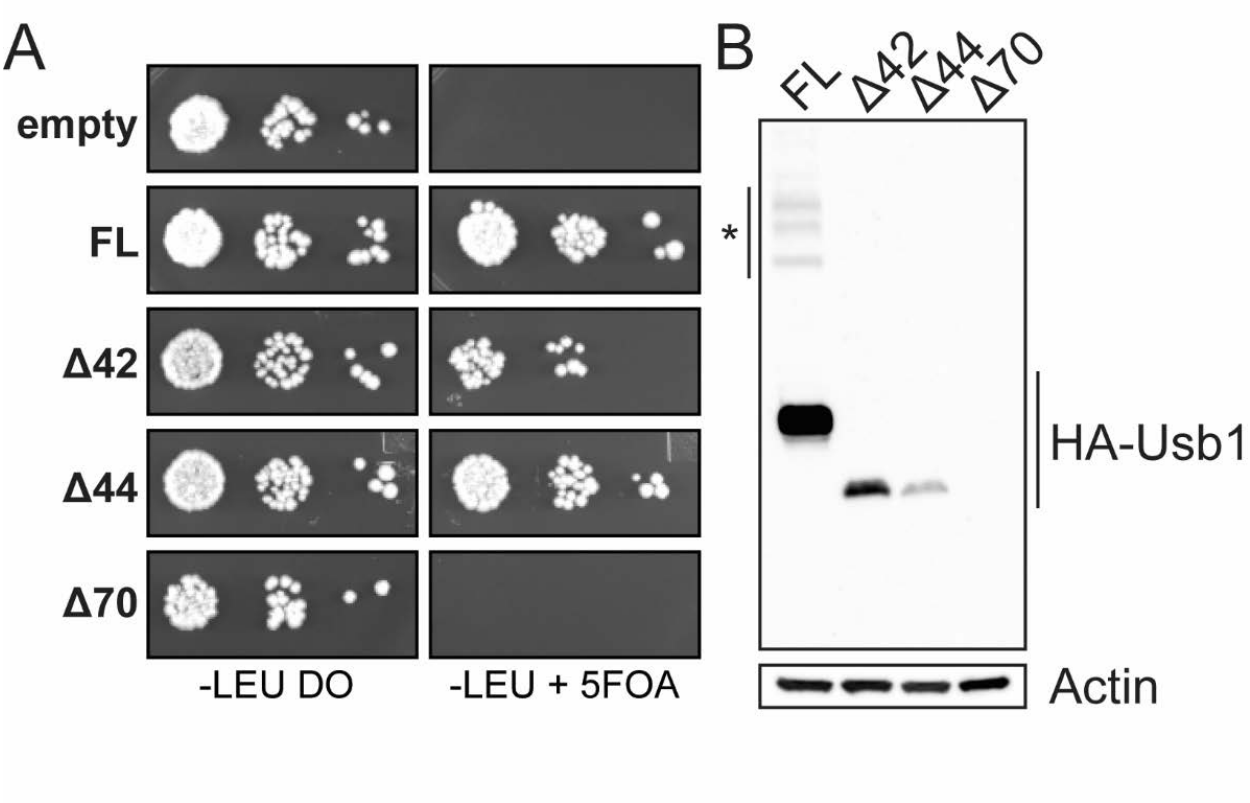

A
S. cerevisiae 1-MEFIS-ADYSSS--DGSDTESESS(17 a.a.)-37
C. glabrata 1-MNILA--EYSSE--SEFESEENLK(34 a.a.)-54
S. pombe 1-MSL---VCYESSSSGEDDDETISD(3 a.a.)-24
E. gossypii 1-MDLVR-ANYAAD--SDSDSQGE--(3 a.a.)-22
A. thaliana 1-MEALR-ASYGDSS-SDSDTDDLSP(11 a.a.)-33
G. max 1-MEALK-ASYGDGS-SDSDSESAHS(8 a.a.)-30
D. melanogaster 1-MAL---VDYGGSSSSASEDEDCTE(6 a.a.)-27
D. rerio 1-MI----VNYSSSS-SEEESGSSSS(28 a.a.)-47
X. laevis 1-MAL---VSYSSS--EEDEGETSEP(3 a.a.)-22
M. musculus 1-MSSAPLVGYSSSG-SEDEAEAVAA(15 a.a.)-38
H. sapiens 1-MSAAPLVGYSSSG-SEDESEDGMR(13 a.a.)-36
<-- YSSS motif -->
S.c. 38-ADSTDLPAIPDSIILKYHIPP 59-(231 a.a.)-290 4/21 (19%)
C.g. 55-DTVIKLPPIPSSVYDKYQIPP 75-(245 a.a.)-320 5/21 (24%)
S.p. 25---MLKVPKLQESFHELYKKKP 43-(222 a.a.)-265 2/19 (11%)
E.g. 23-----S-CPLPEYISHMFRRAA 38-(232 a.a.)-270 2/16 (13%)
A.t. 34--DSISLPPPPLALLDSIVSTG 53-(232 a.a.)-285 4/20 (20%)
G.m. 31--VFTPLPPPPISLLDPPSILD 50-(229 a.a.)-279 7/20 (35%)
D.m. 28-LITLKRPALPKAATLLGPKKP 48-(210 a.a.)-258 3/21 (14%)
D.r. 48-SHLTPRLPLPESVKEMFRDSE 68-(208 a.a.)-276 3/21 (14%)
X.l. 23---RLPPLPPPTTVLRMFQDME 41-(209 a.a.)-250 5/19 (26%)
M.m. 38-PVPSEKLPVPDSVLSMFPSTE 59-(208 a.a.)-267 5/21 (24%)
H.s. 37-PLPRQRFPVPDSVLNMFPGTE 57-(208 a.a.)-265 5/21 (24%)
<Proline-rich region>
B
+yUsb1 FL
+yUsb1 71-290
Time (min)
OH 1080 5 15 30 60 120 240 1080 1080 5 15 30 60 120 240 1080 OH
Lane 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
C
empty
FL
Δ42
Δ44
Δ70
-TRP DO
-TRP + 5FOA
D
empty
yUsb1
hUsb1
hUsb1 Δ78
hUsb1 H84F
hUsb1 H84F Δ78
-LEU DO
-LEU + 5FOA

Figure 3-9. The N-terminal domain is important for yUsb1 expression or stability.

A) Serial dilutions of yeast containing full-length (FL) or N-terminally truncated (Δ42, Δ44, and Δ70) yUsb1 on a high-copy p425-GPD plasmid. yUsb1 is tagged with a single N-terminal HA tag. B) SDS-PAGE and western blotting analysis of Usb1 protein expression detected using a primary antibody to the genetically encoded N-terminal HA epitope tag. Actin was used as a loading control.



3.3.7 Usb1 processing directly controls formation of U6 RNPs

We next investigated the impact of U6 3' end processing on the affinity of U6 snRNA 3' end binding proteins. We tested binding of Lhp1 and Lsm2-8 using a fragment of the U6 3' end (U6 95-112) with either a cis diol or a 3' phosphate. As expected, Lhp1 greatly prefers to bind a cis diol, with a K_d of 43 nM, with essentially no specific binding to a 3' phosphate (**Figure 3-10A** and **Table 3-3**)[32]. In contrast, Lsm2-8 can bind both a cis diol or a 3' phosphate, but preferentially binds a 3' phosphate (K_d of 85 vs. 10 nM for U6 95-112 with a cis diol or 3' phosphate, respectively) (**Figure 3-10B** and **Table 3-3**). Thus, exchange of a cis diol for a 3' phosphate eliminates Lhp1 binding[32] and improves Lsm2-8 binding affinity for U6. In humans, Lsm2-8 has been reported to preferentially bind a cyclic phosphate over a cis diol[15], suggesting that the Lsm2-8 complex in different organisms has evolved to bind the product of Usb1 processing.

Yeast U6 snRNA possesses an oligouridylate tail of heterogeneous length[33] and we therefore compared the affinity of both Lhp1 and Lsm2-8 for U6 95-112 oligonucleotides with or without an additional terminal uridine. As expected, Lhp1 binds U6 95-112 with similar affinity to U6 95-112+1U (**Figure 3-10A** and **Table 3-3**). In contrast, we find that Lsm2-8 binds U6 95-112+1U >4-fold tighter than U6 95-112 regardless of the 3' end modification (**Figure 3-10B** and **Table 3-3**). An extra uridine in the context of full-length U6 RNA also enhances Lsm2-8 binding (**Figure 3-11** and **Table 3-3**). Thus, of the RNAs tested, Lsm2-8 has the highest affinity for U6 with 5 uridine residues and a 3' phosphate modification.

Next, we directly monitored the effect of Usb1 processing on Lhp1 and Lsm2-8 binding to oligonucleotides and full-length U6 RNA. Usb1 processing of oligonucleotides produces binding profiles for Lhp1 that mirror the results with chemically defined 3' ends (**Figure 3-10C,D**), except for the fact that additional shorter products are also formed (**Figure 3-1B** and **Figure 3-10E**) that reduce Lsm2-8 affinity, as Lsm2-8 does not efficiently bind an RNA with only three terminal uridines[34]. The effect of Usb1 processing on U6-binding proteins also extends to full-length U6. Using full-length U6 with an additional uridine at the 3' end (U6 1-112+1U), we tested the affinity

of Lhp1, Lsm2-8, and Prp24 before and after Usb1 processing via native gel shift (**Figure 3-10F**). Lsm2-8 can bind both before and after Usb1 treatment (**Figure 3-10F**, lanes 3-5 vs. 6-8), but binding is slightly reduced due to the reduction in the U-tail length (**Figure 3-10B**). Lhp1 binding is most sensitive to Usb1 processing, with tight binding before processing and virtually no binding after (**Figure 3-10F**, lanes 9-11 vs. 12-14). Prp24 affinity is unchanged (**Figure 3-10F**, lanes 15-17 vs. 18-20), as expected from the structure of the Prp24-U6 RNA complex[35]. These data clearly demonstrate that the modification at the 3' end is the most important determinant for Lhp1 and Lsm2-8 binding and that processing by Usb1 controls the transition from immature (Lhp1-bound) to mature (Lsm2-8 bound) U6.

Figure 3-10. Usb1 processing influences RNP formation.

A) Fluorescence polarization binding data comparing Lhp1 binding to U6 95-112 with a cis diol (black, filled circles), U6 95-112+1U with a cis diol (black, open circles), U6 95-112 with a 3' phosphate (grey, filled circles), and U6 95-112+1U with a 3' phosphate (grey, open circles). Plotted data points represent the average of three technical replicates ± s.d. for (a)-(d).

B) Fluorescence polarization binding data comparing Lsm2-8 binding to U6 95-112 with a cis diol (black, filled circles), U6 95-112+1U with a cis diol (black, open circles), U6 95-112 with a 3' phosphate (grey, filled circles), and U6 95-112+1U with a 3' phosphate (grey, open circles).

C) Fluorescence polarization binding data comparing Lhp1 binding to U6 95-112+1U before (black) and after (grey) Usb1 processing.

D) Fluorescence polarization binding data comparing Lsm2-8 binding to U6 95-112+1U before (black) and after (grey) Usb1 processing.

E) Denaturing gel comparing RNAs used in Figures (C) and (D).

F) The affinities of Lsm2-8 and Lhp1 for full-length U6 are influenced by Usb1 processing. Native gel analysis comparing Lsm2-8, Lhp1, and Prp24 affinity for U6 RNA before and after treatment with Usb1. Usb1 processing does not change the mobility of U6 on a native gel (lanes 1 vs. 2). Lsm2-8 binds similarly before (lanes 3-5) and after (lanes 6-8) Usb1 processing. Lhp1 binding (lanes 9-11) is negligible after Usb1 processing (lanes 12-14). Prp24 binding is unchanged before (lanes 15-17) and after Usb1 processing (lanes 18-20).

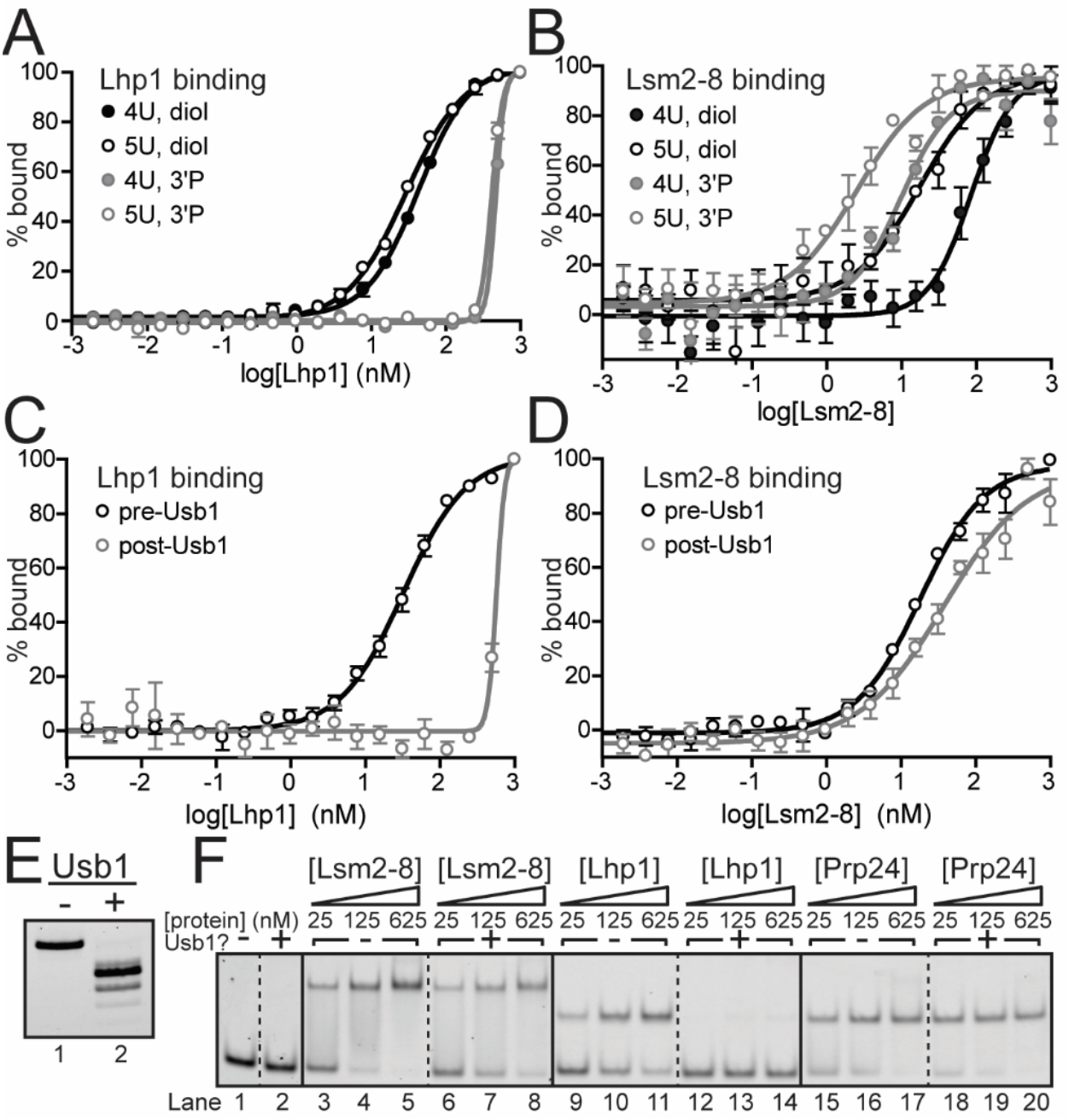

A
Lhp1 binding
4U, diol
5U, diol
4U, 3'P
5U, 3'P
% bound
log[Lhp1] (nM)
B
Lsm2-8 binding
4U, diol
5U, diol
4U, 3'P
5U, 3'P
% bound
log[Lsm2-8]
C
Lhp1 binding
pre-Usb1
post-Usb1
% bound
log[Lhp1] (nM)
D
Lsm2-8 binding
pre-Usb1
post-Usb1
% bound
log[Lsm2-8] (nM)
E
Usb1
- +
1 2
F
[Lsm2-8] [Lsm2-8] [Lhp1] [Lhp1] [Prp24] [Prp24]
[protein] (nM) 25 125 625 25 125 625 25 125 625 25 125 625 25 125 625 25 125 625
Usb1? - + - + - + - + - + - +
Lane 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Figure 3-11. Lsm2-8 preferentially binds full-length U6 RNA with an additional uridine.

A) Native gel analysis comparing Lsm2-8 affinity to full-length fluorescently labeled U6 1-112 with a 2',3'-cis diol versus U6 1-112 with an additional uridine at the 3' end (U6 1-112+1U) and a 2',3'-cis diol.

B) Quantification of Lsm2-8 binding to U6 in (A). Plotted data points represent the average of three technical replicates ± s.d.

C) Native gel analysis comparing Lhp1 affinity to RNAs as described in (A).

D) Quantification of Lhp1 binding to U6 in (C). Plotted data points represent the average of three technical replicates ± s.d.

E) Native gel analysis comparing Lsm2-8 affinity to full-length fluorescently labeled U6 with an additional uridine at the 3' end (U6 1-112+1U) and different 3' modifications. Lsm2-8 binds U6 with a 2',3'-cis diol better than U6 with a 2',3'-cyclic phosphate.

F) Quantification of Lsm2-8 binding to U6 in (E). Plotted data points represent the average of three technical replicates ± s.d.

G) Native gel analysis comparing Lhp1 affinity to RNAs as described in (E). Lhp1 binds U6 with a 2',3'-cis diol, but not a 2',3'-cyclic phosphate.

H) Quantification of Lhp1 binding to U6 in (G). Plotted data points represent the average of three technical replicates ± s.d.

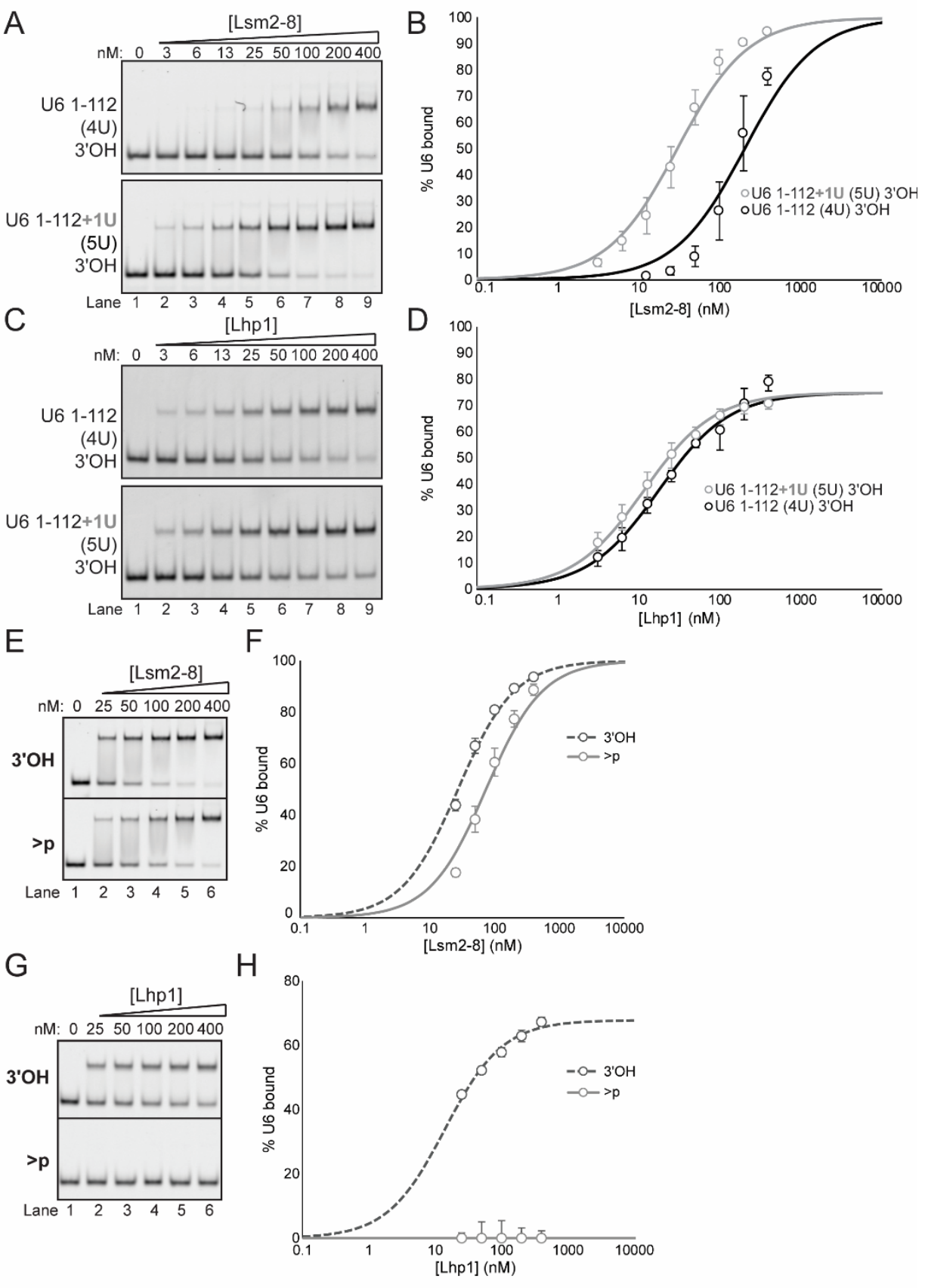

A
[Lsm2-8]
nM: 0 3 6 13 25 50 100 200 400
U6 1-112 (4U) 3'OH
U6 1-112+1U (5U) 3'OH
Lane 1 2 3 4 5 6 7 8 9
B
% U6 bound
U6 1-112+1U (5U) 3'OH
U6 1-112 (4U) 3'OH
[Lsm2-8] (nM)
C
[Lhp1]
nM: 0 3 6 13 25 50 100 200 400
U6 1-112 (4U) 3'OH
U6 1-112+1U (5U) 3'OH
Lane 1 2 3 4 5 6 7 8 9
D
% U6 bound
U6 1-112+1U (5U) 3'OH
U6 1-112 (4U) 3'OH
[Lhp1] (nM)
E
[Lsm2-8]
nM: 0 25 50 100 200 400
3'OH
>p
Lane 1 2 3 4 5 6
F
% U6 bound
3'OH
>p
[Lsm2-8] (nM)
G
[Lhp1]
nM: 0 25 50 100 200 400
3'OH
>p
Lane 1 2 3 4 5 6
H
% U6 bound
3'OH
>p
[Lhp1] (nM)

Table 3-3. Lsm2-8 and Lhp1 binding for U6 RNA with different 3' ends.

	Lhp1			Lsm2-8		
RNA	K_d (nM)	Hill coefficient	R^2	K_d (nM)	Hill coefficient	R^2
U6 3' tail						
U6 95-112 3'OH	43 ± 1	1.27 ± 0.05	0.99	85 ± 1	1.6 ± 0.3	0.89
U6 95-112+1U 3'OH	29 ± 1	1.11 ± 0.04	0.99	17 ± 1	1.0 ± 0.2	0.87
U6 95-112 3'P	467 ± 1	8 ± 2*	0.99	10 ± 1	1.3 ± 0.3	0.92
U6 95-112+1U 3'P	411 ± 1	5.9 ± 0.5*	0.99	3 ± 1	1.0 ± 0.2	0.90
Full-length U6	K_d (nM)	B_{max} (% bound)	R^2	K_d (nM)	B_{max} (% bound)	R^2
U6 1-112 3'OH	18 ± 1	75 ± 1	0.97	210 ± 90	100 ± 20	0.89
U6 1-112+1U 3'OH	12 ± 1	75 ± 1	0.98	30 ± 3	100 ± 3	0.98
U6 1-112+1U >p	n.d.	n.d.	n.d.	70 ± 10	100 ± 7	0.95

*The high Hill coefficient for Lhp1 binding to RNAs with 3' phosphate ends likely indicates non-specific aggregation.

3.3.8 Ordered binding in the U6 snRNP assembly pathway

It is well established that Lhp1 binds to newly synthesized U6 RNA[36,37]. Previous work on the subsequent steps in the U6 lifecycle has established that Prp24 and Lsm2-8 cooperatively bind U6 RNA[15] and that Prp24 and Lsm2-8 directly interact via the C-terminus of Prp24[18]. However, the effect of Prp24 on binding of Lhp1 has not been investigated. We tested co-binding of Prp24 and either Lhp1 or Lsm2-8 (**Figure 3-12A**). Prp24 binds U6 with a K_d that is 10-fold lower than Lhp1 (Didychuk *et al.* 2016; **Figure 3-12A**, lanes 2-6 and 7-11). When U6 RNA is incubated with equimolar amounts of Prp24 and Lhp1, it is preferentially bound by Prp24 and a ternary complex is not visible (lanes 12-16). For example, Lhp1 alone is mostly bound to U6 at 160 µM, but in the presence of Prp24 no Lhp1 binding is observed until 640 µM (**Figure 3-12A**, compare lanes 4 and 16). Formation of a U6-Lhp1-Prp24 ternary complex occurs only at high concentrations (lane 16). These data suggest that Prp24 binding is anti-cooperative with binding of Lhp1. In contrast, Prp24 and Lsm2-8 bind cooperatively and efficiently form a ternary complex (**Figure 3-12A**, lanes 22-26). While Lsm2-8 binds U6 with a cis diol relatively weakly by itself (**Figure 3-12A**, lanes 17-21), inclusion of Prp24 strongly promotes formation of a ternary complex (**Figure 3-12A**, lanes 22-26) even at the lowest concentration tested and despite the lack of a phosphoryl group on U6 to promote Lsm2-8 binding.

The apparent negative cooperativity of Prp24 and Lhp1 binding is intriguing, since known Prp24 and Lhp1 binding sites on U6 RNA are presumed to be non-overlapping[35,38]. We tested the affinity of Prp24 or Lhp1 for U6 pre-saturated with the opposing binding partner (**Figure 3-12B**). When U6 is pre-bound by Lhp1, Prp24 displaces U6 from U6-Lhp1 (**Figure 3-12B**, lanes 9-14). Consistent with these data, U6 pre-bound by Prp24 is not released from U6-Prp24 by the addition of excess Lhp1 (**Figure 3-12B**, lanes 23-28). In both cases, ternary complex formation is inefficient and occurs only at high concentration, suggesting that Prp24 and Lhp1 are near-mutually exclusive for binding. From these data we propose a model for U6 snRNP assembly (**Figure 3-12C**). Synthesis by Pol III followed by binding of Lhp1 protects nascent U6 from 3'

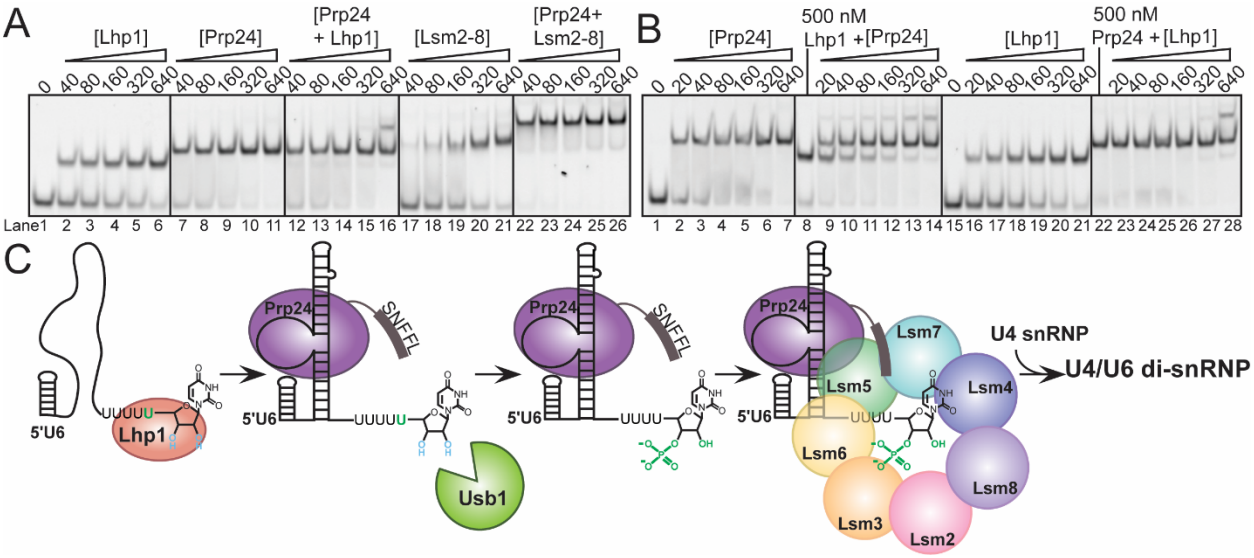
exonucleases. Binding of U6 by Prp24 is anti-cooperative with binding of Lhp1, allowing the freed 3' tail can then be recognized by Usb1 to produce a 3' phosphate. The presence of a 3' phosphate group prevents re-binding of Lhp1 and allows for recruitment of Lsm2-8, whose interaction is stabilized by recognition of the 3' phosphate (**Figure 3-10**) and by interactions with the Prp24 C-terminal SNFFL box[15,18]. The U6 snRNP can then join with the U4 snRNP[17] for efficient incorporation into the tri-snRNP and the spliceosome.

Figure 3-12. The U6 snRNP assembly pathway.

A) Native gel analysis of U6 binding partners. Lhp1 and Prp24 bind U6 1-112 with a cis diol tightly (lanes 2-6 and 7-11). Inclusion of equimolar amounts of Lhp1 and Prp24 does not promote formation of a ternary complex except at the highest concentration (lanes 12-16). In contrast, Lsm2-8 binds U6 relatively weakly (17-21), but upon inclusion of Prp24 (lanes 22-26), Lsm2-8 efficiently forms a co-complex of U6, Lsm2-8, and Prp24.

B) Prp24 binds naked U6 1-112 with a cis diol (lanes 2-7) and U6 pre-saturated with Lhp1 (lanes 9-14) tightly. Prp24 abstracts U6 from U6-Lhp1 much more efficiently than it forms a U6/Prp24/Lhp1 complex. Lhp1 binds naked U6 (lanes 21), but cannot bind or release U6 from pre-formed U6-Prp24.

C) Model of U6 snRNP assembly. U6 is synthesized by RNA polymerase III and initially bound by Lhp1. Binding of Prp24 weakens Lhp1 affinity for the 3' tail of U6, allowing Usb1 to remove a uridine and leave a 3' phosphate modified tail. Lsm2-8 recognizes the 3' tail of U6 and interacts with Prp24 to form the U6 snRNP, which can then be assembled into the spliceosome via the U4/U6 di-snRNP.



3.4 Discussion

Members of the 2H phosphodiesterase superfamily are found in viruses, Archaea, bacteria, and eukaryotes, and are involved in diverse pathways of RNA metabolism. 2H phophodiesterase enzymes are highly divergent in sequence outside of the active site H-X-S/T motif[28], yet structurally homologous. A DALI search[39] (**Table 3-4**) reveals that the closest structural homologs of yUsb1 (after hUsb1) include putative 2'-5' ligases[40,41], 2',3'-cyclic nucleotide 3'-phosphodiesterases (such as the well-characterized enzyme ThpR)[42,43], 2'-5' phosphodiesterases[44,45], and several proteins of unknown activity[46,47]. Thus, the fold of Usb1 is more reminiscent of a 2',3'-cyclic phosphodiesterase than an exoribonuclease, and the closest homologs tend to open cyclic phosphates to produce a 2' phosphate instead of a 3' phosphate. Here, we have demonstrated that two members of the 2H phosphodiesterase superfamily, yUsb1 and hUsb1, are mechanistically distinct yet are involved in the same step of U6 biogenesis in their respective organisms. Yeast Usb1 possesses two RNA processing activities (3'-5' exonuclease activity and 2',3'-cyclic nucleotide 2' phosphodiesterase activity), while hUsb1 lacks CPDase activity. Our results suggest that while the active sites of the 2H phosphodiesterase family are highly similar, residues immediately outside of the active site likely play a previously unappreciated role in RNA recognition and catalysis. Indeed, three highly structurally homologous enzymes - ThpR, yUsb1, and hUsb1- possess virtually superimposable H-X-S/T active sites, but have unique 2',3'-cyclic phosphodiesterase activities (catalyzing formation of 2' phosphate[43], 3' phosphate (this work), and 2',3'-cyclic phosphate[11] products, respectively). Enzyme active sites are dynamic and highly sensitive to sub-angstrom scale differences, making prediction of structure-activity relationships difficult. Further study of other 2H phosphodiesterases may reveal additional diversity of substrate specificities and activities for this superfamily.

The 3' phosphate terminus generated by yUsb1 inhibits subsequent exonucleolytic processing, whereas hUsb1 leaves a 2',3'-cyclic phosphate terminus which is a substrate for an additional exonucleolytic step, resulting in successive trimming of U6[11]. Humans also possess a

counteracting TUTase that extends the 3' tail of U6[48,49]. *S. cerevisiae* has no identifiable TUTase and thus the autoinhibition of yUsb1 by its own product is necessary to prevent over-processing of U6. The combined dual exonuclease and CPDase activity of Usb1 accomplishes three important functions to initiate U6 snRNP biogenesis: it prevents further 3' end processing, improves Lsm2-8 affinity for U6 RNA, and, perhaps more importantly, prevents Lhp1 from rebinding. A 3' phosphate might encounter steric clash with nearby loop residues of Usb1.

The evolutionary divergence in Lsm2-8 binding affinity appears to have co-evolved with the activity of Usb1. Yeast Lsm2-8 tightly binds RNAs with a 3' phosphate, while human Lsm2-8 prefers to bind RNAs with a 2',3'-cyclic phosphate[15]. Interestingly, yeast Lsm2-8 appears to bind a 2',3'-cyclic phosphate even worse than an unmodified cis diol or a non-cyclic phosphate. Therefore, yeast Lsm2-8 and human Lsm2-8 significantly differ in their modes of 3' end recognition and apparently have evolved to bind the product of their cognate Usb1. Along with this observation, we also demonstrate that Lsm2-8 binds the oligonucleotide U6 95-112 with a cis-diol >25-fold worse than a U6 95-112+1U oligonucleotide with 3' phosphate (**Table 3-3**). This model substrate, with five terminal uridines and a phosphate end, is very similar to the major mature form of human U6, which contains five terminal uridines and a cyclic phosphate end[13]. While there is a co-crystal structure of yeast Lsm2-8 with RNA, it was crystallized with an oligonucleotide containing a cis diol rather than a 3' phosphate and containing four terminal uridines instead of five, making it difficult to identify the crucial interactions that promote correct 3' end recognition[34]. Our data indicate that both a terminal 3' phosphate and the length of the oligouridylate tail are important binding determinants for Lsm2-8. It would be interesting to correlate evolutionary changes in Usb1 and Lsm2-8 with the binding mode of the Lsm2-8 complex from yeast and humans to the appropriate biological 3' end of U6.

Studies in fission yeast and human cell lines reveal that U6 (and U6atac in humans) is the major substrate of Usb1[13]. This, along with the observation that overexpression of U6 can rescue loss of Usb1 in *S. cerevisiae*[11], suggests that Usb1 is an exonuclease that is highly specific for

U6 RNA. How is such stunning specificity accomplished within the cell, where many other RNA polymerase III transcripts also terminate in a polyuridine stretch with a 2',3'-cis diol? Usb1 is similarly active on short RNA oligonucleotides and on full-length U6 snRNA, excluding specific recognition of U6 secondary structure by Usb1. We hypothesize that specificity stems from the lack of accessible substrates for Usb1 due to binding of Lhp1 on the majority of $(U)_n$ cis diol-containing RNAs. Binding of Prp24, however, is highly specific for the asymmetric bulge in U6 snRNA[17,35]. We have demonstrated that Prp24 binding to U6 lowers the apparent affinity of Lhp1 (**Figure 3-12A,B**); this, in turn, would allow for Usb1 to modify the free 3' end and thereby promote binding of Lsm2-8. The mechanism for handoff of U6 from Lhp1 to Prp24 is currently unknown, but may be caused by partially overlapping binding sites of Prp24 and Lhp1, occlusion of Lhp1 via electrostatic repulsion, or induction of RNA folding that is unfavorable for Lhp1 association. Future studies aimed at determining the origin of the observed anti-cooperative binding could reveal fundamental principles of ordered, multi-step pathways of RNP assembly.

This handoff of U6 RNA, from Lhp1 to Prp24 to Prp24/Lsm2-8, ensures that full-length and properly 3'-end modified U6 is guarded from exonucleases in the cell. Presence of a terminal 3' phosphate may partially protect U6 RNA from degradation by the exosome, as Rrp6 is inactive on a 3' phosphate-terminated RNA[50], but Rrp44/Dis3 is active on such substrates[51]. This work illustrates how Usb1 processing initiates U6 snRNP formation through a series of protein-RNA interactions that have evolved to protect U6 snRNA and chaperone it into the active site of the spliceosome.

Table 3-4. DALI search of unique structural homologs to yUsb1 catalytic domain.

PDB ID	Z-score	rmsd	% identity	Name	Activity
4H7W	16.2	3.1	14	*Hs*Usb1	RNA exonuclease, 3'-5' phosphodiesterase, leaves >p
1VDX	12	3.9	13	*Ph*RNL	unknown, putative RNA 2'-5' ligase/2',3' CPDase
4QAK	12	3.3	11	*Ec*ThpR	2',3' CPDase, leaves 2'P
5JJ2	11.6	3.8	11	*Hs*AKAP18	binds NMPs, unknown activity, not a 3',5' CPDase
2VFL	11.5	4	10	*Rn*AKAP18	binds NMPs, unknown activity, not a 3',5' CPDase
1IUH	11.3	3.6	12	*Tt*RNL	unknown, putative RNA 2'-5' ligase/2',3' CPDase
2D4G	10.7	4.3	13	*Bs*YjcG	unknown, putative 2'-5' ligase/2',3' CPDase
2FYH	9.8	3.9	14	PF0027	putative RNA 2'-5' ligase/known 2',3' CPDase, leaves 2'P
4Z5V	9.4	3.9	16	MHV ns2	2'-5' phosphodiesterase (cleaves 5'AMP from 2' termini of 2-5A)
2FSQ	9	4.4	5	ATU0111	unknown
5AF2	8.7	3.3	11	VP3	2'-5' phosphodiesterase (cleaves 5'AMP from 2' termini of 2-5A)
1JH7	8.6	4.5	9	*At*CNPase	ADP-ribose 1'',2'' CPDase, leaves PO_3 on 1'' position
2YP0	5.6	3.8	9	*Mm*CNPase	2',3'-cyclic nucleotide 3'- phosphodiesterase, leaves 2'P
1WOJ	5.3	3.9	9	*Hs*CNPase	2',3'-cyclic nucleotide 3'- phosphodiesterase, leaves 2'P
2ILX	5.2	5.1	13	*Rn*CNPase	2',3'-cyclic nucleotide 3'- phosphodiesterase, leaves 2'P
2I3E	4.3	4.4	11	*Ca*RICH	2',3' CPDase, leaves 2'P

3.5 Materials and methods

3.5.1 Protein expression and purification

The Usb1 coding sequence was PCR amplified from *S. cerevisiae* genomic DNA with primers to introduce flanking BamHI and XhoI sites and subcloned into a pET28b plasmid, which encodes an N-terminal hexahistidine tag followed by a TEV cleavage site. The resulting cloning scar was removed using the inverse PCR method with Phusion DNA polymerase (New England Biolabs). All primer sequences are listed in **Table 3-5**. PCR products were DpnI treated, ligated using T4 PNK and T4 DNA ligase, and transformed into *E. coli* NEB 5α competent cells (New England Biolabs). Expression plasmids contained Usb1 residues 1-290 (full-length) or 71-290 (catalytic domain). Mutants of these plasmids were obtained using inverse PCR as described above. Resulting clones were expressed in *E. coli* BL21 STAR (DE3) pLysS cells (Invitrogen) in LB at 37°C with late-log phase induction by addition of 1 mM IPTG and subsequent growth for 3 hours at 37°C (for Usb1 71-290) or 16°C (for Usb1 1-290 and all other proteins used in this study). Cells were collected by centrifugation, resuspended in IMAC buffer (500 mM NaCl, 50 mM HEPES acid, 50 mM sodium HEPES base, 15 mM imidazole base, 10% glycerol, 1 mM TCEP-HCl) supplemented with DNase I, lysozyme, and protease inhibitors (EMD Millipore) and lysed via sonication. Insoluble material was removed by centrifugation. Usb1 was purified by Ni-NTA agarose chromatography by step elution with IMAC buffer + 500 mM imidazole. The eluate was dialyzed at 4°C overnight into IEX-HEPES buffer (100 mM NaCl, 10 mM HEPES acid, 10 mM sodium HEPES base, 10% glycerol, 1 mM TCEP-HCl, pH ~7.0) for Usb1 71-290 or IEX-bis-tris buffer (100 mM NaCl, 20 mM bis-tris, 10 mM HCl, 10% glycerol, 1 mM TCEP-HCl, pH ~6.2) for Usb1 1-290 with 1 mg TEV protease. Precipitated protein was removed, then protein was further purified via cation-exchange chromatography (HiTrap S, GE Healthcare) in IEX buffer with gradient elution against IEX buffer containing 2 M NaCl.

The coding sequence for *Homo sapiens* Usb1 was codon optimized for expression in *E. coli*. The sequence of the synthetic hUsb1 gene is shown in **Table 3-6**. For functional assays, this

sequence was cloned via NdeI and BamHI restriction cloning into a modified pET3a plasmid (Novagen) containing an N-terminal octahistidine tag, GST, and a TEV cleavage site. Human Usb1 (residues 79-265) used in crystallography experiments was similarly cloned into a modified pET3a plasmid (Novagen) containing an N-terminal octahistidine tag, MBP, and a TEV cleavage site. Protein was expressed and purified as described above using Ni-NTA agarose chromatography and cation-exchange chromatography with IEX-HEPES buffer.

The coding sequence for *S. cerevisiae* Lhp1 protein (residues 1-275) was codon optimized for expression in *E. coli.* The sequence of the synthetic Lhp1 gene is shown in **Table 3-6**. A pET3a plasmid (Novagen) was modified using inverse PCR as described above to encode an N-terminal octahistidine tag followed by a TEV cleavage site. The coding sequence for Lhp1 was cloned into this plasmid using the NdeI and BamHI sites. Protein was expressed and purified as described above using Ni-NTA agarose chromatography and cation-exchange chromatography with IEX-HEPES buffer.

The Prp24 coding sequence was PCR amplified from *S. cerevisiae* genomic DNA with primers to introduce flanking NdeI and XhoI sites and subcloned into a pET21b plasmid, which encodes a C-terminal hexahistidine tag. Prp24 was expressed as described above in Terrific Broth. Prp24 was purified by Ni-NTA agarose chromatography as described above using Ni-NTA agarose chromatography with IMAC buffer containing 50 mM imidazole, dialyzed without TEV protease, and purified via heparin chromatography (HiTrap Heparin, GE Heathcare) in IEX-HEPES as described above with the addition of 1 mM sodium azide.

S. cerevisiae Lsm2-8 lacking the C-terminus of Lsm4 was expressed from pQLink-Lsm2-8[34] in *E.coli* BL21 STAR (DE3) pLysS cells in Terrific Broth as described above. Lsm2-8 was first purified via Ni-NTA agarose chromatography via a TEV-labile polyhistidine tag on Lsm8 in IMAC buffer containing 50 mM imidazole base, then dialyzed overnight into IEX-HEPES buffer. After removal of precipitated protein, Lsm2-8 was purified via a TEV-labile GST tag on Lsm6 and glutathione agarose chromatography with step elution in IEX-HEPES supplemented with 10 mM

reduced glutathione, 50 mM HEPES acid, and 50 mM sodium HEPES base. The polyhistidine and GST tags were removed by the addition of 1 mg TEV protease during overnight dialysis at room temperature into IEX-HEPES. Lsm2-8 was then purified by anion-exchange chromatography (HiTrap Q, GE Healthcare) with gradient elution in IEX-HEPES supplemented with 2 M NaCl. Lsm2-8 was then diluted five-fold against IEX-bis-tris containing 1 mM sodium azide and further purified via heparin chromatography (HiTrap Heparin, GE Healthcare) in IEX-bis-tris containing 1 mM sodium azide with gradient elution in buffer supplemented with 2 M NaCl.

A truncated variant of AtRNL containing only the kinase and 2',3'-cyclic phosphate 3'-phosphodiesterase domains (residues 677-1104) was modified from pET28-Smt3-AtRNL(1-1104) (a kind gift from Stewart Shuman[52]) via inverse PCR to remove the N-terminal ligase domain and install a TEV cleavage site after the octahistidine tag. Overexpression and purification was essentially as described above using Ni-NTA agarose chromatography with IMAC buffer with 50 mM imidazole and cation-exchange chromatography using IEX-HEPES. All protein samples were analyzed by SDS-PAGE to assess purity.

3.5.2 Crystallization and structure determination

Crystals of truncated yUsb1 (residues 71-290) were obtained by hanging drop vapor diffusion with 2 µL concentrated protein (9 mg mL^{-1}) and 2 µL of crystallization solution (0.1 M sodium acetate, pH 4.5, 2.0 M ammonium sulfate) with equilibration against 500 µL of crystallization solution at 20°C. Crystals grew as needles of approximate dimensions 50 x 50 x 100 µm over 1-3 days. Crystals were cryoprotected by addition of 10 uL of a solution containing 15% v/v glycerol, 20% w/v PEG 20,000, 0.1 M sodium acetate pH 4.6 to the crystallization drop. A heavy atom derivative was produced as above using a cryoprotectant solution that was also saturated with uranyl acetate and incubation for 1 min prior to freezing. Diffraction data were collected at 100 K on beamlines 21-ID-F or 24-ID-C at the Advanced Photon Source. Data were integrated using *XDS*[53]. Space group determination and scaling were performed in *POINTLESS*[54] and *AIMLESS*[55], respectively. *Phenix.xtriage* was used to assay potential twinning in the

diffraction data[56]. Initial phases could not be determined using molecular replacement with hUsb1 78-265 (PDB 4H7W), therefore phases were determined by the method of Single Isomorphous Replacement with Anomalous Scattering, using initial heavy atom site identification, map calculation and density modification in the *SHELXC/D/E* pipeline[57] as implemented in *HKL2Map*[58]. Automated model building was accomplished with *RESOLVE*[56,59], with subsequent refinement via iterative rounds of manual model building in *Coot*[60] and automated refinement in *PHENIX*[56,61].

Attempts to co-crystallize or soak in nucleotides into yUsb1 were unsuccessful, likely due to the stringent requirement for ammonium sulfate in the crystallization conditions. Crystals of truncated hUsb1 (residues 79-265) were grown by hanging drop vapor diffusion in 1.4 M sodium potassium phosphate pH 5.6 at 289 K and then transferred into a solution containing 20% w/v PEG 20,000, 20% v/v glycerol, 100 mM bis-tris base, 50 mM HCl, and 10 mM disodium 5' UMP and allowed to incubate overnight. Data were collected on beamline 24-ID-C at the Advanced Photon Source and processed as above. Initial phases were determined by molecular replacement using *Phaser*[62] with PDB entry 4H7W[11], and refinement accomplished by the iterative process above. Data collection and refinement statistics are given in **Table 3-1**. All figures were generated with PyMOL (http://www.pymol.org).

3.5.3 RNA production

RNA oligonucleotides containing *S. cerevisiae* U6 nucleotides 95-112 with one, three, or six additional 3' uridine nucleotides and containing a 5' 6-FAM moiety and a 2',3'-cis diol or phosphate were purchased from Integrated DNA Technologies. Sequences of all synthetic RNA oligonucleotides are listed in **Table 3-7**. RNAs were purified via 20% 19:1 acrylamide:bisacrylamide PAGE containing 8 M urea, 89 mM Tris borate, and 2 mM EDTA and extracted into 0.3 M NaOAc/1 mM EDTA. Oligonucleotides were then further purified via anion exchange (HiTrap Q column, GE Healthcare Life Sciences) using 100 mM NaCl, 20 mM bis-tris pH 6.5, 1 mM EDTA and elution with 2 M NaCl, 20 mM bis-tris pH 6.5, 1 mM EDTA. RNAs were concentrated and stored in 50 mM NaCl, 20 mM bis-tris pH 6.5, 1 mM EDTA.

To obtain an RNA with a cis diol, cyclic phosphate, 2' phosphate, or 3' phosphate, a longer RNA oligomer comprising *S. cerevisiae* U6 nucleotides 84-112 with three additional 3' uridine nucleotides was produced by splinted ligation. U6 95-112(UUU) was produced via *in vitro* transcription from a plasmid containing a 5' hammerhead ribozyme (HH) and a 3' hepatitis delta virus ribozyme (HDV). This plasmid was generated via inverse PCR by modification of a plasmid containing a 5' hammerhead ribozyme, full length U6 1-112, and a 3' HDV ribozyme (a kind gift from Kiyoshi Nagai). U6 95-112(UUU), which contained a 5'OH group and 2',3'-cyclic phosphate group due to HH/HDV cleavage, was 5' phosphorylated with ATP by either T4 PNK, T4 PNK (3' phosphatase minus) (both New England Biolabs) or truncated AtRNL. The RNA was then phenol/chloroform/isoamyl alcohol extracted to remove enzyme, then ligated to 5' 6-FAM U6 84-94 (Integrated DNA Technologies) with T4 RNA ligase 2 at 37°C for two hours using a DNA splint that was complementary to the entire RNA product. The ratio of U6 95-112(UUU):FAM U6 84-94:DNA splint was 1:1.5:2. The ligation product was then gel purified, extracted, and purified via ion exchange as described above. Fluorescent U6 84-112(UUU) with a 3' phosphate was prepared synthetically by Integrated DNA Technologies.

Full-length Cy3-labeled U6 (nucleotides 1-112) and full-length U6 containing one additional 3' uridine nucleotide (U6 1-112+1U) were produced by splinted ligation as described above using a 5'-Cy3 U6 1-12 RNA oligonucleotide (Integrated DNA Technologies) and U6 13-112 that was produced via *in vitro* transcription from a modified plasmid containing a 5' hammerhead ribozyme and a 3' HDV ribozyme (a kind gift from Kiyoshi Nagai).

3.5.4 Nuclear magnetic resonance spectroscopy

Samples of 1 mM uridine-2',3'-cyclic monophosphate sodium salt (2',3'-cUMP; Santa Cruz Biotechnology) in 50 mM NaCl, 20 mM bis-tris base, 10 mM HCl, 30 µM DSS, >95% D_2O were incubated with either buffer (100 mM NaCl, 10 mM HEPES base, 10 mM HEPES acid, 10% v/v glycerol, 1 mM TCEP pH ~7.2), 8 µM AtRNL, or 75 µM Usb1 71-290 overnight at room temperature. All spectra were obtained on a Bruker Avance III 600 MHz spectrometer with a 5

mm $^{1}H(^{13}C/^{15}N/^{31}P)$ cryogenic probe at the National Magnetic Resonance Facility at Madison (NMRFAM). ^{1}H chemical shifts were directly referenced to DSS and ^{31}P chemical shifts were indirectly referenced to DSS. The 3' phosphate left after Usb1 treatment was assigned based on chemical shift in 1D ^{31}P spectra and comparison to 3'UMP and 2',3'-cUMP controls. Presence of a 3' phosphate was further confirmed via ^{1}H-^{31}P HMBC and ^{1}H-^{1}H COSY experiments.

3.5.5 Statistical analysis

Unless otherwise noted, all activity and binding assays were carried out in triplicate. The resulting data points were averaged before being used to calculate rate or binding constants. Plotted data points represent these averages ± s.d. Errors in rate or binding constants represent the errors generated by the fits to these data using GraphPad Prism 4 software.

3.5.6 Exoribonuclease assays

Exoribonuclease assays were performed in 10 µL reactions by mixing equal volumes of Usb1 (1 µM) in 100 mM KCl, 20 mM bis-tris, 10 mM HCl, 1 mM TCEP-HCl, 20% w/v sucrose, 0.01% v/v Triton X-100, 0.2 mg mL^{-1} BSA and RNA substrate (200 nM) in 100 mM NaCl, 20 mM bis-tris, 10 mM HCl, 1 mM TCEP-HCl, 1 mM EDTA, 10% v/v glycerol. Final reaction conditions included 500 nM Usb1 and 100 nM RNA substrate in 50 mM NaCl, 50 mM KCl, 20 mM bis-tris, 10 mM HCl, 1 mM TCEP-HCl, 10% w/v sucrose, 5% v/v glycerol, 0.5 mM EDTA, 0.005% v/v Triton X-100, 0.1 mg mL^{-1} BSA. Samples were incubated at room temperature for the indicated time and reactions were quenched by the addition of an equal volume of 100% deionized formamide. Samples were resolved on a 20% 19:1 acrylamide:bisacrylamide PAGE gel containing 8 M urea, 89 mM Tris borate, and 2 mM EDTA. The gels were imaged directly through low fluorescence glass plates (CBS Scientific) using a Typhoon FLA 9000 (GE Healthcare Life Sciences). Alkaline hydrolysis ladders were produced by incubating 5 µL of RNA substrate (200 nM) in buffer with 5 µL of 50 mM bicarbonate buffer pH 9.2, 1 mM EDTA at 95°C for 10 minutes.

Samples were treated with CIP or T4 PNK by addition of “Cutsmart” or “PNK” buffer from New England Biolabs and 10 units of CIP or T4 PNK and incubation at 37°C for 15 minutes. Mock treated samples contained only Cutsmart buffer and water in lieu of CIP or T4 PNK.

For time course experiments, the percent processed was calculated using the ratio of product(s) to total signal at 0, 5, 15, 30 and 60 minute time points. Resulting data were fit to a one-phase exponential association equation ($Y = Y_0 + (Y_{max} - Y_0)*(1-e^{(-k*x)}$ Y is the % of the substrate processed, x is time, and *k* is the rate constant (GraphPad Prism 4).

3.5.7 Gel shift assay

Gel shift assays were performed by incubating Cy3-labeled U6 with increasing concentrations of protein. Cy3-U6 RNA (5 nM) was heated to 90°C for 2 minutes in RNA binding buffer (100 mM KCl, 20% sucrose, 20 mM bis-tris, 10 mM HCl, 1 mM EDTA, 1 mM TCEP-HCl, 0.01% Triton X-100, 0.2 mg mL^{-1} tRNA, 0.02 mg mL^{-1} sodium heparin), then snap cooled on wet ice. Proteins were prepared as 2x stocks in protein binding buffer (100 mM KCl, 20% sucrose, 20 mM bis-tris, 10 mM HCl, 1 mM EDTA, 1 mM TCEP-HCl, 0.01% Triton X-100, 0.2 mg mL^{-1} BSA). Final binding reactions contained 100 mM KCl, 20% sucrose, 20 mM bis-tris, 10 mM HCl, 1 mM EDTA, 1 mM TCEP-HCl, 0.01% Triton X-100, 0.1 mg mL^{-1} tRNA, 0.01 mg mL^{-1} sodium heparin, and 0.1 mg mL^{-1} BSA. Samples were incubated at room temperature for 20 minutes prior to electrophoresis on a 6% polyacrylamide gel (29:1 acrylamide:bis-acrylamide, 89 mM Tris borate, 2 mM EDTA pH 8.0) for 2-3 hours at 3W at 4°C. Gels were imaged directly through low fluorescence glass plates (CBS Scientific) using a Typhoon FLA 9000 (GE Healthcare Life Sciences). Results were analyzed using ImageJ software and binding curves were fit using nonlinear regression in GraphPad Prism 4 to the Hill equation: $\%\ bound = (B_{max} * [protein]^H) / (K_d^H + [protein]^H)$. B_{max} was restrained to be between 0 and 100%, and K_d and H were restrained to be > 0.

3.5.8 Fluorescence polarization binding experiments

Fluorescence polarization binding reactions were performed by mixing 100 µL of 2x RNA in buffer (100 mM NaCl, 20 mM bis-tris, 10 mM HCl, 1 mM TCEP-HCl, 5% v/v glycerol, 1 mM EDTA) containing either 0.2 mg mL^{-1} (for Lsm2-8 binding) or 0.02 mg mL^{-1} (for Lhp1 binding) yeast tRNA (Roche Diagnostics) and 100 µL of protein in buffer containing 0.2 mg mL^{-1} BSA (Ambion) in black 96 well microplates (Greiner Bio-One). Final RNA concentrations for Lsm2-8, Lhp1, and before/after Usb1 binding experiments were 0.25 nM, 1 nM, and 0.5 nM respectively. Final protein concentrations were 0.001-1000 nM. For the before/after Usb1 binding experiment, FAM-U6 95-112+1U (100 nM) with a cis-diol was incubated with Usb1 1-290 (500 nM) as described above for one hour before dilution and use in the binding experiments.

Fluorescence polarization was measured on a Tecan Infinite M1000Pro using an excitation wavelength of 470 nm, emission wavelength of 519 nm. Gain was optimized for each microplate. Fluorescence polarization was measured in triplicate for each condition (using 500-1000 flashes) and averaged. Data were normalized to the values for 0 nM protein (smallest value) and to the highest value, then averaged between three technical replicates. Binding curves were fit using nonlinear regression in GraphPad Prism 4 to a four-parameter logistic equation: % bound = FP_{min} + (FP_{max} - FP_{min})/(1 + 10^((logK_d - log[protein])*H)), where FP_{min} and FP_{max} are the normalized minimum and maximum % bound, K_d is the binding dissociation constant, and H is the Hill coefficient. FP_{min}, FP_{max}, K_d, and H were restrained to be >0.

3.5.9 pH dependence exoribonuclease assay

We found that Usb1 activity is strongly inhibited by the presence of either sulfate or phosphate. Thus, we sought to make a mixed buffer for pH-varied activity assays containing 200 mM each of sodium acetate, bis-tris, sodium HEPES base, Tris base, and CHES (pH ~9). The pH of this 1M solution was adjusted up or down with 5 M NaOH or 5 M HCl respectively. Aliquots were removed at every 0.5 pH unit step and kept as a 50X stock of buffer. Protein and RNA were diluted to 1 µM or 200 nM, respectively, in buffer containing 100 mM NaCl, 10% w/v sucrose, 0.1 mM EDTA, 0.1 mM TCEP, 0.1 mg mL^{-1} BSA, 0.01% v/v Triton X-100, and 20 mM of the composite

buffer. Protein and RNA were mixed together in equal volumes (for final concentrations of 500 nM and 100 nM respectively) for the indicated times. Reactions were quenched by addition of an equal volume of 100% formamide and samples were resolved as described above.

3.5.10 Yeast strains and plasmids

The *USB1* coding region and 500 bp up and downstream of the coding region was amplified from genomic DNA from the BJ2168 (*MATa leu2 trp1 ura3-52 prb1-1122 pep4-3 prc1-407 gal2*) strain via PCR, which created an upstream BamHI site and a downstream XhoI site. The PCR product was digested with BamHI and XhoI (New England Biolabs) and ligated into BamHI/XhoI-cut pRS416 and pRS414. Point mutations in pRS414-Usb1 were generated using inverse PCR as described above.

BJ2168 was transformed with pRS416-Usb1 using the lithium acetate method[63]. The Usb1 disruption strain (yTJC0700) was constructed by transformation of BJ2168 pRS416-Usb1 with linear DNA amplified from pAG32 (Euroscarf) containing the *hph* gene flanked by regions homologous to 500 nucleotides up and downstream of the *USB1* gene. Cells were grown on YPD for one day, then replica plated onto media containing 200 μg mL^{-1} hygromycin. Individual colonies were screened and deletion of *USB1* was confirmed via PCR.

Overexpression plasmids containing Usb1 were generated by PCR amplification of ORFs using primers that added an upstream BamHI and a downstream SalI site. The coding sequence for hUsb1 was codon optimized for expression in *S. cerevisiae* and the nucleotide sequence is listed in **Table 3-6**. This PCR product was digested with BamHI and SalI (New England Biolabs) and ligated into BamHI/SalI-cut p425-GPD (ATCC 87359). All mutations, including addition of a single N-terminal HA tag, were generated using inverse PCR as described above.

3.5.11 *USB1* complementation assays

yTJC0700 was transformed with variants of pRS414-Usb1 or p425-GPD plasmids using the lithium acetate method[63]. Growth phenotypes were assessed by spotting 10-fold serial dilutions (OD_{600} = 0.5, 0.05, 0.005) onto solid medium lacking tryptophan (for pRS414 plasmids)

or leucine (for p425-GPD plasmids) and containing 1 mg mL^{-1} 5-fluoroorotic acid (5-FOA). Plates were incubated at 30°C for 3 days.

3.5.12 Western blotting

Yeast were grown in selective media and total protein was isolated by trichloroacetic acid (TCA) precipitation[64]. Protein concentration was normalized by A280 measurement and equivalent amounts were separated on a 4-20% Criterion TGX midi protein gel (200 V for 1 h; Bio-Rad) and subsequently transferred to a nitrocellulose membrane (30 min, 100 V, 4◦C). The membrane was blocked using 5% (w/v) non-fat dry milk and probed using an HA antibody conjugated to horseradish peroxidase used at 1:5000 dilution (Sigma Aldrich 11667475001). Mouse α-actin (AMB Millipore MAB1501) was used at 1:5000 dilution and goat α-mouse HRP antibodies was used at 1:10000 dilution (Biorad 1706515). Blots were developed using Clarity Western ECL substrate (Biorad) and imaged using an ImageQuant LAS 4000 Imager (GE Healthcare Life Sciences).

3.5.13 Data availability

Coordinates and structure factors have been deposited in the Protein Data Bank with accession codes 5UQJ and 5V1M. Other data supporting the findings of this manuscript are available from the corresponding author upon reasonable request.

Table 3-5. List of synthetic DNA oligonucleotides.

Oligos to make pET28b-HT-yUsb1 1-290, truncation, and mutations	
BamHI-pET28b-USB1	5'-CGGGATCCAATGGAATTCATATCTGCAGACTATTC-3'
XhoI-pET28b-USB1	5'-TGCTCGAGTTAGTTAAAAGGGATTCTAATGGAGTG-3'
Usb1-1-for	5'-ATGGAATTCATATCTGCAGACTATTCTAGCAGTGACGG-3'
Usb1-71-for	5'-ATGAGTCGATTTTGGCGTTCGTTCACATATTTTGAGTG GC-3'
His-TEV-rev	5'-GCCCTGGAAGTACAGGTTCTCGCTGCTGTGATGATGA TGATGATGGC-3'
Usb1_F78A_for	5'-GCGACATATTTTGAGTGGCGTCCGACTCCAGCG-3'
Usb1_F78H_for	5'-CATACATATTTTGAGTGGCGTCCGACTCCAGCG-3'
Usb1_F78_rev	5'-CGAACGCCAAAATCGACTCATATTCATATCC-3'
Oligos to make p3HT-HT-hUsb1 79-265	
hUsb1-D78N-for	5'-GGAAACTGGGCTACACATGTGTACGTTCCT-3'
hUsb1-TEV-rev	5'-CATATGAGCGCCCTGGAAATACAGGTTTTC-3'
Oligos to make pET28b-Smt3-TEV-AtRNL(677-1104)	
Trunc-AtRNL-for	5'-GCTAGCGAAAACCTGTATTTCCAGGGCGCTGTGAAGG AAGCTGTCCAAAAGGATGAG-3'
Trunc-AtRNL-rev	5'-GTGATGATGATGATGATGATGATGATGGCCCATGGT ATATCTCCTTCTTAAAGTTAA-3'
Oligos to amplify *hph* gene flanked by Usb1 up and downstream region for *USB1* deletion strain:	
Usb1_deletion_for	5'-CGAACCCAAACTATCTGTGACATCAGTTGCTGATGCTC TTCTTGGTCCCTTATTTGCGGTTTCAGAGACGTCCAGGC GCGCCAGATCTGTTTAGCTTGCCTTG-3'
Usb1_deletion_rev	5'-CGGGTGGGGGTATGGTAGGATTGACATTATCATGATG TTGAACATATCCTCATCTTGAGTGTCTTTCGCGAAGCGA GCTCGTTTTCGACACTGGATGGC-3'
Usb1_del_con_for	5'-GTTTCCTTAACAAAGGGAATCTC-3'
Usb1_del_con_rev	5'-TACCTAACAAGTTATCGAGG-3'
Oligos to create pRS414-Usb1 and pRS416-Usb1:	
BamHI-Usb1+500-for:	5'-ACGGGATCCATAGTTGCACACCCGCTACTCCCTCATT GACAGC-3'
XhoI-Usb1+500-rev	5'-CAGCTCGAGCTGAGGGATTTATATACCACGGATCTAC-3'
Cloning oligos to introduce mutations into pRS414-Usb1:	
Usb1_W75A_for	5'-GCGCGTTCGTTCACATATTTTGAGTGGCGTCCG-3'
Usb1_W75A_rev	5'-AAATCGACTCATATTCATATCCTGGTGCTCG-3'
Usb1_F78A_for	5'-GCGACATATTTTGAGTGGCGTCCGACTCCAGCG-3'
Usb1_F78H_for	5'-CATACATATTTTGAGTGGCGTCCGACTCCAGCG-3'
Usb1_F78_rev	5'-CGAACGCCAAAATCGACTCATATTCATATCC-3'
Usb1_Y80A_for	5'-GCTTTTGAGTGGCGTCCGACTCCAGCGATTC-3'
Usb1_Y80A_rev	5'-TGTGAACGAACGCCAAAATCGACTCATATTC-3'
Usb1_H133A_for	5'-GCGGTTTCCTTAACTCGATCGTTGTTGTTTGAAAC-3'
Usb1_H133N_for	5'-AACGTTTCCTTAACTCGATCGTTGTTGTTTGAAAC-3'
Usb1_H133_rev	5'-TAAAGGTTTGGGGGCTCCCAAGTGTGAGATGAAC-3'
Usb1_H231A_for	5'-GCGGTTTCAATTGCCATCGCTAGCAACCCTTCAAAAG-3'
Usb1_H231N_for	5'-AACGTTTCAATTGCCATCGCTAGCAACCCTTCAAAAG-3'
Usb1_H231_rev	5'-TAAATTTTGCCTACTTACAATTAAATCCTGATAG-3'
Usb1+500_START-rev	5'-CATAGGCCAGTCTTCTTCTAAGCGATTGGTAACC-3'

Usb1+500_SV40-rev	5'-AACTTTTCTTTTTTTTTTTGGTGGCATAGGCCAGTCTTCT TCTAAGCGATTGGTAACC-3'
Usb1-21_for	5'-AGTAACAAAAGTGAGGTCCAAATCGAATACACG-3'
Usb1-42_for	5'-ACTGATTTACCTGCAATACCGGATTCAATTATATTG AAG-3'
Usb1-43_for	5'-GATTTACCTGCAATACCGGATTCAATTATATTG-3'
Usb1-44_for	5'-TTACCTGCAATACCGGATTCAATTATATTGAAG-3'
Usb1-45_for	5'-CCTGCAATACCGGATTCAATTATATTGAAGTAC-3'
Usb1-46_for	5'-GCAATACCGGATTCAATTATATTGAAGTACCAC-3'
Usb1-47_for	5'-ATACCGGATTCAATTATATTGAAGTACCACATC-3'
Usb1-48_for	5'-CCGGATTCAATTATATTGAAGTACCACATCCCT-3'
Usb1-49_for	5'-GATTCAATTATATTGAAGTACCACATCCCTCCG-3'
Usb1-50_for	5'-TCAATTATATTGAAGTACCACATCCCTCCGAAT-3'
Usb1-51_for	5'-ATTATATTGAAGTACCACATCCCTCCGAATTTAC-3'
Usb1-71_for	5'-ATGAGTCGATTTTGGCGTTCGTTCACATATTTTGAGT GGC-3'
Oligos to make p425-GPD-yUsb1 and hUsb1, truncations, and mutations	
BamHI-yUsb1-for	5'-CGGGATCCAATGGAATTCATATCTGCAGACTATTC-3'
SalI-yUsb1-rev	5'-GGCGTCGACTTAGTTAAAAGGGATTCTAATGGAGTGA-3'
BamHI-hUsb1-for	5'-ACGGGATCCATATGAGTGCAGCCCCTTTAGTAGGG-3'
SalI-hUsb1-rev	5'-GGCGTCGACTCATTTCAGCGGCATGGAAAAAAATTTG-3'
p425-up-HA-r	5'-AGCGTAATCTGGTACGTCGTATGGGTACATATGGAT CCACTAGTTCTAGAATCCGTCG-3'
Usb1-1-for	5'-ATGGAATTCATATCTGCAGACTATTCTAGCAGTGACGG-3'
Usb1-43_for	5'-GATTTACCTGCAATACCGGATTCAATTATATTG-3'
Usb1-45_for	5'-CCTGCAATACCGGATTCAATTATATTGAAGTAC-3'
Usb1-71_for	5'-ATGAGTCGATTTTGGCGTTCGTTCACATATTTTGAGT GGC-3'
hUsb1-H84F-for	5'-TTCGTCTACGTCCCGTATGAAGCCAAGGAGGAA-3'
hUsb1-H84-rev	5'-AGTTGCCCAGTTCCCACGCTCGTGAGGAAA-3'
hUsb1-wt-79-for	5'-GGGAACTGGGCAACTCATGTCTACGTCCCG-3'
hUsb1-H84F-79-for	5'-GGGAACTGGGCAACTTTCGTCTACGTCCCG-3'
Oligos to make HH-U6 95-112+3U-HDV plasmid	
HH-U6_95-112(UUU)-for	5'-AAACACGTTCGCGTGTCAGAGATTTATTTCGTTTTTTTG GGTCGGCATGGCATCTCCACCTCCTCGCGGTCCG-3'
HH-U6_95-112(UUU)-rev	5'-CGGCCTAACGGCCTCATCAGAGAGATTTATTCTCCCTAT AGTGAGTCGTATTAGAAGAGC-3'
Oligos to make HH-U6 13-112-HDV plasmid	
U6_13-112_for	5'-CGAAACACGTTCGCGTGTCACCCTTCGTGGACATTTGGT CAATTTGAAACAATACAGAGA-3'
HH-U6_13-112_rev	5'-GCCTAACGGCCTCATCAGACCCTTCGTGGCTCCCTATAGT GAGTCGTATTAAAG-3'

Table 3-6. Nucleotide sequence of synthetic genes.

hUsb1, codon optimized for *S. cerevisiae* (used for *in vivo* complementation assays)	5'-GGATCCATATGAGTGCAGCCCCTTTAGTAGGGTACAGCTCAA GCGGGAGTGAAGATGAGAGCGAGGACGGAATGAGGACCAGACCGG GTGATGGAAGTCATAGAAGAGGCCAGAGCCCCTTGCCGAGGCAAC GTTTCCCCGTACCCGACAGTGTTTTGAATATGTTCCCAGGTACCGAA GAGGGACCCGAAGATGACTCTACCAAACACGGAGGGAGGGTCCGT ACTTTTCCTCACGAGCGTGGGAACTGGGCAACTCATGTCTACGTCC CGTATGAAGCCAAGGAGGAATTTTTGGATCTGTTGGATGTGCTACTT CCCCACGCCCAAACCTACGTACCGAGACTGGTAAGGATGAAAGTCT TCCACCTGAGTCTGAGTCAGAGCGTTGTGCTGAGGCATCACTGGAT ACTGCCTTTCGTACAGGCGCTTAAGGCTCGTATGACCAGTTTTCACA GGTTCTTTTTTACTGCCAATCAGGTGAAGATATATACAAACCAAGAAA AAACAAGGACCTTTATTGGCCTAGAGGTTACTTCCGGTCACGCACAA TTTTTGGACCTAGTCTCCGAGGTTGACCGTGTGATGGAAGAATTTAA CCTGACCACTTTTTATCAAGACCCTTCATTCCATCTGTCACTTGCATG GTGTGTCGGAGATGCTCGTCTGCAACTAGAAGGGCAATGTCTTCAA GAGCTGCAAGCCATAGTCGATGGTTTCGAGGATGCTGAGGTCCTGC TTAGAGTCCACACAGAGCAGGTGCGTTGCAAGAGTGGCAACAAATT TTTTTCCATGCCGCTGAAATGAGTCGAC-3'
hUsb1, codon optimized for *E. coli* (used for protein expression and crystallization)	5'-CATATGTCAGCTGCACCATTAGTCGGGTATTCATCATCGGGTTC TGAGGATGAGTCCGAGGACGGGATGCGTACACGTCCTGGCGACGG ATCGCACCGTCGTGGGCAGTCCCCATTGCCTCGCCAGCGCTTCCCT GTGCCCGATTCGGTCCTGAACATGTTTCCAGGGACCGAGGAAGGAC CGGAAGATGATAGCACTAAGCATGGAGGTCGCGTACGCACTTTCCC ACACGAGCGCGGAAACTGGGCTACACATGTGTACGTTCCTTACGAA GCGAAGGAAGAATTCCTGGATTTGCTTGATGTCTTGCTTCCACATGC GCAGACTTACGTGCCGCGCTTAGTCCGCATGAAAGTCTTTCACCTG TCCTTGTCCCAATCGGTAGTTCTTCGCCATCATTGGATTTTGCCTTTT GTGCAGGCGTTGAAGGCCCGTATGACATCGTTCCACCGCTTTTTTTT CACCGCCAATCAGGTAAAGATTTATACAAATCAAGAGAAGACACGCA CATTCATTGGCTTGGAGGTAACCTCAGGGCACGCTCAATTTTTGGAC TTAGTAAGTGAAGTAGACCGCGTGATGGAGGAGTTCAACCTTACAA CTTTCTATCAAGATCCGTCCTTCCACCTGAGCTTAGCCTGGTGCGTG GGAGATGCTCGCCTTCAGTTGGAAGGCCAATGCCTTCAAGAACTGC AAGCAATCGTAGACGGGTTCGAGGACGCCGAGGTACTGCTTCGTGT ACACACCGAGCAGGTTCGCTGCAAATCAGGTAATAAGTTTTTCTCCA TGCCACTTAAATGACTCGAGGATCC-3'
Lhp1 from *S. cerevisiae*, codon optimized for *E. coli* (used for protein expression)	5'-CATATGTCTGAAAAACCGCAGCAGGAAGAACAGGAAAAACCGC AGTCTCGTCGTAACTCTTTCGCTGTTATCGAATTCACCCCGGAAGTT CTGGACCGTTGCCTGAAACAGGTTGAATTCTACTTCTCTGAATTCAA CTTCCCGTACGACCGTTTCCTGCGTACCACCGCTGAAAAAAACGAC GGTTGGGTTCCGATCTCTACCATCGCTACCTTCAACCGTATGAAAAA ATACCGTCCGGTTGACAAAGTTATCGAAGCTCTGCGTTCTTCTGAAA TCCTGGAAGTTTCTGCTGACGGTGAAAACGTTAAACGTCGTGTTCCG CTGGACCTGACCGCTGCTCGTAACGCTCGTATCGAACAGAACCAGC GTACCCTGGCTGTTATGAACTTCCCGCACGAAGACGTTGAAGCTTCT CAGATCCCGGAACTGCAGGAAAACCTGGAAGCTTTCTTCAAAAAACT GGGTGAAATCAACCAGGTTCGTCTGCGTCGTGACCACCGTAACAAA AAATTCAACGGTACCGTTCTGGTTGAATTCAAAACCATCCCGGAATG CGAAGCTTTCCTGAAATCTTACTCTAACGACGACGAATCTAACGAAA

	TCCTGTCTTACGAAGGTAAAAAACTGTCTGTTCTGACCAAAAAACAG TTCGACCTGCAGCGTGAAGCTTCTAAATCTAAAAACTTCTCTGGTCG TTCTCGTTCTTTCAACGGTCACAAAAAAAAAAACCTGCCGAAATTCC CGAAAAACAAAAAAAAAAACGGTAAAGAAGAATCTAAAGAAGACTCT TCTGCTATCGCTGACGACGACGAAGAACACAAAGAATGATGACTCG AGGATCC-3'

Table 3-7. List of synthetic RNA oligonucleotides.

FAM-U6 95-112	5'-/56-FAM/rArGrArGrArUrUrUrArUrUrUrCrGrUrU rUrU-3'
FAM-U6 95-112-3P	5'-/56-FAM/rArGrArGrArUrUrUrArUrUrUrCrGrUrU rUrU/3Phos/-3'
FAM-U6 95-112+1U	5'-/56-FAM/rArGrArGrArUrUrUrArUrUrUrCrGrUrU rUrUrU-3'
FAM-U6 95-112+1U-3P	5'-/56-FAM/rArGrArGrArUrUrUrArUrUrUrCrGrUrU rUrUrU/3Phos/-3'
FAM-U6 95-112+3U	5'-/56-FAM/rArGrArGrArUrUrUrArUrUrUrCrGrUrU rUrUrUrUrU-3'
FAM-U6 95-112+3U-3P	5'-/56-FAM/rArGrArGrArUrUrUrArUrUrUrCrGrUrU rUrUrUrUrU/3Phos/-3'
FAM-U6 95-112+6U	5'-/56-FAM/rArGrArGrArUrUrUrArUrUrUrCrGrUrU rUrUrUrUrUrUrU-3'
FAM-U6 95-111dUrU	5'-/56-FAM/rArGrArGrArUrUrUrArUrUrUrCrGrUrU rU/ideoxyU/rU-3'
FAM-U6 84-94	5'-/56-FAM/rCrCrGrUrUrUrUrArCrArA-3'
FAM-U6 84-112+3U-3P	5'-/56-FAM/rCrCrGrUrUrUrUrArCrArArArGrArGrA rUrUrUrArUrUrUrCrGrUrUrUrUrUrU/3Phos/-3'
5'Cy3-U6 1-12	5'-/5Cy3/rGrUrUrCrGrCrGrArArGrUrA-3'

3.6 Acknowledgements and funding

We thank Kiyoshi Nagai, Yigong Shi, and Stewart Shuman for plasmids and Jill Wildonger for the actin antibody. We thank Elsebet Lund and Jim Dahlberg for careful reading of the manuscript. We thank members of the Butcher, Brow, and Hoskins labs for helpful discussions. We thank Marco Tonelli and Mark Anderson for assisting with NMR data collection. This work was supported by grants from the US National Institutes of Health (R01 GM065166 to S.E.B. and D.A.B., R35 GM118131 to S.E.B. and R35 GM118075 to D.A.B., R00 GM086471 and R01 GM112735 to A.A.H., T32-GM08349 to T.J.C.). A.A.H. was supported by the Shaw Scientist Award, the Beckman Young Investigator Award, and startup funding from the University of Wisconsin-Madison, Wisconsin Alumni Research Foundation (WARF), and the Department of Biochemistry. A.L.D. was supported by the University of Wisconsin-Madison Louis and Elsa Thomsen Wisconsin Distinguished Graduate Fellowship. T.J.C. was supported by the William H. Peterson Fellowship. A.T.D. was supported by the Hilldale Undergraduate Research Fellowship. S. L. was supported by a Sophomore Research Fellowship. This study made use of the National Magnetic Resonance Facility at Madison, which is supported by NIH grant P41GM103399 (NIGMS), old number: P41RR002301. Equipment was purchased with funds from the University of Wisconsin-Madison, the NIH P41GM103399, S10RR02781, S10RR08438, S10RR023438, S10RR025062, S10RR029220), the NSF (DMB-8415048, OIA-9977486, BIR-9214394), and the USDA. Use of the Advanced Photon Source, an Office of Science User Facility operated for the US Department of Energy (DOE) Office of Science by Argonne National Laboratory, was supported by the US DOE under contract no. DE-AC02-06CH11357. Use of the NE-CAT Sector 24 was supported by NIH grant P41 GM103403. Fluorescence polarization data were obtained at the University of Wisconsin-Madison Biophysics Instrumentation Facility, which was established with support from the University of Wisconsin-Madison and grants BIR-9512577 (NSF) and S10RR13790 (NIH).

3.7 References

1. Fica, S.M. et al. RNA catalyses nuclear pre-mRNA splicing. *Nature* 503, 229-34 (2013).

2. Kunkel, G.R., Maser, R.L., Calvet, J.P. & Pederson, T. U6 small nuclear RNA is transcribed by RNA polymerase III. *Proc Natl Acad Sci U S A* 83, 8575-9 (1986).

3. Reddy, R., Henning, D., Das, G., Harless, M. & Wright, D. The capped U6 small nuclear RNA is transcribed by RNA polymerase III. *J Biol Chem* 262, 75-81 (1987).

4. Moenne, A. et al. The U6 gene of Saccharomyces cerevisiae is transcribed by RNA polymerase C (III) in vivo and in vitro. *EMBO J* 9, 271-7 (1990).

5. Hamada, M., Sakulich, A.L., Koduru, S.B. & Maraia, R.J. Transcription termination by RNA polymerase III in fission yeast. A genetic and biochemically tractable model system. *J Biol Chem* 275, 29076-81 (2000).

6. Rinke, J. & Steitz, J.A. Association of the lupus antigen La with a subset of U6 snRNA molecules. *Nucleic Acids Res* 13, 2617-29 (1985).

7. Stefano, J.E. Purified lupus antigen La recognizes an oligouridylate stretch common to the 3' termini of RNA polymerase III transcripts. *Cell* 36, 145-54 (1984).

8. Lund, E. & Dahlberg, J.E. Cyclic 2',3'-phosphates and nontemplated nucleotides at the 3' end of spliceosomal U6 small nuclear RNA's. *Science* 255, 327-30 (1992).

9. Mroczek, S. et al. C16orf57, a gene mutated in poikiloderma with neutropenia, encodes a putative phosphodiesterase responsible for the U6 snRNA 3' end modification. *Genes Dev* 26, 1911-25 (2012).

10. Shchepachev, V., Wischnewski, H., Missiaglia, E., Soneson, C. & Azzalin, C.M. Mpn1, mutated in poikiloderma with neutropenia protein 1, is a conserved 3'-to-5' RNA exonuclease processing U6 small nuclear RNA. *Cell Rep* 2, 855-65 (2012).

11. Hilcenko, C. et al. Aberrant 3' oligoadenylation of spliceosomal U6 small nuclear RNA in poikiloderma with neutropenia. *Blood* 121, 1028-38 (2013).

12. Glatigny, A. et al. An in silico approach combined with in vivo experiments enables the identification of a new protein whose overexpression can compensate for specific respiratory defects in Saccharomyces cerevisiae. *BMC Syst Biol* 5, 173 (2011).

13. Shchepachev, V., Wischnewski, H., Soneson, C., Arnold, A.W. & Azzalin, C.M. Human Mpn1 promotes post-transcriptional processing and stability of U6atac. *FEBS Lett* 589, 2417-23 (2015).

14. Walne, A.J., Vulliamy, T., Beswick, R., Kirwan, M. & Dokal, I. Mutations in C16orf57 and normal-length telomeres unify a subset of patients with dyskeratosis congenita, poikiloderma with neutropenia and Rothmund-Thomson syndrome. *Hum Mol Genet* 19, 4453-61 (2010).

15. Licht, K., Medenbach, J., Luhrmann, R., Kambach, C. & Bindereif, A. 3'-cyclic phosphorylation of U6 snRNA leads to recruitment of recycling factor p110 through LSm proteins. *RNA* 14, 1532-8 (2008).

16. Achsel, T. et al. A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *EMBO J* 18, 5789-802 (1999).

17. Didychuk, A.L., Montemayor, E.J., Brow, D.A. & Butcher, S.E. Structural requirements for protein-catalyzed annealing of U4 and U6 RNAs during di-snRNP assembly. *Nucleic Acids Res* 44, 1398-410 (2016).

18. Rader, S.D. & Guthrie, C. A conserved Lsm-interaction motif in Prp24 required for efficient U4/U6 di-snRNP formation. *RNA* 8, 1378-92 (2002).

19. Galej, W.P. et al. Cryo-EM structure of the spliceosome immediately after branching. *Nature* 537, 197-201 (2016).

20. Nguyen, T.H. et al. Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 A resolution. *Nature* 530, 298-302 (2016).

21. Wan, R., Yan, C., Bai, R., Huang, G. & Shi, Y. Structure of a yeast catalytic step I spliceosome at 3.4 A resolution. *Science* 353, 895-904 (2016).

22. Wan, R. et al. The 3.8 A structure of the U4/U6.U5 tri-snRNP: Insights into spliceosome assembly and catalysis. *Science* 351, 466-75 (2016).

23. Yan, C. et al. Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* 349, 1182-91 (2015).

24. Yan, C., Wan, R., Bai, R., Huang, G. & Shi, Y. Structure of a yeast activated spliceosome at 3.5 A resolution. *Science* 353, 904-11 (2016).

25. Zhao, Z.Y. et al. Nucleobase participation in ribozyme catalysis. *J Am Chem Soc* 127, 5026-7 (2005).

26. Schleich, T., Blackburn, B.J., Lapper, R.D. & Smith, I.C. A nuclear magnetic resonance study of the influence of aqueous sodium perchlorate and temperature on the solution conformations of uracil nucleosides and nucleotides. *Biochemistry* 11, 137-45 (1972).

27. Lapper, R.D. & Smith, I.C. A 13 C and 1 H nuclear magnetic resonance study of the conformations of 2',3'-cyclic nucleotides. *J Am Chem Soc* 95, 2878-80 (1973).

28. Mazumder, R., Iyer, L.M., Vasudevan, S. & Aravind, L. Detection of novel members, structure-function analysis and evolutionary classification of the 2H phosphoesterase superfamily. *Nucleic Acids Res* 30, 5229-43 (2002).

29. Buchan, D.W., Minneci, F., Nugent, T.C., Bryson, K. & Jones, D.T. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res* 41, W349-57 (2013).

30. Rost, B., Yachdav, G. & Liu, J. The PredictProtein server. *Nucleic Acids Res* 32, W321-6 (2004).

31. Zarrinpar, A., Bhattacharyya, R.P. & Lim, W.A. The structure and function of proline recognition domains. *Sci STKE* 2003, RE8 (2003).

32. Terns, M.P., Lund, E. & Dahlberg, J.E. 3'-end-dependent formation of U6 small nuclear ribonucleoprotein particles in Xenopus laevis oocyte nuclei. *Mol Cell Biol* 12, 3032-40 (1992).

33. Brow, D.A. & Guthrie, C. Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature* 334, 213-8 (1988).

34. Zhou, L. et al. Crystal structures of the Lsm complex bound to the 3' end sequence of U6 small nuclear RNA. *Nature* 506, 116-20 (2014).

35. Montemayor, E.J. et al. Core structure of the U6 small nuclear ribonucleoprotein at 1.7-A resolution. *Nat Struct Mol Biol* 21, 544-51 (2014).

36. Pannone, B.K., Kim, S.D., Noe, D.A. & Wolin, S.L. Multiple functional interactions between components of the Lsm2-Lsm8 complex, U6 snRNA, and the yeast La protein. *Genetics* 158, 187-96 (2001).

37. Pannone, B.K., Xue, D. & Wolin, S.L. A role for the yeast La protein in U6 snRNP assembly: evidence that the La protein is a molecular chaperone for RNA polymerase III transcripts. *EMBO J* 17, 7442-53 (1998).

38. Teplova, M. et al. Structural basis for recognition and sequestration of UUU(OH) 3' temini of nascent RNA polymerase III transcripts by La, a rheumatic disease autoantigen. *Mol Cell* 21, 75-85 (2006).

39. Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res* 38, W545-9 (2010).

40. Kato, M. et al. Crystal structure of the 2'-5' RNA ligase from Thermus thermophilus HB8. *J Mol Biol* 329, 903-11 (2003).

41. Rehse, P.H. & Tahirov, T.H. Structure of a putative 2'-5' RNA ligase from Pyrococcus horikoshii. *Acta Crystallogr D Biol Crystallogr* 61, 1207-12 (2005).

42. Myllykoski, M., Raasakka, A., Han, H. & Kursula, P. Myelin 2',3'-cyclic nucleotide 3'-phosphodiesterase: active-site ligand binding and molecular conformation. *PLoS One* 7, e32336 (2012).

43. Remus, B.S., Jacewicz, A. & Shuman, S. Structure and mechanism of E. coli RNA 2',3'-cyclic phosphodiesterase. *RNA* 20, 1697-705 (2014).

44. Sui, B. et al. Crystal structure of the mouse hepatitis virus ns2 phosphodiesterase domain that antagonizes RNase L activation. *J Gen Virol* 97, 880-6 (2016).

45. Brandmann, T. & Jinek, M. Crystal structure of the C-terminal 2',5'-phosphodiesterase domain of group A rotavirus protein VP3. *Proteins* 83, 997-1002 (2015).

46. Bjerregaard-Andersen, K., Ostensen, E., Scott, J.D., Tasken, K. & Morth, J.P. Malonate in the nucleotide-binding site traps human AKAP18gamma/delta in a novel conformational state. *Acta Crystallogr F Struct Biol Commun* 72, 591-7 (2016).

47. Gold, M.G., Smith, F.D., Scott, J.D. & Barford, D. AKAP18 contains a phosphoesterase domain that binds AMP. *J Mol Biol* 375, 1329-43 (2008).

48. Trippe, R. et al. Identification, cloning, and functional analysis of the human U6 snRNA-specific terminal uridylyl transferase. *RNA* 12, 1494-504 (2006).

49. Trippe, R., Sandrock, B. & Benecke, B.J. A highly specific terminal uridylyl transferase modifies the 3'-end of U6 small nuclear RNA. *Nucleic Acids Res* 26, 3119-26 (1998).

50. Burkard, K.T. & Butler, J.S. A nuclear 3'-5' exonuclease involved in mRNA degradation interacts with Poly(A) polymerase and the hnRNA protein Npl3p. *Mol Cell Biol* 20, 604-16 (2000).

51. Zinder, J.C., Wasmuth, E.V. & Lima, C.D. Nuclear RNA Exosome at 3.1 A Reveals Substrate Specificities, RNA Paths, and Allosteric Inhibition of Rrp44/Dis3. *Mol Cell* 64, 734-745 (2016).

52. Remus, B.S. & Shuman, S. A kinetic framework for tRNA ligase and enforcement of a 2'-phosphate requirement for ligation highlights the design logic of an RNA repair machine. *RNA* 19, 659-69 (2013).

53. Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* 66, 125-32 (2010).

54. Evans, P.R. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr D Biol Crystallogr* 67, 282-92 (2011).

55. Evans, P.R. & Murshudov, G.N. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* 69, 1204-14 (2013).

56. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66, 213-21 (2010).

57. Sheldrick, G.M. A short history of SHELX. *Acta Crystallogr A* 64, 112-22 (2008).

58. Pape, T. & Schneider, T.R. HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *Journal of Applied Crystallography* 37, 843-844 (2004).

59. Terwilliger, T.C. SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol* 374, 22-37 (2003).

60. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66, 486-501 (2010).

61. Afonine, P.V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* 68, 352-67 (2012).

62. McCoy, A.J. et al. Phaser crystallographic software. *J Appl Crystallogr* 40, 658-674 (2007).

63. Gietz, R.D. & Woods, R.A. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol* 350, 87-96 (2002).

64. Carrocci, T.J., Zoerner, D.M., Paulson, J.C. & Hoskins, A.A. SF3b1 mutations associated with myelodysplastic syndromes alter the fidelity of branchsite selection in yeast. *Nucleic Acids Res* (2017).

Chapter 4: Structural requirements for protein-catalyzed annealing of U4 and U6 RNAs during di-snRNP assembly

This chapter is published in the following form:

Didychuk, A.L., Montemayor, E.J., Brow, D.A., Butcher, S.E. (2016) Structural requirements for protein-catalyzed annealing of U4 and U6 RNAs during di-snRNP assembly. *Nucleic Acids Research* 44(3):1398-1410.

Author contributions:

A.L.D. performed all experiments. E.J.M. assisted with assay development and protein purification. A.L.D. and S.E.B. wrote the paper with input from E.J.M. and D.A.B.

4.1 Abstract

Base-pairing of U4 and U6 snRNAs during di-snRNP assembly requires large-scale remodeling of RNA structure that is chaperoned by the U6 snRNP protein Prp24. We investigated the mechanism of U4/U6 annealing *in vitro* using an assay that enables visualization of ribonucleoprotein complexes and faithfully recapitulates known *in vivo* determinants for the process. We find that annealing, but not U6 RNA binding, is highly dependent on the electropositive character of a 20 Å-wide groove on the surface of Prp24. During annealing, we observe the formation of a stable ternary complex between U4 and U6 RNAs and Prp24, indicating that displacement of Prp24 *in vivo* requires additional factors. Mutations that stabilize the U6 “telestem” helix increase annealing rates by up to 15-fold, suggesting that telestem formation is rate-limiting for U4/U6 pairing. The Lsm2-8 complex, which binds adjacent to the telestem at the 3’ end of U6, provides a comparable rate enhancement. Collectively, these data identify domains of the U6 snRNP that are critical for one of the first steps in assembly of the megaDalton U4/U6.U5 tri-snRNP complex, and lead to a dynamic model for U4/U6 pairing that involves a striking degree of evolved cooperativity between protein and RNA.

4.2 Introduction

Proteins that stimulate the annealing of RNA to target nucleic acids are ubiquitously important in biology, with examples including the Argonaute family proteins (1,2), the CRISPR-Cas systems (3), and Hfq (4,5). Additionally, RNA remodeling proteins are a broad class of proteins that chaperone the proper folding of RNAs, which have a high propensity to misfold into stable alternative structures (6). A striking example of both protein-mediated RNA remodeling and annealing occurs during spliceosome assembly. Spliceosome assembly requires complex conformational rearrangements among a set of five small nuclear RNAs (U1, U2, U4, U5 and U6 snRNAs) and many proteins (7,8). Once assembled, the spliceosome catalyzes precursor-messenger RNA (pre-mRNA) splicing, an essential process in all eukaryotes.

Structural rearrangements in U6 snRNA are particularly dynamic, involving the unwinding and re-annealing of the U6 internal stem loop (ISL) (9), which coordinates catalytic magnesium ions in the active site of fully-assembled spliceosomes (10). Thus, the spliceosome is an "RNPzyme" with an RNA active site that is assembled and organized by proteins (11,12). Prior to incorporation into the spliceosome, yeast U6 exists in the form of the U6 snRNP, containing U6 snRNA, the tetra-RRM (RNA recognition motif) protein Prp24 and the heteroheptameric Lsm2-8 protein ring (13-21). In addition to being a stable component of the U6 snRNP, Prp24 acts as an RNA chaperone to catalyze unwinding of the U6 ISL and base pairing of the U4/U6 di-snRNA (15,16,22). Despite the stability of the U6 ISL (23), Prp24 accomplishes this task in an ATP-independent manner.

Our recent crystal structure of the U6•Prp24 complex revealed an extensive RNA-protein interface (24). Three of Prp24's four RRMs encircle a large loop in U6 RNA, generating a novel interlocked conformation in which U6 RNA-mediated RRM2-oRRM4 interactions result in topologically interlocked "rings" of RNA and protein (**Figure 4-1A**). Thus, U6 RNA and Prp24 must co-fold around each other to form the U6 snRNP core. In the U6 snRNP core, U6 RNA adopts a secondary structure consisting of the ISL (nucleotides 59-88), the large asymmetric bulge

(nucleotides 41-58) bound by Prp24, and an additional helix known as the telestem (nucleotides 30-40 and 91-101) (**Figure 4-1B**). The structure also revealed a 20 Å-wide electropositive groove within Prp24, composed of RRMs 1, 2 and oRRM4, which does not contact U6 within the complex, but binds duplex RNA from a neighboring complex in the crystal lattice. We previously hypothesized that this groove promotes U4/U6 annealing through stabilization of product duplex in the U4/U6 di-snRNA (24).

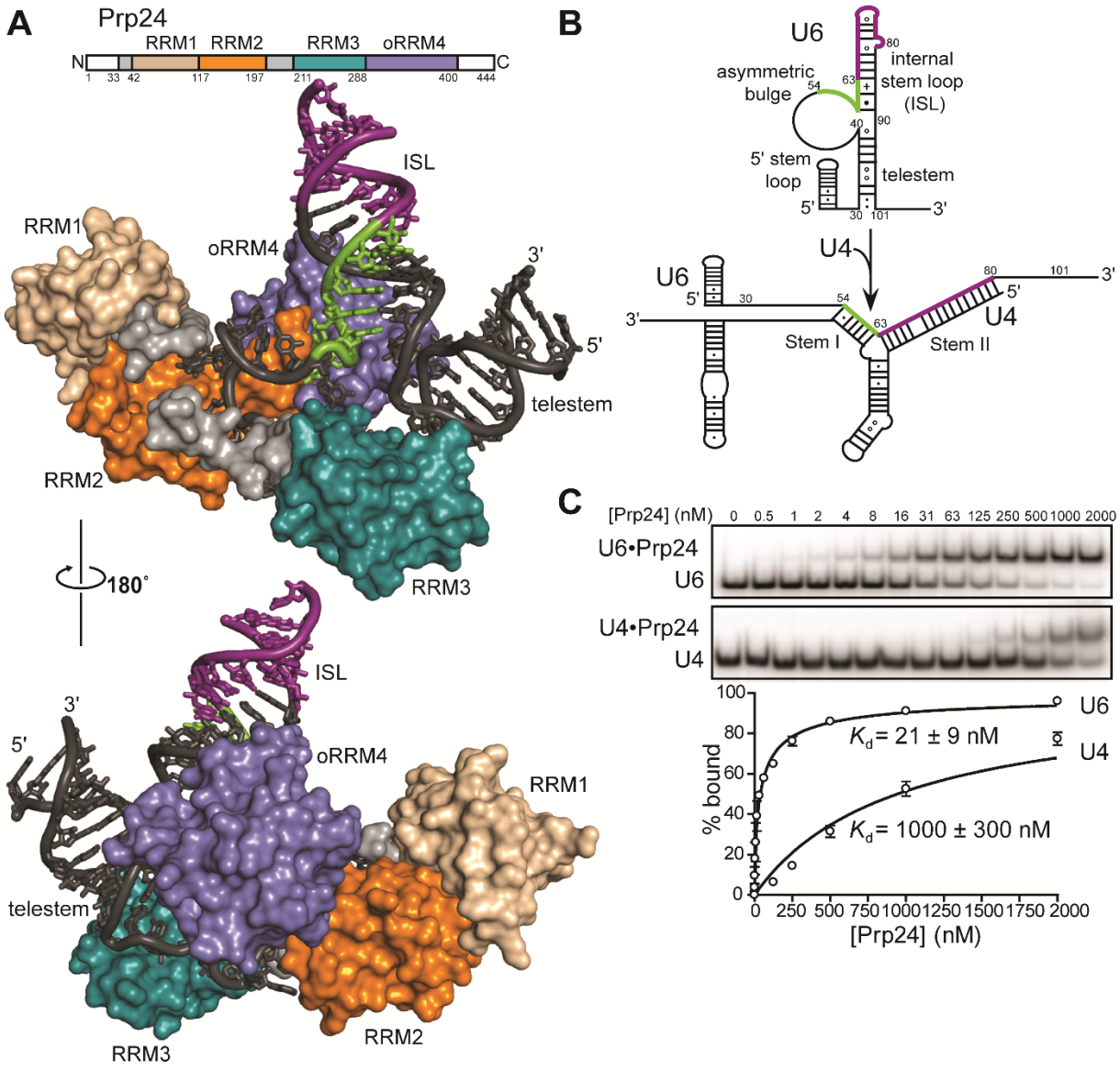
In this study, we analyzed the mechanism of U4/U6 annealing by tracking the formation of all RNA-protein (RNP) species, quantifying their binding affinities, and measuring annealing rates *in vitro*. These approaches, in combination with extensive mutagenesis of both RNA and protein, provided significant insight into the annealing mechanism. Importantly, our *in vitro* system faithfully recapitulates a previously observed *in vivo* defect associated with a mutation in the U6 ISL, as well as its suppression by a second-site mutation (9). We found that Prp24 remained bound to the product U4/U6 in a stable ternary complex, indicating that additional factors are likely required to displace Prp24 from annealed U4/U6 *in vivo*, and showed that the core regions of U6 and Prp24 present in our crystal structure are optimal for annealing. We verified that the electropositive groove on the surface of Prp24 is indeed critical for annealing, but is not required for binding U6. Stabilization of the U6 telestem helix significantly increases annealing rates, as does binding of the Lsm2-8 complex adjacent to the telestem. Remarkably, a specific mutation in the telestem induces Prp24-independent RNA annealing. Based on these findings, we propose a dynamic model for the U4/U6 annealing pathway.

Figure 4-1. Prp24 binds U6 RNA with high affinity and specificity.

A) Top: Primary structure of Prp24. White regions are disordered and were deleted from the crystallization construct as previously described (24). RRM4 is an occluded RRM (oRRM), with terminal α-helices masking its β-sheet face (22,24). Unless indicated otherwise, all assays herein used full-length protein and RNA. Bottom: molecular architecture of the U6 snRNP core. Regions of U6 RNA (black) that form U4/U6 Stem I and Stem II are highlighted in green and purple, respectively.

B) Schematic of U6 RNA annealing to U4 RNA to form U4/U6.

C) Native gel analysis of full-length Prp24 binding to U6 (top) and U4 (bottom) RNAs.


A
Prp24
RRM1 RRM2 RRM3 oRRM4
N C
1 33 42 117 197 211 288 400 444
ISL
RRM1
oRRM4
3'
5'
telestem
RRM2
RRM3
180˚
ISL
3'
5'
oRRM4
RRM1
telestem
RRM3
RRM2
B
U6
80
internal
stem loop
(ISL)
asymmetric
bulge
54
63
40
90
5' stem
loop
telestem
5'
3'
30 101
U4
U6
5'
30
54
80
101
3'
3'
5'
U4
63
Stem I
Stem II
C
[Prp24] (nM) 0 0.5 1 2 4 8 16 31 63 125 250 500 1000 2000
U6•Prp24
U6
U4•Prp24
U4
100
80
60
40
20
0
% bound
U6
U4
K_d = 21 ± 9 nM
K_d = 1000 ± 300 nM
0 250 500 750 1000 1250 1500 1750 2000
[Prp24] (nM)

4.3 Materials and methods

4.3.1 Overexpression and purification of proteins

S. cerevisiae Prp24 protein was recombinantly overexpressed in *E. coli* using a pET15b plasmid (EMD Millipore) modified to replace the N-terminal hexahistidine tag with a decahistidine tag and the N-terminal thrombin cleavage site with a TEV cleavage site. Expression plasmids contained Prp24 residues 1-444 (full-length), 34-400 (1234), 1-400 (N1234), or 34-444 (1234C). Mutants of this plasmid were generated using inverse PCR with Phusion DNA polymerase (New England Biolabs); PCR products were DpnI treated, self-ligated using T4 DNA ligase and T4 PNK, and transformed into *E. coli* NEB 5-alpha competent cells (New England Biolabs). Clones were isolated by plasmid minipreps (Qiagen) and the identity of each verified by Sanger sequencing. Resulting clones were transformed into *E. coli* STAR pLysS cells (Invitrogen) for protein overexpression and purification essentially as described (24), with the exception that dialysis buffer and cation-exchange chromatography buffers did not contain EDTA, and that 1 mg of TEV protease was added during dialysis into cation-exchange chromatography buffer. All protein samples were analyzed by SDS-PAGE to assess their purity.

Recombinant *S. cerevisiae* Lsm2-8 complex was expressed in *E. coli* from a pQLink vector containing all seven yeast *Lsm* genes (a kind gift from Yigong Shi, Tsinghua University) and purified essentially as described (25) but was not subjected to gel filtration.

4.3.2 RNA synthesis

In vitro transcription was used to synthesize RNAs corresponding to *S. cerevisiae* U6 nucleotides 30-101 with a U100C/U101C double mutation, full-length wild-type and mutant U6 (nucleotides 1-112), full-length U4 (nucleotides 1-160), and a portion of human U1 (nucleotides 1-143). The U6 (30-101) U100C/U101C construct was transcribed from synthetic DNA oligonucleotide templates (Integrated DNA Technologies). Full-length U4 (nucleotides 1-160) was transcribed off a modified pUC118 plasmid template (a kind gift from P. Fabrizio) containing the T7 polymerase promoter sequence, two additional G nucleotides at the beginning of the transcript

(for efficient T7 transcription) and a BsaI restriction site at the end of the transcript to allow for run-off transcription. Full-length U6 (nucleotides 1-112) was transcribed off a modified pUC57 plasmid containing the T7 polymerase promoter sequence with an additional G nucleotide at the beginning of the transcript (for efficient T7 transcription) and an HDV ribozyme sequence at the end of the transcript to confer a homogeneous 3' terminus, followed by a BamHI restriction site to allow run-off of the polymerase. A portion of human U1 RNA (nucleotides 1-143) was transcribed off a pUC18 plasmid (a kind gift from A. Hoskins) containing the T7 polymerase promoter sequence, two additional G nucleotides at the beginning of the transcript, an HDV ribozyme sequence, and a BamHI restriction site for run-off transcription. All plasmid templates were linearized with either BsaI or BamHI (New England Biolabs) prior to transcription.

All RNAs were transcribed *in vitro* using recombinant His_6-tagged T7 RNA polymerase (26,27) in 40 mM TrisCl pH 8.0, 1 mM spermidine, 0.01% Triton X-100, 38 mM $MgCl_2$, 5 mM DTT, and 5 mM each of ATP, CTP, GTP and UTP. RNAs were purified from abortive transcripts, linearized plasmid, and the HDV ribozyme using an 8% 29:1 acrylamide: bis-acrylamide denaturing gel containing 8M urea, 89 mM Tris borate, 2 mM EDTA. RNA was visualized using UV shadowing and extracted from the gel by passive diffusion into 0.3 M sodium acetate pH 5.2. RNA was ethanol precipitated and resuspended in water.

4.3.3 RNA labeling

RNAs were dephosphorylated prior to 5' end labeling by incubating 10 pmol of RNA at 37°C with 10 units of CIP (New England Biolabs) in 50 mM potassium acetate, 20 mM Tris-acetate pH 7.9, 10 mM magnesium acetate, 100 µg/mL BSA. This reaction was extracted with phenol:chloroform:isoamyl alcohol 25:24:1 and ethanol precipitated. After precipitation, RNAs were pelleted by centrifugation and resuspended in the 5' end labeling reaction consisting of 10 units of T4 PNK, 70 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 5 mM DTT, and 0.03 µCi of [γ-^{32}P] ATP (3,000 Ci/mmol). The reaction was stopped by the addition of an equal volume of urea loading

dye and purification on an 8% denaturing gel. RNAs were extracted from the gel by passive diffusion into 300 mM sodium acetate pH 5.2 and ethanol precipitated.

Fluorescently-labeled U4 and U6 were made via ligation of an *in vitro* transcribed RNA to a synthesized fluorescently labeled RNA (IDT) using T4 RNA ligase 2. An RNA oligonucleotide consisting of nucleotides 1-12 of U6 with a 5'-Cy3 label was ligated to nucleotides 13-112 of U6 transcribed from a plasmid containing a 5' hammerhead ribozyme (to confer a homogeneous 5' terminus for ligation) and a 3' HDV ribozyme (modified from a kind gift by Kiyoshi Nagai). Fluorescent U4 was created by ligation of a 5'-Cy5 labeled RNA consisting of U4 nucleotides 1-13 to an *in vitro* transcribed U4 containing nucleotides 14-160 (modified from the pUC118 plasmid described previously). After *in vitro* transcription, U4(14-160) was modified as previously described using CIP and T4 PNK (NEB). U6(13-112) was modified using only T4 PNK due to the presence of a 5'-OH group in the original transcript. RNAs were ligated using DNA splints and T4 RNA ligase 2 and purified by urea PAGE as above (28).

4.3.4 Binding and annealing buffer preparation

A stock buffer containing 100 mM KCl, 20% glycerol,10 mM HEPES acid, 10 mM sodium HEPES base, 1 mM EDTA acid, 1 mM TCEP-HCl, 0.01% Triton X-100, pH~ 7 was prepared. A “2x protein dilution buffer” was made by the addition of BSA to 0.2 mg/mL, and a “2x RNA dilution buffer” was made by the addition of yeast tRNA to 0.2 mg/mL and sodium heparin to 0.02 mg/mL. All buffer stock solutions were passed through a HiTrapSP cation exchange column (GE Healthcare) to remove trace levels of ribonuclease activity. The heparin component of the RNA dilution buffer was added after the buffer had been passed though the column. The final “binding buffer” containing 1x RNA dilution buffer and 1x protein dilution buffer was used for all binding and annealing experiments.

4.3.5 Gel shift assay

Binding of RNAs with Prp24 was performed with trace (<1 nM) [^{32}P]-labeled RNA with variable concentrations of protein. RNAs were heated to 90°C for 2 minutes in RNA binding buffer,

then snap cooled on wet ice. Proteins were prepared as 2x stocks in protein binding buffer. The affinity of Prp24 for RNAs was determined using a 10 µL binding reaction prepared at room temperature containing 5 µL of RNA and 5 µL of 2x protein stock. Samples were incubated at room temperature for 20 minutes prior to loading onto a 16.5 cm x 22 cm 6% polyacrylamide gel (29:1 acrylamide:bis-acrylamide, 89 mM Tris borate, 2 mM EDTA pH 8.0). Samples were electrophoresed for 2-3 hours at 150V at 4°C. Radioactive gels were dried on BioRad filter paper, exposed to a PhosphorImager screen, and imaged on a Typhoon FLA 9000 biomolecular imager. Results were analyzed using ImageJ software and binding curves were fit using nonlinear regression in GraphPad Prism 4 to the Hill equation: % bound = (B_{max} * $[Prp24]^H$) / (K_d^H + $[Prp24]^H$). B_{max} was restrained to be between 0 and 100%, and the H (Hill coefficient) and K_d were restrained to be greater than 0. Binding affinities are reported for three technical replicates.

Protein activity was determined as previously described (29) using 5 nM 5'-Cy3 labeled U6 RNA supplemented with 500 nM unlabeled full-length U6. Stoichiometric amounts of protein were added and binding reactions were treated as described above. Fluorescent gels were imaged directly through low fluorescence glass plates (CBS Scientific) on a Typhoon FLA 9000. The percent of labeled U6 in U6/Prp24 was analyzed using ImageJ software and the binding curve of protein concentration vs. percentage bound was plotted (GraphPad Prism 4). The linear region of this curve (protein:RNA stoichiometry < 2) was fit to a linear regression. The activity of the protein is expressed as the slope of this line.

4.3.6 U4/U6 annealing assay

Annealing reactions were carried out at 30°C in 10 µL reactions in binding buffer (as described above) containing <1 nM ^{32}P-labeled U4 RNA, 25 nM U6 RNA and 250 nM Prp24 protein. Reactions were stopped by the addition of 2 µL of proteinase K buffer (0.5% SDS, 0.3 mg/mL tRNA, 5 mM $CaCl_2$, 30 mM HEPES pH 7.0, 0.2 mg/mL proteinase K) or by separation on a 6% polyacrylamide gel (29:1 acrylamide:bis-acrylamide, 89 mM Tris borate, 2 mM EDTA pH 8.0) that had been pre-run at 150V for at least 30 minutes. Samples were electrophoresed for 3

hours at 150 V at 4°C, dried, and analyzed as above. Annealing rates were calculated using the ratio of free U4 to U4/U6 in Proteinase K-treated lanes at 0, 15, 30, 60, and 90 minutes; resulting data were then fit to a one-phase exponential association equation (GraphPad Prism 4). Annealing rates are reported for three technical replicates.

4.4 Results

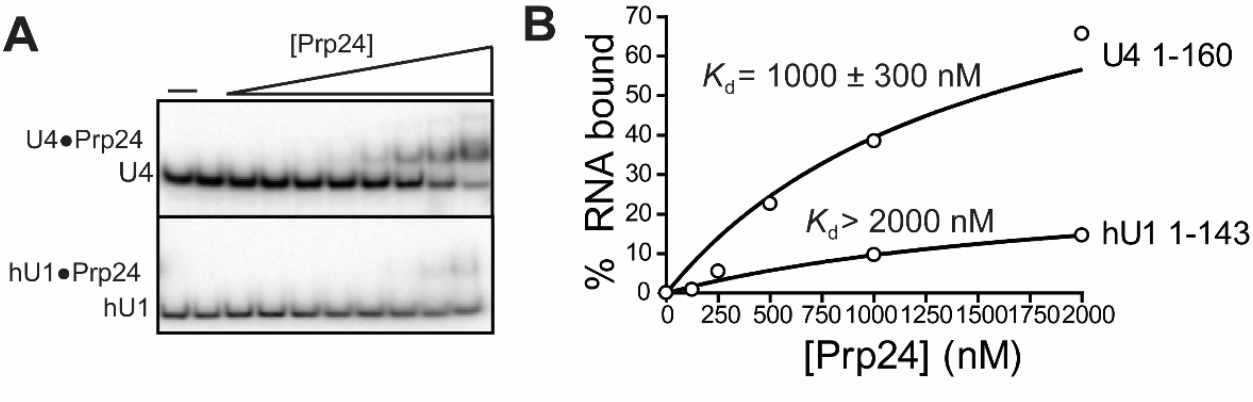
4.4.1 Prp24 binds U6 snRNA with much higher affinity than U4 snRNA

Prp24 protein is capable of binding both U6 and U4 RNAs (**Figure 4-1C**), as previously observed (16). Previous studies of U6-Prp24 binding via gel shift showed the appearance of higher order complexes at high Prp24 concentrations, presumably due to non-specific binding events (18). We observed formation of multimeric Prp24 binding is suppressed by the inclusion of tRNA, heparin, and BSA in the buffer conditions (29). Using these conditions, we demonstrated that full-length Prp24 has a K_d of 21 ± 9 nM for full-length U6 RNA, similar to the value of 43 ± 11 nM reported previously (18) (**Figure 4-1C**). The tight binding of Prp24 to U6 RNA is in agreement with the large number of protein-RNA contacts present in the crystal structure of the U6•Prp24 complex, as twenty nucleotides in U6 are directly contacted by Prp24 (24). Interestingly, Prp24 binds U6 with a Hill coefficient of 0.50 ± 0.07. This may be due to structural heterogeneity in the RNA population, with one conformation having a higher affinity for Prp24 than others. Prp24 has a ~50-fold lower affinity of 1000 ± 300 nM for full-length U4 RNA under these conditions, with a Hill coefficient near unity (**Figure 4-1C**). Although Prp24 does display relatively weak binding to U4, this binding is not nonspecific, as Prp24 displayed tighter binding to U4 than to a control RNA of similar size (human U1 snRNA nucleotides 1-143) (**Figure 4-2**).

Figure 4-2. Prp24 binds U4 with greater affinity than a control RNA of similar length.

A) Native gel analysis comparing Prp24 binding to U4 RNA (top) and a non-specific control, human U1 RNA (bottom).

B) Quantitation of Prp24 binding to U4 and hU1 RNAs. Prp24 binds U4 with a K_d of 1000 nM, while the binding of Prp24 to hU1 is significantly weaker and cannot be accurately quantified.



4.4.2 An annealing assay that preserves assembled RNPs

Previous gel-shift based approaches for determining U4/U6 annealing rates used treatment with proteinase K in order for U4/U6 RNA complexes to enter the gel, and so did not directly assess RNP formation (15,16). To define the RNPs formed in the annealing reaction, we employed fluorescently labeled U4 and U6 RNAs and our optimized conditions for specific binding. We observed formation of U6•Prp24 and U4•Prp24 complexes, as well as a U4/U6•Prp24 ternary complex when both RNAs were present (**Figure 4-3A**). Upon treatment of the annealing reaction with proteinase K, the U4•Prp24 and U6•Prp24 binary complexes resolve into free U4 and U6 RNA, and the U4/U6•Prp24 complex resolves into U4/U6 di-snRNA, showing that the U4 and U6 RNAs are base-paired in the U4/U6•Prp24 complex.

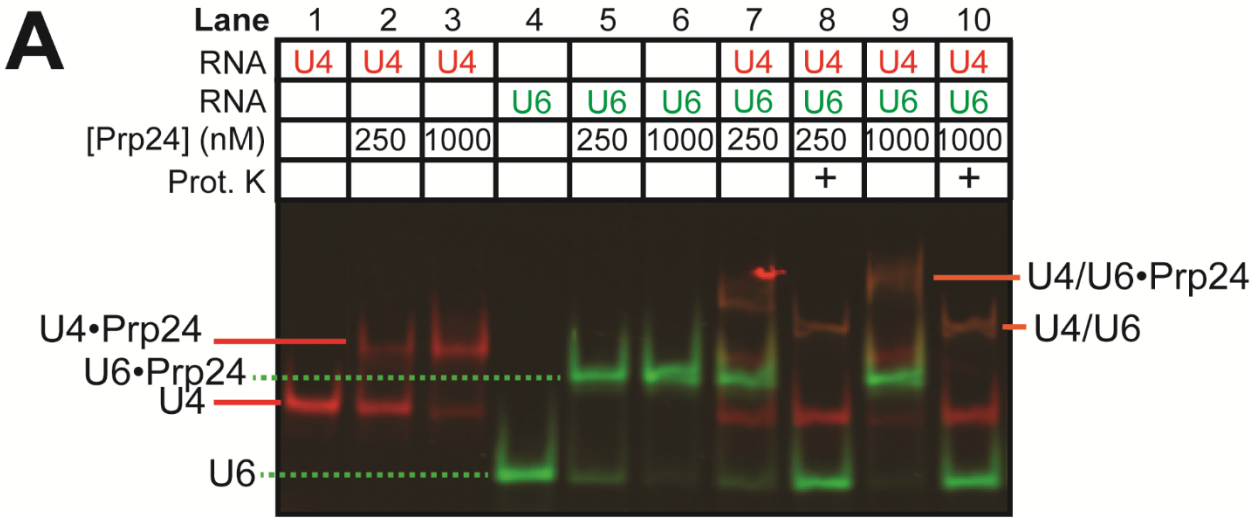
Having established the identity of the U4/U6•Prp24 complex, we followed annealing kinetics using radiolabeled U4 RNA and excess unlabeled U6 RNA (**Figure 4-3B**). In our assay, U6 was pre-bound by excess Prp24 and the annealing reaction was started by the addition of U4. Under these conditions, we found that Prp24 accelerates the rate of U4/U6 annealing at 30°C at least 20-fold (**Figure 4-3C**). In non-deproteinized lanes, we observed formation of a small amount U4•Prp24 binary complex, which is expected based on the observed K_d for U4 and the fact that Prp24 is in excess in the assay. Formation of a Prp24•U4 binary complex is not a prerequisite for annealing *in vitro*, as titration of Prp24 displayed a $K_{1/2}$ for annealing of 20 nM (**Figure 4-4**), almost two orders of magnitude lower than the K_d for U4•Prp24 binding (**Figure 4-1C**), and consistent with the K_d for U6. Interestingly, at 10 nM Prp24 and without proteinase K digestion, about half of the product U4/U6 RNA complex is bound to Prp24 (**Figure 4-4**). Thus, the K_d of Prp24 for U4/U6 under these conditions must be about 10 nM, close to the 18 nM affinity observed by Ghetti *et al*. (16).

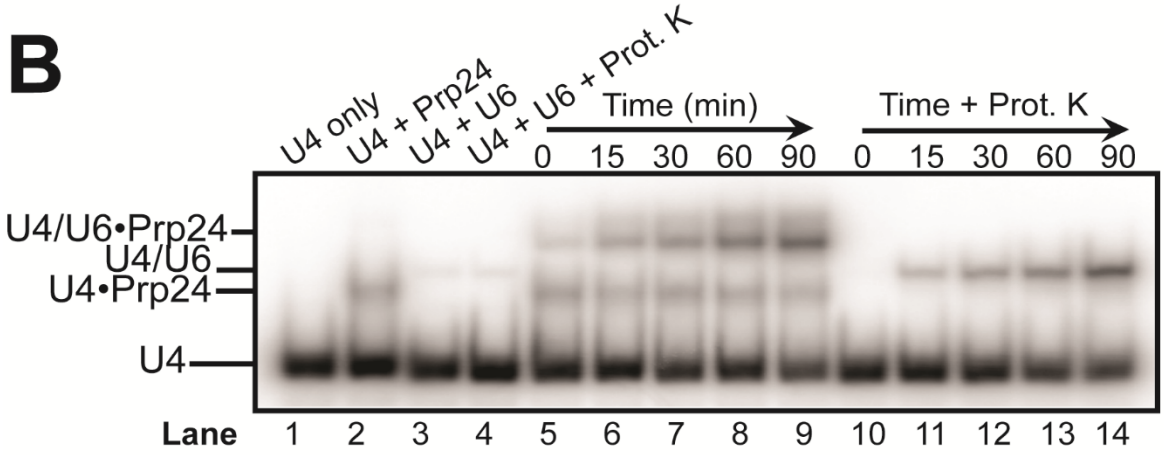
Figure 4-3. Prp24 catalyzes annealing of U4 and U6 RNAs, and remains bound to product di-RNA.

A) Two-color gel demonstrating tight binding of Prp24 to Cy3-U6 (lanes 5 and 6) and weaker binding of Prp24 to Cy5-U4 (lanes 2 and 3). Co-localization of Cy3 and Cy5 fluorescence in the presence of U4 and U6 (lanes 7 and 9) shows that the slowest-migrating species (orange) contains both RNAs, and the increased mobility upon treatment with proteinase K (lanes 8 and 10) shows that the di-snRNA retains bound Prp24. Annealing reactions were incubated at 30°C for 90 minutes prior to loading.

B) Time-dependent formation of U4/U6, using radiolabeled U4 snRNA and unlabeled U6 and Prp24. Control reactions in lanes 1-4 were incubated for 90 minutes.

C) Quantification of Prp24-dependent annealing from proteinase K treated lanes (10-14) in (B) compared to protein-independent annealing (lane 4 in B).


A
Lane 1 2 3 4 5 6 7 8 9 10
RNA U4 U4 U4 U4 U4 U4
RNA U6 U6 U6 U6 U6 U6
[Prp24] (nM) 250 1000 250 1000 250 250 1000 1000
Prot. K + +
U4•Prp24
U6•Prp24
U4
U6
U4/U6•Prp24
U4/U6


B
U4 only
U4 + Prp24
U4 + U6
U4 + U6 + Prot. K
Time (min)
0 15 30 60 90
Time + Prot. K
0 15 30 60 90
U4/U6•Prp24
U4/U6
U4•Prp24
U4
Lane 1 2 3 4 5 6 7 8 9 10 11 12 13 14

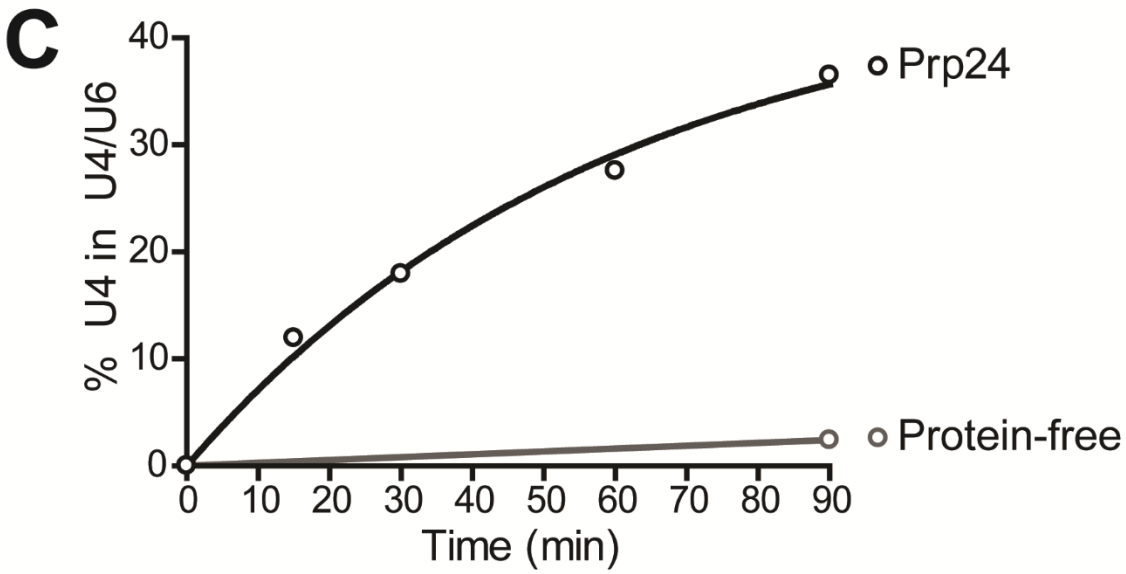
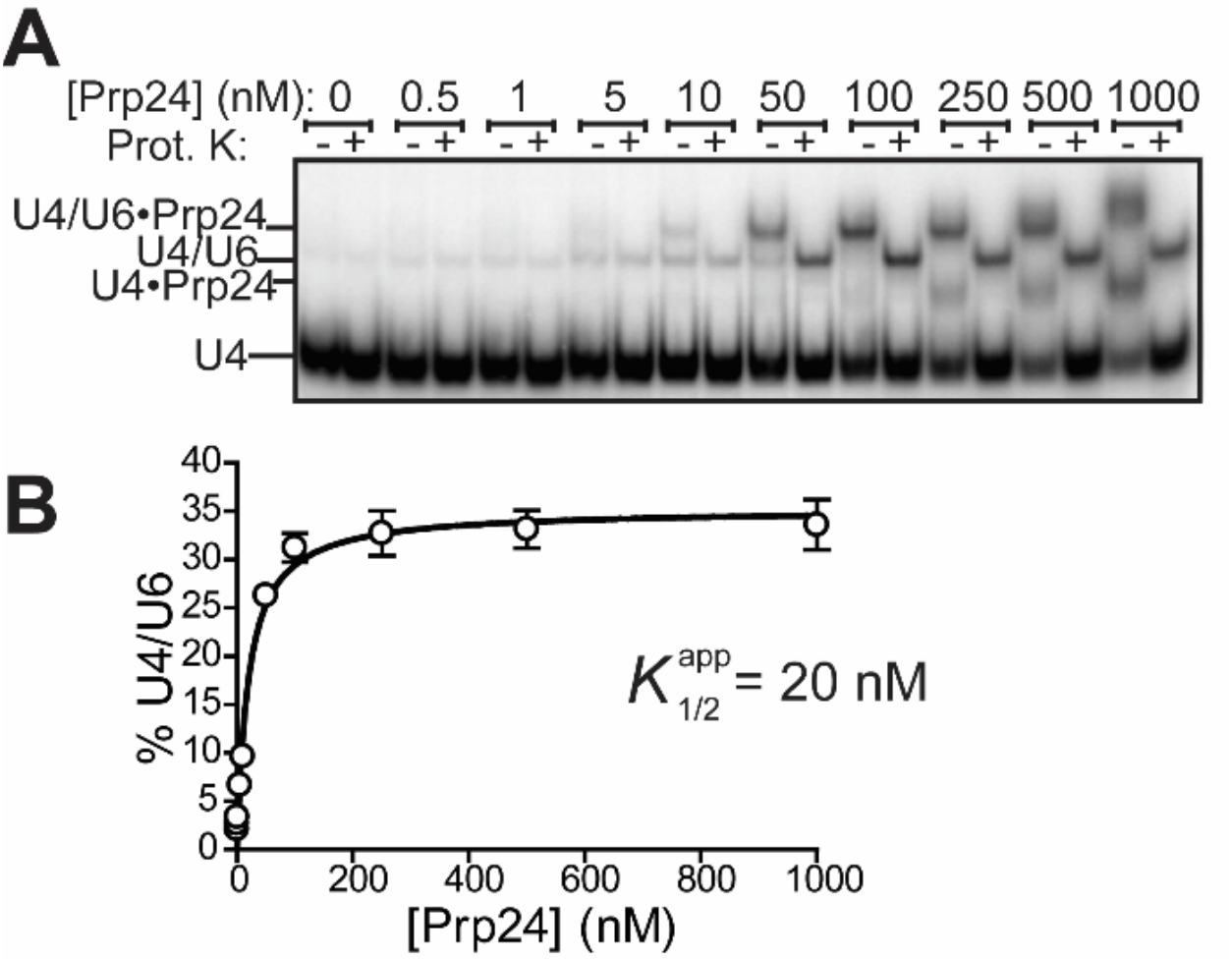

C
% U4 in U4/U6
40
30
20
10
0
0 10 20 30 40 50 60 70 80 90
Time (min)
Prp24
Protein-free

Figure 4-4. Effect of Prp24 concentration on the rate of annealing.

A) Native gel showing the formation of U4/U6 at a single time point (90 minutes) in the presence of increasing concentrations of full-length Prp24. Samples were divided into intact ribonucleoprotein (minus lanes) and proteinase K-treated (plus lanes) samples. B) Quantification of U4/U6 formation after 90 minutes for varying concentrations of Prp24.



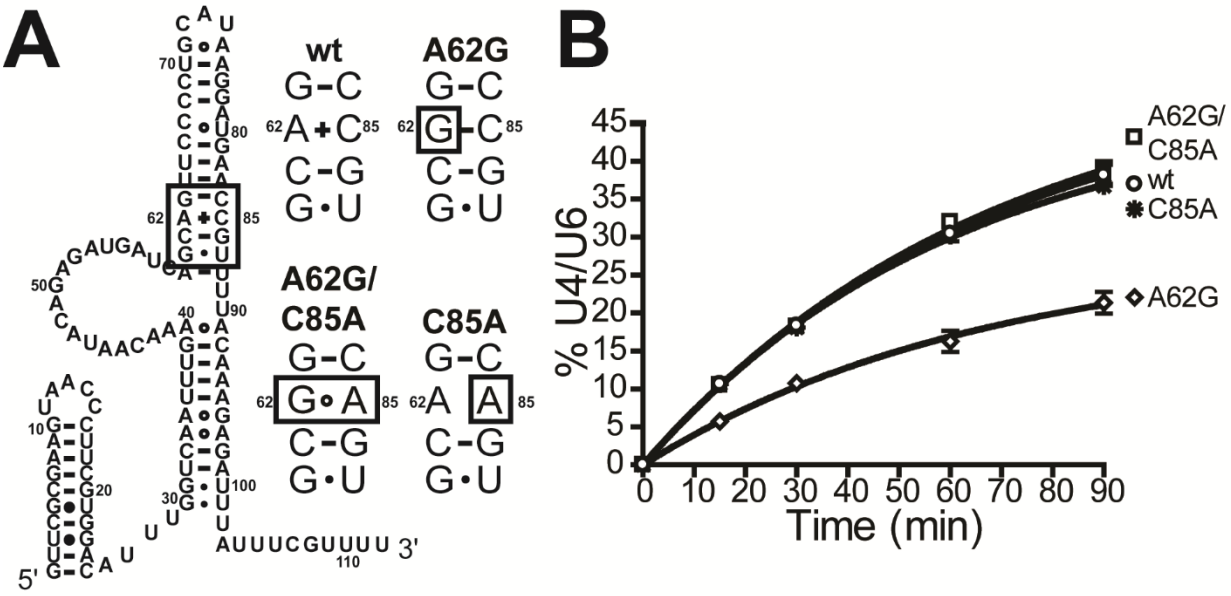
4.4.3 The *in vitro* annealing assay faithfully recapitulates *in vivo* phenotypes of U6 ISL mutations

Mutation of U6-A62 to G was previously reported to reduce the levels of U4/U6 RNA complex in yeast cells grown at 30°C or 18°C and cause a cold-sensitive growth defect (9). This mutation changes an A-C mismatch at the base of ISL into a stable G-C base pair (**Figure 4-5A**). The A62G annealing and growth defects are corrected by *cis*-acting suppressors that reintroduce a mismatch at this position and restore the stability of the ISL to near-wild type levels (9). We tested U6-A62G and one of its cis-acting suppressors, U6-C85A, which converts the G-C pair formed by U6-A62G into a G-A mismatch and was previously demonstrated to reverse the *in vivo* U4/U6 annealing defect and cold sensitivity. Both the A62G annealing defect and its suppression by C85A were recapitulated in the *in vitro* assay (**Figure 4-5B**). U6-A62G results in a 2-fold decrease in the rate of annealing, and this annealing defect is corrected by inclusion of U6-C85A. U6-C85A alone did not change the annealing rate in comparison to wild-type U6. Thus, mutations that induce the formation of a stable base pair at the base of the U6-ISL are deleterious to Prp24-catalyzed U4/U6 annealing, and can be rescued by suppressors that disrupt stable pairing, both *in vitro* and *in vivo*.

Figure 4-5. The *in vitro* annealing assay recapitulates *in vivo* phenotypes of U6 RNA substitutions.

A) Secondary structure of U6 RNA bound to Prp24 (24), with the boxed region of the ISL shown with substitutions at right.

B) Time course of U4/U6 annealing for reactions containing the indicated substitutions in U6.



4.4.4 The components of the U6 snRNP core are sufficient for annealing to U4 snRNA

We sought to determine if the truncated forms of U6 and Prp24 present in the crystallized U6 snRNP core structure are sufficient for U4/U6 annealing. The crystallized complex contains Prp24 residues 34-400, including all four RNA recognition motifs but lacking the unstructured N- and C-terminal domains (22,30), and U6 nucleotides 30-101, with stabilizing mutations A62G, U100C, and U101C (24). Since the A62G substitution is detrimental to annealing (**Figure 4-5**) and not required for crystallization (our unpublished data), this substitution was not included in our analysis. Truncation of Prp24 to remove the N- and C-terminal domains (Prp24ΔN,C) had no effect on U4/U6 annealing (**Figure 4-6A,C**). The U4•Prp24 complex is more pronounced in annealing reactions containing the truncated protein because truncation of Prp24 increased its affinity for U4 approximately 3-fold (**Figure 4-7**). Interestingly, truncation and mutation of U6 RNA stimulated annealing (**Figure 4-6B,C**).

To investigate the basis of the rapid annealing of U6 30-101 (U100C/U101C), we tested the separate contributions of terminal truncation and telestem stabilization to the annealing rate. Truncation of the 3' end of U6 RNA increased the annealing rate 2- to 3-fold, while 5' truncations have little or no effect (**Figure 4-8**). In contrast, stabilization of the telestem within full-length U6 via the U100C/U101C substitutions accelerated annealing approximately ten-fold, a rate enhancement similar to that observed in the annealing reaction using the crystal construct (**Figure 4-9**). Since the U100C/U101C mutation is expected to increase telestem stability by replacing two terminal G-U wobble pairs with two G-C pairs, we hypothesized that stabilization of the telestem may be responsible for increasing the annealing rate. Consistent with this hypothesis, destabilization of the telestem with mutation U37C, which changes a U-A base-pair into a C-A mismatch, reduced the rate of annealing approximately two-fold (**Figure 4-9**).

Figure 4-6. The components of the U6 snRNP core efficiently promote annealing.

A) Full-length (residues 1-444) and truncated (residues 34-400) Prp24 catalyze U4/U6 annealing with similar rates. Lane 1 contains U4 RNA with full-length Prp24 and lane 2 contains U4 with truncated Prp24. Lanes 3-18 contain full-length U6 RNA (unlabeled), U4 RNA, and the indicated form of Prp24 incubated 0-90 minutes at 30°C. Samples in lanes 7-10 and 15-18 were treated with proteinase K before electrophoresis.

B) Truncated U6 RNA (30-101) with two stabilizing mutations (U100C/U101C) anneals more rapidly than full length U6 (1-112). Lanes 1 and 2 contain only U4 and full-length U6, while lanes 3 and 4 contain U4 and truncated U6. Lanes 2 and 4 are proteinase K treated. Lanes 5-20 correspond to lanes 3-18 of Panel A, but with full length Prp24 and full length or truncated U6 as indicated.

C) Annealing timecourses for experiments shown in Panels A and B.

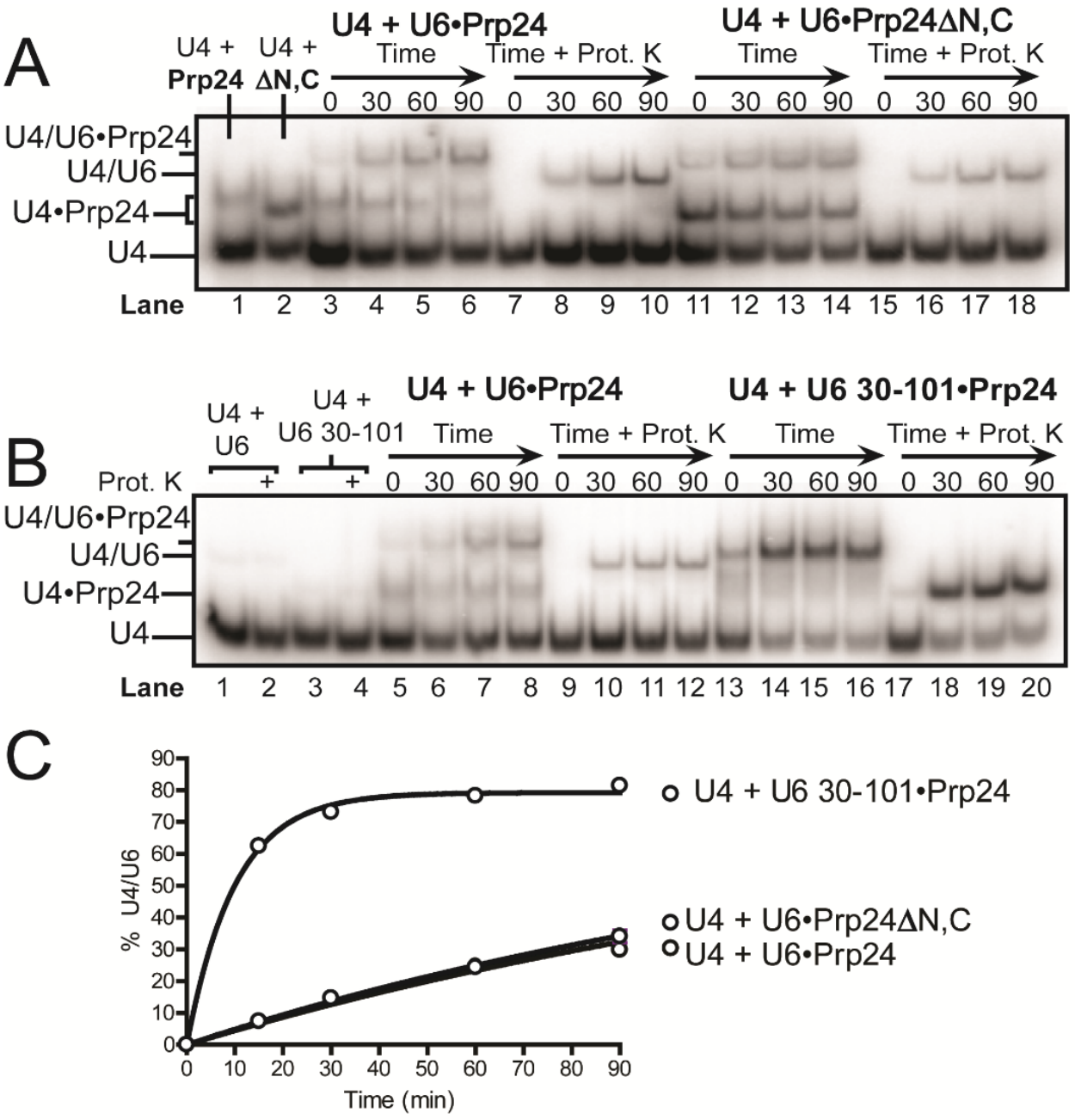

A
U4 + U6•Prp24
U4 + U6•Prp24ΔN,C
U4 + Prp24
U4 + ΔN,C
Time
Time + Prot. K
0 30 60 90
U4/U6•Prp24
U4/U6
U4•Prp24
U4
Lane 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
B
U4 + U6•Prp24
U4 + U6 30-101•Prp24
U4 + U6
U4 + U6 30-101
Prot. K
+
Lane 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
C
% U4/U6
Time (min)
U4 + U6 30-101•Prp24
U4 + U6•Prp24ΔN,C
U4 + U6•Prp24

Figure 4-7. Truncation of Prp24 protein enhances binding to U4 RNA.

A) Native gel analysis of U4-Prp24 binding. Full-length Prp24 (top) and truncated Prp24 (bottom) both bind U4 RNA.

B) Truncated Prp24 (ΔN,C) binds U4 with a K_d of 320 nM, while full-length Prp24 binds with a K_d of 1000 nM.

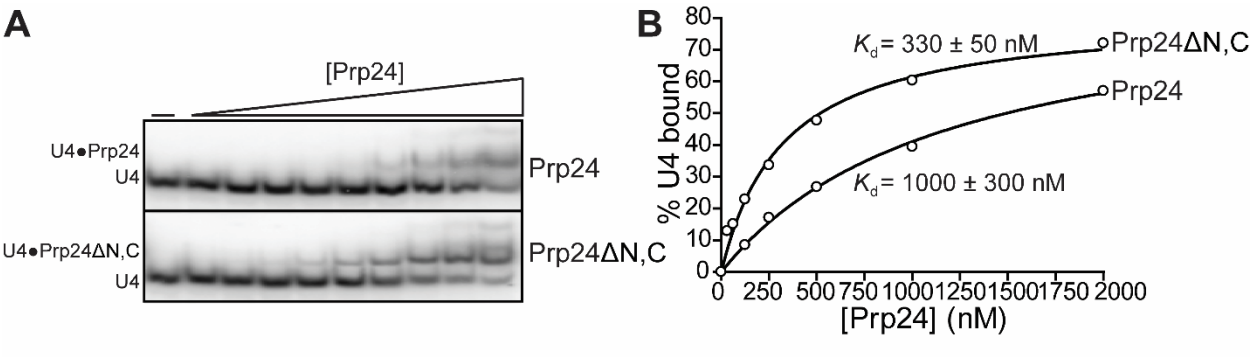


Figure 4-8. Truncation of U6 RNA does not impede U4/U6 annealing.

A) Secondary structure of U6 showing the regions deleted.

B) Quantification of annealing rates for truncations of U6.

C-F) Annealing gels comparing wild-type U6 to various truncations of U6. In all gels, a control lane containing the annealing reaction at 90 minutes that has not been proteinase K-treated precedes a time course of annealing reactions treated with proteinase K.

C) Truncation of U6 to nucleotides 13-112, disrupting the 5' stem-loop.

D) Truncation of U6 to nucleotides 22-112, removing the 5' stem-loop.

E) Truncation of U6 to nucleotides 30-112, removing the 5' stem-loop and linker region.

F) Truncation of U6 to nucleotides 1-101, removing the 3' single stranded region.

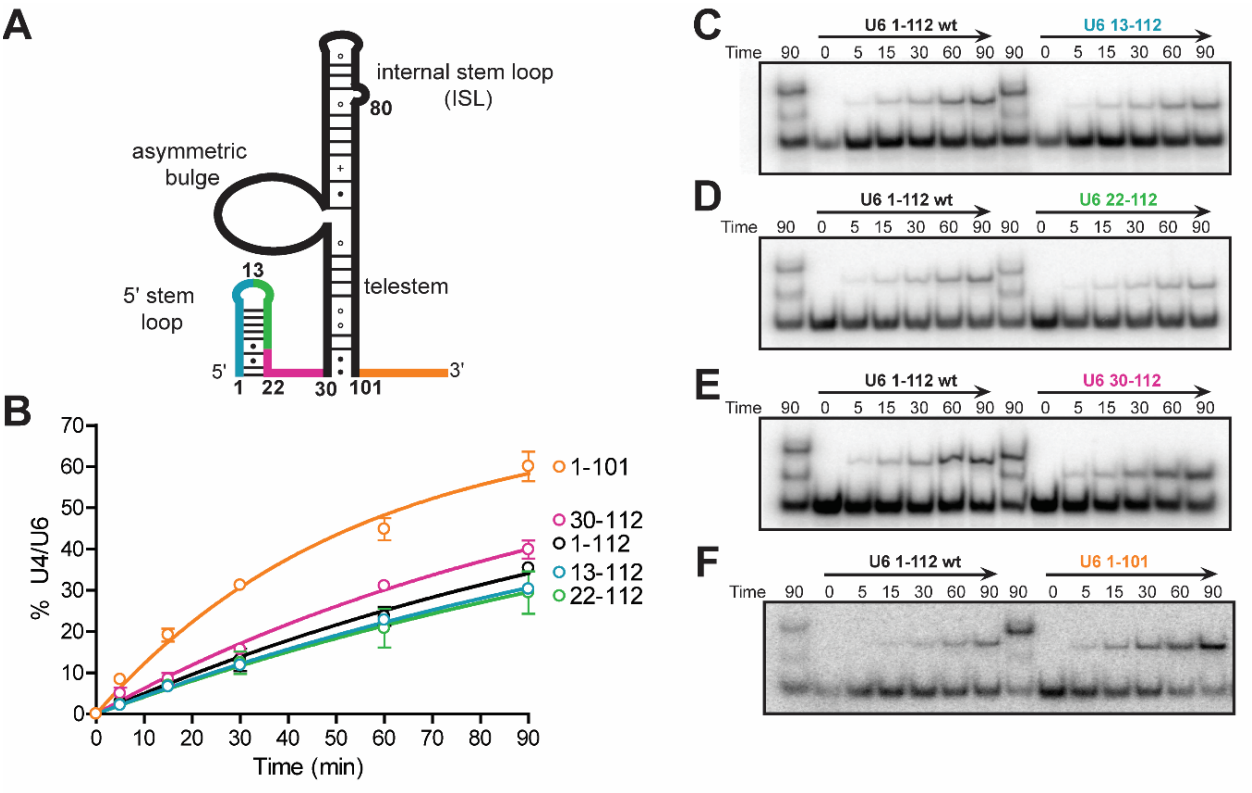
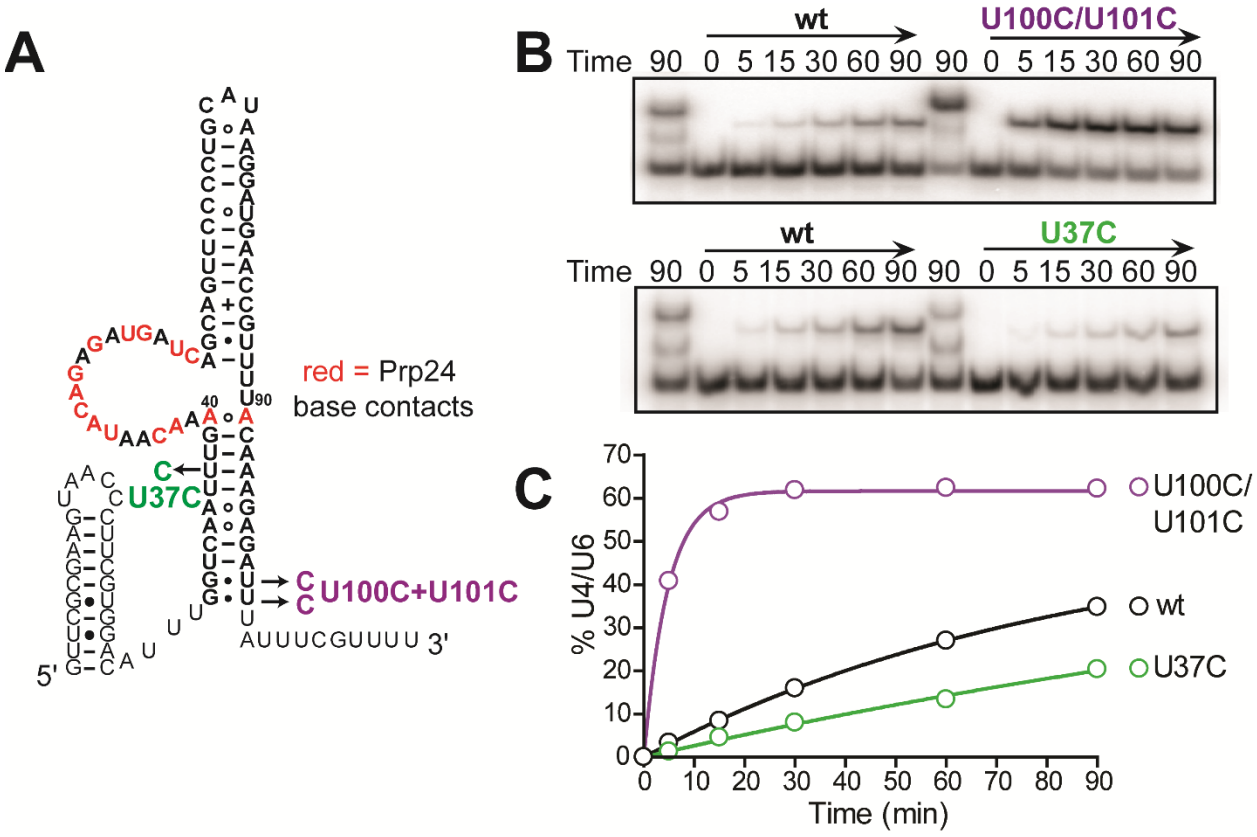


Figure 4-9. Mutations within the telestem affect annealing rate.

A) Secondary structure of U6 showing the position of telestem-stabilizing (U100C/U101C) or destabilizing (U37C) mutations. Base specific contacts with Prp24 are highlighted in red.

B) Annealing gel comparing full-length wild-type U6 to U6-U100C/U101C (top) or U6-U37C (bottom). A control lane containing the annealing reaction at 90 minutes that was not proteinase K-treated precedes each time course of annealing reactions treated with proteinase K.

C) Annealing timecourses for reactions containing wild-type, U100C/U101C, and U37C variants of full-length U6.



4.4.5 Both U6 telestem stability and sequence influence the U4/U6 annealing rate

In order to determine if telestem stability generally correlates with annealing rates, we systematically altered the pairing potential of the six base pairs at the base of the telestem (**Figure 4-10A**). A clear general trend is that stabilization of the telestem results in a faster annealing rate (**Figure 4-10B**).

In addition, we found that mutations U100G/U101G in combination with any nucleotides at positions 30 and 31 of U6 - either stabilizing or destabilizing - increase the rate of annealing (**Figure 4-10B and C**, asterisks). Intriguingly, these mutations substantially increased the rate of annealing even in the absence of Prp24 (**Figure 4-10C**), thus accounting for the anomalous kinetic behavior of apparent outlier mutants 4 and 7. RNA secondary structure prediction of the U100G/U101G mutations via MFold analysis (31) suggests that the U100G/U101G mutations induce an alternative fold of U6 RNA that destabilizes the ISL, which could account for the observed increase in protein-free annealing rates (**Figure 4-11A, B**). In this potential alternate fold, nucleotides 64-68 are single stranded in the loop of a very short helix. These nucleotides form part of U4/U6 Stem II and in wild-type U6 are base-paired in the U6 ISL. Destabilization of the ISL via stabilization of this alternate fold may drive the annealing reaction forward in the absence of Prp24 by exposing nucleotides involved in U4/U6 base pairing. In the presence of Prp24, these mutations decreased the electrophoretic mobility of U4/U6•Prp24 (**Figure 4-10D**). This difference in mobility is likely due to multiple copies of Prp24 bound to U4/U6, as increasing the concentration of Prp24 caused the formation of discrete higher bands in U4/U6-U100G/U101G, but not wild-type U4/U6 (**Figure 4-11C**). Despite the difference in U4/U6-Prp24 binding, the U100G/U101G mutation did not change U6-Prp24 binding affinity and retained a Hill coefficient of about one-half, similar to wild-type RNA (**Figure 4-11D,E**).

The behavior of the U100G/U101G-containing mutants is in stark contrast to the behavior of a hyperstabilized U6 mutant 17, in which the last 6 base pairs of the telestem are all G-C or C-G (**Figure 4-10A**). These mutations accelerated the rate of Prp24-mediated annealing

approximately ten-fold, but no enhancement of protein-independent annealing was observed and the resulting U4/U6-Prp24 ternary complex retained wild-type electrophoretic mobility (**Figure 4-10D,** see lane marked 17). Interestingly, U6 mutant 17 RNA exhibited a ~10-fold tighter K_d for Prp24 binding than wild-type RNA (**Figure 4-12, Table 4-1**) with a Hill coefficient near unity (H = 1.3 ± 0.1 versus H = 0.50 ± 0.07 for wild-type) (**Figure 4-12A,B**). U6 mutant 17 also displayed an approximately 5-fold slower off-rate of Prp24 (**Figure 4-12C,D, Table 4-1**). The slower off-rate for mutant 17 largely accounts for the lower K_d and is consistent with the observed interlocked topology of the complex and the hypothesis that the telestem must unfold in order for Prp24 to dissociate (24). The mutant 17-Prp24 complex also had a lower electrophoretic mobility than wild-type U6-Prp24 (**Figure 4-12A,C**). A possible explanation for this behavior is that the wild-type RNA may be in conformational exchange with an alternatively folded, higher mobility species.

From the measured off-rates and dissociation constants, we can calculate that the apparent on-rates are approximately 5 orders of magnitude slower than diffusion (**Table 4-1**). The slow on-rates are also consistent with the interlocked topology of the RNP, which requires the intricate co-folding of RNA and protein (24).

Figure 4-10. Stabilization of the U6 telestem results in significant rate enhancement of U4/U6 annealing.

A) Secondary structure of the lower telestem, with tested mutations boxed.

B) U4/U6 annealing rates of mutant U6 RNAs. Mutations predicted to destabilize the telestem (numbers 1-6) or to be isoenergetic with wild type RNA (nos. 7-9) had little effect on annealing rate, while those predicted to be slightly stabilizing (nos. 10-12) or stabilizing (nos. 13-17) significantly increased the annealing rate.

C) Protein-free annealing of U4 and U6 RNAs at 90 minutes. The percentage of U4 incorporated into U4/U6 is shown. Mutations that enhance the protein-free annealing rate are marked with an asterisk.

D) Electrophoretic mobility of U4/U6•Prp24 species after 90 minutes of annealing.

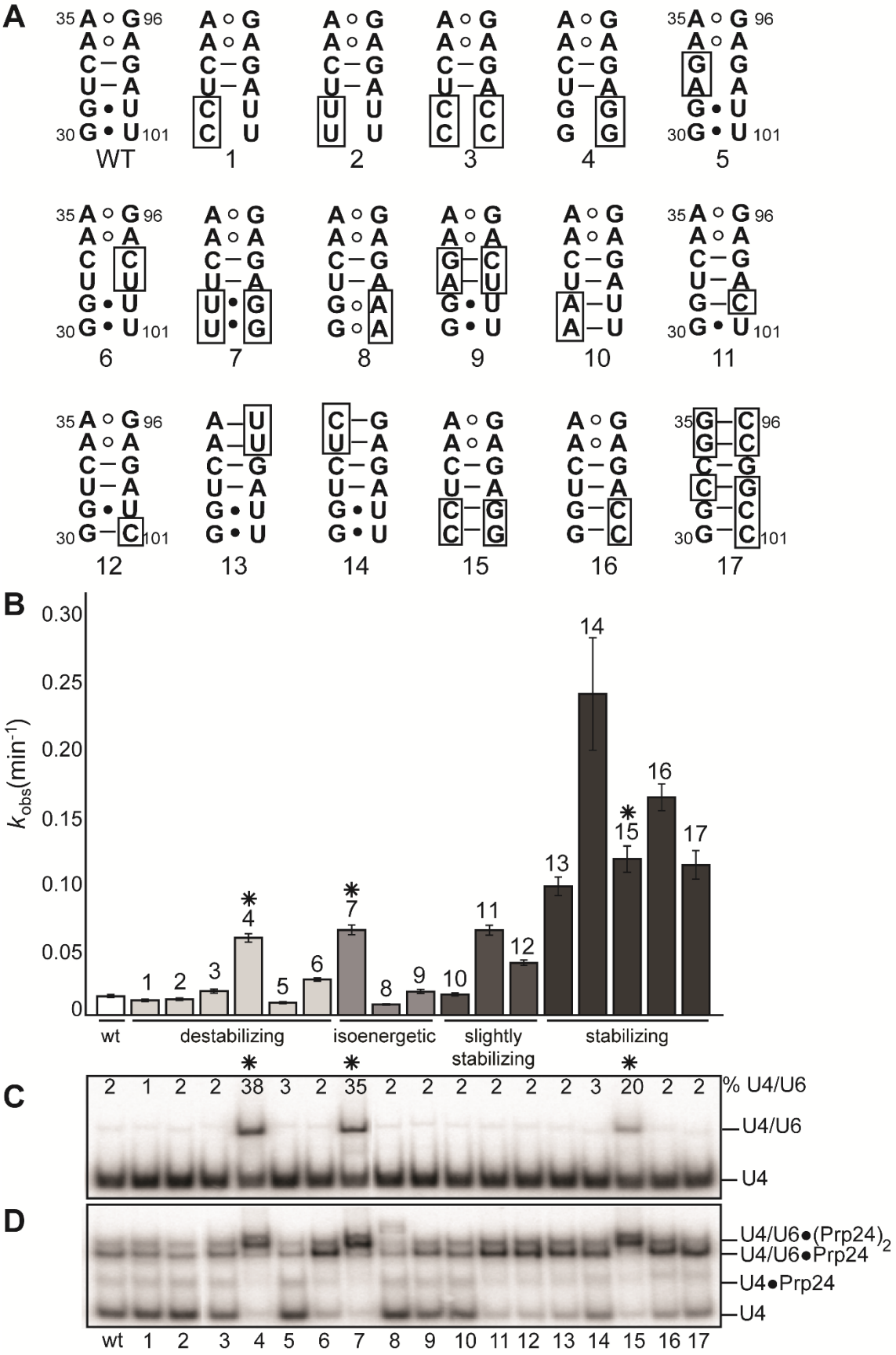

A
WT
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
B
k_{obs}(min^{-1})
0.30
0.25
0.20
0.15
0.10
0.05
0
wt
destabilizing
isoenergetic
slightly stabilizing
stabilizing
C
% U4/U6
U4/U6
U4
D
U4/U6•(Prp24)$_2$
U4/U6•Prp24
U4•Prp24
U4
wt 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

Figure 4-11. U100G/U101G mutations may stabilize an alternative fold of U6.

A) Secondary structure of wild-type RNA (left) is predicted to be in equilibrium with a more stable alternative secondary structure in the absence of Prp24. MFold predicted free energies are indicated below secondary structures. The ISL region is shown in red.

B) The U100G/U101G mutations are predicted to further stabilize the putative alternative fold, which may account for the observed protein-free annealing rates.

C) Titration of Prp24 reveals multiple binding sites on U4/U6-U100G/U101G (mutant #4). U4 and U6 (wt vs. U100G/U101G) were incubated for 90 minutes with increasing concentrations of Prp24. In lanes 17 and 18, 250 nM Prp24 was incubated with U4 and U6 for 90 minutes and then Proteinase K treated.

D) Native gel analysis of U6-Prp24 binding comparing wild type (left) and U6-U100G/U101G (right).

E) Prp24-binding curves of wild-type and mutant U6 RNA.

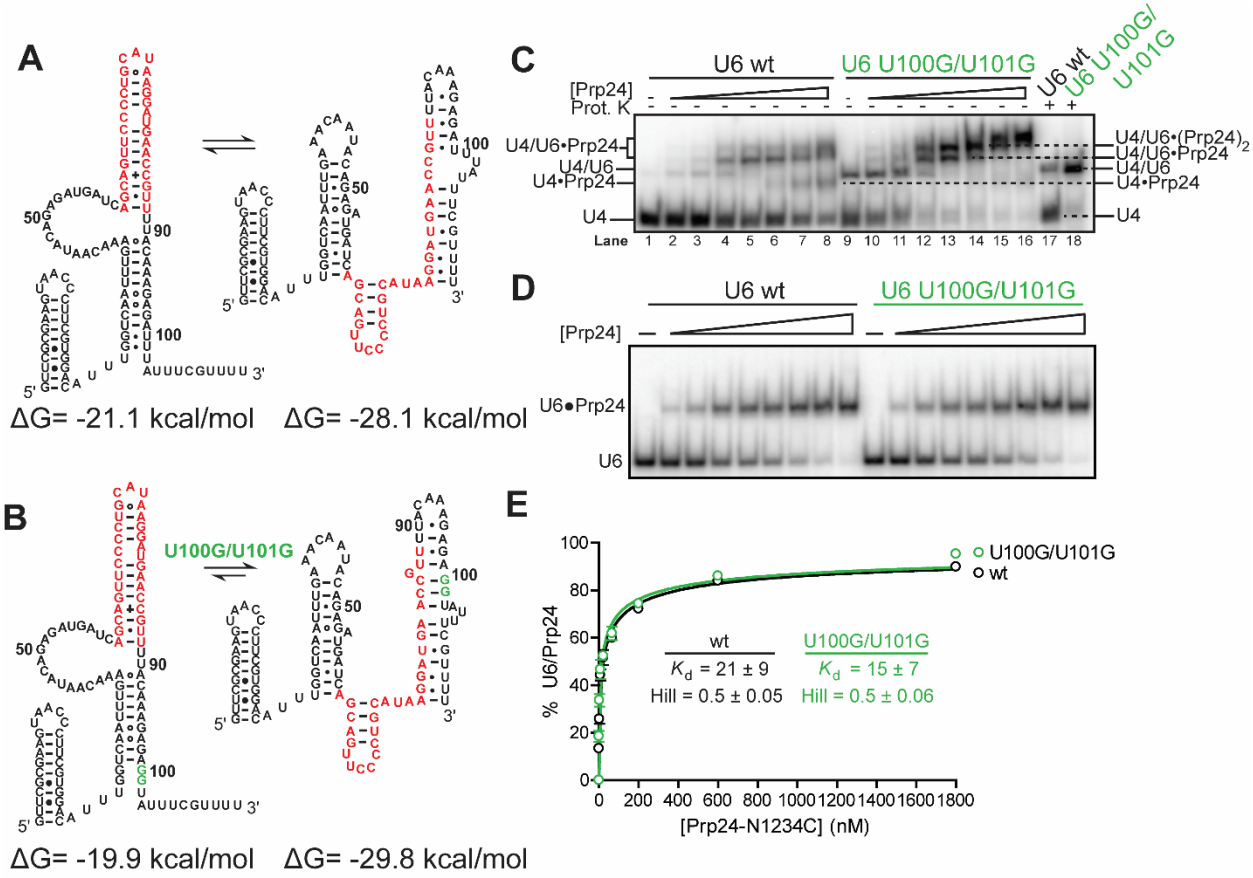

A
ΔG= -21.1 kcal/mol
ΔG= -28.1 kcal/mol
B
U100G/U101G
ΔG= -19.9 kcal/mol
ΔG= -29.8 kcal/mol
C
U6 wt
U6 U100G/U101G
U6 wt
U6 U100G/ U101G
[Prp24]
Prot. K
U4/U6•Prp24
U4/U6
U4•Prp24
U4
Lane 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
U4/U6•(Prp24)2
U4/U6•Prp24
U4/U6
U4•Prp24
U4
D
U6 wt
U6 U100G/U101G
[Prp24]
U6•Prp24
U6
E
% U6/Prp24
U100G/U101G
wt
wt
Kd = 21 ± 9
Hill = 0.5 ± 0.05
U100G/U101G
Kd = 15 ± 7
Hill = 0.5 ± 0.06
[Prp24-N1234C] (nM)

Figure 4-12. Hyperstabilization of the telestem strengthens U6-Prp24 binding.

A) Native gel analysis of U6-Prp24 binding for wild-type (left) and mutant 17 (right) RNAs.

B) Binding curves of wild-type and mutant RNA.

C) Native gel analysis of off-rates for wild-type and mutant RNA. Radiolabeled U6-Prp24 was formed with 50 nM Prp24, then challenged with 200 nM unlabeled U6 (a > 200-fold excess) for increasing lengths of time.

D) Quantification of off-rate gel. Data was fit with a one phase exponential decay equation.

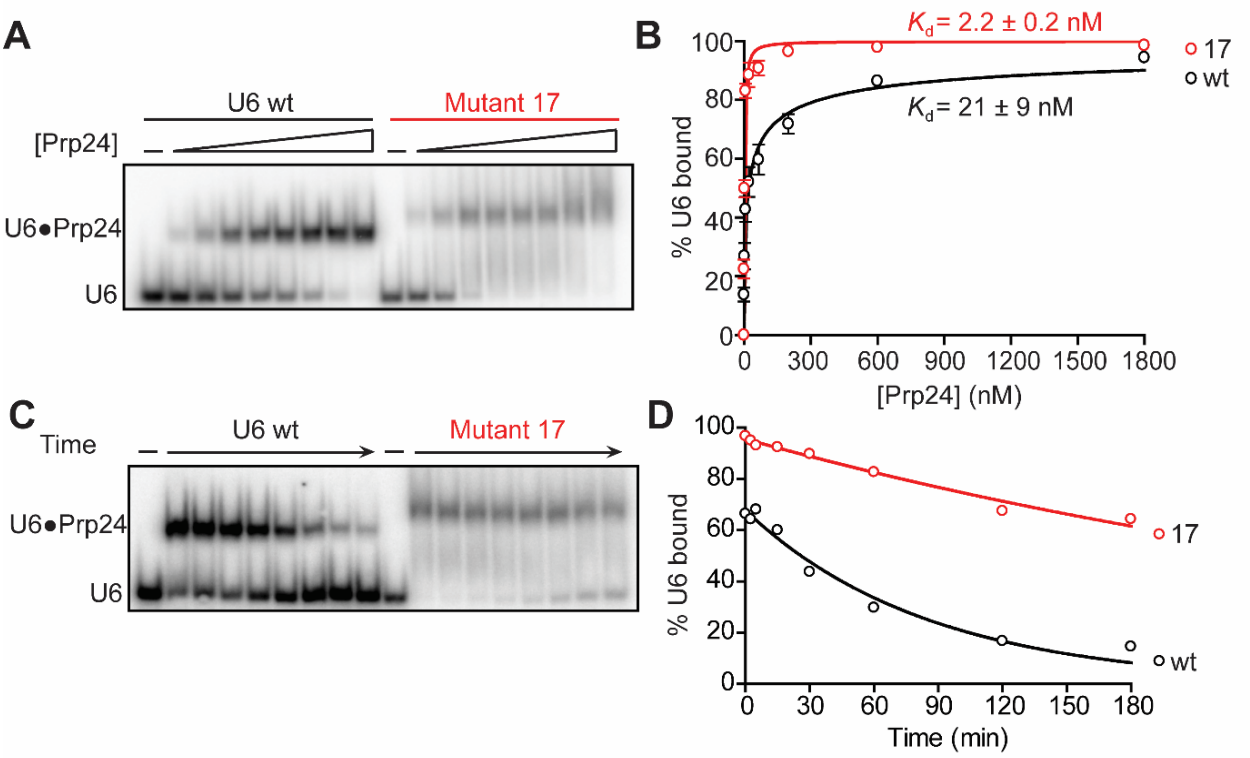


Table 4-1. U6-Prp24 binding properties for wild-type U6 and Mutant 17.

Property	**U6 wt**	**Mutant 17**
K_d (nM)	21 ± 9	2.2 ± 0.2
B_{max} (% bound)	100 ± 8	96 ± 1
Hill coefficient	0.49 ± 0.07	1.3 ± 0.1
k_{off} x 10^{-5} (sec^{-1})	20 ± 2	4 ± 0.3
Calculated k_{on} x 10^3 ($M^{-1}sec^{-1}$)	9.5	18

4.4.6 Reduction of net positive charge in the electropositive groove of Prp24 inhibits annealing

The electropositive groove of Prp24 is comprised of RRMs 1 and 2 and oRRM4 (24). In the crystal structure of U6•Prp24, this region does not bind any nucleotides of U6 within the crystallographic asymmetric unit. However, the ISL of a neighboring U6•Prp24 complex occupies the groove. This groove is well-suited for binding double-stranded RNA, as it has a width of approximately the diameter of double stranded RNA (20 Å) and many positively charged residues for interacting with the negatively charged phosphate backbone of RNA. To test whether this groove may be the active site for U4/U6 annealing, we mutated positively charged residues within the groove and observed the effect on annealing rate. Arginines (R81, R131, and R134) and lysines (K50, K77, and K78) in the electropositive groove were substituted with alanine (**Figure 4-13A**). In addition, the amido groups of N53 and Q54 were mutated to carboxylate groups. Five combinations of mutations in these eight residues were tested (**Figure 4-13B**). Mutagenesis of the selected residues does not affect protein production or purity (**Figure 4-14**).

When tested for their effect on annealing, all combinations of mutations within the groove significantly decreased the rates of annealing (**Figures 7B and C**). Both positive to neutral mutations (arginine and lysine to alanine) and neutral to negative mutations (asparagine and glutamine to aspartate and glutamate, respectively) inhibit annealing to a similar extent. The inhibitory effect is correlated with the number of mutations, where the sextuple mutation is most deleterious. In order to exclude the possibility that the inhibitory effect of the electropositive groove mutations is simply due to an overall reduction in charge of the protein, surface arginine and lysines outside of the electropositive groove (**Figure 4-13A**) were also mutated and tested in the annealing assay. Similarly to the electropositive groove mutations, these mutations did not affect protein purification or stability (**Figure 4-14**). All four sets of mutations outside the groove, including a quadruple mutant, did not significantly affect annealing rate (**Figure 4-13A, B, C**).

To verify that the mutations in the electropositive groove were not deficient for annealing due to a reduced binding affinity for U6, the activity and affinity of wild-type and sextuple mutant Prp24 for U6 RNA were determined. The preparations of both of these proteins used for binding and annealing studies displayed the same binding activity (**Figure 4-13D**), where both proteins appeared to be ~45% active for U6 binding. These proteins were then used in a native gel assay to determine the affinity for U6 RNA (**Figure 4-13E**). Wild-type protein displayed a K_d of 21 ± 9 nM for U6 RNA, while mutant protein displayed a K_d of 28 ± 7 (**Figure 4-13F**). Both proteins bound to U6 with a Hill coefficient of less than unity. In contrast, mutations within the electropositive groove had a measurable effect on the affinity of Prp24 for U4, as the K_d of Prp24 for U4 was at least doubled from 1000 nM to >2000 nM for the mutant protein (**Figure 4-13E, F**). These results indicate that the electropositive groove substitutions do not alter Prp24's affinity for U6 RNA, but do diminish its affinity for U4 RNA, consistent with binding of U4 to the groove prior to annealing with U6.

Figure 4-13. Reduction of net positive charge in the electropositive groove decreases the rate of U4/U6 annealing without affecting U6 RNA-binding.

A) Electrostatic surface of Prp24, contoured from +8 *kT/e* (blue) to -8 *kT/e* (red). Positions of mutations within the electropositive groove (blue text) and outside the groove (green text) are indicated.

B) Annealing gel showing a single time point (90 minutes) for wild-type Prp24 and each of the nine mutants. Samples were divided into native (odd lanes) and proteinase K-treated (even lanes). Substitutions in the electropositive groove (blue text) reduce the amount of U4/U6 after 90 minutes, while substitutions outside this region (green text) do not.

C) Rate of U4/U6 formation over time for wild-type (black), constructs with an electropositive groove mutation (blue text), and constructs with mutations outside of this region (green text).

D) Wild-type and “6mut” Prp24 are equally active for U6 binding. Labeled U6 (5 nM Cy5-U6) was supplemented with 500 nM unlabeled U6, and the binding of stoichiometric amounts of Prp24 was monitored.

E) Native gel analysis of U6-Prp24 binding (top) and U4-Prp24 binding (bottom) using wild-type and mutant full-length Prp24. “6mut” refers to the presence of K50A/K77A/K78A/R81A/R131A/R134A mutations.

F) Binding curves (simple one site binding model) of wild-type versus mutant protein with full-length U6 and U4.

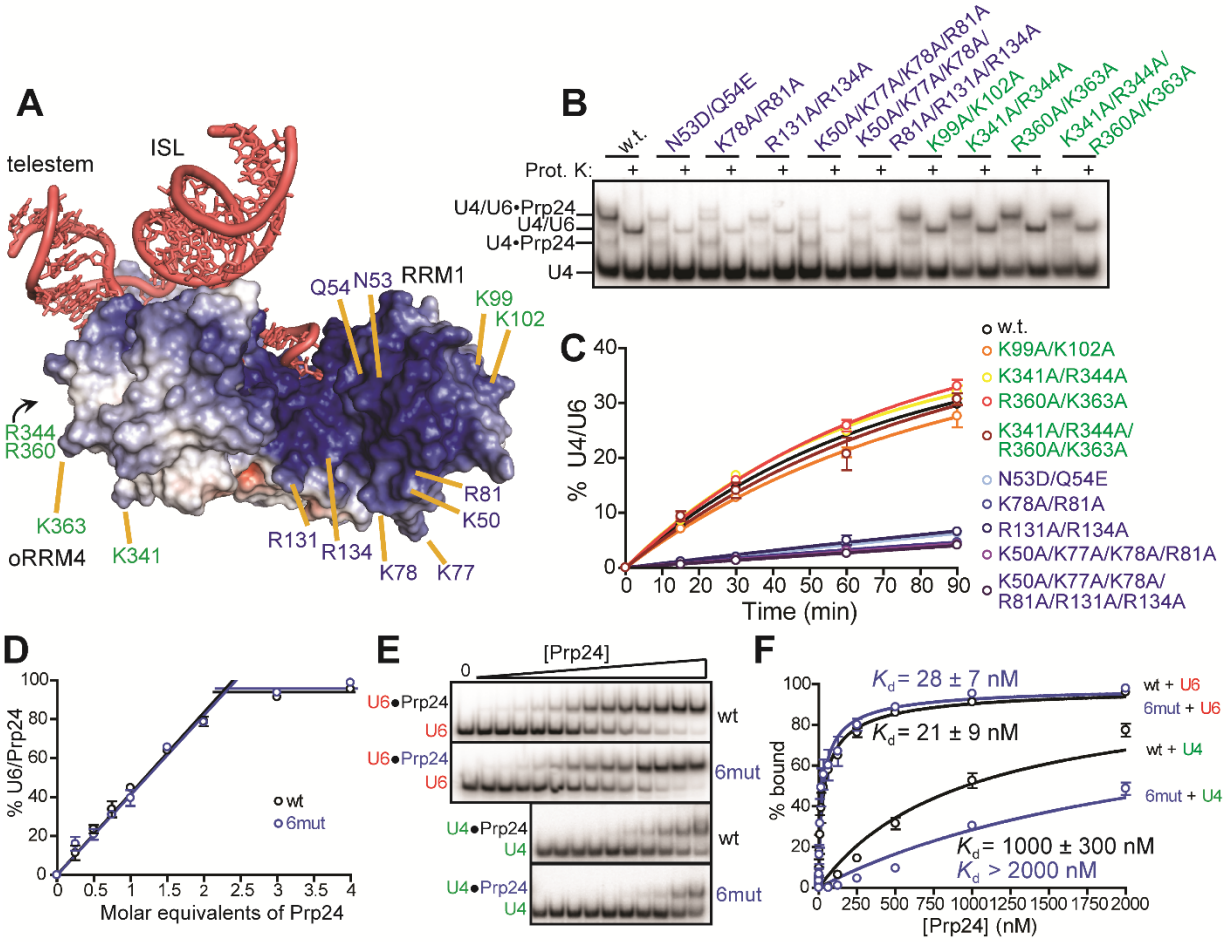
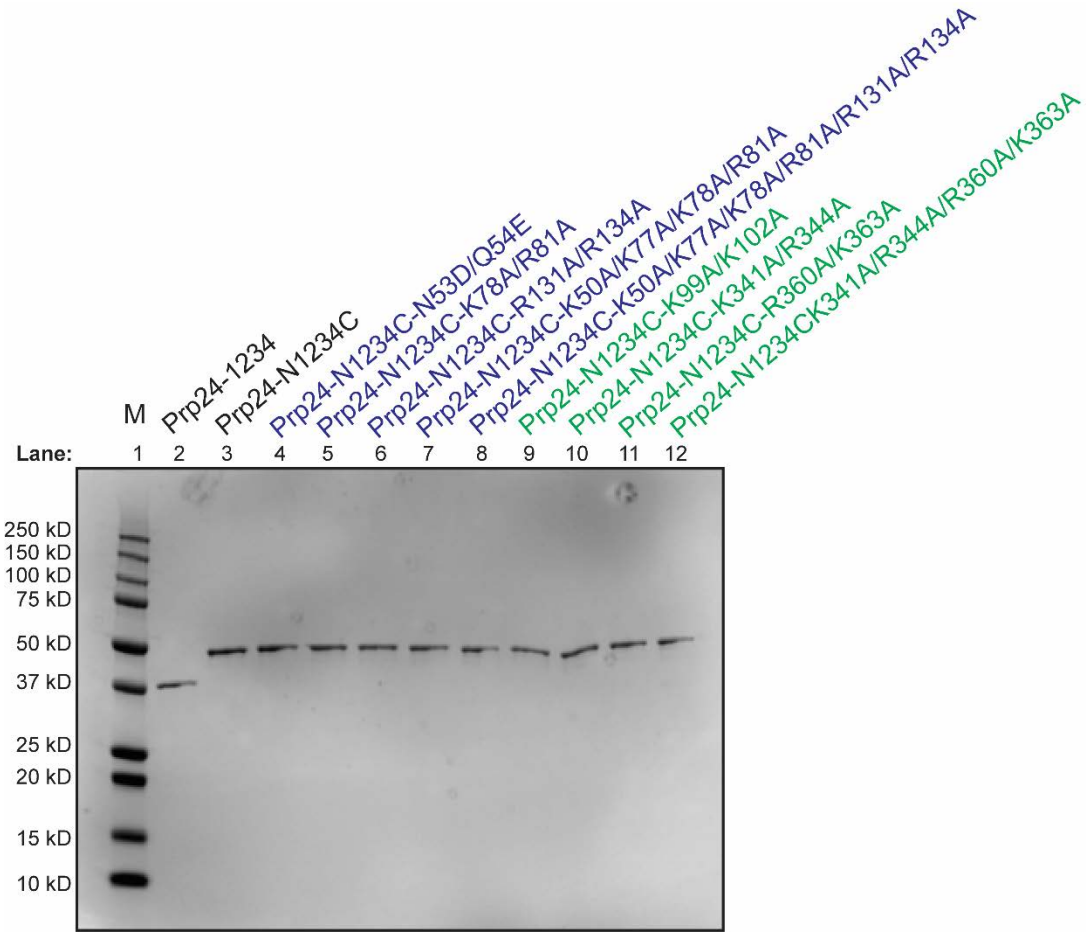

A
telestem
ISL
Q54
N53
RRM1
K99
K102
R344
R360
K363
oRRM4
K341
R131
R134
K78
K77
R81
K50
B
w.t.
N53D/Q54E
K78A/R81A
R131A/R134A
K50A/K77A/K78A/R81A
K50A/K77A/K78A/
R81A/R131A/R134A
K99A/K102A
K341A/R344A
R360A/K363A
K341A/R344A/
R360A/K363A
Prot. K:
U4/U6•Prp24
U4/U6
U4•Prp24
U4
C
% U4/U6
Time (min)
w.t.
K99A/K102A
K341A/R344A
R360A/K363A
K341A/R344A/
R360A/K363A
N53D/Q54E
K78A/R81A
R131A/R134A
K50A/K77A/K78A/R81A
K50A/K77A/K78A/
R81A/R131A/R134A
D
% U6/Prp24
Molar equivalents of Prp24
wt
6mut
E
[Prp24]
U6•Prp24
U6
wt
6mut
U4•Prp24
U4
F
% bound
[Prp24] (nM)
K_d = 28 ± 7 nM
K_d = 21 ± 9 nM
K_d = 1000 ± 300 nM
K_d > 2000 nM
wt + U6
6mut + U6
wt + U4
6mut + U4

Figure 4-14. SDS-PAGE gel of purified recombinant Prp24 constructs.



4.4.7 The Lsm2-8 ring enhances Prp24-mediated U4/U6 annealing *in vitro*

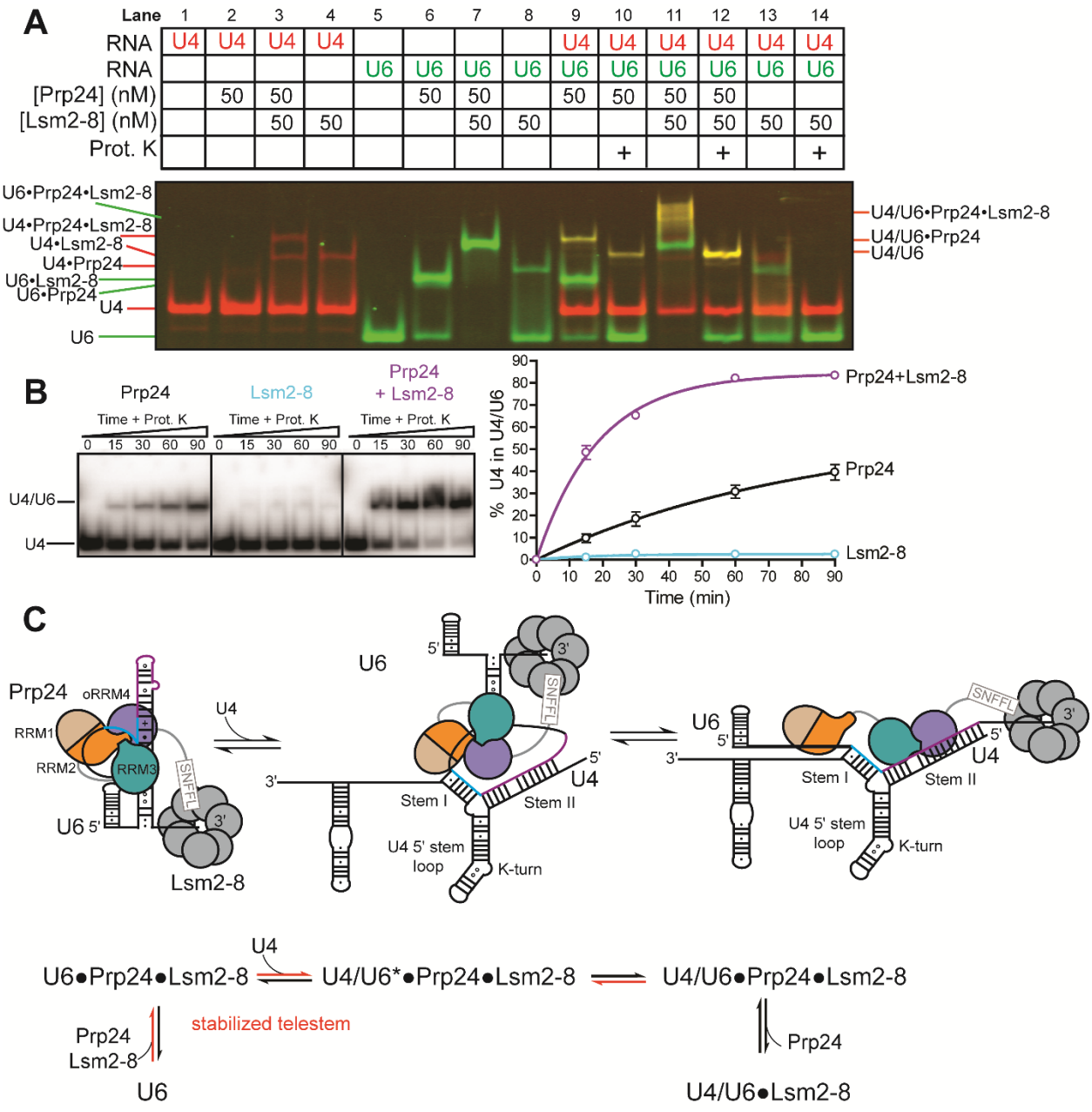
It was previously shown that deletion of the C-terminal decapeptide of Prp24, which binds to the Lsm2-8 complex, decreases the efficiency of Prp24-mediated U4/U6 annealing in extracts (32). Using recombinant Prp24 and Lsm2-8, we directly tested the influence of the Lsm2-8 ring on U4 and U6 annealing *in vitro*. We again used a two-color gel system to monitor the formation of RNP species that contribute to annealing (**Figure 4-15A**). This system reveals that a productive U4/U6•Prp24•Lsm2-8 complex is formed that results in a 4-fold increase in the rate of U4/U6 annealing compared to Prp24 alone (**Figure 4-15B**). We also observe that although Lsm2-8 can bind to both U6 and U4 *in vitro*, it cannot anneal the RNAs in the absence of Prp24 (**Figure 4-15A, lanes 4, 8, 13-14**) even under saturating binding conditions (data not shown). The kinetic profile obtained upon addition of the Lsm2-8 proteins is highly similar in terms of rate enhancement to the moderately stabilizing telestem mutations (11, 12, and 13) (**Figure 4-10**).

Figure 4-15. The Lsm2-8 ring enhances U4/U6 annealing.

A) Two-color native gel with Cy5-U4 (red) and Cy3-U6 (green). The Lsm2-8 ring alone binds U4 and U6 RNAs (lanes 4 and 8), but exhibits strong cooperative binding with Prp24 only on U6 RNA (cf. lanes 3 and 7). Lsm2-8 enhances Prp24-mediated annealing (cf. lanes 10 and 12; yellow shows colocalized RNAs), but does not catalyze annealing alone (lane 14). Annealing reactions were incubated at 30°C for 90 minutes prior to loading.

B) Time-dependent formation of U4/U6, using radiolabeled U4 snRNA and unlabeled U6, Prp24 (50 nM), and Lsm2-8 (50 nM).

C) Model for the rate enhancement conferred by stabilizing the telestem or by inclusion of the Lsm2-8 ring. The C-terminal decapeptide of Prp24 (the SNFFL box) interacts with Lsm2-8.


A
Lane 1 2 3 4 5 6 7 8 9 10 11 12 13 14
RNA U4 U4 U4 U4 U4 U4 U4 U4
RNA U6 U6 U6 U6 U6 U6 U6 U6
[Prp24] (nM) 50 50 50 50 50 50
[Lsm2-8] (nM) 50 50 50 50 50 50
Prot. K + + + +
U6•Prp24•Lsm2-8
U4•Prp24•Lsm2-8
U4•Lsm2-8
U4•Prp24
U6•Lsm2-8
U6•Prp24
U4
U6
U4/U6•Prp24•Lsm2-8
U4/U6•Prp24
U4/U6
B
Prp24
Lsm2-8
Prp24 + Lsm2-8
Time + Prot. K
0 15 30 60 90
U4/U6
U4
% U4 in U4/U6
Prp24+Lsm2-8
Prp24
Lsm2-8
Time (min)
C
Prp24
oRRM4
RRM1
RRM2
RRM3
SNFFL
U6 5'
3'
Lsm2-8
U4
U6
5'
3'
Stem I
Stem II
U4 5' stem loop
K-turn
U6
5'
3'
U4
5'
Stem I
Stem II
U4 5' stem loop
K-turn
U4
U6•Prp24•Lsm2-8
U4/U6*•Prp24•Lsm2-8
U4/U6•Prp24•Lsm2-8
Prp24
Lsm2-8
stabilized telestem
U6
Prp24
U4/U6•Lsm2-8

4.5 Discussion

A critical step in spliceosome biogenesis is U4/U6 RNA annealing, which involves large-scale rearrangement of RNA structure. Prp24 accelerates the extensive remodeling of RNA structure in the absence of ATP hydrolysis. ATP-dependent RNA/RNP remodeling is required for later steps in spliceosome assembly, activation and disassembly, controlled in part by the eight DExD/H helicases in the spliceosome (33). Since Prp24 only uses binding free energy to stimulate annealing, the process is likely driven forward by the favorable free energy of U4/U6 base pairing. For example, the predicted standard free energies for the U6 ISL vs. U4/U6 stem II are -6.7 and -28.1 kcal/mol, respectively (31). Additionally, ours and other's data (16) indicate that Prp24 binds more tightly to the annealed U4/U6 complex than U6, which is likely to further stabilize the annealed product and help drive the process forward. Therefore, it is likely that additional factors are required to displace Prp24 from U4/U6 upon formation of the U4/U6 di-snRNP, which does not contain Prp24 (14,17). Interestingly, human Prp24 does remain associated with U4/U6 (34), but is displaced upon U4/U6.U5 tri-snRNP assembly (35,36).

4.5.1 *In vitro* reconstitution of snRNP assembly

Our *in* vitro reconstitution of U4/U6 annealing faithfully recapitulates known genetic interactions in the U6 ISL and thus shows promise for dissecting the complex series of conformational changes that drive U4/U6 di-snRNP assembly *in vivo*. Inhibition of U4/U6 annealing by the U6-A62G mutation and correction of this defect by the C85A mutation indicate that the stability of the 62-85 base pair at the base of the ISL is critical for annealing. Since nucleotide 85 is outside of the U4 binding region of U6, we hypothesize that U6 nucleotide A62 must be made available for pairing to U4 at a relatively early stage in the annealing process. Indeed, A62 is at the 5' terminal end of the U6 ISL, and is adjacent to the single stranded region of U6 that is positioned most closely to the electropositive groove in our crystal structure (24). We did not observe correction of the A62G annealing defect *in vitro* in the presence of substitutions in Prp24 that suppress the U6-A62G cold-sensitive growth defect, or suppressor substitutions in

U6 that lie in the Prp24•U6 interface rather than in the ISL (24) (data not shown). These results suggest that the latter group of suppressors act at a stage of di-snRNP assembly not assayed in our *in vitro* system, most likely displacement of Prp24 from U4/U6. We predict that, as we add additional splicing factors to our di-snRNP reconstitution assay, we will be able to detect the effect of this latter group of suppressor substitutions.

4.5.2 An active site for U4/U6 annealing

Due to the interlocked topology of the U6 snRNP core, it is unlikely that the electropositive groove is completely formed in the absence of U6 RNA. However, U4 RNA still binds to Prp24 with some specificity, albeit with low (K_d = 1 μM) affinity. It is likely that U4 RNA binding to the pre-formed U6 snRNP core is a much higher affinity interaction, as reflected in the observed 20 nM $K_{1/2}$ of annealing. Mutation of the electropositive groove impairs U4 RNA binding and annealing, consistent with a model in which U4 RNA binds to the electropositive groove of the pre-formed U6•Prp24 complex to initiate annealing. The U6 nucleotide that is closest to the electropositive groove is A51, which stacks on A53 and is therefore proximal to the U6 nucleotides that participate in formation of U4/U6 Stem I (U6 54-63). Thus, we hypothesize that U4 localization to the electropositive groove positions U4 in close proximity to U6 nucleotides 54-63 to form Stem I (**Figure 4-1A,B**). Based on our crystal structure and prior NMR data (37), we previously proposed a model in which the single stranded nucleotides 54-60 of U6 dynamically reorient into the electropositive groove to initiate annealing (24). Our proposal that the electropositive groove stabilizes the nascent annealing of U4/U6 is consistent with the observation that mutations that reduce the positive charge within this groove have a strong deleterious effect on annealing efficiency. This proposal is also consistent with the observation that electropositive groove mutations do not affect U6-Prp24 binding, implying that the electropositive groove functions in the annealing process subsequent to U6•Prp24 formation.

4.5.3 The telestem is a prerequisite for efficient U4/U6 annealing

It has previously been demonstrated that truncations to the 5' and 3' ends of U6 had only a marginal effect on splicing activity in vitro (38), and our data indicate that these regions are not important for Prp24-mediated U4/U6 annealing. Although hydroxyl radical probing and chemical crosslinking results suggest that Prp24 interacts with and protects the 5' SL and linker region in U6 (39), we conclude these interactions are not necessary for efficient annealing *in vitro*. Truncation of the 3' end may increase annealing rate by biasing formation of the telestem, or the 3’ end may otherwise occupy the electropositive groove in the absence of the Lsm ring, competing with annealing.

There is a clear correlation between the presence of stabilizing mutations in the telestem and an increased Prp24-mediated annealing rate (**Figure 4-10**), leading us to conclude that formation of the telestem is important for the annealing process. Interestingly, there appears to be additional sequence requirements involving the 3' side of the telestem, as the U100G/U101G mutation increases both the Prp24-catalyzed annealing rate and the protein-free annealing rate. These mutations may stabilize an alternate fold involving the 3’ end of U6 that invades and destabilizes the ISL, while preserving the Prp24 binding site 5’ of the ISL, thereby increasing the annealing rate regardless of the presence of Prp24. The U100G/U101G mutation also stabilizes a second binding site for Prp24, observed as a U4/U6•$(Prp24)_2$ complex with decreased gel mobility (**Figures 6 and S5**). We have no evidence for the biological relevance of this secondary binding site, but hypothesize it is due to sequence similarities between the primary binding site for Prp24 RRM2 (nucleotides 46-54: UACAGAGAU) (24) and the 3’ side of the telestem (nucleotides 90-100: UACAAAGAGAU). Note that the underlined sequences are identical. We hypothesize that formation of the U4/U6 complex disrupts the telestem and exposes this 3’ sequence, allowing for the binding of a second Prp24 molecule.

In the U6•Prp24 complex, the telestem is sandwiched between RRM3 and oRRM4 (**Figure 4-1A**). We propose that the telestem acts as a linchpin to stabilize the “interlocked ring” topology

of U6•Prp24 required for formation of the electropositive groove (24). In support of this model, we find that the most stable telestem mutant (#17) has a lower K_d and slower off-rate for Prp24 (**Figure 4-12**).The fact that we observe a general correlation between telestem stability and annealing activity for a large number of mutants (18 total, **Figures 5 and 6**) with different RNA sequences suggests that the observed rate effect is a function of the thermodynamic stability of the telestem. A model that reconciles our data is shown in **Figure 4-15C**. Stabilization of the telestem may be achieved either via mutation of the telestem or by binding of the Lsm2-8 ring to the 3' end of U6. Thus, we propose that an important function of the Lsm2-8 ring is to ensure proper folding of the U6 telestem region without over-stabilizing the RNA fold. Once U4/U6 annealing has occurred, coaxial stacking of U4/U6 Stems I and II, as observed in the tri-snRNP structure (40), likely splays apart the ends of U6 RNA to destabilize the telestem.

The ability of the components of the crystallized U6 snRNP core (truncated U6 and truncated Prp24) to efficiently anneal, combined with the observation that the U100C/U101C mutations greatly accelerate annealing, strongly suggests that the crystal structure represents an on-pathway conformation in the annealing process. As both Prp24 and the Lsm ring can crosslink to the base of the telestem (39), it is possible that the protein components of the U6 snRNP stabilize the telestem and exert allosteric control over U4/U6 di-snRNP assembly. Future studies will seek to address the mechanism by which Prp24 is displaced from U4/U6 *in vivo* (14).

4.5.4 Conclusions and prospects

The U4/U6 di-snRNA annealing mechanism illustrates the high level of specificity and complexity required for spliceosome assembly and catalysis. Unlike other nucleic acid annealing proteins, Prp24 has evolved a high degree of specificity for its snRNA substrates. This specificity is illustrated by the interlocked ring topology of U6 and Prp24, which is an extreme (and thus far, apparently unique) example of RNP co-folding. The mechanism of U4/U6 annealing is also intriguing, given that Prp24 must resolve considerable secondary structure in U6 in order to form even more extensive intermolecular base pairing in the U4/U6 di-snRNA. Annealers of short

RNAs, such as Argonaute and Hfq, typically do not need to resolve such extensive secondary structure in order to anneal their target RNAs. However, the mechanism of Prp24-mediated annealing is not without analogy to these other annealers, as they all share a common feature in the presence of a positively charged region that is critical for the annealing process. In Hfq, an "arginine patch" along the rim of the protein is essential for annealing sRNAs to their mRNA targets (4). Like the electropositive groove in Prp24, mutation of the arginine patch in Hfq reduces annealing but does not prevent sRNA/mRNA binding to Hfq. Similarly, Argonaute also contains a deep electropositive cleft in which the guide RNA/target RNA duplex is annealed (41).

Like the ribosome, the spliceosome is a key supramolecular enzyme of gene expression. The largest pre-assembled spliceosome particle is the U4/U6.U5 tri-snRNP, which has a maximum dimension of 31 nM, larger than both the post-catalytic *S. pombe* spliceosome (27 nm) (12) and the fully assembled 70S ribosome (24 nm) (42). Assembly pathways for the small ribosomal subunit were initially mapped out in the late 1960's (43-45), and mechanistic descriptions of this process are still being elucidated (46,47). In comparison to the ribosome, relatively little is understood about the assembly mechanism of the spliceosome. The first structures of the U4/U6.U5 tri-snRNP and spliceosome have recently emerged into view (12,40) and the astonishing complexity of these large structures suggests an equally complex assembly pathway. Here we elucidate the molecular requirements for one of the first steps in tri-snRNP biogenesis, formation of the U4/U6 di-snRNA. This work provides a quantitative and structural foundation for understanding the pathway and kinetics of tri-snRNP assembly, starting from the U6 snRNP. Important future directions include understanding the contributions of U4 snRNP proteins to U4/U6 di-snRNP assembly (and Prp24 release), and to understand how the U5 snRNP joins the U4/U6 di-snRNP to form the tri-snRNP.

4.6 Acknowledgements and funding

We thank Aaron Hoskins, Kiyoshi Nagai, and Yigong Shi for plasmids, Aaron Hoskins for critical reading of the manuscript, and members of the Butcher, Brow, and Hoskins labs for helpful discussions. This work was supported by the National Institutes of Health [grant number GM065166] to S.E.B. and D.A.B.

4.7 References

1. Ameres, S.L., Martinez, J. and Schroeder, R. (2007) Molecular basis for target RNA recognition and cleavage by human RISC. *Cell*, **130**, 101-112.
2. Herzog, V.A. and Ameres, S.L. (2015) Approaching the Golden Fleece a Molecule at a Time: Biophysical Insights into Argonaute-Instructed Nucleic Acid Interactions. *Mol Cell*, **59**, 4-7.
3. Jiang, F. and Doudna, J.A. (2015) The structural biology of CRISPR-Cas systems. *Curr Opin Struct Biol*, **30**, 100-111.
4. Panja, S., Schu, D.J. and Woodson, S.A. (2013) Conserved arginines on the rim of Hfq catalyze base pair formation and exchange. *Nucleic acids research*, **41**, 7536-7546.
5. Rajkowitsch, L. and Schroeder, R. (2007) Dissecting RNA chaperone activity. *Rna*, **13**, 2053-2060.
6. Woodson, S.A. (2010) Taming free energy landscapes with RNA chaperones. *RNA biology*, **7**, 677-686.
7. Brow, D.A. (2002) Allosteric cascade of spliceosome activation. *Annual review of genetics*, **36**, 333-360.
8. Wahl, M.C., Will, C.L. and Luhrmann, R. (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell*, **136**, 701-718.
9. Fortner, D.M., Troy, R.G. and Brow, D.A. (1994) A stem/loop in U6 RNA defines a conformational switch required for pre-mRNA splicing. *Genes & development*, **8**, 221-233.
10. Fica, S.M., Tuttle, N., Novak, T., Li, N.S., Lu, J., Koodathingal, P., Dai, Q., Staley, J.P. and Piccirilli, J.A. (2013) RNA catalyses nuclear pre-mRNA splicing. *Nature*, **503**, 229-234.
11. Hang, J., Wan, R., Yan, C. and Shi, Y. (2015) Structural basis of pre-mRNA splicing. *Science*.
12. Yan, C., Hang, J., Wan, R., Huang, M., Wong, C.C. and Shi, Y. (2015) Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science*.
13. Mayes, A.E., Verdone, L., Legrain, P. and Beggs, J.D. (1999) Characterization of Sm-like proteins in yeast and their association with U6 snRNA. *EMBO J*, **18**, 4321-4331.
14. Shannon, K.W. and Guthrie, C. (1991) Suppressors of a U4 snRNA mutation define a novel U6 snRNP protein with RNA-binding motifs. *Genes Dev*, **5**, 773-785.
15. Raghunathan, P.L. and Guthrie, C. (1998) A spliceosomal recycling factor that reanneals U4 and U6 small nuclear ribonucleoprotein particles. *Science*, **279**, 857-860.
16. Ghetti, A., Company, M. and Abelson, J. (1995) Specificity of Prp24 binding to RNA: a role for Prp24 in the dynamic interaction of U4 and U6 snRNAs. *RNA*, **1**, 132-145.
17. Jandrositz, A. and Guthrie, C. (1995) Evidence for a Prp24 binding site in U6 snRNA and in a putative intermediate in the annealing of U6 and U4 snRNAs. *EMBO J*, **14**, 820-832.

18. Kwan, S.S. and Brow, D.A. (2005) The N- and C-terminal RNA recognition motifs of splicing factor Prp24 have distinct functions in U6 RNA binding. *Rna*, **11**, 808-820.
19. Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M. and Luhrmann, R. (1999) A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *EMBO J*, **18**, 5789-5802.
20. Pannone, B.K., Xue, D. and Wolin, S.L. (1998) A role for the yeast La protein in U6 snRNP assembly: evidence that the La protein is a molecular chaperone for RNA polymerase III transcripts. *The EMBO journal*, **17**, 7442-7453.
21. Salgado-Garrido, J., Bragado-Nilsson, E., Kandels-Lewis, S. and Seraphin, B. (1999) Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin. *The EMBO journal*, **18**, 3451-3462.
22. Martin-Tumasz, S., Richie, A.C., Clos, L.J., 2nd, Brow, D.A. and Butcher, S.E. (2011) A novel occluded RNA recognition motif in Prp24 unwinds the U6 RNA internal stem loop. *Nucleic acids research*, **39**, 7837-7847.
23. Sashital, D.G., Allmann, A.M., Van Doren, S.R. and Butcher, S.E. (2003) Structural basis for a lethal mutation in U6 RNA. *Biochemistry*, **42**, 1470-1477.
24. Montemayor, E.J., Curran, E.C., Liao, H.H., Andrews, K.L., Treba, C.N., Butcher, S.E. and Brow, D.A. (2014) Core structure of the U6 small nuclear ribonucleoprotein at 1.7-A resolution. *Nat Struct Mol Biol*, **21**, 544-551.
25. Zhou, L., Hang, J., Zhou, Y., Wan, R., Lu, G., Yin, P., Yan, C. and Shi, Y. (2014) Crystal structures of the Lsm complex bound to the 3' end sequence of U6 small nuclear RNA. *Nature*, **506**, 116-120.
26. Milligan, J.F., Groebe, D.R., Witherell, G.W. and Uhlenbeck, O.C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic acids research*, **15**, 8783-8798.
27. Milligan, J.F. and Uhlenbeck, O.C. (1989) Synthesis of small RNAs using T7 RNA polymerase. *Methods in enzymology*, **180**, 51-62.
28. Nandakumar, J., Ho, C.K., Lima, C.D. and Shuman, S. (2004) RNA substrate specificity and structure-guided mutational analysis of bacteriophage T4 RNA ligase 2. *The Journal of biological chemistry*, **279**, 31337-31347.
29. Ryder, S.P., Recht, M.I. and Williamson, J.R. (2008) Quantitative analysis of protein-RNA interactions by gel mobility shift. *Methods in molecular biology*, **488**, 99-115.

30. Bae, E., Reiter, N.J., Bingman, C.A., Kwan, S.S., Lee, D., Phillips, G.N., Jr., Butcher, S.E. and Brow, D.A. (2007) Structure and interactions of the first three RNA recognition motifs of splicing factor prp24. *Journal of molecular biology*, **367**, 1447-1458.
31. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research*, **31**, 3406-3415.
32. Rader, S.D. and Guthrie, C. (2002) A conserved Lsm-interaction motif in Prp24 required for efficient U4/U6 di-snRNP formation. *RNA*, **8**, 1378-1392.
33. Cordin, O., Hahn, D. and Beggs, J.D. (2012) Structure, function and regulation of spliceosomal RNA helicases. *Current opinion in cell biology*, **24**, 431-438.
34. Bell, M., Schreiner, S., Damianov, A., Reddy, R. and Bindereif, A. (2002) p110, a novel human U6 snRNP protein and U4/U6 snRNP recycling factor. *The EMBO journal*, **21**, 2724-2735.
35. Fabrizio, P., Esser, S., Kastner, B. and Luhrmann, R. (1994) Isolation of S. cerevisiae snRNPs: comparison of U1 and U4/U6.U5 to their human counterparts. *Science*, **264**, 261-265.
36. Liu, S., Rauhut, R., Vornlocher, H.P. and Luhrmann, R. (2006) The network of protein-protein interactions within the human U4/U6.U5 tri-snRNP. *Rna*, **12**, 1418-1430.
37. Martin-Tumasz, S., Reiter, N.J., Brow, D.A. and Butcher, S.E. (2010) Structure and functional implications of a complex containing a segment of U6 RNA bound by a domain of Prp24. *Rna*, **16**, 792-804.
38. Ryan, D.E., Stevens, S.W. and Abelson, J. (2002) The 5' and 3' domains of yeast U6 snRNA: Lsm proteins facilitate binding of Prp24 protein to the U6 telestem region. *Rna*, **8**, 1011-1033.
39. Karaduman, R., Fabrizio, P., Hartmuth, K., Urlaub, H. and Luhrmann, R. (2006) RNA structure and RNA-protein interactions in purified yeast U6 snRNPs. *Journal of molecular biology*, **356**, 1248-1262.
40. Nguyen, T.H., Galej, W.P., Bai, X.C., Savva, C.G., Newman, A.J., Scheres, S.H. and Nagai, K. (2015) The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature*, **523**, 47-52.
41. Schirle, N.T. and MacRae, I.J. (2012) The crystal structure of human Argonaute2. *Science*, **336**, 1037-1040.
42. Selmer, M., Dunham, C.M., Murphy, F.V.t., Weixlbaumer, A., Petry, S., Kelley, A.C., Weir, J.R. and Ramakrishnan, V. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science*, **313**, 1935-1942.
43. Hosokawa, K., Fujimura, R.K. and Nomura, M. (1966) Reconstitution of functionally active ribosomes from inactive subparticles and proteins. *Proc Natl Acad Sci U S A*, **55**, 198-204.

44. Traub, P. and Nomura, M. (1968) Structure and function of Escherichia coli ribosomes. I. Partial fractionation of the functionally active ribosomal proteins and reconstitution of artificial subribosomal particles. *Journal of molecular biology*, **34**, 575-593.

45. Traub, P. and Nomura, M. (1968) Structure and function of E. coli ribosomes. V. Reconstitution of functionally active 30S ribosomal particles from RNA and proteins. *Proc Natl Acad Sci U S A*, **59**, 777-784.

46. Kim, H., Abeysirigunawarden, S.C., Chen, K., Mayerle, M., Ragunathan, K., Luthey-Schulten, Z., Ha, T. and Woodson, S.A. (2014) Protein-guided RNA dynamics during early ribosome assembly. *Nature*, **506**, 334-338.

47. Mulder, A.M., Yoshioka, C., Beck, A.H., Bunner, A.E., Milligan, R.A., Potter, C.S., Carragher, B. and Williamson, J.R. (2010) Visualizing ribosome biogenesis: parallel assembly pathways for the 30S subunit. *Science*, **330**, 673-677.

Chapter 5: Structural analysis of multi-helical RNAs by NMR-SAXS/WAXS: Application to the U4/U6 di-snRNA

This chapter is published in the following form:

Cornilescu, G.*****, Didychuk, A.L.*****, Rodgers, M.L., Michael, L.A., Burke, J.E., Montemayor, E.J., Hoskins, A.A., Butcher, S.E. (2016) Structural analysis of multi-helical RNAs by NMR-SAXS/WAXS: Application to the U4/U6 di-snRNA. *Journal of Molecular Biology* 428:777-789.

*****co-first authors

Author contributions:

C.G. and A.L.D. calculated structures and analyzed data. A.L.D., L.A.M., and J.E.B. made samples for NMR-SAXS/WAXS analysis and analyzed data. E.J.M. helped make figures. M.L.R. performed all smFRET experiments and analysis. S.E.B. wrote the paper with help from A.L.D., G.C., and A.A.H.

5.1 Abstract

NMR and SAXS/WAXS are highly complementary approaches for the analysis of RNA structure in solution. Here we describe a streamlined NMR-SAXS/WAXS approach for structural investigation of multi-helical RNAs. We illustrate this approach by determining the overall fold of a 92-nucleotide 3-helix junction from the U4/U6 di-snRNA. The U4/U6 di-snRNA is conserved in eukaryotes and is part of the U4/U6.U5 tri-snRNP, a large ribonucleoprotein complex that comprises a major subunit of the assembled spliceosome. Helical orientations can be determined by X-ray scattering data alone, but the addition of NMR RDC restraints improves the structure models. RDCs were measured in 2 different external alignment media and also by magnetic susceptibility anisotropy. The resulting alignment tensors are collinear, which is a previously noted problem for nucleic acids. Including WAXS data in the calculations produces models with significantly better fits to the scattering data. In solution, the U4/U6 di-snRNA forms a 3-helix junction with a planar Y-shaped structure and has no detectable tertiary interactions. Single molecule FRET data support the observed topology. A comparison with the recently determined cryo-EM structure of the U4/U6.U5 tri-snRNP illustrates how proteins scaffold the RNA and dramatically alter the geometry of the U4/U6 3-helix junction.

5.2 Introduction

Nearly two decades have passed since the first high-resolution view of an RNA structure over 100 nucleotides (nts) [1]. Since this time, great progress has been made in structure determination of RNAs and RNA-protein complexes (RNPs). However, RNA-containing structures are still under-represented in the RCSB Protein Data Bank [2], with currently less than 3% of entries containing RNA. Given the biological significance of RNA and its complexes, and the weak predictive power of *ab initio* modeling of RNA tertiary structure, it is important to continue to develop efficient methods for deriving accurate structure models from experimental data. It is also important to understand structural intermediates in RNA folding pathways, how helical junction topology determines the conformational space of RNA folding [3], and how protein binding influences RNA folding to stabilize biologically relevant conformations.

Remarkable advancements have been made in biomolecular structure determination over the past few years. Solution NMR can now be used to determine RNA structure models over 100 nts [4-6], with the largest RNA NMR structure currently at 155 nts [5]. Small angle X-ray scattering (SAXS), although an intrinsically low-resolution method, has become a valuable tool for analyzing molecular structure in solution. The SAXS region of X-ray scattering, defined as momentum transfer (q) between 0 and 0.3 $Å^{-1}$, provides sufficient information for resolving molecular features on the order of 20 Å, such as the width of the A-form helix, and provides useful information on overall molecular size and shape [7-10]. Wide angle X-ray scattering (WAXS), where $q > 0.3$ $Å^{-1}$, contains even finer structural information such as nucleic acid helical groove width [7] and in favorable cases can even resolve the spacing of phosphate groups between base pairs [11]. Additionally, significant advances have been made in computational modeling of RNA structure, including the validation and improvement of high resolution models [12], generation of models from chemical mapping [13], and de novo structure prediction approaches [14]. Until recently, X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy were the only experimental methods capable of elucidating molecular structures at the atomic level. However,

cryo-electron microscopy (cryoEM) has made striking advances towards atomic resolution, with two recent structures reported at near-atomic, sub-3 Å resolution [15, 16], although many cryoEM structures still depend upon the fitting of models derived from X-ray crystallography and/or NMR. Hybrid methods are also starting to emerge, such as combining cryoEM and cryo-electron tomography with NMR [17, 18] and NMR-SAXS [4, 6, 9, 10, 19, 20].

We previously described a method for determining the overall structures of large RNAs from sparse NMR and SAXS data, and applied this approach to a 111 nucleotide U2-U6 spliceosomal RNA complex [6]. Our previous method involved generation of 2500 all-atom models using the MC-Sym pipeline [21], which were then filtered and sorted based on goodness of fit (χ^2 agreement) to the individual SAXS and RDC data sets. The models that showed the best agreement to both data sets were then subsequently refined using the XPLOR-NIH structure determination program [22] in order to jointly optimize the agreement with SAXS and NMR data. Wang and coworkers have described a different approach to generate initial models consistent with SAXS data using the G2G program [23]. The initial G2G models can then be subsequently refined in XPLOR-NIH [7]. Here we describe the direct calculation of RNA structure models in XPLOR-NIH using NMR and SAXS/WAXS data in a single step, without the need for generation of initial models and subsequent refinement.

Residual dipolar couplings (RDCs) can, in principle, provide the long-range restraints needed to precisely determine interhelical orientations. However, RDCs have four-fold degenerate solutions and are fit equally well with 180° rotations around the three principal axes of the alignment tensor [24]. In the case of a multi-helical RNA with N helices, the degeneracy of the RDC can be satisfied by 4^{N-1} conformations [19]. A potential solution to this problem is to measure RDCs in two or more different alignment media that result in different alignment tensors [24]. However, in practice this has not been achieved because alignment media tend to sterically and electrostatically align nucleic acids with their longest helical axis along the *z*-component of the alignment tensor [25]. Although very few RNAs have been analyzed in this way, Pardi and

coworkers showed that the 6 most commonly used alignment media resulted in very similar alignment tensors for $tRNA^{Val}$ [25], and thus do not solve the RDC degeneracy problem. For a model RNA (TAR), analysis of the alignment tensors from a variety of alignment media showed that RDCs measured in negatively charged stretched gels appeared significantly different from those measured in Pf1 phage ($R^2 = 0.65$)[26].

As an alternative to external alignment media, magnetic alignment of RNA can be achieved via the intrinsic magnetic susceptibility anisotropy (MSA) of the RNA nucleobases, which imparts a small degree of alignment at high magnetic fields [27]. Magnetic alignment of RNA has been investigated for few RNAs, and in the case of $tRNA^{Val}$, the magnetic alignment tensor was indeed found to be different from those produced by external alignment media [25]. However, magnetic alignment did not completely alleviate the aforementioned 4-fold degeneracy as the z component of the magnetic alignment tensor was rotated by ~90 degrees, resulting in a tensor with an overall similar (collinear) orientation to those of external alignment media. Therefore, the intrinsic MSA RDCs only eliminated 2 out of 4 degenerate solutions for $tRNA^{Val}$ [25].

The U4/U6 di-snRNA secondary structure is conserved in eukaryotes [28-31]. Formation of the U4/U6 di-snRNA is one of the first steps in assembly of the U4/U6.U5 tri-snRNP [32], the largest pre-assembled spliceosome complex. The tri-snRNP subsequently joins the U1 and U2 small nuclear ribonucleoprotein particles (snRNPs), along with the nineteen nineteen complex (NTC) and nineteen related complex (NTC) to form the assembled spliceosome [33]. The U4/U6 di-snRNA is then unwound by the helicase Brr2 during spliceosome active site formation (activation) [34-38], resulting in U4 release and incorporation of U6 into the catalytic center [39]. The spliceosome catalyzes pre-mRNA splicing using RNA-mediated metal ion catalysis in a manner identical to group II self-splicing intron catalysis [40, 41]. Since the spliceosomal RNAs form extensive interactions with many spliceosomal proteins [32, 39], it is unlikely that the structures of free spliceosomal RNAs will resemble their protein-bound conformations. However, a comparison of the free and protein-bound spliceosomal RNA structures provides unique insights

into the degree to which proteins remodel and regulate RNA structures and the conformational dynamics that are likely required during RNP assembly.

Here we determine the overall conformation of a 92 nucleotide (nt) U4/U6 di-snRNA using NMR-SAXS/WAXS. Structure models were calculated in a single step using the XPLOR-NIH program [22, 42]. The impact of SAXS, WAXS, and sparse NMR restraints (NOEs and RDCs) on the structure models is evaluated. The overall structure of the RNA is well defined by SAXS/WAXS data alone, but further benefits from NMR data. A comparison of the free U4/U6 di-snRNA conformation with the recently determined tri-snRNP structure [32] illustrates how extensive RNA-protein interactions remodel the RNA fold and stabilize coaxial stacking of helices.

5.3 Results

The *S. cerevisiae* U4 and U6 snRNAs are 160 and 112 nt, respectively (**Figure 5-1A**). In order to facilitate structural analysis of the di-snRNA helical junction, we created a 92 nt linked U4-U6 RNA construct that spans the entire base-paired region between U4 and U6 (**Figure 5-1B**). NMR data indicate that the RNA is well-folded in a single major conformation (**Figures 1C** and **Figure 5-3**). The individual helical domains were investigated by 2D NOESY in order to facilitate assignment of base-paired imino resonances of the larger RNA construct (**Figure 5-2**). The NOEs observed for the individual helical domains are apparent as sub-spectra of the 2D NOESY data of the intact RNA (**Figure 5-2**), suggesting a lack of tertiary interactions between helices in the intact RNA. Nearly all base-paired imino proton and nitrogen resonances could be assigned by 2D NOESY and ^{1}H-^{15}N HSQC-TROSY (**Figures 1C** and **Figure 5-3**). In addition, five adenine H2 resonances with disperse chemical shifts could be readily assigned via intense NOE cross-peaks to uracil imino resonances, owing to the short 2.7 Å distance between these protons when in a Watson-Crick base pair (**Figure 5-4**). Additionally, the upfield chemical shift of the A75 H2 proton further confirms its assignment, owing to the unique chemical shift associated with the first A in the sequence 5'-UAA-3' that occurs when this motif is Watson-Crick base-paired [43].

Residual dipolar couplings (RDCs) were measured in 2 different external alignment media, Pf1 phage [44] and negatively charged stretched polyacrylamide gels [45], and also by magnetic susceptibility anisotropy [27]. 29 total RDCs were measured from the resolved (non-overlapped) imino (^{1}H-^{15}N) and adenine (^{1}H-^{13}C) correlations. The degree of alignment from MSA is small and hence the magnitudes of the measured RDCs are on the order of a few Hz. Such small dipolar couplings can be difficult to accurately measure for large molecular weight (MW) complexes with broad linewidths. In order to accurately measure small dipolar couplings in a high MW RNA, we used the ARTSY method [46], which quantitatively measures the dipolar coupling by measuring differences in TROSY peak intensities recorded with different dephasing delays. Thus, ARTSY

overcomes many of the problems associated with measuring dipolar couplings in high MW systems. Plotting the measured RDCs against each other produces no apparent correlation (**Figure 5-5**). However, analysis of the scalar product of the tensors [47] shows that they are indeed highly correlated as the magnitudes of the scalar products are above 0.9 for most calculated structures (data not shown).

Structures were calculated using X-ray scattering data, including both SAXS and WAXS data. Sparse NMR restraints included 29 RDCs, 47 experimentally determined inter-residue NOEs derived from the imino resonances (**Figures 1C** and **Figure 5-4**). Non-experimentally derived restraints were employed to maintain idealized A-form helical geometry in the Watson-Crick paired regions where experimentally determined NOEs indicated that base-pairs are indeed formed (**Table 5-1**). The use of idealized A-form helical geometries in fitting both RDC and SAXS data has been previously validated [7, 20, 48]. Additionally, non-experimental restraints were used for the tetraloop [49, 50] and kink-turn structures [51]. The 10 lowest energy structures (out of 100 calculated) are shown in **Figure 5-6**. The overall rmsd of the structure models is 2.4 Å (Table 5-1). The molecule forms a mostly planar Y-shaped 3-helix junction (**Figure 5-6**).

The agreement between the SAXS/WAXS data and the predicted scattering from the structure models is excellent (χ^2 = 0.089) (**Figure 5-7A**). Inclusion of WAXS data significantly improves the agreement between the models and the scattering data. When structures are refined against only data from the SAXS region (defined as $0 < q < 0.3$ Å^{-1}), the agreement is worse over the SAXS region (χ^2 = 0.13) and significantly worse over the full scattering curve (χ^2 = 0.31)(**Figure 5-8**). When structure models are calculated using only the RDC and SAXS data (no WAXS), there are large deviations between the models and the experimental data in the region q = 0.35-0.45 Å^{-1}. The high residuals in this region are therefore a direct consequence of truncating the SAXS data to $q < 0.30$. The region q = 0.35-0.45 Å^{-1}. corresponds to interatomic distances on the order of 14-18 Å, and visual inspection of the structure models show that both the short Stem

I and 5' stem-loop helices are less well-ordered and more proximal when WAXS data are omitted from the calculations (**Figure 5-9**). When RDCs are the only source of long-range restraints and no SAXS/WAXS data are used in the calculation, the resulting agreement is very poor (χ^2 = 18.6) (**Figure 5-8**). The impact of the various restraints on the structure models is shown visually in **Figure 5-9**. When compared to structure calculations lacking RDC/SAXS/WAXS data, it appears that inclusion of RDCs in the absence of SAXS/WAXS data increases both the overall model rmsd (**Figure 5-9**) and lowers the agreement between the structure models and experimental SAXS/WAXS data (**Figure 5-8**, compare bottom panels). This is likely due to the overfitting inherent to structure calculations with sparse data, where the apparent decrease in model agreement simply reflects a reduction in overfitting of sparse experimental restraints. While inclusion of RDC restraints does not improve the agreement between the structure models and experimental SAXS/WAXS data (compare **Figure 5-7** and **Figure 5-8** top left panel), we do not expect it to, as the contribution of RDC restraints to the overall shape of the molecule are negligible when compared to the contribution of SAXS/WAXS.

The agreements between measured and predicted RDCs are also excellent (**Figure 5-7B-D**). Furthermore, the multiple sets of RDCs allow the calculation of R-factors (**Table 5-2**). Standard jack-knife validation methods are problematic with sparse data, so we chose to randomly remove only 20% of the RDCs and recalculate 48 structures with all other restraints unchanged (**Table 5-2**). The SVD fits to the lowest energy structure with the excluded RDCs show good correlations (**Figure 5-10**). We also used the 'structural noise Monte-Carlo' method in PALES [52] to estimate the propagation of RDC experimental errors in the orientation of the alignment tensor axes. The results for the NH RDCs in the 3 media using 1000 SVD calculations with an amplitude of Gaussian random NH structural variation in a cone matched to the experimental error of the RDCs is shown in **Figure 5-11**. Note that due to the high rhombicity of the Pf1 tensor (0.66), the directions of two approximately equal components S_{zz} and S_{yy} appear occasionally permuted, which is an artifact of the $|S_{zz}| > |S_{yy}|$ convention. The magnitude and

rhombicity of the alignment tensors and the RDC agreement with the NMR structures are included in Table 5-2 and the structure statistical quality indicators are reported in Table 5-1.

We note that the orientations of the 3 helices appear to be relatively fixed with respect to each other, despite a lack of coaxial stacking. Thus the sparse NMR and SAXS/WAXS data do not provide evidence for large-scale interhelical dynamics. However, it is possible that dynamic motions are averaged out during the ~200-500 milliseconds of data acquisition required for the NMR and X-ray scattering experiments. In an effort to detect alternate conformations or dynamics that may be occurring in only a small subset of U4/U6 molecules, we used single molecule Förster Resonance Energy Transfer (smFRET). We prepared U4 RNAs for smFRET by ligating two RNA fragments together to create a single RNA spanning U4 nt 1-83. This U4 fragment also contained a 3' biotin for surface immobilization and a Cy3 FRET donor fluorophore located at either U5 or U38. These U4 RNAs were then annealed to U6 fragments (nt 49-88) containing Cy5 FRET acceptor fluorophores located at either U54 or U74. By combining U4 and U6 RNAs labeled at different positions, we were able to prepare different di-RNAs for reporting on stem I/stem II, stem I/U4 5' stem-loop, or stem II/U4 5' stemloop dynamics (**Figure 5-12 A,D,G**). We could not detect any transitions on the 0.1-100s timescale between different FRET states for any of these di-RNAs (**Figure 5-12 B,E,H**). Furthermore, histograms of calculated FRET efficiencies for each di-RNA showed narrow distributions consistent with a single predominant conformation (**Figure 5-12 C,F,I**). While we cannot exclude either exceedingly rare transitions or the presence of very unstable structures, the smFRET data are consistent with a predominant U4/U6 di-RNA conformation in which stem I, stem II, and the U4 5' stem-loop remain relatively fixed in their orientation.

We observe FRET efficiencies of 0.52 (Stem I-Stem II, **Figure 5-12A-C**), 0.64 (Stem I- U4 5´SL, **Figure 5-12D-F**) and 0.53 (Stem II-U4 5´SL). It is difficult to quantitatively interpret FRET efficiencies in terms of distances, in part because the fluorophores are connected to the RNA via 10 atom (~14 Å) flexible linkers. Assuming a Förster radius of R_0 = 60 Å [53], a FRET efficiency

of 0.52 for Stem I-Stem II, for example, should correspond to a distance of approximately 59 Å. The measured Stem I-Stem II interhelical distance corresponding to the approximate sites of fluorophore attachment in our structures is approximately 67 Å (this distance is approximate because the NMR-SAXS/WAXS structures have a tetraloop at the position corresponding to the Stem I fluorophore). Therefore, we cannot interpret the observed FRET efficiencies in terms of precise distances but note they are in reasonable agreement with the structures, given the uncertainty in the measurement.

Figure 5-1. Secondary structure of U4/U6.

A) Proposed secondary structure diagram of the S. cerevisiae U4/U6 di-snRNA.

B) NMR construct of U4/U6. Numbering corresponds to the yeast numbering in (A). Only the base pairs that could be experimentally determined by NMR are shown as lines for Watson-Crick pairs and dots for wobble pairs.

C) 2D NOESY with NOE walk color coded to match (B).

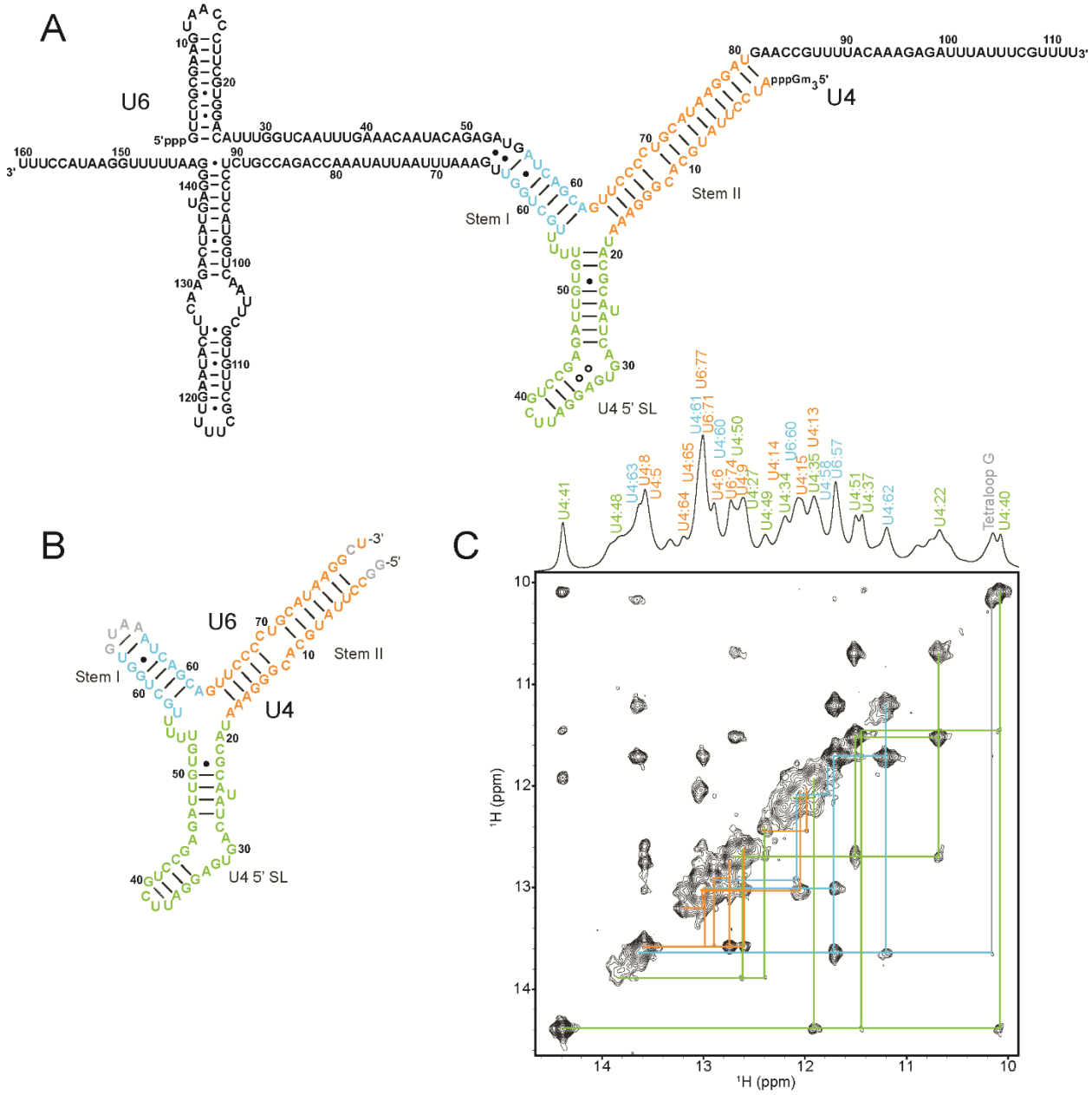


Figure 5-2. NMR of individual helical domains.

A), Individual domain constructs.

B) 2D NOESY of Stem I (blue) overlayed over the 2D NOESY from the entire 92 nucleotide U4-U6 NMR construct (black).

C) 2D NOESY of Stem II (orange) overlayed over the 2D NOESY from the entire 92 nucleotide U4-U6 NMR construct (black). Note that some resonances are shifted due to addition of a GUAA tetraloop.

D) 2D NOESY of the 5' stem-loop (green) overlayed over the 2D NOESY from the entire 92 nucleotide U4-U6 NMR construct (black).

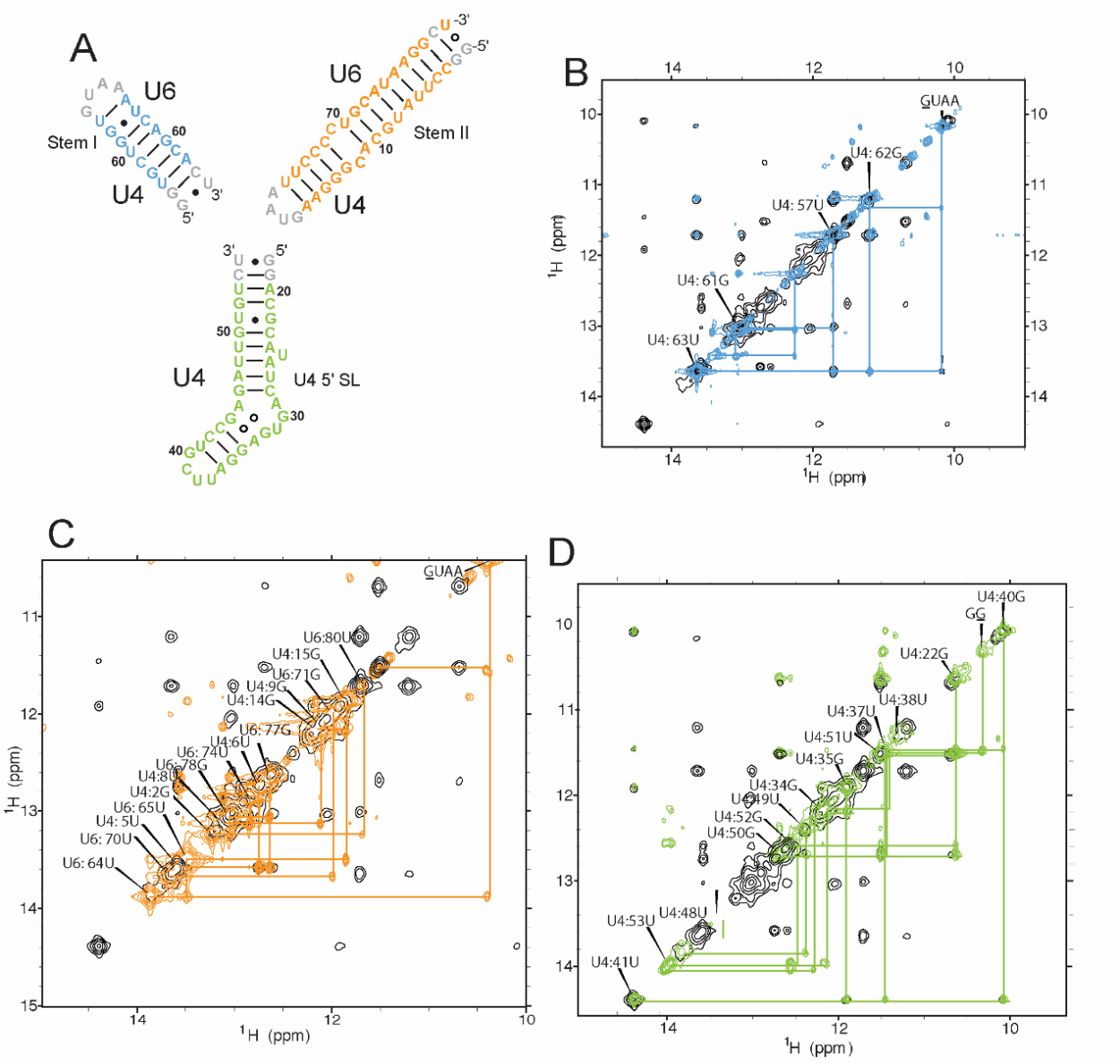

A
U6
Stem I
U4
5'
3'
U6
Stem II
U4
3'
5'
U4
U4 5' SL
B
GUAA
U4: 62G
U4: 57U
U4: 61G
U4: 63U
^{1}H (ppm)
C
GUAA
U6:80U
U4:15G
U6:71G
U4:9G
U4:14G
U6: 77G
U4:6U
U6: 74U
U6: 78G
U4:8U
U4:2G
U6: 65U
U4: 5U
U6: 70U
U6: 64U
^{1}H (ppm)
D
U4:40G
GG
U4:22G
U4:38U
U4:37U
U4:51U
U4:35G
U4:34G
U4:49U
U4:52G
U4:50G
U4:53U
U4:48U
U4:41U
^{1}H (ppm)

Figure 5-3. 2D ^{1}H-^{15}N HSQC-TROSY spectrum of the U4/U6 imino correlations.

Assignments are indicated and color coded to match **Figure 5-1**.

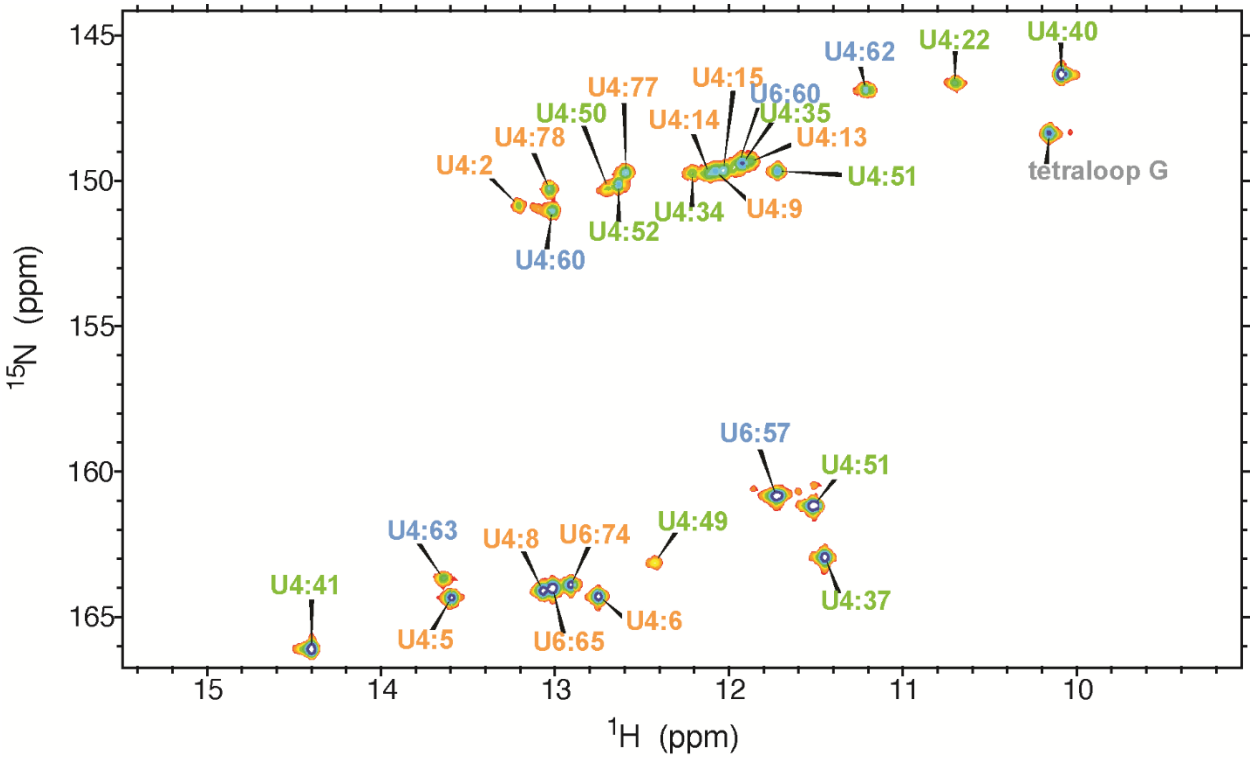


Figure 5-4. Assignment of adenine H2 and C2 resonances for the 92 nucleotide U4-U6 RNA. Left, 2D NOESY showing NOEs between imino and adenine H2 proton resonances. Right, 2D ^{1}H-^{13}C HSQC TROSY of adenine C2-H2 correlations.

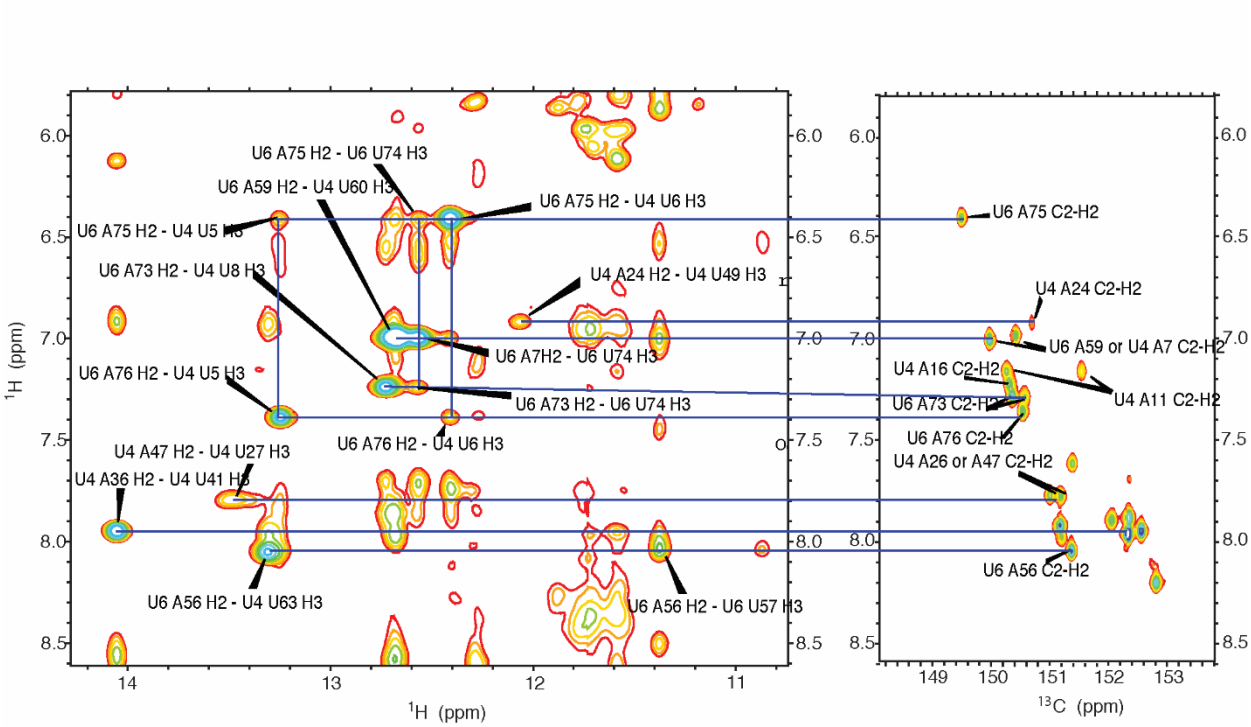


Figure 5-5. Scatter plots of RDCs.

Scatter plots of RDCs measured 3 different ways (Pf1 phage, stretched polyacrylamide gel, and magnetic susceptibility anisotropy (MSA). The MSA RDCs are upscaled in order to compare to the larger magnitude RDCs from Pf1 and stretched polyacrylamide gel.

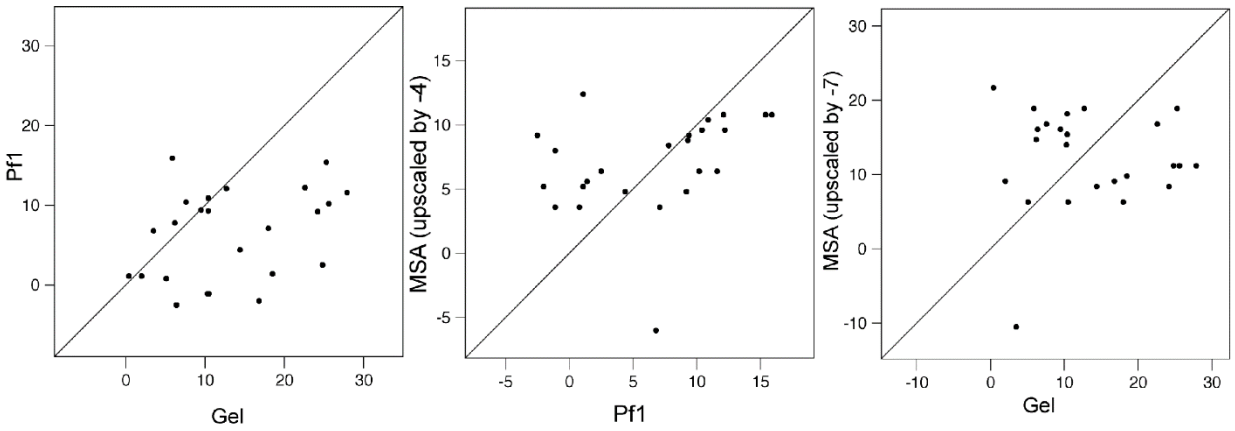


Figure 5-6. NMR-SAXS/WAXS structure of U4/U6.

The ensemble of the 10 lowest energy structures (out of 100 calculated) are shown.

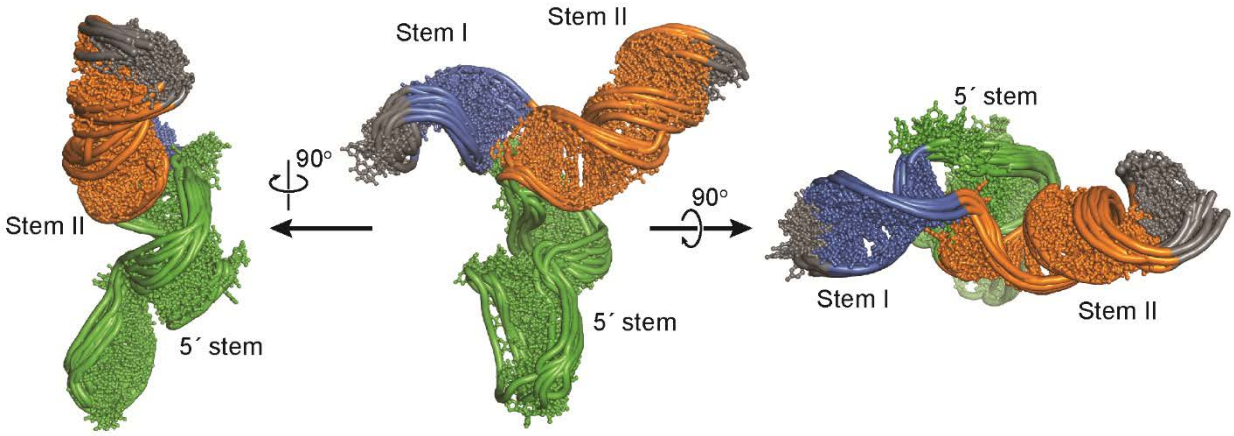


Figure 5-7. Agreement between structure models and the experimental SAXS/WAXS and RDC data.

A) Experimental SAXS/WAXS data (gray) were merged from 10 individual 0.5 second exposures and are plotted with error bars. Back calculated SAXS/WAXS data from the models are shown as colored lines. Residuals are plotted above, in red.

B-D) Agreement between measured and predicted RDC for Pf1 phage (B), negatively charged stretched polyacrylamide gels (C) and magnetic susceptibility anisotropy (D).

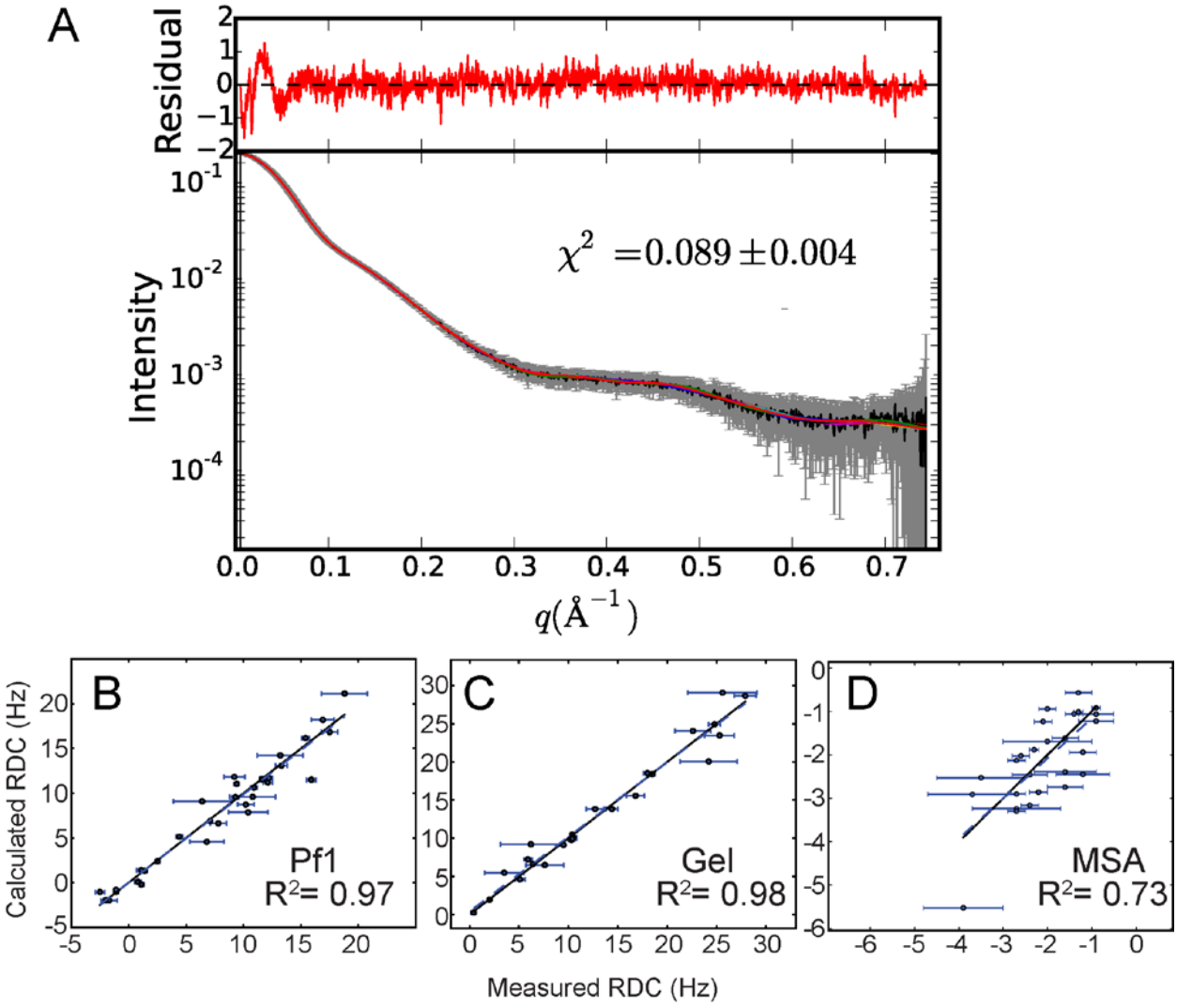


Figure 5-8. Agreement between experimental and predicted scattering data.

Agreement between experimental SAXS/WAXS data (gray) and predicted scattering data from the calculated models (colored lines). Residuals are shown in red. The corresponding structure models are shown in Figure 5.

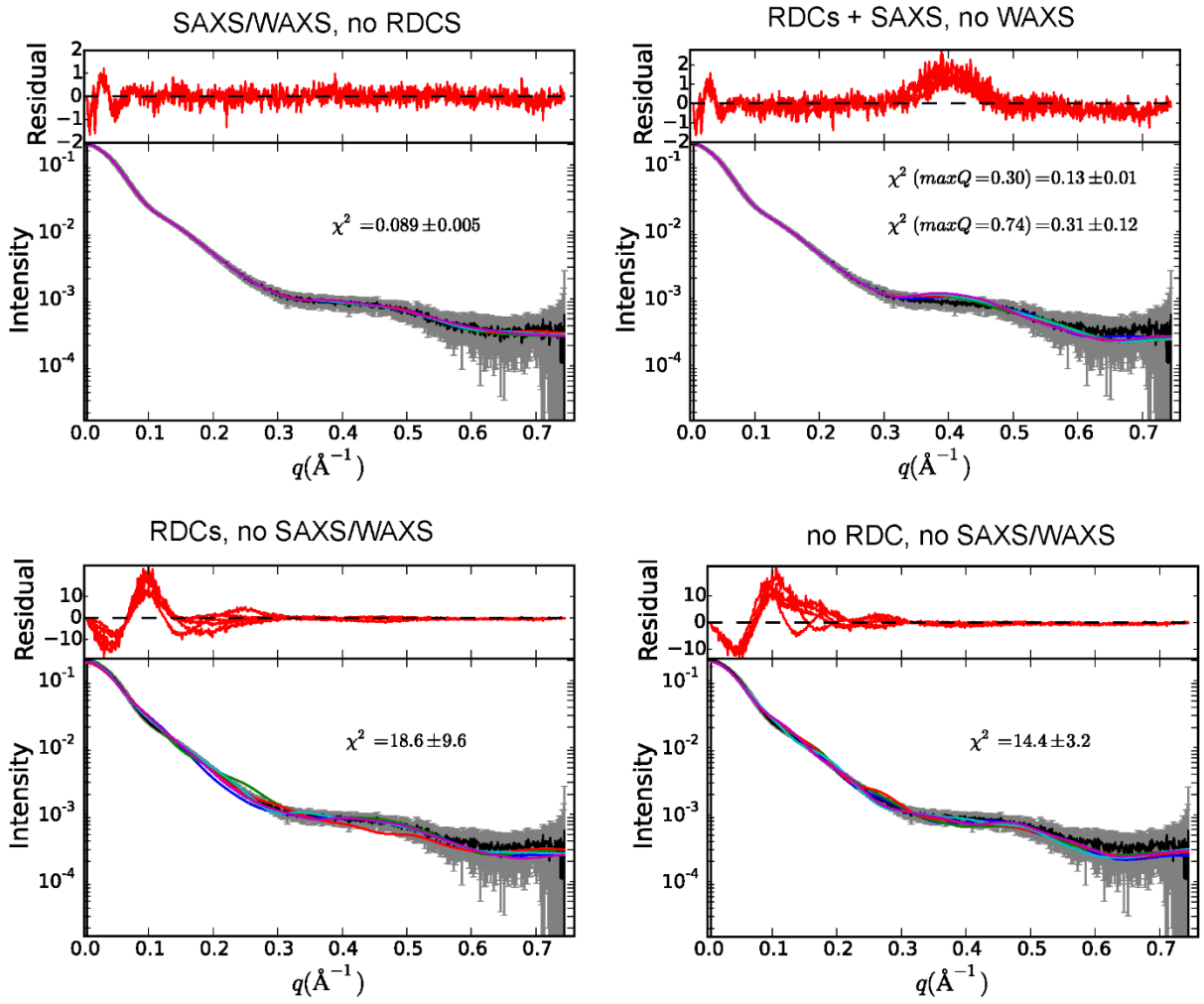
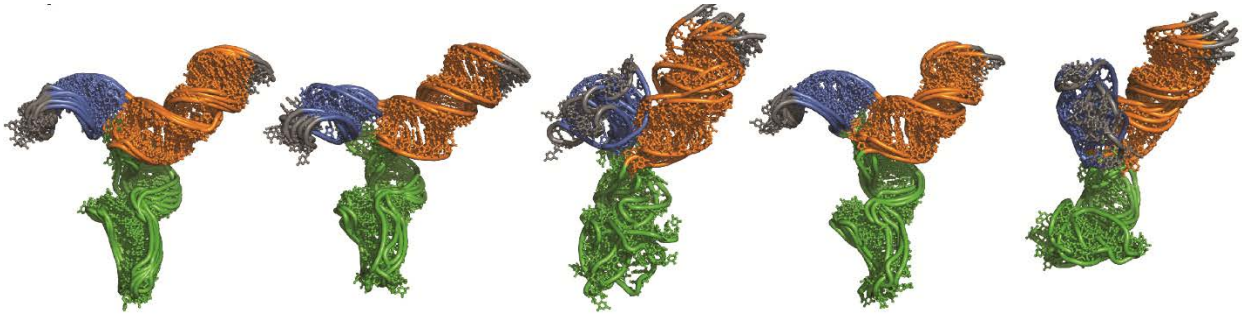


Figure 5-9. Impact of SAXS, WAXS and NMR restraints on structure models.

The lowest 5 energy models out of 48 calculated are shown. RMSDs for the ensembles are indicated.



NOE + RDC + SAXS/WAXS	without WAXS	without SAXS/-WAXS	without RDCs	without RDCs or SAXS/WAXS
RMSD: 2.4 Å	3.3 Å	7.6 Å	3.2 Å	5.7 Å

Figure 5-10. Jack-knife validation.

SVD fit of approx. 20% of RDCs to the lowest energy structure (out of 48) recalculated with the respective set (red) excluded from the experimental restraints.

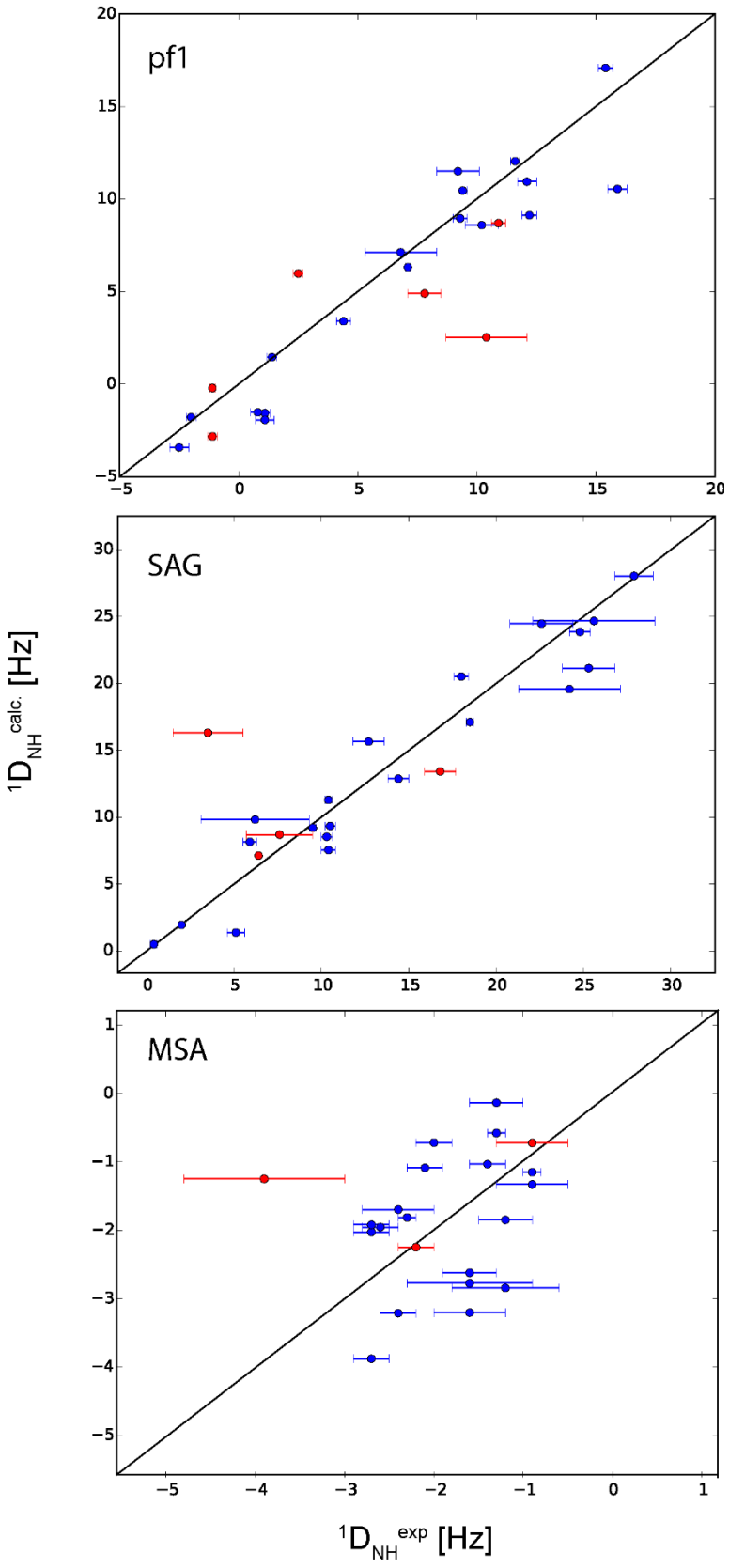

pf1
SAG
MSA
$^1D_{NH}^{calc.}$ [Hz]
$^1D_{NH}^{exp}$ [Hz]

Figure 5-11. Validation using the ‘structural noise Monte-Carlo’ method.

Validation using the ‘structural noise Monte-Carlo’ method in PALES [52]. 1000 SVD calculations were performed with an amplitude of Gaussian random NH structural variation in a cone matched to the RDCs’ experimental error. Sanson-Flamsteed (sinusoidal) representation of alignment tensor axes, blue/green/red represent the orientation of the $S_{zz}/S_{yy}/S_{xx}$ axes, respectively. Gridlines are spaced by 20°.

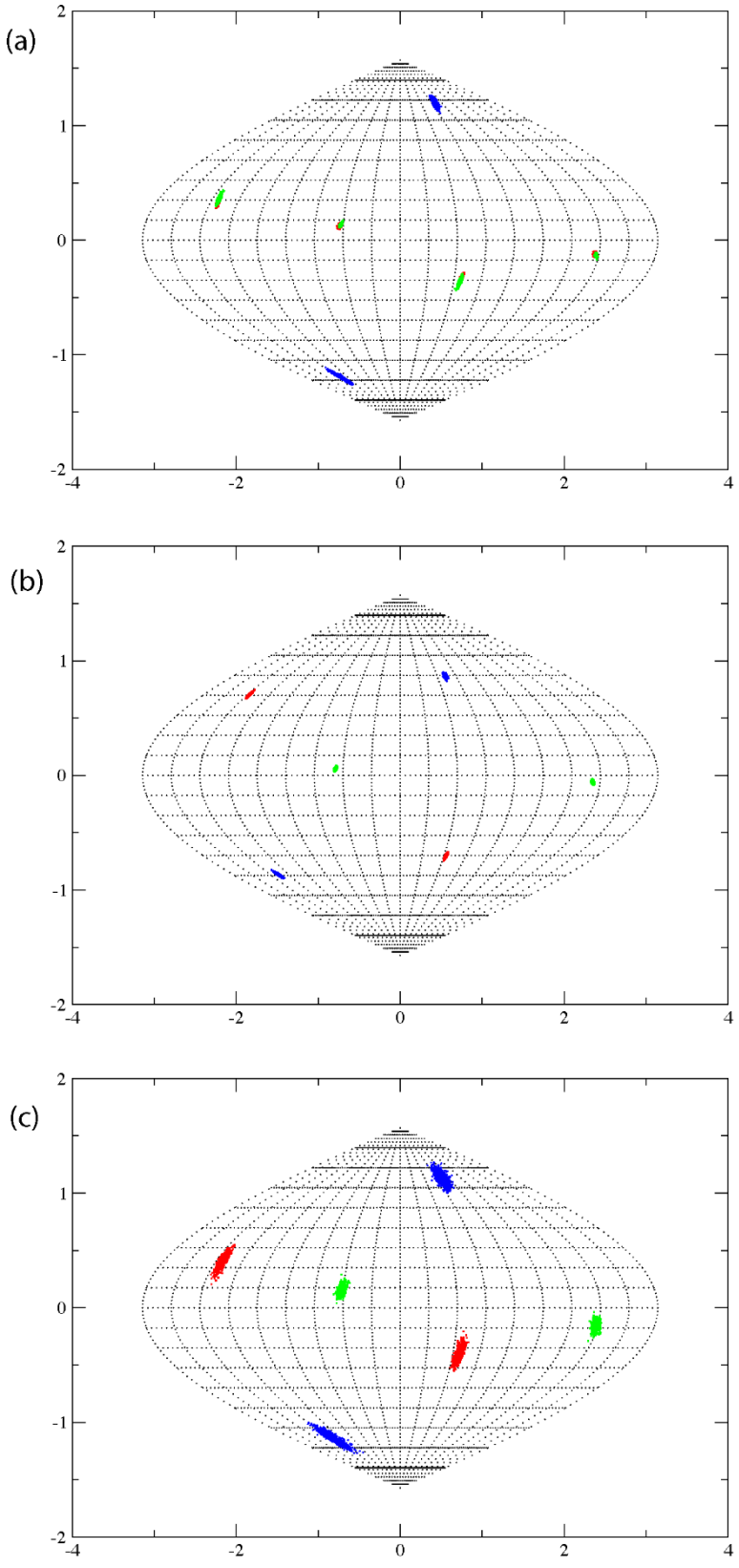

(a)
(b)
(c)
2
1
0
-1
-2
-4
-2
0
2
4

Figure 5-12. smFRET data from U4/U6 di-RNAs containing fluorophores in either U4/U6 stem I, U4/U6 stem II, or the U4 5' stemloop.

A) Diagram of a U4/U6 di-RNA for reporting on dynamics between U4/U6 stem I and U4/U6 stem II (double-headed arrow). The U4 RNA (green) contains a Cy3 FRET-donor fluorophore (green star) and a biotin tether for surface immobilization ("B"). The U6 RNA (red) contains a Cy5 FRET-acceptor fluorophore.

B) FRET efficiency (E_{FRET}) trajectory for a single molecule the di-RNA shown in (A).

C) Histogram of E_{FRET} calculated from N = 76 molecules of the di-RNA shown in (A).

D) Diagram of a U4/U6 di-RNA for reporting on dynamics between U4/U6 stem I and the U4 5' stemloop (double-headed arrow).

E) E_{FRET} trajectory for a single molecule the di-RNA shown in (D).

F) Histogram of E_{FRET} calculated from N = 84 molecules of the di-RNA shown in (D).

G) Diagram of a U4/U6 di-RNA for reporting on dynamics between U4/U6 stem II and the U4 5' stemloop (double-headed arrow).

H) E_{FRET} trajectory for a single molecule the di-RNA shown in (G).

I) Histogram of E_{FRET} calculated from N = 98 molecules of the di-RNA shown in (G). Red lines in panels C,F, and I represent the fitting results of each histogram to single exponential Gaussian functions.

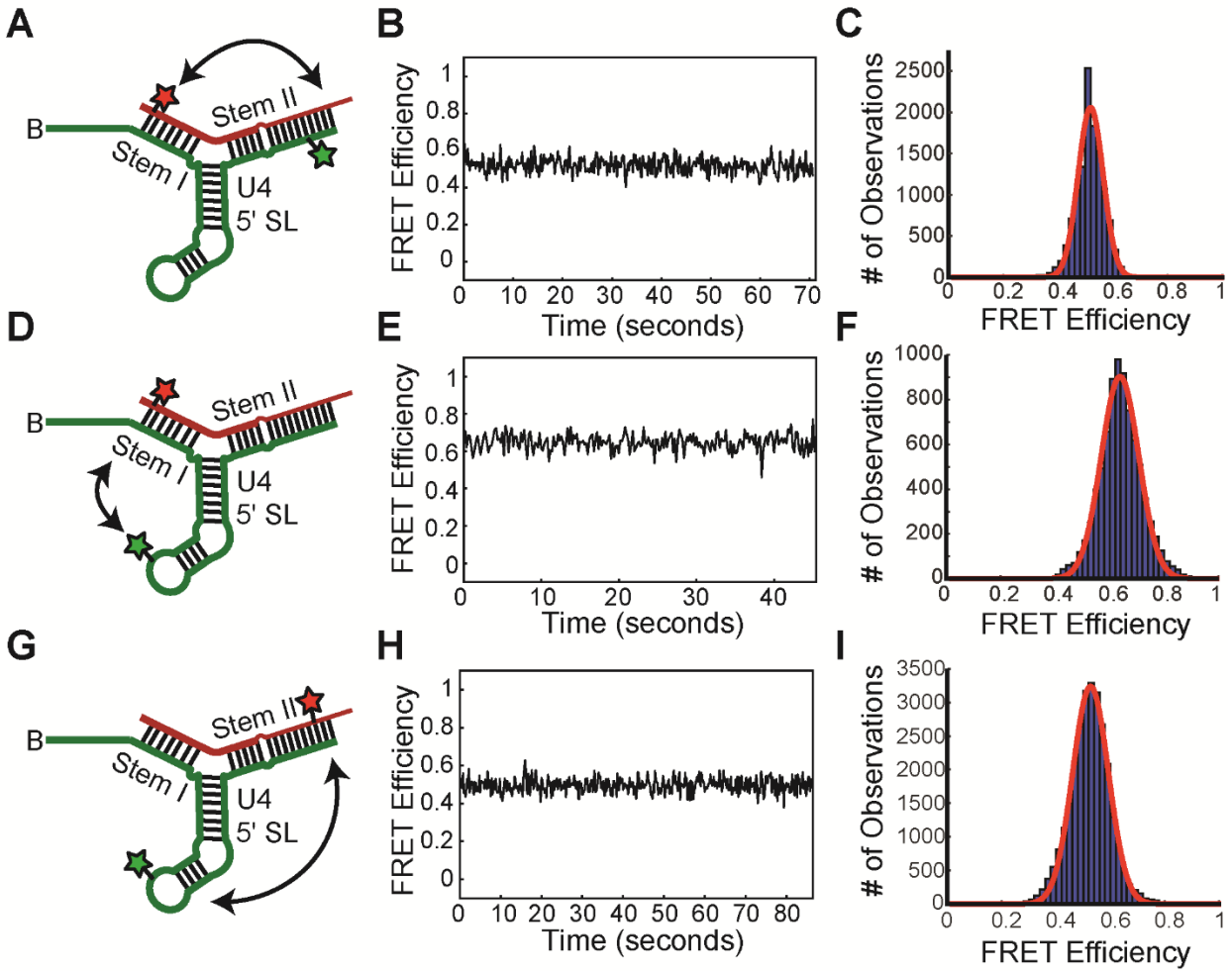

A
B
Stem II
Stem I
U4
5' SL
B
FRET Efficiency
Time (seconds)
C
# of Observations
FRET Efficiency
D
E
F
G
H
I

Table 5-1. **NMR and refinement statistics for U4/U6.**

	U4/U6 10 lowest energy structures
NMR distance and dihedral constraints	
Distance restraints	
Total NOE	685
Intra-residue	640
Inter-residue (experimental)	47
Sequential (|i-j| = 1)	33
Non-sequential (|i-j| > 1)	12
Hydrogen bonds	191
P-P distance	99
Total dihedral angle restraints	615
RDCs (experimental)	29
Structure statistics	
RMSD (dev)	
Distance constraints (Å)	0.05±0.01
Hydrogen bond constraints (Å)	0.06±0.01
Dihedral angle constraints (°)	2.2±0.3
Deviations from idealized geometry	
Bond lengths (Å)	0.006±0.000
Bond angles (°)	0.79±0.02
Impropers (°)	0.54±0.03
Average pairwise r.m.s.d.[a] (Å)	
All RNA heavy	2.41

[a]Pairwise r.m.s.d. was calculated among 10 lowest energy out of 100 refined structures.

Table 5-2. Agreement of the U4/U6 NMR Structures with Experimental RDCs.

RDC (medium/count)	RDC Cross Validation[a]				
	R[b] (std. dev.)/ Pearson's R[d]	$R_{infinity}$[b] (std. dev.)	R_{free}[c] / Pearson's R[c]	Da[b] (std. dev.) [Hz]	Rho[b] (std. dev.)
$^1D_{NH}$ (Pf1/24) **$^1D_{CH}$ (Pf1/8)[f]**	3.1 (0.5) / 97.9%	7.4 (0.9)	17.4 / 72.2% 10.8 / 87.1%[e]	-19.5 (0.6)	0.66 (0.01)
$^1D_{NH}$ (SAG/24)	0.6 (0.1) / 98.2%	1.6 (0.2)	14.6 / 0.1% 4.6 / 99.1%[e]	-45.1 (1.6)	0.49 (0.02)
$^1D_{NH}$ (MSA/22) **$^1D_{CH}$ (MSA/4)[f]**	7.0 (0.8) / 70.4%	15.7 (1.9)	24.9 / 26.6% 2.1 / 99.9%[e]	4.6 (0.1)	0.52 (0.03)

[a]The 10 NMR structures deposited in PDB were used for the analysis (except for R_{free})

[b]R-factor, $R_{infinity}$ (in percent), the magnitude (Da) and Rho(mbicity) of the alignment tensors are defined as in XPLOR-NIH [67].

[c]R_{free} defines how well ~20% of the RDCs fit the lowest energy structure (out of 48) recalculated with the respective set of RDCs excluded from the experimental restraints.

[d]SVD fitted to the lowest energy structure (out of 100).

5.4 Discussion

The NMR-SAXS/WAXS approach described here can be used to determine the overall orientation of helical junctions with a precision (rmsd) of < 3 Å. However, the nts at the interface of the junction are not well determined by this approach. More precise structural information on single stranded regions that may form potential tertiary interactions requires additional NMR data and more expensive labeling strategies such as ^{2}H [5] or selective ^{13}C,^{15}N labeling [54]. NMR-SAXS/WAXS is an efficient and cost-effective strategy for determining the overall fold of an RNA and may inform the decision as to whether more extensive NMR measurements should be pursued. In this study, the agreement between the NMR data for the individual helices and the intact RNA suggest that stable tertiary interactions are unlikely to form across the junction, a conclusion that is corroborated by the structure models and the smFRET experiments. Our data show that the overall fold of the U4/U6 RNA is fairly well determined by SAXS/WAXS data (3.2 Å rmsd) when the base pairs are restrained to A-form geometry. In this example, the base pairing has been confirmed by experimental NMR data (NOEs), but this information may also be inferred from secondary structure predictions with approximately 73% accuracy depending upon the RNA length and sequence [55]. We show that addition of sparse RDC data lowers the rmsd of the ensembles from 3.2 to 2. 4 Å (**Figure 5-9**). The inclusion of sparse RDCs clearly makes a useful contribution to the global accuracy of the orientations of bonds and secondary structure elements.

Often, the topology of an RNA 3-helix junction can be predicted based on the number of nts in the strands that connect the helices, where helices with no intervening nts are often observed to coaxially stack [56]. The topology of the U4/U6 3-helix junction is not easily predicted, as the strands connecting the helices are nearly equivalent [56]. Moreover, the exact base pairing of the terminal base pairs at the ends of the helices adjacent to the central 3-helix junction cannot be readily determined from the NMR data, because the imino protons at the helical termini rapidly exchange with water and are not observed. The observed Y-shaped conformation is very similar

to helices 20, 21 and 22 in 16S rRNA, which form a Y-shaped 3-helix junction in isolation but fold into a coaxially stacked conformation upon binding the S15 protein [57].

The NMR-SAXS/WAXS and smFRET data show that the 3-helix junction is not coaxially stacked and adopts a single conformation with no detectable helical reorientations or alternate conformations occurring on the ms to second timescale. The lack of observeable interhelical dynamics in free U4/U6 RNA was also recently reported by Hardin *et al.* [58]. However, Hardin et al. detected the presence of multiple but non-interconverting FRET states, which were attributed to different conformations of the 5' stem-loop. In contrast, our smFRET histograms show single Gaussian distributions. Potentially, this discrepancy may be related to differences in RNA construct design and fluorophore location. Using fluorophores tethered at the very ends of the helices, Hardin *et al*. measured a Stem I-Stem II FRET efficiency of 0.2 and concluded that Stem I and II span a distance of approximately 76 Å and are coaxially stacked. Although our structure models show that Stems I and II are not coaxially stacked, the maximum distance spanned by these helices in our structures is 79 Å, which is in excellent agreement with the FRET efficiency observed by Hardin *et al*. Thus it is difficult to infer coaxial stacking of multi-helical RNAs from smFRET data alone, as different helical configurations can span very similar distances.

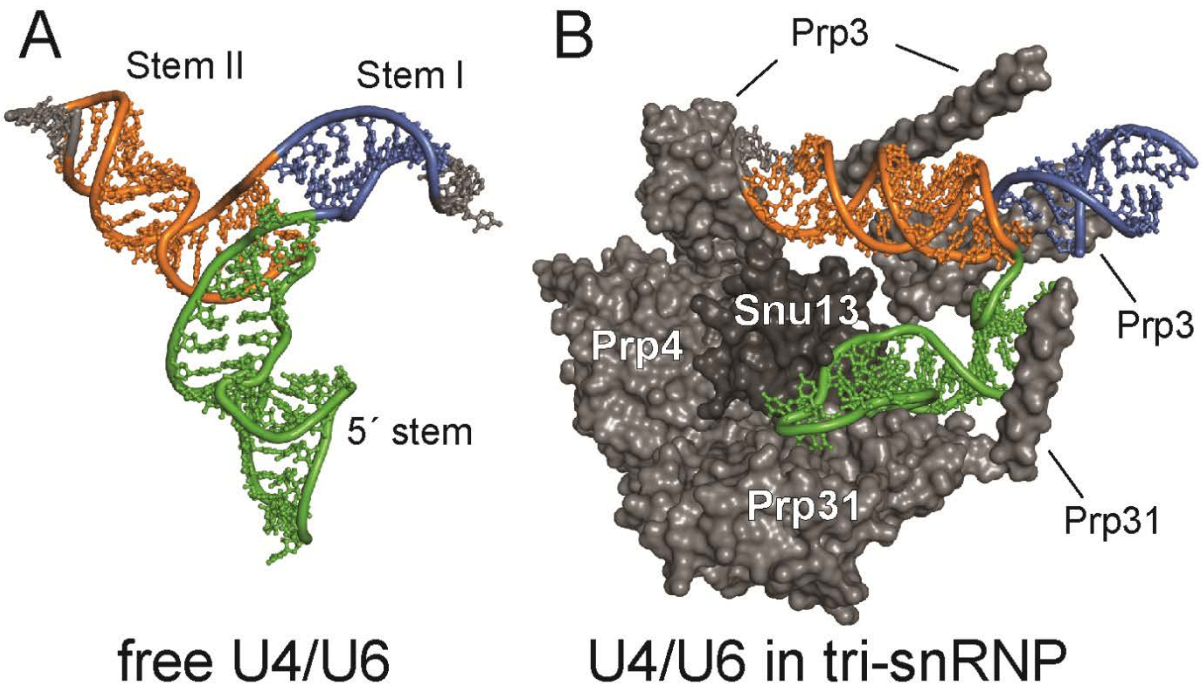
Cryo-electron microscopy was recently used to determine the structure of the yeast U4/U6.U5 tri-snRNP [32], providing a view of the U4/U6 di-snRNA and the interaction between its core base pairing region and the proteins Snu13, Prp3, Prp4 and Prp31. A comparison of this structure with free U4/U6 (**Figure 5-13**) reveals that U4/U6 stems I and II form a coaxial stack and extensively interact with the protein Prp3. Other significant differences in interhelical angles also likely arise from extensive protein interactions. For example, the angle between the 5´ stem-loop and Stem II decreases from approximately 120° in the free RNA to approximately 60° in tri-snRNP. Additionally, the pitch of U4/U6 Stem II is offset by a full half turn, as is the 5´ stem-loop. These differences can be potentially explained by remodeling of the di-snRNA by the proteins: Snu13, Prp3 and Prp31 all bind at or near to the three-helix junction in U4/U6. The U4/U6 proteins

are thought to assemble onto RNA sequentially, where binding of Snu13 and Prp31 to the K-turn motif in free U4 [51, 59-62] allows protein-protein mediated recruitment of Prp3 and Prp4, which are then poised to interact with structural elements present only after annealing of U4 with U6 [32, 61, 63]. In the tri-snRNP structure, Prp3 contacts U4/U6 stem II and appears to contain an alpha helix that reaches across the stem I-stem II junction (**Figure 5-13**). Therefore, the Prp3 binding site on U4/U6 does exist until after U4 and U6 anneal and the stem I and II helices are formed. The initial binding of Snu13 and Prp31 to the U4 5´ stem-loop serves to recruit Prp 3 and Prp4, which remodel the U4/U6 interhelix junction by stabilizing the bound form of the RNA, including the coaxial stacking of stems I and II.

Figure 5-13. Structural comparison of U4/U6 in the presence and absence of spliceosomal proteins.

A) NMR/SAXS structure of free U4/U6.

B) Cryo-EM structure of U4/U6 in the yeast tri-snRNP. Associated protein cofactors in tri-snRNP promote a conformation that is not favored in the free RNA, consistent with their essential role in driving RNA structural rearrangements during spliceosome assembly and disassembly.



5.5 Materials and methods

5.5.1 RNA synthesis and sample preparation

RNA samples were produced via *in vitro* transcription using His_6-tagged T7 RNA polymerase [64] in 40 mM TrisCl pH 8.0, 1 mM spermidine, 0.01% Triton X-100, 38 mM $MgCl_2$, 5 mM DTT, 5.9 mM ATP, 5.4 mM CTP, 7.1 mM GTP, and 7.6 mM UTP. ^{13}C-^{15}N labeled samples were synthesized using ^{13}C-^{15}N labeled nts (Cambridge Isotope Laboratories). The U4-U6 92 nt RNA was transcribed from a modified pUC19 plasmid containing the T7 polymerase promoter sequence, two G nts for efficient transcription, *S. cerevisiae* snRNA U4 nts 3-63, a GUAA linker, *S. cerevisiae* snRNA U6 nts 56-80 with an A79C mutation to stabilize the U4/U6-Stem II helix, then a BsaI restriction site to allow for run-off transcription. RNAs were purified from abortive transcripts and linearized plasmid using an 8% 29:1 acrylamide:bis-acrylamide denaturing gel containing 8 M urea, 89 mM Tris borate, 2 mM EDTA. RNA was visualized by UV shadowing and extracted from the gel by passive diffusion into 0.3 M sodium acetate pH 5.2. Samples were further purified by DEAE anion exchange using low salt-buffer (20 mM Tris pH 7.6, 0.1 mM EDTA, 20 mM NaCl) to load and wash the RNA and a high-salt buffer (20 mM Tris pH 7.6, 0.1 mM EDTA, 1.5 M NaCl) to elute RNA. Samples were ethanol precipitated and resuspended in water. RNA was then diluted to <20 µM in 310 mM KCl, 11.25 mM KPO_4 pH 7.0, 5 mM EDTA and refolded by heating to 90°C for 5 minutes then snap cooling on ice. Samples were concentrated in 3 kDa-cutoff spin concentrators and dialyzed for 24 hours into 20 mM potassium phosphate pH 7.0, 20 mM KCl (NMR samples) or 50 mM Tris pH 7.0, 150 mM KCl (SAXS samples).

5.5.2 SAXS data collection and analysis

SAXS data were obtained at the Advanced Photon Source, Sector 12 at Argonne National Laboratory. Measurements were carried out in 50 mM Tris, pH 7.0, 150 mM KCl as previously described [6, 65].

5.5.3 NMR data collection and analysis

All NMR data were collected on Bruker Avance or Varian NMR spectrometers equipped with cryogenic single z-axis gradient HCN probes at the National Magnetic Resonance Facility at Madison. Resonances were assigned using ^{1}H-^{1}H 2D NOESY with a mixing time of 100 ms and ^{1}H-^{15}N 2D TROSY-HSQC experiments. Isotropic NH and CH couplings were acquired on Bruker and Varian spectrometers equipped with cryogenic probes in symmetrical 5mm Shigemi microtubes at 25°C on 300 µl samples containing 0.6mM ^{15}N-GU labeled and 0.7 mM ^{13}C-A labeled U4-U6, in 20 mM potassium phosphate pH 7.0, 20 mM KCl with 0.15% NaN_3 and 7% $^{2}H_2O$. The samples were used to measure $^{1}D_{NH}$ and $^{1}D_{CH}$ residual dipolar couplings (RDCs) by magnetic field alignment at 900, 800, 750, 600 and 500 MHz, by adding 5 mg/ml Pf1 filamentous bacteriophage, or by hydrating a negatively charged stretched (from 5.4 to 4.2 mm) acrylamide gel (SAG).[45] $^{1}D_{NH}$ and $^{1}D_{CH}$ couplings were measured from 2D ARTSY[46] spectra. Data were processed using NMRPipe package [66].

5.5.4 Structure calculations

All structure calculations were performed with XPLOR-NIH [22, 42, 67]. Non-experimentally derived distance restraints for the U4 kink-turn region (nts 28-35 and 42-46 of U4) were modified from restraints for the NMR structure of the human U4 k-turn (PDB 2XEB) [51]. Restraints for the UUCG tetraloop (nts 37-40 of U4) were obtained from the NMR structure of the P1 helix (PDB 1HLX) [49], and restraints for the GUAA tetraloop between U4 nucleotide 63 and U6 nucleotide 56 was modified from the structure of a GCAA tetraloop [50]. Structures were calculated with XPLOR-NIH [68] with radius of gyration, NOEs, inferred hydrogen bonds and dihedral restraints for A-form helical regions, RNA database orientational restraints, base pair planarity restraints, inter phosphorus distance restraints (global and local), SAXS, and RDC constraints. PyMol (Delano Scientific, LLC) and VMD-XPLOR [69] were used to analyze the structures.

5.5.5 Preparation of fluorophore-labeled RNAs for smFRET

RNAs (**Table 5-3**) were purchased from IDT containing amino allyl modified nucleotides at the positions indicated. RNAs (5 nmol) were fluorescently labeled with mono-reactive NHS ester Cy3 or Cy5 dyes (40 nmol, GE) by overnight incubation at room temperature in labeling buffer (33% DMSO v/v, 100mM sodium bicarbonate pH 8.5). Excess dye was removed using an Illustra microspin G-25 column (GE) prior to gel purification using 12% denaturing polyacrylamide.

5.5.6 Ligation of U4 RNAs for smFRET experiments

U4 RNAs encompassing nt 1-83 were prepared by splinted ligation of U4_1 and U4_2 (**Table 5-3**) containing the appropriate aminoallyl derivatives. Prior to ligation, U4_1 (60 pmol) was phosphorylated with T4 polynucleotide kinase (20U, NEB) for 30 mins at 37^{o}C. U4_2 (120 pmol) and a DNA splint (5' CAC AAT CTC GGA CGA ATC CTC ACT GAT ATG CGT ATT TCC CGT GCA TAA GGA T – 3', 100 pmol) were then added and the oligos annealed by heating to 95°C for 5 mins followed by slow cooling to 25°C over 30 min. Following annealing, ligation was carried out by addition of T4 RNA ligase II (10 U, NEB) and incubation at 37^{o}C for 30 min. Ligation products were purified by 12% denaturing polyacrylamide gel electrophoresis (PAGE).

5.5.7 Heat annealing of U4 and U6 RNAs

U4/U6 di-snRNAs were prepared by heat annealing U6 (3 µM) and U4 (200 nM) RNAs in annealing buffer (50mM Tris pH 7.4, 400 mM NaCl) at 95^{o}C for 5 min followed by slow cooling to room temperature over 30 min.

5.5.8 Single-molecule FRET data collection and analysis

Quartz slides were first passivated with a mixture of polyethylene glycol (PEG) and biotin-derivatized PEG as previously described [70]. Streptavidin (Prozyme) was then added to the slide, incubated, and unbound proteins removed by copious washing with PBS. Heat annealed U4/U6 RNAs were then diluted to 25pM in annealing buffer, flowed onto the slide, and immobilized. Unbound RNAs were then removed by washing with imaging buffer containing an oxygen

scavenging system [71] (50 mM Tris pH 7.4, 400 mM NaCl, 450 µg/µl glucose, 40 U/mL glucose oxidase, 1500 U/mL catalase, and 3 mM Trolox).

Molecules were imaged using a home-built prism-based TIRF microscope [70] using 532 nm and 640 nm lasers for excitation. Images (200 ms exposure) were collected for ~3 minutes by first imaging Cy5 fluorophores with the 640 nm laser, followed by observation of FRET with the 532 nm laser, and ending with imaging of the Cy5 fluorophores with the 640 nm laser. Data were recorded from both the Cy3 and Cy5 emission channels simultaneously on an EM-CCD camera (Andor) using a DualView apparatus (Photometrics) and a 630dcxr dichroic mirror (Chroma).

Well-resolved single-molecules were selected in the red channel (>630 nm) and mapped onto the corresponding molecule in the green channel (<630 nm) using a mathematical mapping function generated for each experiment using fluorescent beads that fluoresce in both channels as fiducial markers. Donor (Cy3) and acceptor (Cy5) intensities were obtained for each frame by integrating the spot intensity in each channel using custom software (Imscroll, [72]) in MatLab (MathWorks). Background subtraction was carried out using Imscroll and photobleaching and blinking were removed with vbFRET [73]. E_{FRET} values for each frame were calculated as $I_A/(I_A+I_D)$, where I_A is the intensity of the acceptor at a particular frame and I_D is the intensity of the donor at a particular frame. E_{FRET} histograms were generated by binning the E_{FRET} values for many individual molecules.

5.5.9 Data availability

Atomic coordinates have been deposited in the Protein Data Bank (PDB ID code 2n7m) and the NMR data in the BioMagResBank (BMRB access code 25811).

Table 5-3. RNA oligonucleotides used for smFRET assays.

Name	Sequence
U6_Stem1[1]	5' - AGA GA**U** GAU CAG CAG UUC CCC UGC AUA AGG AUG AAC CGU U - 3'
U6_Stem2_InternalLabel[1]	5' - AGA GAU GAU CAG CAG UUC CCC UGC A**U**A AGG AUG AAC CGU U - 3'
U4_1[1]	5' - AUC C**U**U AUG CAC GGG AAA UAC GCA UAU CAG UGA- 3'
U4_2[1]	5' - GGA U**U**C GUC CGA GAU UGU GUU UUU GCU GGU UGA AAU UUA AUU AUA AAC C - Biotin 3'

[1]Red nucleotides indicate the position of aminoallyl incorporation.

5.6 Acknowledgements and funding

We thank the Advanced Photon Source (APS) staff for technical support, Charles Schwieters for technical assistance with structure calculations, Ad Bax for helpful comments, and Kiyoshi Nagai for model coordinates of the U4/U6.U5 tri-snRNP. This study made use of the National Magnetic Resonance Facility at Madison, which is supported by NIH grants P41 GM103399 (NIGMS) and P41GM66326 (NIGMS). Additional equipment was purchased with funds from the University of Wisconsin, the NIH (RR02781, RR08438), the NSF (DMB-8415048, OIA-9977486, BIR-9214394), and the USDA. This work was supported by the NIH (NIGMS) grant GM065166 to S.E.B and R00 GM086471 and R01 GM112735 to AAH. AAH and MLR were also supported by the Arnold and Mabel Beckman Foundation, the Shaw Scientist Program of the Greater Milwaukee Foundation, and startup funding from U. Wisconsin-Madison, WARF, and the Department of Biochemistry. MLR is supported by the Molecular Biophysics Training Program (T32-GM08293).

5.7 References

[1] Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, et al. Crystal structure of a group I ribozyme domain: principles of RNA packing. Science. 1996;273:1678-85.

[2] Rose PW, Prlic A, Bi C, Bluhm WF, Christie CH, Dutta S, et al. The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. Nucleic Acids Res. 2015;43:D345-56.

[3] Mustoe AM, Brooks CL, 3rd, Al-Hashimi HM. Topological constraints are major determinants of tRNA tertiary structure and dynamics and provide basis for tertiary folding cooperativity. Nucleic Acids Res. 2014;42:11792-804.

[4] Zuo X, Wang J, Yu P, Eyler D, Xu H, Starich MR, et al. Solution structure of the cap-independent translational enhancer and ribosome-binding element in the 3' UTR of turnip crinkle virus. Proc Natl Acad Sci U S A. 2010;107:1385-90.

[5] Keane SC, Heng X, Lu K, Kharytonchyk S, Ramakrishnan V, Carter G, et al. RNA structure. Structure of the HIV-1 RNA packaging signal. Science. 2015;348:917-21.

[6] Burke JE, Sashital DG, Zuo X, Wang YX, Butcher SE. Structure of the yeast U2/U6 snRNA complex. RNA. 2012;18:673-83.

[7] Fang X, Wang J, O'Carroll IP, Mitchell M, Zuo X, Wang Y, et al. An unusual topological structure of the HIV-1 Rev response element. Cell. 2013;155:594-605.

[8] Fang X, Stagno JR, Bhandari YR, Zuo X, Wang YX. Small-angle X-ray scattering: a bridge between RNA secondary structures and three-dimensional topological structures. Curr Opin Struct Biol. 2015;30:147-60.

[9] Grishaev A, Tugarinov V, Kay LE, Trewhella J, Bax A. Refined solution structure of the 82-kDa enzyme malate synthase G from joint NMR and synchrotron SAXS restraints. Journal of biomolecular NMR. 2008;40:95-106.

[10] Zuo X, Wang J, Foster TR, Schwieters CD, Tiede DM, Butcher SE, et al. Global molecular structure and interfaces: refining an RNA:RNA complex structure using solution X-ray scattering data. J Am Chem Soc. 2008;130:3292-3.

[11] Zuo X, Cui G, Merz KM, Jr., Zhang L, Lewis FD, Tiede DM. X-ray diffraction "fingerprinting" of DNA structure in solution for quantitative evaluation of molecular dynamics simulation. Proc Natl Acad Sci U S A. 2006;103:3534-9.

[12] Jain S, Richardson DC, Richardson JS. Computational Methods for RNA Structure Validation and Improvement. Methods Enzymol. 2015;558:181-212.

[13] Cheng CY, Chou FC, Kladwang W, Tian S, Cordero P, Das R. Consistent global structures of complex RNA states through multidimensional chemical mapping. Elife. 2015;4:e07600.

[14] Cruz JA, Blanchet MF, Boniecki M, Bujnicki JM, Chen SJ, Cao S, et al. RNA-Puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction. RNA. 2012;18:610-25.
[15] Fischer N, Neumann P, Konevega AL, Bock LV, Ficner R, Rodnina MV, et al. Structure of the E. coli ribosome-EF-Tu complex at <3 A resolution by Cs-corrected cryo-EM. Nature. 2015;520:567-70.
[16] Bartesaghi A, Merk A, Banerjee S, Matthies D, Wu X, Milne JL, et al. 2.2 A resolution cryo-EM structure of beta-galactosidase in complex with a cell-permeant inhibitor. Science. 2015;348:1147-51.
[17] Gong Z, Schwieters CD, Tang C. Conjoined use of EM and NMR in RNA structure refinement. PLoS One. 2015;10:e0120445.
[18] Miyazaki Y, Irobalieva RN, Tolbert BS, Smalls-Mantey A, Iyalla K, Loeliger K, et al. Structure of a conserved retroviral RNA packaging element by NMR spectroscopy and cryo-electron tomography. J Mol Biol. 2010;404:751-72.
[19] Grishaev A, Ying J, Canny MD, Pardi A, Bax A. Solution structure of tRNAVal from refinement of homology model against residual dipolar coupling and SAXS data. Journal of biomolecular NMR. 2008;42:99-109.
[20] Wang YX, Zuo X, Wang J, Yu P, Butcher SE. Rapid global structure determination of large RNA and RNA complexes using NMR and small-angle X-ray scattering. Methods. 2010;52:180-91.
[21] Parisien M, Major F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. Nature. 2008;452:51-5.
[22] Schwieters CD, Kuszewski JJ, Clore GM. Using Xplor-NIH for NMR molecular structure determination. Prog Nucl Mag Res Sp. 2006;48:47-62.
[23] Wang J, Zuo X, Yu P, Xu H, Starich MR, Tiede DM, et al. A method for helical RNA global structure determination in solution using small-angle x-ray scattering and NMR measurements. J Mol Biol. 2009;393:717-34.
[24] Al-Hashimi HM, Valafar H, Terrell M, Zartler ER, Eidsness MK, Prestegard JH. Variation of molecular alignment as a means of resolving orientational ambiguities in protein structures from dipolar couplings. J Magn Reson. 2000;143:402-6.
[25] Latham MP, Hanson P, Brown DJ, Pardi A. Comparison of alignment tensors generated for native tRNA(Val) using magnetic fields and liquid crystalline media. Journal of biomolecular NMR. 2008;40:83-94.
[26] Bardaro MF, Jr., Varani G. Independent alignment of RNA for dynamic studies using residual dipolar couplings. Journal of biomolecular NMR. 2012;54:69-80.

[27] Ying J, Grishaev A, Latham MP, Pardi A, Bax A. Magnetic field induced residual dipolar couplings of imino groups in nucleic acids from measurements at a single magnetic field. Journal of biomolecular NMR. 2007;39:91-6.
[28] Rinke J, Appel B, Digweed M, Luhrmann R. Localization of a base-paired interaction between small nuclear RNAs U4 and U6 in intact U4/U6 ribonucleoprotein particles by psoralen cross-linking. J Mol Biol. 1985;185:721-31.
[29] Hashimoto C, Steitz JA. U4 and U6 RNAs coexist in a single small nuclear ribonucleoprotein particle. Nucleic Acids Res. 1984;12:3283-93.
[30] Zucker-Aprison E, Thomas JD, Blumenthal T. C. elegans snRNAs: a model for U4/U6 base pairing. Nucleic Acids Res. 1988;16:7188.
[31] Brow DA, Guthrie C. Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. Nature. 1988;334:213-8.
[32] Nguyen TH, Galej WP, Bai XC, Savva CG, Newman AJ, Scheres SH, et al. The architecture of the spliceosomal U4/U6.U5 tri-snRNP. Nature. 2015;523:47-52.
[33] Fabrizio P, Dannenberg J, Dube P, Kastner B, Stark H, Urlaub H, et al. The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. Mol Cell. 2009;36:593-608.
[34] Maeder C, Kutach AK, Guthrie C. ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C terminus of Prp8. Nat Struct Mol Biol. 2009;16:42-8.
[35] Mozaffari-Jovin S, Santos KF, Hsiao HH, Will CL, Urlaub H, Wahl MC, et al. The Prp8 RNase H-like domain inhibits Brr2-mediated U4/U6 snRNA unwinding by blocking Brr2 loading onto the U4 snRNA. Genes Dev. 2012;26:2422-34.
[36] Mozaffari-Jovin S, Wandersleben T, Santos KF, Will CL, Luhrmann R, Wahl MC. Inhibition of RNA helicase Brr2 by the C-terminal tail of the spliceosomal protein Prp8. Science. 2013;341:80-4.
[37] Raghunathan PL, Guthrie C. RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. Curr Biol. 1998;8:847-55.
[38] Zhang L, Li X, Hill RC, Qiu Y, Zhang W, Hansen KC, et al. Brr2 plays a role in spliceosomal activation in addition to U4/U6 unwinding. Nucleic Acids Res. 2015;43:3286-97.
[39] Hang J, Wan R, Yan C, Shi Y. Structural basis of pre-mRNA splicing. Science. 2015.
[40] Keating KS, Toor N, Perlman PS, Pyle AM. A structural analysis of the group II intron active site and implications for the spliceosome. RNA. 2010;16:1-9.
[41] Fica SM, Mefford MA, Piccirilli JA, Staley JP. Evidence for a group II intron-like catalytic triplex in the spliceosome. Nat Struct Mol Biol. 2014;21:464-71.

[42] Schwieters CD, Clore GM. Using small angle solution scattering data in Xplor-NIH structure calculations. Prog Nucl Magn Reson Spectrosc. 2014;80:1-11.
[43] Lu K, Heng X, Garyu L, Monti S, Garcia EL, Kharytonchyk S, et al. NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging. Science. 2011;334:242-5.
[44] Hansen MR, Mueller L, Pardi A. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. Nat Struct Biol. 1998;5:1065-74.
[45] Ulmer TS, Ramirez BE, Delaglio F, Bax A. Evaluation of backbone proton positions and dynamics in a small protein by liquid crystal NMR spectroscopy. J Am Chem Soc. 2003;125:9179-91.
[46] Ying J, Wang J, Grishaev A, Yu P, Wang YX, Bax A. Measurement of (1)H-(15)N and (1)H-(13)C residual dipolar couplings in nucleic acids from TROSY intensities. Journal of biomolecular NMR. 2011;51:89-103.
[47] Sass J, Cordier F, Hoffmann A, Cousin A, Omichinski JG, Lowen H, et al. Purple membrane induced alignment of biological macromolecules in the magnetic field. Journal of the American Chemical Society. 1999;121:2047-55.
[48] Musselman C, Pitt SW, Gulati K, Foster LL, Andricioaei I, Al-Hashimi HM. Impact of static and dynamic A-form heterogeneity on the determination of RNA global structural dynamics using NMR residual dipolar couplings. Journal of biomolecular NMR. 2006;36:235-49.
[49] Allain FH, Varani G. Structure of the P1 helix from group I self-splicing introns. J Mol Biol. 1995;250:333-53.
[50] Jucker FM, Heus HA, Yip PF, Moors EH, Pardi A. A network of heterogeneous hydrogen bonds in GNRA tetraloops. J Mol Biol. 1996;264:968-80.
[51] Falb M, Amata I, Gabel F, Simon B, Carlomagno T. Structure of the K-turn U4 RNA: a combined NMR and SANS study. Nucleic Acids Res. 2010;38:6274-85.
[52] Zweckstetter M, Bax A. Evaluation of uncertainty in alignment tensors obtained from dipolar couplings. Journal of biomolecular NMR. 2002;23:127-37.
[53] Murphy MC, Rasnik I, Cheng W, Lohman TM, Ha T. Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. Biophys J. 2004;86:2530-7.
[54] Liu Y, Holmstrom E, Zhang J, Yu P, Wang J, Dyba MA, et al. Synthesis and applications of RNAs with position-selective labelling and mosaic composition. Nature. 2015;522:368-72.
[55] Mathews DH, Turner DH. Prediction of RNA secondary structure by free energy minimization. Curr Opin Struct Biol. 2006;16:270-8.

[56] Lescoute A, Westhof E. Topology of three-way junctions in folded RNAs. RNA. 2006;12:83-93.
[57] Batey RT, Williamson JR. Effects of polyvalent cations on the folding of an rRNA three-way junction and binding of ribosomal protein S15. RNA. 1998;4:984-97.
[58] Hardin JW, Warnasooriya C, Kondo Y, Nagai K, Rueda D. Assembly and dynamics of the U4/U6 di-snRNP by single-molecule FRET. Nucleic Acids Res. 2015.
[59] Liu S, Ghalei H, Luhrmann R, Wahl MC. Structural basis for the dual U4 and U4atac snRNA-binding specificity of spliceosomal protein hPrp31. RNA. 2011;17:1655-63.
[60] Mougin A, Gottschalk A, Fabrizio P, Luhrmann R, Branlant C. Direct probing of RNA structure and RNA-protein interactions in purified HeLa cell's and yeast spliceosomal U4/U6.U5 tri-snRNP particles. J Mol Biol. 2002;317:631-49.
[61] Nottrott S, Urlaub H, Luhrmann R. Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. EMBO J. 2002;21:5527-38.
[62] Vidovic I, Nottrott S, Hartmuth K, Luhrmann R, Ficner R. Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. Mol Cell. 2000;6:1331-42.
[63] Liu S, Mozaffari-Jovin S, Wollenhaupt J, Santos KF, Theuser M, Dunin-Horkawicz S, et al. A composite double-/single-stranded RNA-binding region in protein Prp3 supports tri-snRNP stability and splicing. Elife. 2015;4.
[64] Milligan JF, Uhlenbeck OC. Synthesis of small RNAs using T7 RNA polymerase. Methods Enzymol. 1989;180:51-62.
[65] Burke JE, Butcher SE. Nucleic acid structure characterization by small angle X-ray scattering (SAXS). Curr Protoc Nucleic Acid Chem. 2012;Chapter 7:Unit7 18.
[66] Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. Journal of biomolecular NMR. 1995;6:277-93.
[67] Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM. The Xplor-NIH NMR molecular structure determination package. J Magn Reson. 2003;160:65-73.
[68] Tian Y, Schwieters CD, Opella SJ, Marassi FM. A practical implicit solvent potential for NMR structure calculation. J Magn Reson. 2014;243:54-64.
[69] Schwieters CD, Clore GM. The VMD-XPLOR visualization package for NMR structure refinement. J Magn Reson. 2001;149:239-44.
[70] Roy R, Hohng S, Ha T. A practical guide to single-molecule FRET. Nat Methods. 2008;5:507-16.
[71] Aitken CE, Marshall RA, Puglisi JD. An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. Biophys J. 2008;94:1826-35.

[72] Friedman LJ, Chung J, Gelles J. Viewing dynamic assembly of molecular complexes by multi-wavelength single-molecule fluorescence. Biophys J. 2006;91:1023-31.

[73] Bronson JE, Fei J, Hofman JM, Gonzalez RL, Jr., Wiggins CH. Learning rates and states from biophysical time series: a Bayesian approach to model selection and single-molecule FRET data. Biophys J. 2009;97:3196-205.

Chapter 6: Conclusions and Future Directions

6.1 Conclusions

My work on the early steps of U6 snRNP assembly and annealing have highlighted an essential spliceosomal assembly process that is poorly understood at the molecular level. Using a combination of structural biology, biochemistry, and genetics, I have helped to define these early steps for yeast U6. I have reconstituted the majority of the early assembly steps *in vitro*: initial binding by Lhp1, Usb1 processing, Prp24 and Lsm2-8 binding, and annealing to U4. The work in this dissertation lays the groundwork for more intricate studies of the roles of Usb1 and Prp24. Continued study of Usb1 is warranted because of its unique and evolutionarily divergent cyclic phosphodiesterase while Prp24-mediated annealing is one of the best examples of an ATP-independent RNA chaperone at work. Both areas of research necessitate the integration of different and creative approaches, and may serve as quintessential examples of important RNA-binding proteins that will aid our understanding of RNA biology.

6.2 Future Directions

Many exciting advances in the splicing field occurred during my graduate career. The cryoEM revolution has brought us near-atomic resolution structures of multiple splicing complexes. Namely, the structures of the tri-snRNP (NGUYEN *et al.* 2016) and B complex (PLASCHKA *et al.* 2017) have shown us the large structural rearrangement that must occur in U6 during B complex formation. Hopefully in the near future, cryoEM will be used to look at the U4/U6 annealing process and allow us molecular-level insight into how U4 is recognized and annealed to U6. Since the annealing process occurs on the minute timescale *in vitro* and because its components can be produced easily, time-resolved cryoEM (FRANK 2017) should be possible and could be applied to U4/U6 annealing.

Although my work has significantly advanced our understanding of U6 snRNP formation and U4/U6 annealing, many unanswered questions remain. In the sections below, I propose new lines of investigation and potential experiments to further our understanding of U6 snRNP

assembly, with a focus on the following unanswered questions that have come from or been highlighted by my work:

1. What is the essential role of the N-terminal domain of Usb1?

2. Why does human Usb1 lack cyclic phosphodiesterase activity (or why does yeast Usb1 possess CPDase activity)? What are structural determinants of CPDase activity? How did Lsm2-8 co-evolve with Usb1 to tightly bind its product?

3. How does 3' end modification affect U6 RNA level and snRNP formation *in vivo*? How does this affect splicing kinetics and fidelity?

4. Why do Prp24 and Lhp1 bind anti-cooperatively? How does the U6 snRNP form *in vivo*?

5. How is the interlocked topology of U6-Prp24 formed? Do U6 and Prp24 co-fold? Is the interlocked topology conserved?

6. What is the role of the electropositive groove in Prp24? How does the Lsm2-8 ring stimulate annealing?

7. How do U4/U6 di-snRNP proteins affect annealing kinetics and Prp24 release? How is the interlocked U6-Prp24 topology resolved after annealing?

6.2.1 Defining the role of the N-terminal domain of Usb1 *in vivo*

The N-terminal domain of Usb1 is essential for yeast viability (HILCENKO *et al.* 2013; DIDYCHUK *et al.* 2017), yet it does not contribute to 3'-5' exonuclease activity (DIDYCHUK *et al.* 2017). Yeast cells are viable when Usb1 is truncated by removing the first 43, but not the first 44 residues (**Appendix 2**). Deletion into the N-terminal domain affects Usb1 levels *in vivo*, suggesting that the N-terminal domain is important for expression or stability. As discussed in **Appendix 2**, whether the N-terminal domain is involved in post-transcriptional or post-translational regulation of Usb1 should be determined first. Provided that regulation occurs at the post-translational level, it is likely that the N-terminal domain plays a role in stabilizing Usb1 by mediating a protein-protein interaction that is lost upon truncation. *In vivo* BPA crosslinking could

be performed using the amber stop codon nonsense suppressor tRNA/tRNA synthetase system (CHIN *et al.* 2003). Using this system, BPA crosslinkers could be inserted at sites between residues 30-60, photocrosslinked, and immunoprecipitated to aid in the identification of interacting partners. U6 snRNA is the primary target of Usb1 (HILCENKO *et al.* 2013; SHCHEPACHEV *et al.* 2015), suggesting that Usb1 is highly specific. This specificity may be determined in part by the N-terminal domain. Interestingly, in human cells, Usb1 also processes Vault RNA 1-1 (SHCHEPACHEV *et al.* 2015), a non-coding RNA of unknown function.

6.2.2 Understanding the determinants of cyclic phosphodiesterase activity

The structures of yeast and human Usb1 are very similar despite low sequence identity, and their active sites are nearly superimposable (DIDYCHUK *et al.* 2017). Additionally, the active site of Usb1 resembles many other enzymes with cyclic phosphodiesterase activity, yet these enzymes leave different products. How is cyclic phosphodiesterase activity determined? Since the active sites are nearly identical in all cases, other factors must necessarily be at play. Residues immediately outside the active site influence catalytic activity in a poorly understood way (DIDYCHUK *et al.* 2017). One potential approach is to structurally and biochemically characterize multiple homologs of Usb1 and correlate structural features (such as loop length, aromatic residues at the position homologous to F78/H84, hydrophobicity of the active site, and active site histidine p*K*a's) with the activity observed *in vitro* (namely whether or not the enzyme produces a 2',3'-cyclic phosphate or non-cyclic phosphate). Since Usb1 is poorly conserved at the sequence level yet it possesses a common protein scaffold, it represents an interesting enzyme evolution case study. Additionally, since Lsm2-8 from yeast and human preferentially binds the product of its corresponding Usb1, Usb1 and Lsm2-8 may have co-evolved to ensure that U6 is properly bound after Usb1 modification.

6.2.3 Defining the role of 3' end modification *in vivo*

Usb1 modification prevents Lhp1 re-binding and promotes Lsm2-8 binding. However, the precise role of Usb1 modification *in vivo* remains to be determined. Deletion of *USB1* can be

rescued by overexpressing U6 snRNA (HILCENKO *et al.* 2013) or by overexpressing other U6 binding factors (**Appendix 4**), including Lhp1. Thus, yeast can survive without a 3' phosphate on U6. Interestingly, when Lhp1 is overexpressed and Usb1 is bypassed, a new, slower-migrating U4/U6 species can be observed on a native gel (**Appendix 4**). It would be interesting to see whether Lhp1 is retained in the U4/U6.U5 tri-snRNP, if it persists into B complex, and if it is displaced by the NTC in a manner similar to Lsm2-8. More generally, the question remains as to whether unmodified U6 affects spliceosome assembly or fidelity. Increasing U6 levels to overcome the decreased degradation observed in Usb1-depleted cells (MROCZEK *et al.* 2012; HILCENKO *et al.* 2013) raises an interesting question of how U6 levels may also affect spliceosome assembly or fidelity. A simple experiment could be to compare splicing fidelity using an *ACT1CUP1* assay with and without a U6 overexpression plasmid (i.e. p424-snR6) with and without Usb1 deleted. Even when Usb1 is present in the cell, overexpression of U6 likely leads to an increase in unmodified U6 which is presumably incorporated into spliceosomes. Determining how the 3' end modification and U6 levels affects splicing fidelity could have important implications for disease, as U6 levels vary significantly in human tissues and could be dysregulated in disease. Indeed, mutations in Usb1 are linked to the disease poikiloderma with neutropenia and myelodysplastic syndromes in humans (WANG *et al.* 2017).

How U6 snRNA is normally degraded is also unknown. *In vitro* work suggests that the nuclear exosome associated exoribonucleases Rrp6 and Rrp44 have different abilities to process substrates with 3' phosphates (ZINDER AND LIMA 2017). Since Rrp44 can degrade 3' phosphate-containing substrates, is U6 normally degraded by Rrp44? When Usb1 is deleted, do U6 half-lives decrease because Rrp6 can now also degrade U6? Do genetic interactions between Usb1 and Rrp6 and Rrp44 exist (i.e. does deletion of Rrp6 rescue Usb1 deletion)? Additionally, how does U6 structure or the presence of U6-binding proteins affect degradation rates?

6.2.4 Defining the order of U6 snRNP assembly

Despite non-overlapping binding sites, Prp24 and Lhp1 bind anti-cooperatively *in vitro* while Prp24 and Lsm2-8 bind cooperatively (DIDYCHUK *et al.* 2017). Why do Lhp1 and Prp24 exhibit this anti-cooperativity? One possibility is that Lhp1 recognizes additional secondary structure that is occluded by Prp24 binding, while another possibility is that Lhp1 and Prp24 have unfavorable electrostatics that preclude co-binding. Lhp1 and Prp24 could be fluorescently labeled and a colocalization single-molecule spectroscopy (CoSMoS) experiment performed with fluorescently labeled U6 attached via the 5' end to the slide surface to observe directly whether Prp24 directly displaces Lhp1 (i.e. Lhp1 is bound, then Prp24 binds and Lhp1 is displaced) or if binding of the two proteins are mutually exclusive. If Lsm2-8 is also fluorescently labeled, one could reconstitute the U6 snRNP assembly pathway on the single molecule level *in vitro*. Although the binding site for Prp24 (the high affinity binding site between nucleotides 40-60) exists before the binding site for either Lhp1 or Lsm2-8 (the 3' tail), it remains to be determined how the U6 snRNP assembles *in vivo* and if Prp24 is responsible for displacing Lhp1.

6.2.5 Understanding the mechanism of U6-Prp24 binding

The interlocked topology observed in the U6-Prp24 snRNP core structure is unprecedented and raises many questions about how such a topology is formed. The RNA and protein could co-fold, or one component could fold around the other. The dynamics of U6 have been measured using single molecule FRET with fluorescent dyes at U54 and U90, which showed that U6 RNA alone has very little structural dynamics, but that U6 in the context of U4/U6 di-snRNA is more dynamic and samples two structural conformations (RODGERS *et al.* 2016). However, these dyes are incompatible with high-affinity Prp24 binding (MONTEMAYOR *et al.* 2014) and report primarily on the structure of the extended ISL. These dyes could be moved to the base of the telestem (i.e. nucleotides U32 and U101) and their dynamics could be measured in the presence and absence of Prp24 using smFRET (**Figure 6-1A**). Prp24 could also be site-specifically labeled with fluorescent dyes using a cysteine-less construct with re-introduced cysteines at desired locations (**Figure 6-1B**). The homolog of Prp24 from *K. lactis* retains high-

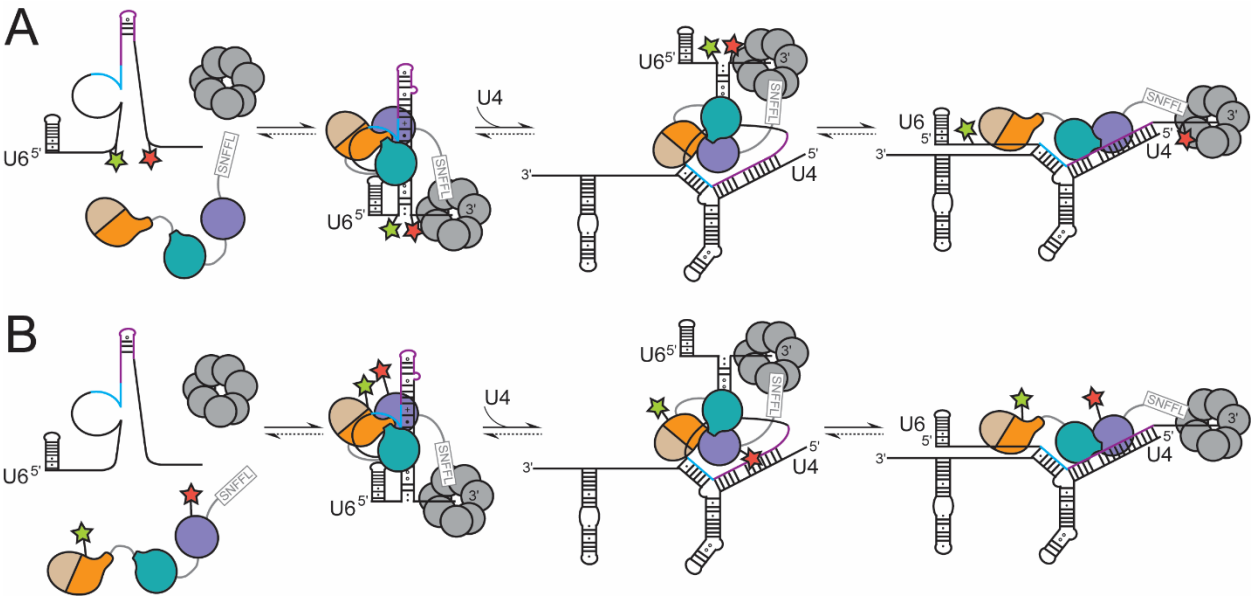
affinity binding behavior when cysteines are removed, suggesting that it could be used for such studies (**Appendix 6**). Defining the dynamics (via smFRET) and the distance between site-specific labels (either via smFRET or electron paramagnetic spectroscopy) for both U6 and Prp24 at different stages in the binding and annealing pathway will be extremely informative and could be used as restraints to model the conformational changes that must occur during binding and annealing.

The interlocked topology of U6-Prp24 also begs the question of prevalence. Many proteins contain multiple RRM domains, yet few structures of multi-RRM RNP complexes exist. Is the utilization of tandem RRMs to form unique topologies common? Additionally, the question remains of whether the interlocked topology is conserved in homologs of U6-Prp24. U6-Prp24 from *K. lactis* shares many biophysical characteristics with *S. cerevisiae*, strongly suggesting that it shares the interlocked topology. The *K. lactis* core snRNP is crystallizable (**Appendix 6**). Determining its structure will be informative as to which protein-RNA (or protein-protein) contacts are essential for the interlocked topology. In sharp contrast, human Prp24 has only two RRMs, suggesting that it must adopt a different binding arrangement or act as a dimer. Determining the structure of U6-Prp24 from multiple homologs could give us insights into how conserved the annealing mechanism is, or reveal how evolution has engineered a new solution to keep U6 protected and chaperoned into the spliceosome.

Figure 6-1. Site-specific labels on U6 and Prp24 could give insights into molecular rearrangements during U6-Prp24 binding.

A) Labeling U6 RNA at the base of the telestem (i.e. nucleotides U32 and U100) will report on telestem dynamics throughout Prp24 binding and annealing.

B) Labeling Prp24 will report on structural dynamics between the four RRMs of Prp24 as it binds U6 and anneals U4 to U6.



6.2.6 Understanding the mechanism of Prp24-mediated annealing

When I began my dissertation work, little was known about the mechanism of Prp24-mediated annealing other than the observations that Prp24 behaved as an ATP-independent chaperone and that inclusion of the Lsm2-8 ring enhanced the annealing rate (GHETTI *et al.* 1995; RAGHUNATHAN AND GUTHRIE 1998; ACHSEL *et al.* 1999; MAYES *et al.* 1999; RADER AND GUTHRIE 2002). By reconstituting the annealing process *in vitro*, I demonstrated that positive charge within the groove was essential for annealing, that the secondary structure of U6 influenced annealing rate, and that the Lsm2-8 ring lacked annealing activity by itself yet stimulated Prp24-mediated annealing. However, we still lack a mechanistic understanding of the annealing process. We propose that both inclusion of the Lsm2-8 ring and stabilization of the telestem promote efficient formation of the electropositive groove, which acts as the "active site" for annealing. However, the precise role of the electropositive groove is not known. It could be important for stabilizing a transiently unwound ISL (MONTEMAYOR *et al.* 2014) or could be required for electrostatic recruitment of U4. In **Appendix 8**, I show that the majority of electropositive groove mutations tested behave *in vivo* as predicted based on their effect on annealing *in vitro* (DIDYCHUK *et al.* 2016). A patch of three residues on one side of the groove appears to be more important than other positively charged residues, suggesting that *in vivo* experiments could be useful to narrow down the role of the groove. These residues interact with the top of the ISL from a neighboring U6 molecule in the structure of the U6/Prp24/Lsm2-8 snRNP, strongly suggesting that they could be involved in RNA binding (MONTEMAYOR *et al.* in prep).

6.2.7 Understanding the contribution of U4/U6 di-snRNP proteins to Prp24-mediated annealing

The annealing assay developed in Didychuk *et al.* 2016 could easily be expanded to test if the U4/U6 di-snRNP proteins Prp3 and Prp4 influence Prp24-mediated annealing. It was previously proposed that Prp3 could enhance annealing by partially unwinding the ISL by binding nucleotides 80-90 (LIU *et al.* 2015). The structure of the U4/U6.U5 tri-snRNP structure revealed

that Prp3 binds the majority of U4/U6 Stem II, which could enhance annealing. Protein-protein contacts (i.e. between Prp3 and Lsm2-8) could also help stabilize U4/U6 and enhance annealing. Prp3 and Prp4 can be expressed in yeast (HARDIN *et al.* 2015) and included in the annealing assay.

Although Prp24 binds tightly to U4/U6 di-snRNA *in vitro* (DIDYCHUK *et al.* 2016), it is likely released before association with U5 to form the tri-snRNP *in vivo*. How is the interlocked topology of U6-Prp24 resolved upon annealing and how is Prp24 released from the product di-snRNA? U4/U6 di-snRNP proteins may play a role in removing Prp24 by occluding its binding site or promoting a conformational change that weakens Prp24's affinity for U6. This could also be tested *in vitro* using the annealing assay.

6.3 References

Achsel, T., H. Brahms, B. Kastner, A. Bachi, M. Wilm *et al.*, 1999 A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. EMBO J 18**:** 5789-5802.

Chin, J. W., T. A. Cropp, J. C. Anderson, M. Mukherji, Z. Zhang *et al.*, 2003 An expanded eukaryotic genetic code. Science 301**:** 964-967.

Didychuk, A. L., E. J. Montemayor, D. A. Brow and S. E. Butcher, 2016 Structural requirements for protein-catalyzed annealing of U4 and U6 RNAs during di-snRNP assembly. Nucleic Acids Res 44**:** 1398-1410.

Didychuk, A. L., E. J. Montemayor, T. J. Carrocci, A. T. DeLaitsch, S. E. Lucarelli *et al.*, 2017 Usb1 controls U6 snRNP assembly through evolutionarily divergent cyclic phosphodiesterase activities. Nat Commun 8**:** 497.

Frank, J., 2017 Time-resolved cryo-electron microscopy: Recent progress. J Struct Biol 200**:** 303-306.

Ghetti, A., M. Company and J. Abelson, 1995 Specificity of Prp24 binding to RNA: a role for Prp24 in the dynamic interaction of U4 and U6 snRNAs. RNA 1**:** 132-145.

Hardin, J. W., C. Warnasooriya, Y. Kondo, K. Nagai and D. Rueda, 2015 Assembly and dynamics of the U4/U6 di-snRNP by single-molecule FRET. Nucleic Acids Res 43**:** 10963-10974.

Hilcenko, C., P. J. Simpson, A. J. Finch, F. R. Bowler, M. J. Churcher *et al.*, 2013 Aberrant 3' oligoadenylation of spliceosomal U6 small nuclear RNA in poikiloderma with neutropenia. Blood 121**:** 1028-1038.

Liu, S., S. Mozaffari-Jovin, J. Wollenhaupt, K. F. Santos, M. Theuser *et al.*, 2015 A composite double-/single-stranded RNA-binding region in protein Prp3 supports tri-snRNP stability and splicing. Elife 4**:** e07320.

Mayes, A. E., L. Verdone, P. Legrain and J. D. Beggs, 1999 Characterization of Sm-like proteins in yeast and their association with U6 snRNA. EMBO J 18**:** 4321-4331.

Montemayor, E. J., E. C. Curran, H. H. Liao, K. L. Andrews, C. N. Treba *et al.*, 2014 Core structure of the U6 small nuclear ribonucleoprotein at 1.7-A resolution. Nat Struct Mol Biol 21**:** 544-551.

Mroczek, S., J. Krwawicz, J. Kutner, M. Lazniewski, I. Kucinski *et al.*, 2012 C16orf57, a gene mutated in poikiloderma with neutropenia, encodes a putative phosphodiesterase responsible for the U6 snRNA 3' end modification. Genes Dev 26**:** 1911-1925.

Nguyen, T. H. D., W. P. Galej, X. C. Bai, C. Oubridge, A. J. Newman *et al.*, 2016 Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 A resolution. Nature 530**:** 298-302.

Plaschka, C., P. C. Lin and K. Nagai, 2017 Structure of a pre-catalytic spliceosome. Nature 546**:** 617-621.

Rader, S. D., and C. Guthrie, 2002 A conserved Lsm-interaction motif in Prp24 required for efficient U4/U6 di-snRNP formation. RNA 8**:** 1378-1392.

Raghunathan, P. L., and C. Guthrie, 1998 A spliceosomal recycling factor that reanneals U4 and U6 small nuclear ribonucleoprotein particles. Science 279**:** 857-860.

Rodgers, M. L., A. L. Didychuk, S. E. Butcher, D. A. Brow and A. A. Hoskins, 2016 A multi-step model for facilitated unwinding of the yeast U4/U6 RNA duplex. Nucleic Acids Res 44**:** 10912-10928.

Shchepachev, V., H. Wischnewski, C. Soneson, A. W. Arnold and C. M. Azzalin, 2015 Human Mpn1 promotes post-transcriptional processing and stability of U6atac. FEBS Lett 589**:** 2417-2423.

Wang, L., C. Clericuzio and L. Larizza, 2017 Poikiloderma with Neutropenia in *GeneReviews(R)*, edited by M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. H. Bean *et al.*, Seattle (WA).

Zinder, J. C., and C. D. Lima, 2017 Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. Genes Dev 31**:** 88-100.

Appendix 1: A protein-protein interaction between Usb1 and Cus2

This work could not have been performed without help and insightful discussions with Tucker Carrocci.

A1.1 Overview

Results in *S. cerevisiae, S. pombe,* and human cell lines suggest that Usb1 is a highly specific exonuclease that is limited to processing U6 snRNA in yeast and U6atac snRNA in higher metazoans (SHCHEPACHEV *et al.* 2012; HILCENKO *et al.* 2013; SHCHEPACHEV *et al.* 2015). This is surprising, considering that the polyuridine tail of U6 is derived from the RNA polymerase III transcription termination site and is present in many other nuclear RNAs. How is Usb1 targeted specifically to U6 snRNA? This specificity may come from interactions between Usb1 and other splicing factors or U6-associated proteins. Our search for protein binding partners of Usb1 led us to a previously annotated (from a high-throughput Y2H screen) but unappreciated interaction with Cus2. This interaction minimally requires the Usb1 catalytic domain and the UHM domain of Cus2. While the biological role of this interaction remains untested, it may reveal unprecedented connections between U6 recycling and other splicing factors.

A1.2 Materials and Methods

A1.2.1 Yeast-two-hybrid (Y2H) assay

The open reading frames (ORFs) for Usb1, Cus2, Lhp1, and Prp19 and were cloned into pGADT7 and pGBKT7 (Clontech) to generate an N-terminal GAL4 activation domain (AD) or binding domain (BD) fusion. ORFs of Prp24, Hsh49, Hsh155, Ysf3, Prp9, Prp11, Prp21, Mud2, Prp5, Clf1, Prp46, Prp2, Prp43, Spo12, and Mss18 were cloned into pGBKT7. Deletions and point mutations within the Usb1 and Cus2 plasmids were generated by inverse PCR as previously described (DIDYCHUK *et al.* 2017) using oligonucleotides listed in **Table A1-1**. Y2H GOLD, a strain which contains the GAL4 UAS upstream of the *HIS3* and *ADE2* loci, was transformed with pGADT7 and pGBKT7 fusions using the lithium acetate method (GIETZ AND WOODS 2002). Interactions were assessed by spotting 10-fold serial dilutions (OD_{600} = 0.5, 0.05, 0.005) onto solid medium lacking tryptophan, leucine (to maintain plasmid selection) and histidine (to assess activation of the *HIS3* gene and infer protein-protein interaction).

A1.2.2 Yeast *USB1* complementation assays

yTJC0700 was transformed with pRS414-Usb1 wild-type, W75A, or W83A using the lithium acetate method (GIETZ AND WOODS 2002). Growth phenotypes were assessed by spotting 10-fold serial dilutions (OD_{600} = 0.5, 0.05, 0.005) onto solid medium lacking tryptophan and containing 1 mg/mL 5-fluoroorotic acid. Plates were incubated at 30°C for 3 days.

Table A1-1. List of synthetic DNA oligonucleotides.

Cus2-D208A-f	5'-GCTCTTTTGGAAGGCTGTGAAGAGATAGGTCAA-3'
Cus2-208-r	5'-TTCCTGAATATCATTTATATCATCATTTGT-3'
Cus2-R251_to_F253_ALA	5'-GCTGCAGCTGATGGGCAAAAGTTGCTTGCTTTCATTTCT-3'
Cus2-R251_to_F253_RRM	5'-ATTACTGTTGATGGGCAAAAGTTGCTTGCTTTCATTTCT-3'
Cus2-251-r	5'-TCCGGTCATGATTTTGCAACATTGCAATGC-3'
Cus2-D282N-f	5'-TTCCACTTCGCTGTCCTCATTCTTG-3'
Cus2-D282-r	5'-AATGACCTTATATAGGGATCCATCG-3'
Cus2-L284F-f	5'-GTCATCTTCCACTTCGCTGTCC-3'
Cus2-L284-r	5'-TTTATATAGGGATCCATCGAGCTCG-3'
Cus2-trun-stop-for	5'-TAGGGATCCATCGAGCTCGAGCTG-3'
Cus2-trun-start-rev	5'-ATGAGCGTAATCTGGTACGTCG-3'
Cus2-275-r	5'-ATTCTTGTCACTTGTGCTTGAAGTGTTCTCATC-3'
Cus2-130-f	5'-CAAAACAAGGAAGGTGACAATATGCACGGGAAAG-3'
Cus2-180-f	5'-AAAGCTGATCGAACGGTGATATTTGCCAAC-3'

A1.3 Results and Discussion

A1.3.1 Usb1 physically interacts with Cus2, but not other splicing factors

We utilized a yeast-two-hybrid (Y2H) assay to screen for Usb1 interacting partners. We fused Usb1 to either the GAL4 DNA binding domain (BD) or GAL4 activation domain (AD) and fused candidate binding partners to the complementary domain (**Figure A1-1A**). We initially tested for interactions between Usb1 and Prp19, Cus2, and Lhp1, as Prp19 and Cus2 were annotated as Usb1-interacting proteins based on high-throughput Y2H screens in the Saccharomyces Genome Database (Yu *et al.* 2008). Lhp1 was also tested as Sro9, a La-motif containing protein, was annotated as interacting with Usb1. We found that Usb1 interacts strongly with Cus2 and weakly with itself (**Figure A1-1B**). No interaction was observed between Usb1 and Prp19 or Lhp1 (**Figure A1-1B**). Importantly, fusion of all proteins to either domain (AD or BD) yielded the same result, and fusing full-length Usb1 to either domain did not exhibit any autoactivation (**Figure A1-1B**). To investigate the self-interaction observed, Usb1 was divided into N-terminal (residues 1-76) and catalytic domain (residues 71-290). As AD-Usb1 1-76 (but not full-length AD-Usb1) exhibited autoactivation, we used BD-Usb1 for studies with Cus2 that required truncation of Usb1.

We next expanded our search for binding partners to other splicing factors and U6-associated factors. We found that Usb1 exhibits a Y2H interaction with the U6 snRNP protein Prp24 (**Figure A1-2**). We also find that Usb1 does not interact via Y2H with a number of other splicing factors tested, including Prp9, Prp11, Prp21, Hsh49, Hsh155, Ysf3, Mud2, Prp5, Prp2, and Prp43 (**Figure A1-2**). Growth on media lacking histidine was observed for the AD-Usb1/BD-Prp9 yeast. Yeast containing the BD-Prp9 vector exhibit autoactivation (data not shown). Interestingly, a Y2H interaction was observed between Usb1 and Prp46, a member of the NTC, but not Clf1 or Prp19 (also components of the NTC) (**Figure A1-2**). This interaction may be a false positive, as the BD-Prp46 plasmid has not been validated to ensure that autoactivation is not an issue.

Figure A1-1. Usb1 interacts with Cus2 in a yeast-two-hybrid (Y2H) assay.

A) Schematic of the Y2H assay. Protein-protein interactions between factors fused to the GAL4 binding domain (BD) or activation domain (AD) bring the BD and AD together to allow for expression of the downstream *HIS3* gene, which allows for growth on media lacking histidine. Growth on -LEU -TRP -HIS (-LWH) plates requires the Y2H interaction or autoactivation.

B) Usb1 does not interact via Y2H with Lhp1 or Prp19, but does interact with Cus2. Fusion of candidate binding partners to the AD or BD does not change the Y2H interaction pattern.

C) The Usb1 self-interaction is mediated through the N-terminal unstructured domain. AD-Usb1 1-76 exhibits autoactivation, precluding its use in further experiments.

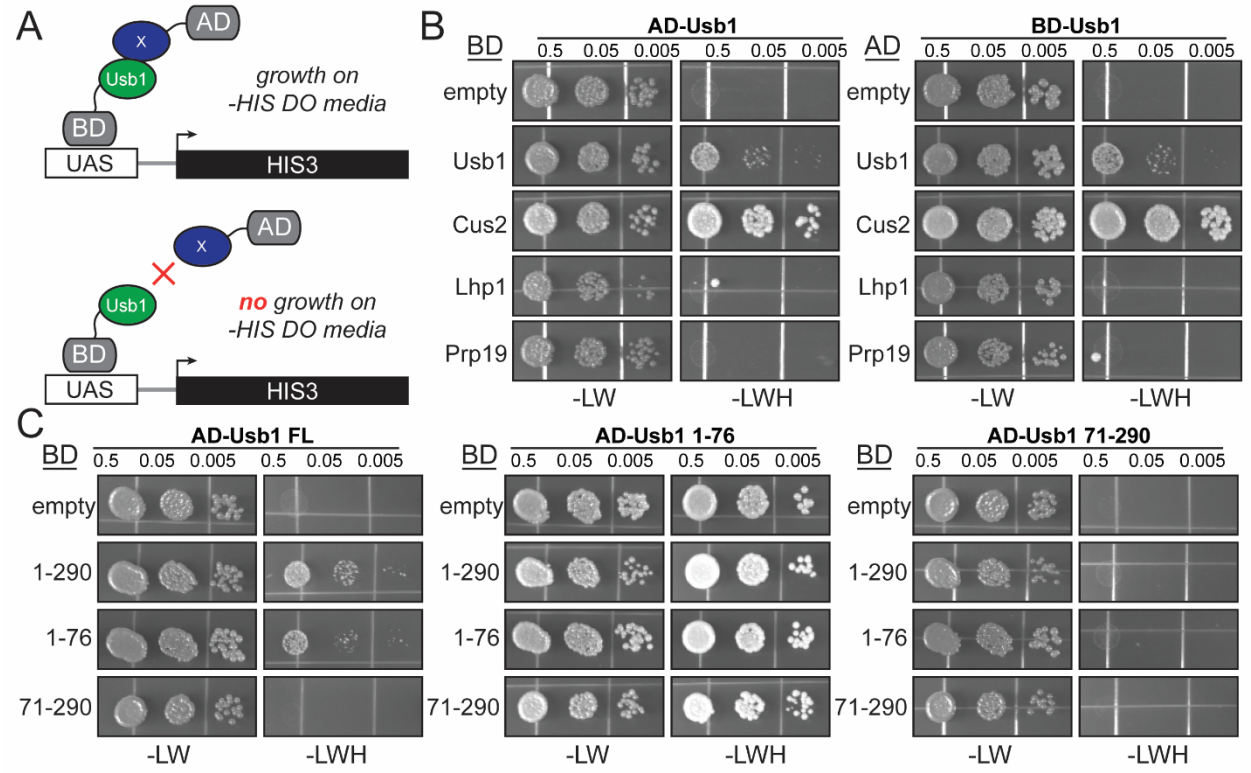
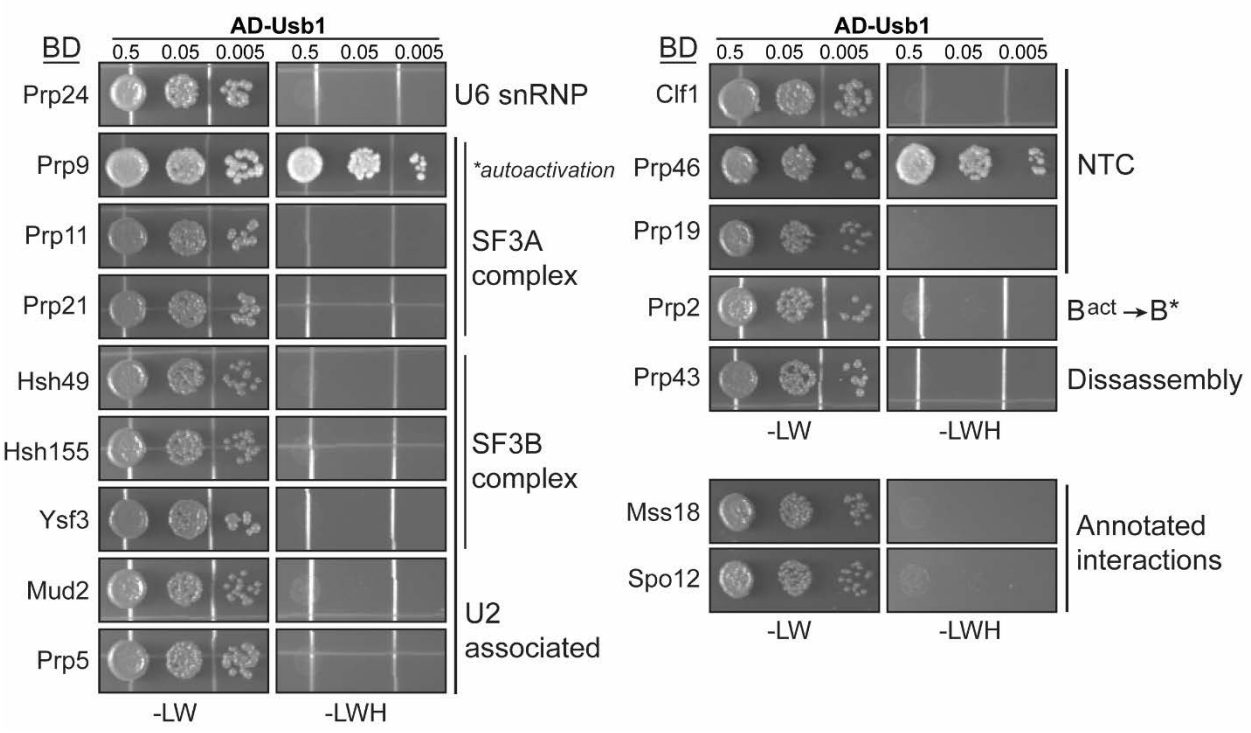


Figure A1-2. Usb1 does not interact with many splicing factors.

Full-length Usb1 was fused to the GAL4 activation domain to produce AD-Usb1. Y2H interactions between AD-Usb1 and BD fusions of candidate splicing factors were assayed by comparing growth on media containing and lacking histidine. The spliceosomal complex or step in which the factors participate is indicated to the right. Two additional factors, Mss18 and Spo12, were tested as they were annotated as physically interacting with Usb1 in the SGD.



A1.3.2 Usb1 may interact with Cus2 via a ULM-UHM interaction

In order to identify how Usb1 and Cus2 interact, we divided both Usb1 and Cus2 into two segments based on predicted domain boundaries (**Figure A1-3A**). As described above, Usb1 was divided into the N-terminal domain and the catalytic domain. Cus2 was divided into a region containing a N-terminal domain and the RNA recognition motif (RRM) (residues 1-130), a region containing the linker, the U2AF homology motif (UHM), and the acidic tail (residues 131-285), and a region lacking the acidic tail (1-275) (**Figure A1-3A**). We find that Usb1 interacts with Cus2 via its catalytic domain and Cus2 interacts with Usb1 via its UHM (U2AF homology motif) domain (**Figure A1-3B**). Mutations within the acidic tail (D282N and L284F) that suppress mutations in U2 (Yan... Ares 1998) had no effect on the Y2H interaction (data not shown).

The UHM domain of Cus2 is a protein-protein interaction domain that interacts with U2 snRNP component Hsh155 (Clara Kielkopf, personal communication). UHM domains typically interact with a ULM (UHM ligand motif) consisting of a short peptide sequence that contains a conserved tryptophan flanked by basic residues (KIELKOPF *et al.* 2004; LOERCH AND KIELKOPF 2016) (**Figure A1-3C**).

Usb1 contains two tryptophan residues (W75 and W83), both of which are present in the minimal Cus2-interaction domain (Usb1 residues 71-290). To test whether Usb1 and Cus2 interact through a canonical UHM-ULM interaction, we mutated Usb1 residues W75 and W83 to alanine in our Y2H assay and found that mutation of W75 to alanine has no effect on the Y2H interaction, but mutation of W83 to alanine severely reduces the interaction in the context of full-length protein (**Figure A1-3D**). The W83A mutation appeared more deleterious in the context of the crystal construct (Usb1 residues 71-290) than in full-length Usb1. This implies that the interaction between Usb1 and Cus2 may depend on regions outside the putative ULM motif in Usb1.

Next, we mutated conserved residues in the UHM-RXF motif. We find that mutation of residues in the UHM-RXF motif (Cus2 residues R251, Y252, and F253) eliminates the Y2H

interaction with Usb1 (**Figure A1-3E**). Mutation of the RXF motif to the corresponding residues in a canonical RRM motif (R251I, Y252T, F253V) also disrupts the interaction (data not shown). Mutation of Cus2-D208 to alanine, which is structurally conserved among UHM domains (KIELKOPF *et al.* 2004) (**Figure A1-3C**), reduced but did not fully ablate the Y2H interaction (**Figure A1-3E**). Mutating UHM elements in the context of full-length Cus2 or the UHM alone (**Figure A1-3E**) did not change the resulting Y2H interactions. Importantly, changes in the observed Y2H interaction of Usb1 with Cus2 mimicked that of Hsh155 and Cus2, a validated interactor of the Cus2 UHM (**Figure A1-3F**).

To test whether the Usb1-Cus2 interaction is required *in vivo*, we introduced wild-type, W75A, or W83A alleles of Usb1 on a shuffle plasmid and assayed for growth (**Figure A1-3G**). Neither the W75A nor W83A mutations resulted in a loss of viability. As some of the Usb1(W83A)-Cus2 interaction is still observed in the context of full-length Usb1 (**Figure A1-3D**), it is possible that this minimal interaction is sufficient for growth. Alternatively, the Usb1-Cus2 may not be necessary for growth under permissive growth conditions, or may result in a splicing fidelity defect that is undetectable in this assay.

Our results suggest that Usb1 may interact with Cus2 through a canonical UHM-ULM interaction. This is surprising, considering how divergent the ULM motif in Usb1 is in comparison to confirmed ULM motifs (LOERCH AND KIELKOPF 2016). In addition to this, W83 is buried in our crystal structure of yeast Usb1 (**Figure A1-3H**). Whether the observed Usb1-Cus2 interaction is functional remains to be investigated.

Cus2 is a U2 snRNP-associated protein that binds U2 snRNA (YAN *et al.* 1998). While Cus2 may persist into the penta-snRNP (STEVENS *et al.* 2002), it is not a component of active spliceosomes (FABRIZIO *et al.* 2009). Though Cus2 may not cross paths with U6 RNA during the chemical steps of splicing, the players involved in recycling U6 from post-catalytic spliceosomes are not fully understood. Data from HeLa cell extracts suggests that U6 may not be processed by Usb1 until *after* it successfully completes a round of splicing (TAZI *et al.* 1993). Cus2 could act as

a “quality control” regulator for U6 and bind functional U6 after completion of one round of splicing, marking U6 for processing by Usb1. This hypothesis does not necessarily counteract our model of snRNP biogenesis (DIDYCHUK *et al.* 2017); a complex relationship between Cus2, Usb1, Lhp1, Prp24, and Lsm2-8 may exist in order to ensure that fully functional U6 is marked and guarded from predatory exonucleases in the cell. Further investigation of the Usb1-Cus2 interaction may reveal new steps in U6 biogenesis or recycling.

Figure A1-3. The Usb1-Cus2 interaction may occur via a ULM-UHM interaction.

A) Secondary structure of Usb1 (top) and Cus2 (bottom). Usb1 consists of an N-terminal unstructured domain (residues 1-71) and the catalytic domain (71-290). Cus2 contains an N-terminal region, an RRM domain (residues 41-130), a UHM domain (residues 182-275) and an acidic tail (residues 275-285).

B) Top: Cus2 interacts with Usb1 residues 71-290, but not residues 1-76. Bottom: Usb1 interacts with Cus2 131-285 (deletion of the RRM domain) and 1-275 (deletion of the acidic tail), but not residues 1-130 (deletion of the UHM domain).

C) Crystal structure of a canonical UHM-ULM interaction from PDB 4OZ1 (LOERCH *et al.* 2014). The alternative splicing factor CAPERα possesses a UHM domain (blue) that interacts with the ULM motif (yellow) of core splicing factor SF3B1, the human homolog of Hsh155. Residues R488, W489, and F490 of CAPERα form the RXF motif. Residue D443 is also conserved in UHM domains. In the ULM, W338 of SF3B1 forms the core of the interaction and is invariant in annotated ULMs.

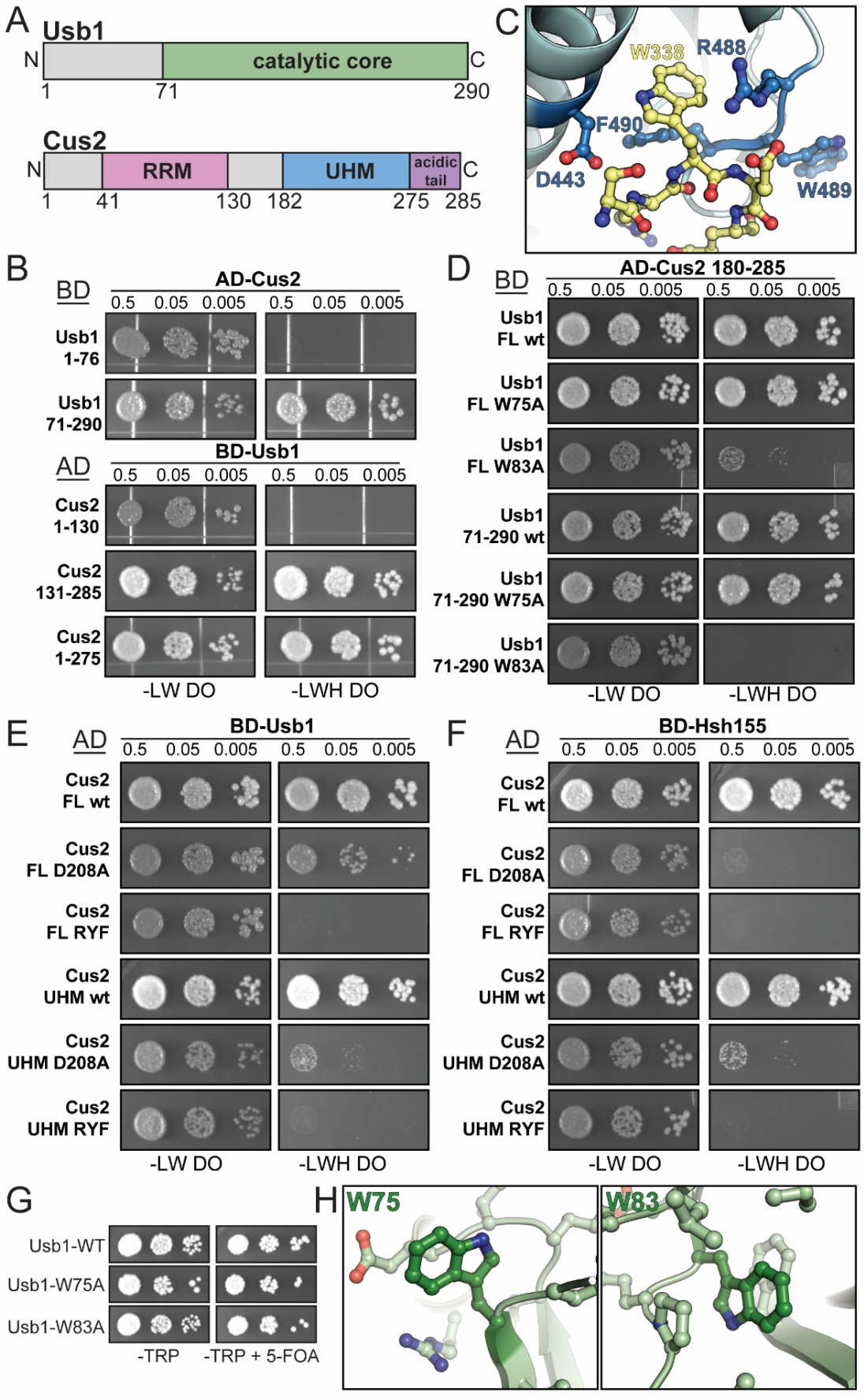
D) Mutating Usb1 residue W83, but not W75, partially disrupts (in full-length Usb1) or fully disrupts (in the minimal catalytic core construct, residues 71-290) the interaction with the Cus2 UHM domain.

E) Mutating conserved UHM motifs in Cus2 (Cus2 D208A and mutation of the R-X-F motif, R251A/Y252A/F253A) in the context of full-length or Cus2 UHM (residues 180-285) partially disrupts (for D208A) or fully disrupts (for the RYF mutation) the interaction with Usb1.

F) Mutating conserved UHM motifs in Cus2 (as in E) disrupts the Y2H interaction with its validated binding partner Hsh155.

G) Mutation of either tryptophan (W75 or W83) to alanine does not have an effect on viability.

H) Both tryptophan residues are within the crystallizable catalytic core of Usb1 (PDB 5UQJ). W75 is surface exposed, while W83 is buried and stacked between F271 and P85.


A
Usb1
N
catalytic core
C
1
71
290
Cus2
N
RRM
UHM
acidic tail
C
1
41
130
182
275
285
C
W338
R488
F490
D443
W489
B
BD
AD-Cus2
0.5 0.05 0.005 0.5 0.05 0.005
Usb1 1-76
Usb1 71-290
AD
BD-Usb1
Cus2 1-130
Cus2 131-285
Cus2 1-275
-LW DO
-LWH DO
D
BD
AD-Cus2 180-285
Usb1 FL wt
Usb1 FL W75A
Usb1 FL W83A
Usb1 71-290 wt
Usb1 71-290 W75A
Usb1 71-290 W83A
-LW DO
-LWH DO
E
AD
BD-Usb1
Cus2 FL wt
Cus2 FL D208A
Cus2 FL RYF
Cus2 UHM wt
Cus2 UHM D208A
Cus2 UHM RYF
-LW DO
-LWH DO
F
AD
BD-Hsh155
Cus2 FL wt
Cus2 FL D208A
Cus2 FL RYF
Cus2 UHM wt
Cus2 UHM D208A
Cus2 UHM RYF
-LW DO
-LWH DO
G
Usb1-WT
Usb1-W75A
Usb1-W83A
-TRP
-TRP + 5-FOA
H
W75
W83

A1.4 References

Didychuk, A. L., E. J. Montemayor, T. J. Carrocci, A. T. DeLaitsch, S. E. Lucarelli *et al.*, 2017 Usb1 controls U6 snRNP assembly through evolutionarily divergent cyclic phosphodiesterase activities. Nat Commun 8**:** 497.

Fabrizio, P., J. Dannenberg, P. Dube, B. Kastner, H. Stark *et al.*, 2009 The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. Mol Cell 36**:** 593-608.

Gietz, R. D., and R. A. Woods, 2002 Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. Methods Enzymol 350**:** 87-96.

Hilcenko, C., P. J. Simpson, A. J. Finch, F. R. Bowler, M. J. Churcher *et al.*, 2013 Aberrant 3' oligoadenylation of spliceosomal U6 small nuclear RNA in poikiloderma with neutropenia. Blood 121**:** 1028-1038.

Kielkopf, C. L., S. Lucke and M. R. Green, 2004 U2AF homology motifs: protein recognition in the RRM world. Genes Dev 18**:** 1513-1526.

Loerch, S., and C. L. Kielkopf, 2016 Unmasking the U2AF homology motif family: a bona fide protein-protein interaction motif in disguise. RNA 22**:** 1795-1807.

Loerch, S., A. Maucuer, V. Manceau, M. R. Green and C. L. Kielkopf, 2014 Cancer-relevant splicing factor CAPERalpha engages the essential splicing factor SF3b155 in a specific ternary complex. J Biol Chem 289**:** 17325-17337.

Shchepachev, V., H. Wischnewski, E. Missiaglia, C. Soneson and C. M. Azzalin, 2012 Mpn1, mutated in poikiloderma with neutropenia protein 1, is a conserved 3'-to-5' RNA exonuclease processing U6 small nuclear RNA. Cell Rep 2**:** 855-865.

Shchepachev, V., H. Wischnewski, C. Soneson, A. W. Arnold and C. M. Azzalin, 2015 Human Mpn1 promotes post-transcriptional processing and stability of U6atac. FEBS Lett 589**:** 2417-2423.

Stevens, S. W., D. E. Ryan, H. Y. Ge, R. E. Moore, M. K. Young *et al.*, 2002 Composition and functional characterization of the yeast spliceosomal penta-snRNP. Mol Cell 9**:** 31-44.

Tazi, J., T. Forne, P. Jeanteur, G. Cathala and C. Brunel, 1993 Mammalian U6 small nuclear RNA undergoes 3' end modifications within the spliceosome. Mol Cell Biol 13**:** 1641-1650.

Yan, D., R. Perriman, H. Igel, K. J. Howe, M. Neville *et al.*, 1998 CUS2, a yeast homolog of human Tat-SF1, rescues function of misfolded U2 through an unusual RNA recognition motif. Mol Cell Biol 18**:** 5000-5009.

Yu, H., P. Braun, M. A. Yildirim, I. Lemmens, K. Venkatesan *et al.*, 2008 High-quality binary protein interaction map of the yeast interactome network. Science 322**:** 104-110.

Appendix 2: *In vivo* and *in vitro* effects of mutations within *USB1*

This work was performed with the help of undergraduate researchers Andrew DeLaitsch and Stefani Lucarelli, as well as the invaluable advice of Tucker Carrocci.

A2.1 Overview

Usb1 belongs to the 2H phosphodiesterase superfamily (MAZUMDER *et al.* 2002; MROCZEK *et al.* 2012; HILCENKO *et al.* 2013). This family employs two pseudosymmetric H-X-S motifs in an active site cleft for catalysis, with the active site histidines acting as a catalytic acid and base and the active site serines helping to coordinate the substrate. However, despite significant research, including crystal structures and biochemical characterizations, how substrate specificity and the stereospecificity of the products is mechanistically determined is poorly understood. Yeast Usb1 (yUsb1) represents an interesting case for study, as it possesses both exonuclease and 2'-cyclic phosphodiesterase activity to produce a 3' non-cyclic phosphate (DIDYCHUK *et al.* 2017). Its human homolog (hUsb1) lacks phosphodiesterase activity (HILCENKO *et al.* 2013), despite having an active site that is nearly identical to yUsb1 (DIDYCHUK *et al.* 2017). Additionally, both enzymes possess a preference for modifying uridine and adenine-containing substrates over substrates containing cytidine and guanine nucleotides (HILCENKO *et al.* 2013, NOMURA *et al.* in prep). However, the determinants of human and yeast shared substrate specificity and divergent catalytic activity are unknown. Understanding the mechanism of yeast and human Usb1 has broad implications for RNA biology and structure-based activity prediction.

A2.2 Materials and Methods

A2.2.1 Protein expression and purification

Recombinant wild-type yUsb1 1-290 and mutants were produced as previously described (DIDYCHUK *et al.* 2017). Briefly, mutations were introduced into the pET28b-HT-yUsb1 1-290 via inverse PCR and ligation. Protein was expressed in *E. coli* BL21 STAR (DE3) pLysS (Invitrogen) and grown at 37°C until an OD_{600} of ~0.6-0.8. Cells were induced by addition of 1 mM IPTG and grown for 20 hours at 16°C. Cells were collected by centrifugation, resuspended in IMAC buffer (500 mM NaCl, 50 mM HEPES acid, 50 mM sodium HEPES base, 15 mM imidazole base, 10% glycerol, 1 mM TCEP-HCl) containing DNase I and lysozyme, sonicated, and centrifuged to remove insoluble material. Protein was purified by Ni-NTA agarose chromatography by step elution with IMAC buffer supplemented with 500 mM imidazole, dialyzed overnight at 4°C into IEX buffer (100 mM NaCl, 20 mM bis-tris, 10 mM HCl, 10% glycerol, 1 mM TCEP, pH ~ 6.2) with 1 mg TEV protease. Precipitated protein was removed by centrifugation and soluble Usb1 was further purified via cation-exchange chromatography (HiTrap S, GE Healthcare) in IEX buffer with gradient elution against IEX buffer containing 2 M NaCl.

A2.2.2 *In vitro* exoribonuclease assay

In vitro exoribonuclease assays were performed as previously described (DIDYCHUK *et al.* 2017). Time course assays used 500 nM yUsb1 and 100 nM FAM-U6 95-112(UUU)-3'OH substrate in 50 mM NaCl, 50 mM KCl, 20 mM bis-tris, 10 mM HCl, 1 mM TCEP-HCl, 10% surcrose, 5% glycerol, 0.5 mM EDTA, 0.005% Triton X-100, 0.1 mg/mL BSA. The CIP/T4 PNK sensitivity experiments were performed by incubating 500 nM yUsb1 and 100 nM FAM-U6 95-112(UUU)-3'OH in 100 mM NaCl, 10% sucrose, 0.1 mM EDTA, 0.1 mM TCEP-HCl, 0.1 mg/mL, 0.01% Triton X-100, and 20 mM of a mixed buffer consisting of sodium acetate, bis-tris, sodium HEPES base, Tris base, and CHES that had been adjusted to the indicated pH (DIDYCHUK *et al.* 2017). After 60 minutes, samples were adjusted to 1X Cutsmart (for CIP) or 1X T4 PNK buffer (for T4 PNK) using 10x stocks and 10 units of CIP or T4 PNK were added (New England Biolabs)

and incubated at 37°C for 15 minutes. All reactions were quenched by the addition of an equal volume of 100% formamide, resolved on a 20% 19:1 acrylamide:bis-acrylamide PAGE gel containing 8 M urea, 89 mM Tris borate, and 2 mM EDTA. Gels were imaged directly through low fluorescence glass plates (CBS Scientific) using a Typhoon FLA 9000 (GE Healthcare Life Sciences). For time course experiments, the percentage of the substrate processed was calculated using the ratio of product(s) to total signal at 0, 5, 15, 30, and 60 minute time points. Data were fit to a one-phase exponential association equation $Y=Y_0+(Y_{max} - Y_0)*(1-e^{-kt})$, where Y is the % of the substrate processed, t is the time, and *k* is the rate constant (GraphPad Prism 4).

A2.2.3 Yeast *USB1* complementation assay

Mutations were introduced into pRS414-Usb1 via inverse PCR and ligation as previously described (DIDYCHUK *et al.* 2017). yTJC0700 was transformed with pRS414-Usb1 and variants using the lithium acetate method (GIETZ AND WOODS 2002). Complementation was assessed by spotting 10-fold serial dilutions (OD_{600} = 0.5, 0.05, 0.005) onto solid medium lacking tryptophan and containing 1 mg/mL 5-FOA. Plates were incubated at 30°C for 3 days.

A2.3 Results and discussion

A2.3.1 Mutation of the active site residues reduce activity *in vitro*

The active site of yUsb1 consists of two pseudosymmetric H-X-S motifs that lie in a cleft in the center of the protein (**Figure A2-1A**) (DIDYCHUK *et al.* 2017). The two catalytic histidines (residues 133 and 231) are essential for yeast viability (HILCENKO *et al.* 2013; DIDYCHUK *et al.* 2017) and activity *in vitro* (**Figure A2-1B**). The active site serines (residues 135 and 233) are not essential but do affect yeast growth, and S135A is more deleterious than S233A (HILCENKO *et al.* 2013). Interestingly, when these residues are mutated to cysteine and assayed for activity *in vitro*, S135C is more deleterious than S233C (**Figure A2-1B**), showing that yeast viability strongly correlates with *in vitro* activity. The precise role of the active site histidines and serines (i.e. unambiguous determination of the catalytic acid and catalytic base, as well as how the active site serines coordinate the substrate) remains to be determined.

The active site of yUsb1 is flanked by several aromatic residues, including F78 and Y187 (**Figure A2-1A**). We demonstrated that mutations at position 78 affect activity *in vitro* and yeast viability (DIDYCHUK *et al.* 2017). However mutation of tyrosine 187 to phenylalanine, which was selected because hUsb1 has a structurally conserved phenylalanine at this location, has little effect *in vitro* (**Figure A2-1C**) and mutation to either alanine or phenylalanine has no effect on yeast viability (**Table A2-1**). Since introducing mutant alleles into yeast and assaying for viability is higher throughput than producing recombinant yUsb1 and assaying for activity *in vitro*, I selected residues based on observed contacts in the crystal structure, proximity to the active site and/or conservation with hUsb1 and tested their ability to complement loss of *USB1*. The majority of these mutations had no effect on yeast viability (**Figure A2-1D** and **Table A2-1**). However, the sensitivity of the yeast viability assay is likely very low, since it was previously observed that only mutations with >~50-fold reduction in yUsb1 activity manifest as viability defects, as in the case of yUsb1-F78A (DIDYCHUK *et al.* 2017). It may be interesting to quantitatively measure the effect

of mutations on activity *in vitro* for residues near the active site such as R138, N229, Y80, and Y178.

Figure A2-1. Mutations in yUsb1 affect activity *in vitro.*

A) Top: mutation of either of the active site histidines (H133 and H231) to asparagine completely ablates *in vitro* 3'-5' exonuclease activity. Bottom: mutation of active site serine S135 to cysteine completely ablates activity, but mutation of the second active site serine S233 to cysteine reduces but does not ablate activity.

B) The active site of yUsb1 consists of residues H133, S135, H231, and S233. Aromatic residues F78 and Y187 flank the active site.

C) Mutation of tyrosine 187 to phenylalanine does not significantly change *in vitro* activity in yUsb1. This position is structurally conserved and is a phenylalanine in hUsb1.

D) The majority of selected residues, when mutated to alanine, have no effect on yUsb1 activity *in vivo*. Residues in yellow were assayed for ability to support yeast growth and their location in the structure are shown. Effect of yUsb1 mutations *in vivo* are summarized in **Table A2-1**.

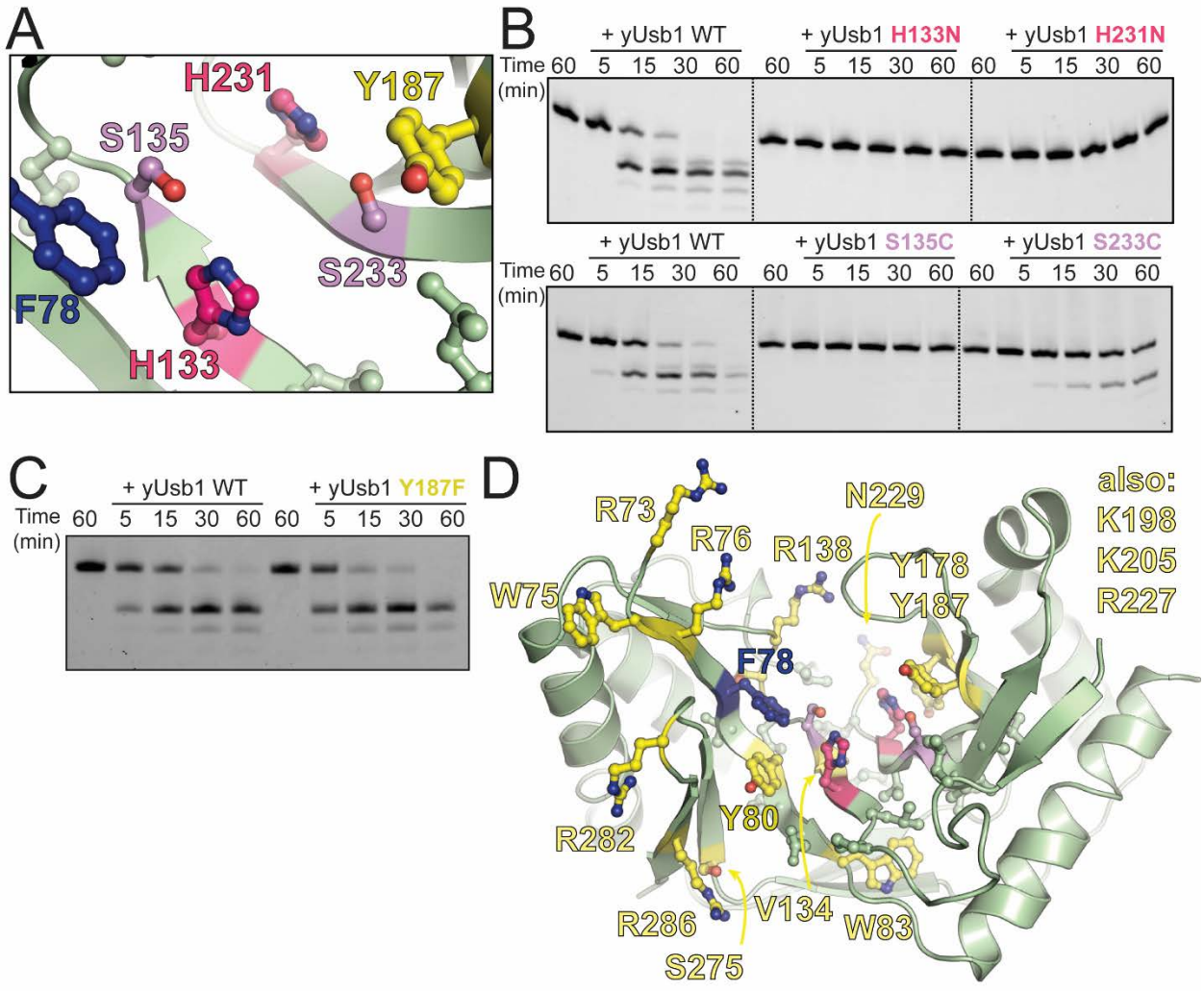

A
H231
Y187
S135
S233
F78
H133
B
+ yUsb1 WT
+ yUsb1 H133N
+ yUsb1 H231N
Time (min)
60 5 15 30 60 60 5 15 30 60 60 5 15 30 60
+ yUsb1 WT
+ yUsb1 S135C
+ yUsb1 S233C
Time (min)
60 5 15 30 60 60 5 15 30 60 60 5 15 30 60
C
+ yUsb1 WT
+ yUsb1 Y187F
Time (min)
60 5 15 30 60 60 5 15 30 60
D
R73
R76
R138
N229
Y178
Y187
W75
F78
also:
K198
K205
R227
Y80
R282
R286
V134
W83
S275

Table A2-1. Summary of the effect of mutations in yUsb1 *in vivo*.

All mutations were introduced into pRS414-Usb1 and plated on media lacking TRP and containing 5-FOA. Ability to complement loss of the *URA3* marked wild-type Usb1 plasmid is listed in the right hand column, where "-" indicates no growth and "+++" indicates growth similar to wild-type Usb1. * indicates strains that were not tested by serial dilution.

***USB1* allele**	***USB1Δ* complementation**
Vector	-
wild type	+++
Point mutations	
T42A	+++
T42D	+++
T42E	+++
D43A	+++
D43N	+++
L44A	+++
D43A/L44A	+++
P45A	+++
S50A	+++
K54A	+++
Y55A	++
H56A	+++
R73A	+++
W75A	+++
R76A	+++
F78A	+
F78H	+++
Y80A	+++
W83A	+++
H133A*	-
H133N*	-
V134L*	+++
T137S*	+++
R138A	+++
Y178A	+++
Y187A	+++
Y187F	+++
K198A	+++
K205A	+++
R227A	+++
N229A	+++
N229D	+++
H231A*	-
H231N*	-
S275A	+++

S275Q	+++
R282A	+++
R286A	+++
Alanine scanning of N-terminal domain residues 42-70	
T42A/D43A/L44A (3Ala1)	+++
P45A/A46/I47A (3Ala2)	+++
P48A/D49A/S50A (3Ala3)	+++
I51A/I52A/L53A (3Ala4)	+++
K54A/Y55A/H56A (3Ala5)	+
I57A/P58A/P59A (3Ala6)	+++
N60A/L61A/Q62A (3Ala7)	+++
K63A/Y64A/E65A (3Ala8)	+++
H66A/Q67A/D68A (3Ala9)	+++
M69A/N70A/M71A (3Ala10)	+++
N-terminal truncations and mutations	
usb1Δ1-20	+++
usb1Δ1-41	+++
usb1Δ1-42	+++
usb1Δ1-43	+++
usb1Δ1-44	-
usb1Δ1-45	-
usb1Δ1-46	-
usb1Δ1-47	-
usb1Δ1-48	-
usb1Δ1-49	-
usb1Δ1-50	-
usb1Δ1-50 + SV40 NLS	-
usb1Δ1-70 + SV40 NLS	-
usb1Δ20 D43A/L44A	-
usb1Δ41 T42A/D43A/L44A	-
usb1Δ42 D43A	++
usb1Δ42 L44A	-
usb1Δ42 D43A/L44A	-
Δ42-45	+++
Loop mutations	
107-115→4ALA	+++
122-131→4ALA	+
123-131→4ALA	+++
123-131→9ALA	+
Δ123-129	++
180-185→HsLoop (S180N/E181Q/R182E/A183K/N184T/T185R)	+++
180-186→4ALA	+++
180-186→7ALA	+++
279-283→4ALA	+

A2.3.2 Mutation of residue 78 in yUsb1 affects activity *in vitro*

We previously observed that phenylalanine 78 in yUsb1 was immediately adjacent to the active site and that there tended to be an aromatic residue at this position in homologs (DIDYCHUK *et al.* 2017). Indeed, mutation of this residue to alanine *in vivo* affects yeast viability and yUsb1 activity *in vitro* (DIDYCHUK *et al.* 2017). Interestingly, when the products of yUsb1-F78A or F78H were incubated with CIP or T4 PNK, they were CIP-insensitive (**Figure A2-2A**), suggesting that mutation of position F78 resulted in a product with a 2',3'-cyclic phosphate akin to the product of hUsb1 (**Figure A2-2B**). This result was exciting, as it suggested that F78 (H84 in hUsb1) is involved in the second catalytic step (2',3'-cyclic phosphodiesterase activity) that differs between yeast and human Usb1. We also previously observed that both yeast and human Usb1 are pH sensitive and the pH optimums of the homologs differ (DIDYCHUK *et al.* 2017). Using the CIP/T4 PNK assay, I tested the products of wild-type and F78H yUsb1 at different pH (**Figure A2-2C**). As predicted, both wild-type and F78H yUsb1 were slower at low (4.5) and high (8.5) pH. Interestingly, wild-type yUsb1 produced a CIP-insensitive product at high pH and a partially CIP-insensitive product at low pH (**Figure A2-2C**). These data suggest that the experiment is capturing an intermediate between the first (exonucleolytic) and second (phosphodiester opening) steps. Reducing the efficiency of the first step, either by mutation of F78 or by moving away from the pH optimum, slows the first step and allows for observation of the 2',3'-cyclic phosphate intermediate.

Although these experiments suggest that residue F78 plays a role in the first step of catalysis (rather than affecting the second step exclusively), how it affects either step has not been determined. Mutation of F78 to tyrosine has only a 3-fold effect on activity (**Figure A2-2D,E**). In contrast, mutation to glutamine has an effect similar to the alanine mutation (**Figure A2-2D,E**). This is surprising, as the homolog of Usb1 from *S. pombe* has a glutamine at this position (DIDYCHUK *et al.* 2017). The effect of different substitutions at position 78 appears to correlate with aromaticity, as mutation to tyrosine has little effect, mutation to histidine has an intermediate effect, and mutation to glutamine or alanine nearly abolishes activity (**Figure A2-2E**). While this

result suggests that position 78 could be important for stacking with the substrate to aid in its coordination in the active site, no contacts between H84 and 5'UMP were observed (DIDYCHUK *et al.* 2017) or are observed in co-structures of hUsb1 with 5'AMP, polyU, or polyA (NOMURA *et al.* in prep). Interestingly, H84 can hydrogen bond to and change the conformation of the active site serine residue S122. In the apo structure of hUsb1, S122 is in two conformations: one in which it hydrogen bonds to H84, and another one in which it assists in coordinating a water molecule along with active site histidine H120 (HILCENKO *et al.* 2013). As we previously proposed, this suggests that F78 influences activity by modulating the conformation and/or p*K*a of active site residues.

Figure A2-2. The role of residue F78 in catalysis.

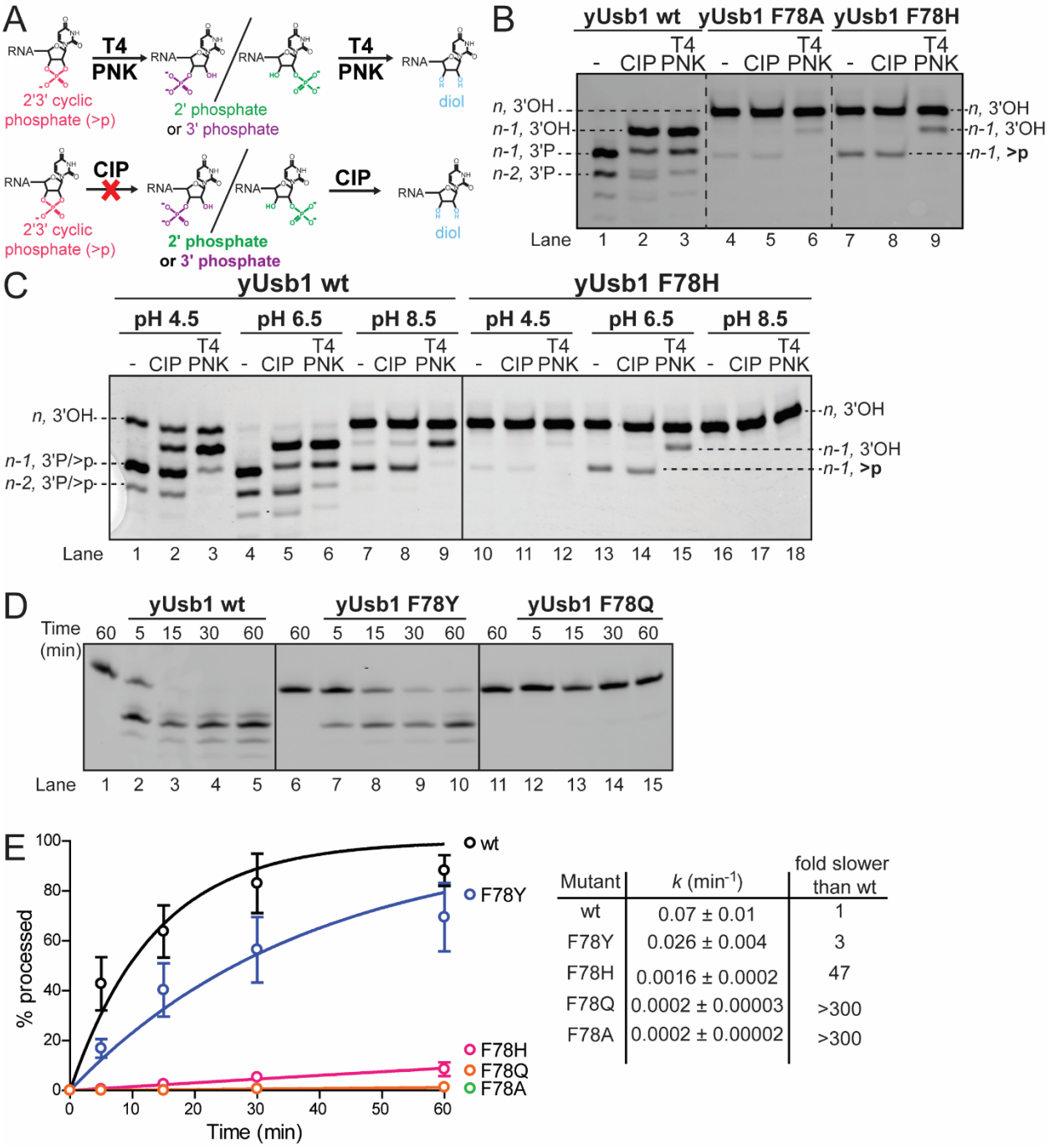
A) T4 PNK can open up a 2',3'-cyclic phosphate and remove non-cyclic phosphates to produce a 2',3'-cis diol. CIP is inactive on 2',3'-cyclic phosphates and can only remove non-cyclic phosphate to produce a 2',3'-cis diol. Sensitivity to T4 PNK with insensitivity to CIP suggests that the product is a 2',3'-cyclic phosphate, while sensitivity to both enzymes suggests that the product is a non-cyclic phosphate.

B) Mutation of position F78 to alanine or histidine reduces the rate of Usb1 activity and changes the reactivity of the product to CIP, suggesting that the product of yUsb1-F78A/H is a cyclic phosphate rather than a 3' non-cyclic phosphate.

C) Yeast Usb1 is less efficient at both low (4.5) and high (8.5) pH than at its pH optimum (pH 6.5). The products of yUsb1 activity at high pH mimic the effect of the F78 mutations in (A) in that they are CIP-insensitive. The F78H mutation is also less efficient at low and high pH.

D) Different substitutions at position 78 have divergent effects. Mutation to tyrosine has little effect on activity, while mutation to glutamine ablates activity.

E) Quantification of the effect of different substitutions in yUsb1 at position F78. The percent of the initial substrate processed over time was fit with a single exponential rate equation and the rate constants are shown in the chart to the right. Mutation to a tyrosine has a ~3-fold effect, while mutation to histidine has a ~50-fold effect. Mutation to a non-aromatic residue (glutamine or alanine) has a larger effect.



Mutant	k (min^{-1})	fold slower than wt
wt	0.07 ± 0.01	1
F78Y	0.026 ± 0.004	3
F78H	0.0016 ± 0.0002	47
F78Q	0.0002 ± 0.00003	>300
F78A	0.0002 ± 0.00002	>300

A2.3.3 Mutation of loops in yUsb1 have different effects *in vivo*

There are several loops near the active of yUsb1 and other important regions (**Figure A2-3A**). These loops are poorly conserved in both length and sequence, and have different effects when mutated *in vivo* (**Figure A2-3B**). While it unlikely that these loops play a direct role in catalysis, they may contribute to substrate specificity, activity, or protein stability.

The first loop (Loop 1), consisting of residues 107-115, is disordered in the crystal structure of yUsb1 (DIDYCHUK *et al.* 2017). Loop 1 could conceivably alter nearby secondary structure or contribute to protein stability. The second loop (Loop 2), consisting of residues 122-131, is nine residues long in yUsb1 but only three residues long in hUsb1 (**Figure A2-3C**). Interestingly, Loop 2 contacts the substrate in a co-structure of hUsb1 with RNA (NOMURA *et al.* in prep). The third loop (Loop 3), consisting of residues 180-186, is very close to the active site. Loop 3 is conserved in length in hUsb1, but is in a slightly different conformation (**Figure A2-3C**). Loop 3 in yUsb1 clashes with 5'UMP in the structure of hUsb1-5'UMP (DIDYCHUK *et al.* 2017) and in the co-structure of hUsb1 with RNA (NOMURA *et al.* in prep), suggesting that it might be flexible in solution and may adopt a different conformation when substrate is present. Although the proximity of this loop to the active site suggested that it may play an important role in Usb1 activity, mutating this loop to its homologous sequence from hUsb1 has no effect on the kinetics of activity *in vitro* (**Figure A2-3E**) or on the 3' end modification produced *in vitro* (data not shown). This loop can also be replaced with four or seven alanines with no effect *in vivo* (**Figure A2-3B**), suggesting that neither sequence nor length greatly affect activity of yUsb1. Loop 4 is distal to the active site, but may be important in substrate binding.

The role of loops 2 and 4 in activity or stability remain to be determined. Loop 4 cannot be replaced with four alanines *in vivo*, although this mutation also shortens the loop by one residue (**Figure A2-3B**). Both the length and sequence of loop 2 also influences viability, as residues 122-131 cannot be replaced with four alanines, yet 123-131 can, suggesting that F122 plays an important role (**Figure A2-3B**). Interestingly, residues 123-131 cannot be replaced by nine

alanines (which maintains the native length of the loop), despite 123-131→4ALA having little effect on growth (**Figure A2-3B**).

Figure A2-3. Mutation of loop regions in yUsb1.

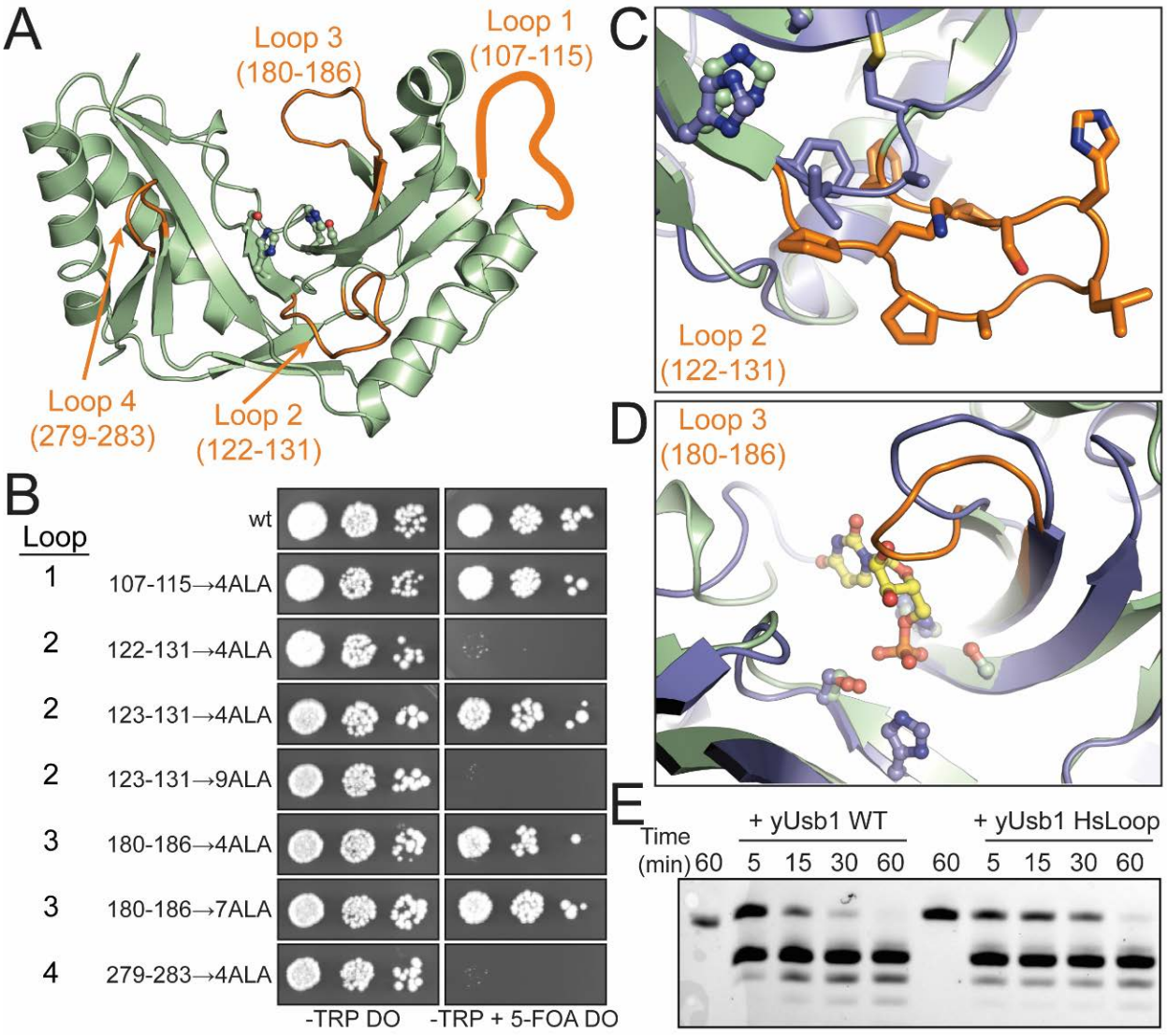
A) Loops (orange) in the catalytic domain of yUsb1 (PDB 5UQJ; green). Loop1 (residues 107-115) is disordered in the crystal structure.

B) Alanine scanning of loop mutations *in vivo*. Mutating and shortening loops 1 and 3 has no effect *in vivo*. Replacing loops 2 and 4 with four alanines severely limits growth *in vivo*. Interestingly, both the length of loop 2 and whether phenylalanine 122 is mutated to an alanine affect viability.

C) Loop 2, with 9 residues, is significantly longer in yUsb1 (green; loop in orange) than the 3 residue loop in hUsb1 (blue).

D) Loop 3 (residues 180-186) is conserved in length in hUsb1 (PDB 5V1M), but likely must move to accommodate the substrate (5'UMP, yellow).

E) Mutating residues 180-185 to the homologous human sequence (S180N/E181Q/ R182E/A183K/N184T/T185R) has little effect on yUsb1 activity *in vitro* or on yeast viability (data not shown).


A
Loop 3
(180-186)
Loop 1
(107-115)
Loop 4
(279-283)
Loop 2
(122-131)
B
Loop
wt
1
107-115→4ALA
2
122-131→4ALA
2
123-131→4ALA
2
123-131→9ALA
3
180-186→4ALA
3
180-186→7ALA
4
279-283→4ALA
-TRP DO
-TRP + 5-FOA DO
C
Loop 2
(122-131)
D
Loop 3
(180-186)
E
Time
(min)
+ yUsb1 WT
+ yUsb1 HsLoop
60 5 15 30 60 60 5 15 30 60

A2.3.4 Towards understanding the essential role of the N-terminal domain

The N-terminal domain of yUsb1 (residue 1-70) is predicted to be largely disordered (LIU AND ROST 2004; BUCHAN *et al.* 2013) (**Figure A2-4A**). This domain does not contribute to catalytic activity *in vitro*, yet N-terminal truncation *in vivo* is not tolerated after residue 43 (DIDYCHUK *et al.* 2017). Removing the first 43 amino acids (Δ43) is viable, but removing an additional N-terminal amino acid (Δ44) is inviable (**Figure A2-4B** and Didychuk 2017). Interestingly, threonine 42 may be phosphorylated (ALBUQUERQUE *et al.* 2008), suggesting that this region may be regulated or have an essential function. This region is also very proline-rich, which may contribute to its function (DIDYCHUK *et al.* 2017).

Mutation of T42, D43, or L44 to alanine in the context of full-length protein has no effect (**Figure A2-4B**), even when all three residues are mutated (**Table A2-1**). However, when L44 (but not D43) is mutated to alanine in the context of Δ42 or Δ20, yeast cannot survive (**Figure A2-4B**). This data suggests that L44 has an important role within Usb1 *in vivo*. Additionally, it suggests that L44 by itself is not essential, but that other N-terminal domain residues may contribute to its function.

The N-end protein degradation rule cannot explain the dependence on the N-terminal domain, as further truncation into the domain changes the N-terminal sequence but does not affect viability (**Figure A2-4B**). Additionally, this region is likely not important for nuclear localization, as inclusion of an N-terminal SV40 nuclear localization signal could not rescue deletion of the first 50 or 70 amino acids (DIDYCHUK *et al.* 2017). However, either expression or stability of Usb1 is affected by N-terminal truncation, as western blotting revealed that virtually no protein was detectable when Δ70 was expressed (DIDYCHUK *et al.* 2017). Furthermore, when Usb1 was expressed from a high copy 2μ plasmid, yeast were able to tolerate larger truncations of the N-terminus (DIDYCHUK *et al.* 2017). This could be due to higher expression levels or due to the N-terminal HA tag changing the local sequence or structure environment. To unambiguously determine if expression or protein stability is affecting protein levels, qRT-PCR should be

performed to assess mRNA levels. If the mRNA for the N-terminal truncations is stable, pulse-chase experiments with ^{35}S-labeled methionine followed by immunoprecipitation and western blotting could be done to determine the half-lives of different forms of Usb1. The half-life of the protein could also be measured by placing *USB1* and truncations therein under control of a galactose-inducible *GAL1* promoter and quantifying protein levels after glucose repression.

To identify other amino acids that are potentially important for Usb1 activity *in vivo*, alanine scanning mutagenesis was performed between residues 42-70 (**Figure A2-4C**). Sets of three residues at a time were mutated to alanine. One set of mutations, which contained the K54A/Y55A/H56A mutations, grew worse than wild-type (**Figure A2-4C**). Of these three mutations, only Y55A had any effect by itself (**Figure A2-4D**). Interestingly, this set of mutations is adjacent to a short (5 residue) predicted α-helix (LIU AND ROST 2004; BUCHAN *et al.* 2013). Mutation of K54, Y55, and H56 to alanine could potentially extend this helix, as alanines have the highest α-helix propensity (PACE AND SCHOLTZ 1998). The catalytic domain of yUsb1 (residues 71-290) is amenable to NMR and backbone assignment is underway (see **Appendix V**). Collecting a ^{1}H-^{15}N-HSQC of the full-length protein could aid in determining whether the N-terminal domain contains significant secondary structure, as chemical shifts can be correlated with protein secondary structure (MIELKE AND KRISHNAN 2009). Determining how the N-terminal domain of Usb1 functions *in vivo* could lead to new insights into how Usb1 is targeted to U6 snRNA. It may aid in protein-protein interactions that target Usb1 to its target RNA, or may aid in localization or RNA binding itself.

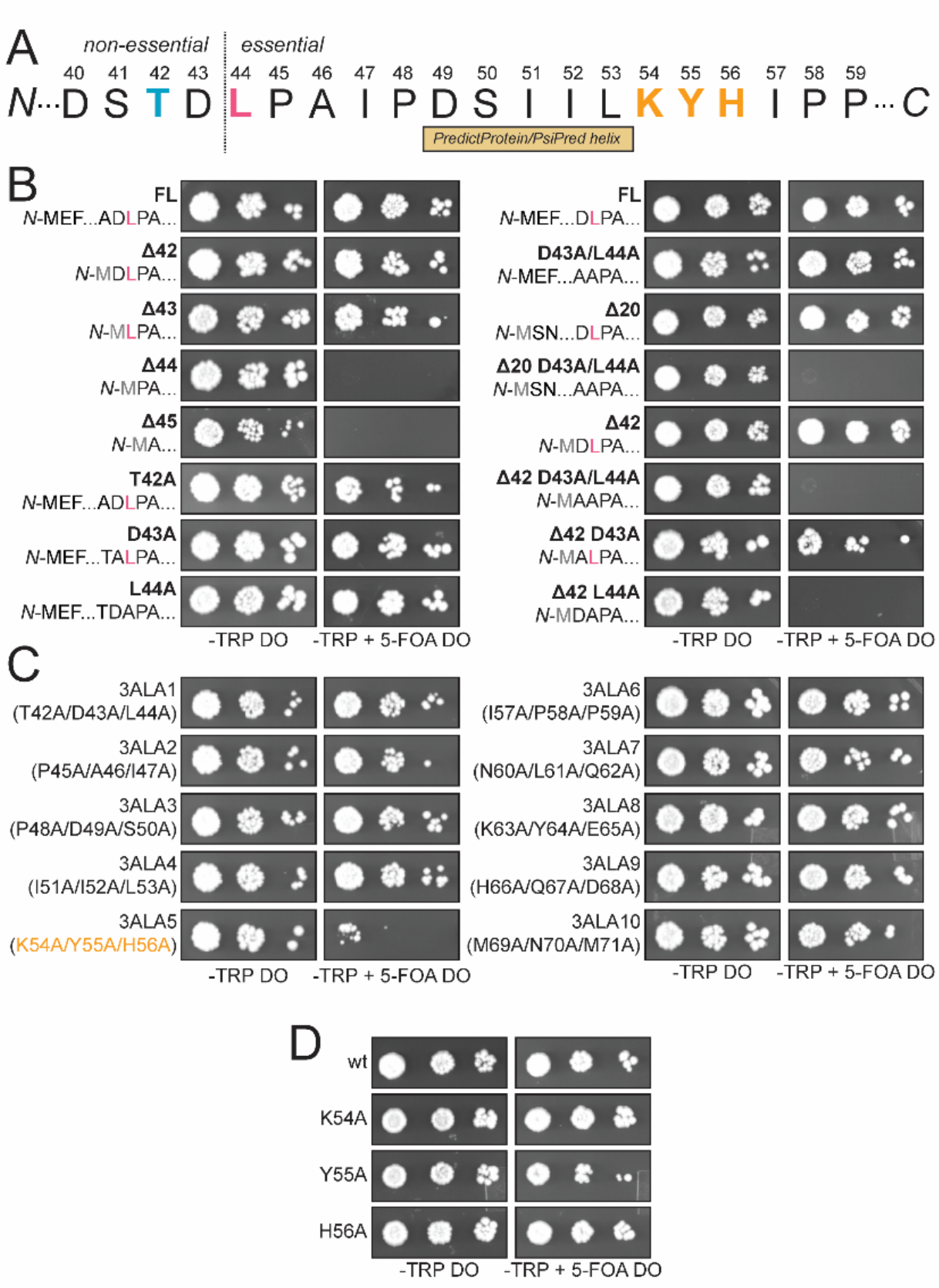
Figure A2-4. Mutations and truncations within the N-terminal domain of Usb1 affect viability.

A) Annotated of a portion of the N-terminal domain of yUsb1. Truncations up to residue 44 are tolerated *in vivo*. Threonine 42 (blue) may be phosphorylated (ALBUQUERQUE *et al.* 2008). Leucine 44 (pink) is essential for viability. Lysine 54, tyrosine 55, and histidine 56 (orange) when mutated to alanine affect viability and are C-terminal to a short predicted α-helix.

B) Truncations and mutations within yUsb1. Truncation up to leucine 44 is viable; however, further truncation is inviable. Mutation of threonine 42, aspartic acid 43, or leucine 44 to alanine is tolerated in the context of full-length yUsb1. However, mutation of leucine 44 in the context of truncation to remove the first 20 or 42 amino acids is inviable. The sequence of the N-terminal of the encoded protein is indicated.

C) Alanine scanning of residues 42-71 of full-length yUsb1. Sets of three residues were mutated to alanine. One set of mutations, 3Ala5 (K54A/Y55A/H56A), affects viability.

D) When lysine 54, tyrosine 55, and histidine 56 are individually mutated to alanine, little phenotypic effect is observed.


A
non-essential
essential
40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59
N···D S T D L P A I P D S I I L K Y H I P P···C
PredictProtein/PsiPred helix
B
FL
N-MEF...ADLPA...
Δ42
N-MDLPA...
Δ43
N-MLPA...
Δ44
N-MPA...
Δ45
N-MA...
T42A
N-MEF...ADLPA...
D43A
N-MEF...TALPA...
L44A
N-MEF...TDAPA...
-TRP DO
-TRP + 5-FOA DO
FL
N-MEF...DLPA...
D43A/L44A
N-MEF...AAPA...
Δ20
N-MSN...DLPA...
Δ20 D43A/L44A
N-MSN...AAPA...
Δ42
N-MDLPA...
Δ42 D43A/L44A
N-MAAPA...
Δ42 D43A
N-MALPA...
Δ42 L44A
N-MDAPA...
-TRP DO
-TRP + 5-FOA DO
C
3ALA1
(T42A/D43A/L44A)
3ALA2
(P45A/A46/I47A)
3ALA3
(P48A/D49A/S50A)
3ALA4
(I51A/I52A/L53A)
3ALA5
(K54A/Y55A/H56A)
-TRP DO
-TRP + 5-FOA DO
3ALA6
(I57A/P58A/P59A)
3ALA7
(N60A/L61A/Q62A)
3ALA8
(K63A/Y64A/E65A)
3ALA9
(H66A/Q67A/D68A)
3ALA10
(M69A/N70A/M71A)
-TRP DO
-TRP + 5-FOA DO
D
wt
K54A
Y55A
H56A
-TRP DO
-TRP + 5-FOA DO

A2.4 References

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

Buchan, D. W., F. Minneci, T. C. Nugent, K. Bryson and D. T. Jones, 2013 Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res 41**:** W349-357.

Didychuk, A. L., E. J. Montemayor, T. J. Carrocci, A. T. DeLaitsch, S. E. Lucarelli *et al.*, 2017 Usb1 controls U6 snRNP assembly through evolutionarily divergent cyclic phosphodiesterase activities. Nat Commun 8**:** 497.

Gietz, R. D., and R. A. Woods, 2002 Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. Methods Enzymol 350**:** 87-96.

Hilcenko, C., P. J. Simpson, A. J. Finch, F. R. Bowler, M. J. Churcher *et al.*, 2013 Aberrant 3' oligoadenylation of spliceosomal U6 small nuclear RNA in poikiloderma with neutropenia. Blood 121**:** 1028-1038.

Liu, J., and B. Rost, 2004 Sequence-based prediction of protein domains. Nucleic Acids Res 32**:** 3522-3530.

Mazumder, R., L. M. Iyer, S. Vasudevan and L. Aravind, 2002 Detection of novel members, structure-function analysis and evolutionary classification of the 2H phosphoesterase superfamily. Nucleic Acids Res 30**:** 5229-5243.

Mielke, S. P., and V. V. Krishnan, 2009 Characterization of protein secondary structure from NMR chemical shifts. Prog Nucl Magn Reson Spectrosc 54**:** 141-165.

Mroczek, S., J. Krwawicz, J. Kutner, M. Lazniewski, I. Kucinski *et al.*, 2012 C16orf57, a gene mutated in poikiloderma with neutropenia, encodes a putative phosphodiesterase responsible for the U6 snRNA 3' end modification. Genes Dev 26**:** 1911-1925.

Nomura, Y., E.J. Montemayor, S.E. Butcher. In preparation.

Pace, C. N., and J. M. Scholtz, 1998 A helix propensity scale based on experimental studies of peptides and proteins. Biophys J 75**:** 422-427.

Appendix 3: Screening for mutations in human Usb1

A3.1 Overview

Usb1 is an essential protein in yeast (GIAEVER *et al.* 2002; MROCZEK *et al.* 2012; HILCENKO *et al.* 2013). Overexpression of human Usb1 in yeast very weakly complements loss of yUsb1 (DIDYCHUK *et al.* 2017). However, human Usb1 mutated to include a phenylalanine instead of a histidine at position 84 permits yeast growth (DIDYCHUK *et al.* 2017). This substitution was hypothesized to make the protein more “yeast-like”. The crystal structures of human and yeast Usb1 revealed that hUsb1-H84 and yUsb1-F78 are structurally homologous and positioned directly adjacent to the active site (DIDYCHUK *et al.* 2017). Therefore, I reasoned that other mutations in human Usb1 could also make its activity more analogous to yeast Usb1. Identifying mutations that allow human Usb1 to complement yeast Usb1 may reveal important determinants of Usb1 activity. I performed an error-prone PCR mutagenesis screen to identify mutations in the catalytic domain of human Usb1 that improved complementation of yeast Usb1 loss. After plasmid rescue and sequencing, several mutations were identified and validated. One such mutation, hUsb1-S126T, complements *USB1* deletion.

A3.2 Materials and Methods

A3.2.1 Error-prone PCR

Using primers upstream of the internal native Kpn1 site in hUsb1 (+162-KpnI-f) and downstream of the SalI site (+190-SalI-r), the coding region of hUsb1 and 196 bases downstream of the coding region was amplified from the p425-GPD-hUsb1 plasmid (DIDYCHUK *et al.* 2017) (**Table A3-1**). This fragment was used in a 50 µL error-prone PCR reaction using a Mutazyme II kit (Agilent Genomics) with either 100 ng or 1000 ng of template. PCR was conducted by heating to 95°C for 2 minutes, then cycling 35 times between 95°C for 30 seconds, 55°C for 30 seconds, and 72°C for 1 minute followed by a final extension time of 10 minutes according to the manufacturer's instructions. Mutagenized fragments were purified using a PCR & DNA Cleanup Kit (New England BioLabs).

A3.2.2 Gap repair and screen

The p425-GPD-hUsb1 plasmid was mutagenized via inverse PCR and subsequent circularization as previously described (DIDYCHUK *et al.* 2017) to replace an internal KpnI site with an AscI site. The resulting plasmid was digested with AscI and Kpn1 overnight and gel purified. yTJC0700 was transformed with 4.2 µg gapped AscI/SalI p425-GPD-hUsb1+AscI alone or with 1.4 µg of WT or H84F insert, or 4.2 µg of EP-PCR insert using the lithium acetate method (GIETZ AND WOODS 2002). After one day, yeast were replica plated onto media lacking leucine and containing 1 mg/mL 5-fluoroorotic acid (5-FOA). After two days, individual colonies from these plates were re-streaked onto -LEU + 5-FOA plates. After two days of growth, individual colonies were inoculated in 3 mL of liquid -LEU drop out media and grown overnight. Plasmids were isolated from 1.5 mL of culture after overnight growth using a Qiagen miniprep kit where the lysis step contained 250 µL glass beads. Isolated plasmids were transformed into DH5α cells. Plasmids were isolated and sequenced using standard techniques. Improved growth was confirmed using isolated and sequenced plasmids re-transformed into yTJC0700. To further validate the screen, selected mutations were reintroduced into p425-GPD-hUsb1 via inverse PCR

and ligation (**Table A3-2**) and assayed for growth on media lacking leucine and containing 1 mg/mL 5-FOA.

A3.2.3 Yeast *USB1* complementation assays

yTJC0700 was transformed with p425-GPD-hUsb1 and variants using the lithium acetate method (GIETZ AND WOODS 2002). Growth phenotypes were assessed by spotting 10-fold serial dilutions (OD_{600} = 0.5, 0.05, 0.005) onto solid medium lacking leucine and containing 1 mg/mL 5-FOA. Plates were incubated at 30°C for 3 days.

Table A3-1. Nucleotide sequence of synthetic genes.

hUsb1, codon optimized for *S. cerevisiae* KpnI sequence replaced with AscI sequence underlined. Region 3' of stop codon that is amplified in gap repair fragment in lower case.	5'-ATGAGTGCAGCCCCTTTAGTAGGGTACAGCTCAAGCGGGAG TGAAGATGAGAGCGAGGACGGAATGAGGACCAGACCGGGTGAT GGAAGTCATAGAAGAGGCCAGAGCCCCTTGCCGAGGCAACGTT TCCCCGTACCCGACAGTGTTTTGAATATGTTCCCAGGTACCGAA GAGGGACCCGAAGATGACTCTACCAAACACGGAGGGAGGGTCC GTACTTTTCCTCACGAGCGTGGGAACTGGGCAACTCATGTCTAC GTCCCGTATGAAGCCAAGGAGGAATTTTTGGATCTGTTGGATGT GCTACTTCCCCACGCCCAAACCTACGTACCGAGACTGGTAAGGA TGAAAGTCTTCCACCTGAGTCTGAGTCAGAGCGTTGTGCTGAGG CATCACTGGATACTGCCTTTCGTACAGGCGCTTAAGGCTCGTAT GACCAGTTTTCACAGGTTCTTTTTTACTGCCAATCAGGTGAAGAT ATATACAAACCAAGAAAAAACAAGGACCTTTATTGGCCTAGAGGT TACTTCCGGTCACGCACAATTTTTGGACCTAGTCTCCGAGGTTG ACCGTGTGATGGAAGAATTTAACCTGACCACTTTTTATCAAGACC CTTCATTCCATCTGTCACTTGCATGGTGTGTCGGAGATGCTCGT CTGCAACTAGAAGGGCAATGTCTTCAAGAGCTGCAAGCCATAGT CGATGGTTTCGAGGATGCTGAGGTCCTGCTTAGAGTCCACACAG AGCAGGTGCGTTGCAAGAGTGGCAACAAATTTTTTTCCATGCCG CTGAAATGAgtcgacctcgagtcatgtaattagttatgtcacgcttacattcacgccctcccc ccacatccgctctaaccgaaaaggaaggagttagacaacctgaagtctaggtccctatttatttttt ttatagttatgttagtattaagaacgttatttatatttcaaatttttctttttttctgtacagacgcgtgtac gc-3'

Table A3-2. List of synthetic DNA oligonucleotides.

Primer name	Primer sequence
GPD-for sequencing primer	5'-CGGTAGGTATTGATTGTAATTCTG-3'
Cyc1-rev sequencing primer	5'-GCGTGAATGTAAGCGTGAC-3'
+162-KpnI-f	5'-ATGAGTGCAGCCCCTTTAGTAGGGTACAGCTCAAGCG-3'
+190-SalI-r	5'-GCGTACACGCGTCTGTACAGAAAAAAAAGAAAAATTTG-3'
Add-Asc1-f	5'-GGCGCGCCGAAGAGGGACCCGAAGATGACTCTACCAAAC-3'
Add-Asc1-r	5'-TGGGAACATATTCAAAACACTGTCGGGTAC-3'
hUsb1-E77K-f	5'-AAACGTGGGAACTGGGCAACTCATGTCTACGTC-3'
hUsb1-E77-r	5'-GTGAGGAAAAGTACGGACCCTCCCTCCGTG-3'
hUsb1-D97G-f	5'-GGTCTGTTGGATGTGCTACTTCCCCACGCCCAAAC-3'
hUsb1-D97-r	5'-CAAAAATTCCTCCTTGGCTTCATACGGGAC-3'
hUsb1-Q125H-f	5'-CATAGCGTTGTGCTGAGGCATCACTGGATACTG-3'
hUsb1-Q125R-f	5'-AGAAGCGTTGTGCTGAGGCATCACTGGATACTG-3'
hUsb1-Q125-r	5'-ACTCAGACTCAGGTGGAAGACTTTCATCCTT-3'
hUsb1-S126T-f	5'-ACTGTTGTGCTGAGGCATCACTGGATACTGCCTTT-3'
hUsb1-S126-r	5'-CTGACTCAGACTCAGGTGGAAGACTTTCATC-3'
hUsb1-T176S-f	5'-TCTTCCGGTCACGCACAATTTTTGGACCTAGTC-3'
hUsb1-T176-r	5'-AACCTCTAGGCCAATAAAGGTCCTTGTTTTTTC-3'
hUsb1-H247R-f	5'-AGAACAGAGCAGGTGCGTTGCAAGAGTGGCAAC-3'
hUsb1-H247L-f	5'-CTTACAGAGCAGGTGCGTTGCAAGAGTGGCAAC-3'
hUsb1-H247-r	5'-GACTCTAAGCAGGACCTCAGCATCCTCGAAA-3'

A3.3 Results and discussion

The observation that mutations at yUsb1-F78/hUsb1-H84 modulate activity *in vivo* and *in vitro* along with the observation that these residues are structurally homologous and near the active site strongly suggested that residues outside the active site modulate the catalytic activity of Usb1 (DIDYCHUK *et al.* 2017). The mechanism by which the hUsb1-H84F mutation allows for yeast growth and affects *in vitro* activity is not known, and therefore identification of other mutations that behave similarly could be informative. I adopted a forward genetics approach where yeast selected hUsb1 variants that could complement deletion of *USB1*. To accomplish this, I used gap repair in yeast to introduce mutations in hUsb1 (**Figure A3-1A**). Error-prone PCR was used to introduce mutations in hUsb1 amino acids 54-265 (**Figure A3-1A**). Gap repair was efficient and the background for this step was low, with ~150 total colonies on the no insert plates and >10 times as many colonies on gap repair plates with insert (**Figure A3-1B**). After one day of growth on media lacking leucine to select for the p425-GPD plasmid, yeast were replica plated onto media lacking leucine and containing 5-FOA, which selects against the pRS416-yUsb1 plasmid. Resulting colonies were further selected on -LEU + 5-FOA, then grown in liquid -LEU culture for plasmid isolation.

Rescued plasmids were transformed into *E. coli* cloning cells (NEB 5α) for amplification and sequencing. Two colonies from the wild-type and hUsb1-H84F controls (gap repair 2 and 3 respectively) were selected for sequencing, which confirmed their identity (**Table A3-3**). Eighteen colonies from the gap repair using EP-PCR inserts (gap repair 4 and 5) were isolated and transformed into *E. coli*. Several of these isolated plasmids did not produce colonies in *E. coli* or were not sequenced using the Cyc1-rev primer. Plasmids that were successfully sequenced are listed in **Table A3-3**. Interestingly, the isolated plasmids from the gap repair that used the EP-PCR insert amplified from 1000 ng template contained no mutations in hUsb1 (**Table A3-3**), suggesting that the mutational frequency was lower than advertised. Isolated plasmids from the 100 ng template EP-PCR insert contained 0-5 mutations (**Table A3-3**). Interestingly, many of the

mutations occurred in nucleotides 1-162 (amino acids 1-54), a region that should have undergone homologous recombination with the gapped plasmid and contain the wild-type sequence (**Table A3-3**; 4.1, 4.5, 4.8, 4.11).

To confirm that the isolated plasmids were the source of improved yeast growth, screen plasmids were transformed into yTJC0700 and assayed for growth (**Figure A3-2A**). Several screen plasmids showed weak complementation (4.5, 4.6, 4.11), but others (4.8 and 4.9) showed a marked improvement in growth. The mutated residues identified in the screen are distributed throughout the coding region of hUsb1, with many falling outside the crystallizable domain (**Figure A3-2B**). To confirm that the mutations identified in the screen resulted in improved growth, site-directed mutagenesis was used to introduce mutations in the p425-GPD-hUsb1 plasmid. Of the tested mutations, hUsb1-S126T (identified in screen 4.9), grew better than hUsb1-H84F and grew similarly to yUsb1 (**Figure A3-2C**). Two screen mutations- S126T and Q125H- occur in amino acids that are directly adjacent to the active site of hUsb1 and near hUsb1-H84F (**Figure A3-2D**). Mutation of S126 to threonine, which adds a bulky methyl group near the protein backbone, could conceivably restrict the conformation of this amino acid and affect the conformation of the strand that leads to the first catalytic histidine, H120. This mutation could also affect the conformation of the nearby residues H84 and Y202 (**Figure A3-2D**).

This screen, conducted to demonstrate proof of principle, has identified one novel mutation (hUsb1-S126T) that influences hUsb1 activity *in vivo*. I have demonstrated that this subtle mutation causes improved yeast growth. The mechanism behind this improved growth (and that of hUsb1-H84F) remains undetermined. Quantitatively determining how these mutations affect *in vitro* activity and confirming that they do not influence protein expression, stability, or localization, will give important clues to how the environment outside the active site influences Usb1 activity.

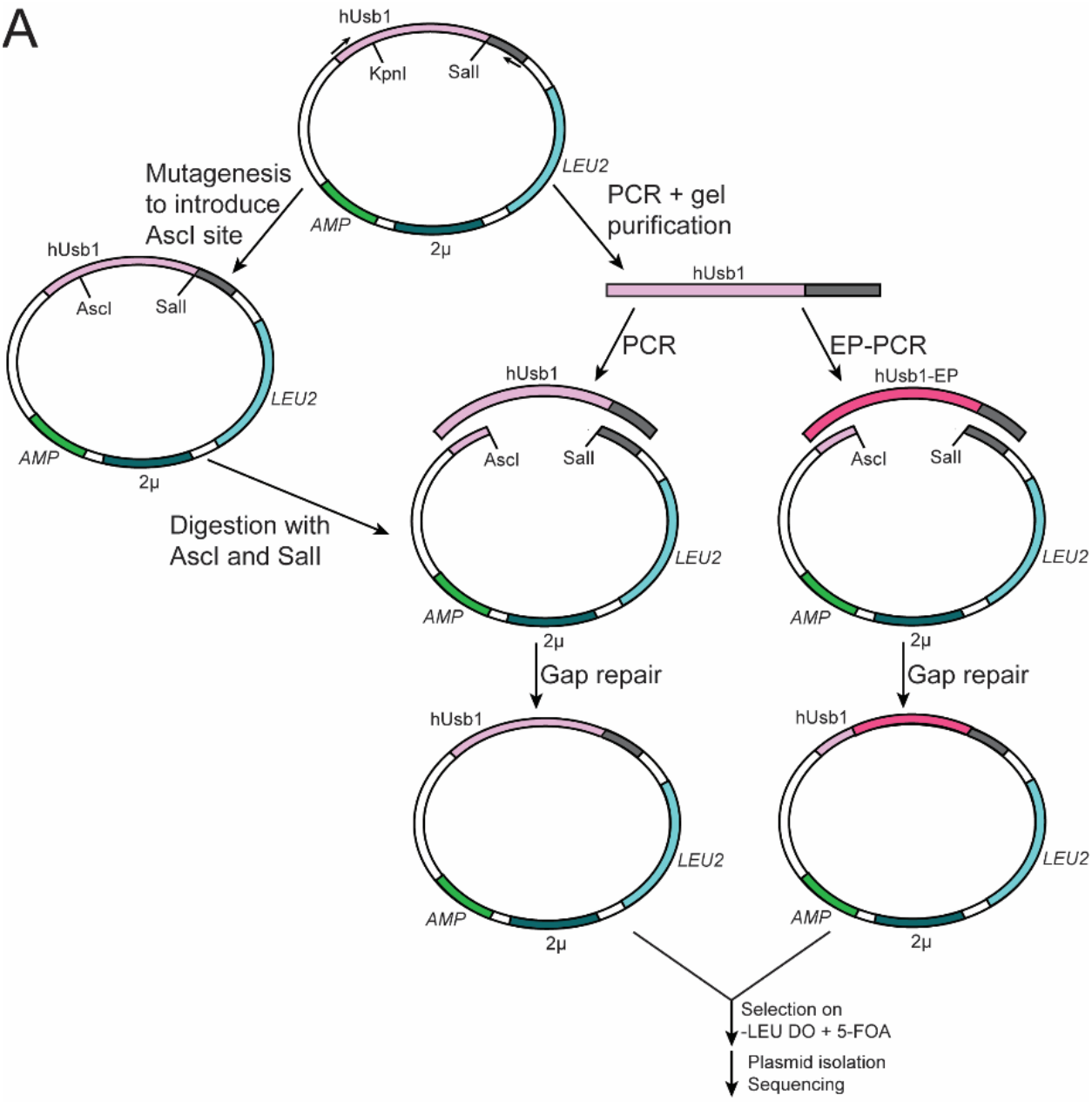
This screen could easily be repeated, expanded, and improved upon to increase the stringency of selection and increase the number of plasmids screened. Typical genetic screens

for suppressor mutations screen >100 colonies, and given the low frequency of complementation success in isolated plasmids (**Figure A3-3A**), more colonies must be screened. Alternatively, a targeted selection could be performed, in which residues near the active site (such as H84, M116, V118, F119, Q125, S126, Y161, Y202, S206, and Q208) could be mutated. A Geneblock with mixed bases for the targeted residues could be used for gap repair in lieu of using error-prone PCR to generate mutations across the span of the coding region. Additionally, the screen could be modified to move away from the p425-GPD backbone, which uses a 2µ promoter. Using a *CEN* promoter could reduce any anomalous results from overexpression of a human protein in yeast.

Figure A3-1. Gap repair workflow and efficiency.

A) Workflow for gap repair and screen. The hUsb1 ORF and 196 bp downstream was amplified using the +162-KpnI-f and +190-SalI-r primers. The resulting 994 nt fragment was gel purified to remove template plasmid. This fragment was then used for EP-PCR. The gapped plasmid was created by first using site-directed mutagenesis to replace an internal KpnI site with an AscI site, digestion with AscI and SalI, and gel purification. Gap repair with the EP-PCR fragment should produce a p425-GPD-hUsb1 plasmid with mutations introduced between amino acids 54-265.

B) As expected, gap repair with no insert (1A) was inefficient. Background colonies likely come from incomplete digestion of the plasmid. Gap repair with wild-type hUsb1 insert (2A) or hUsb1-H84F insert (3A) were efficient, as was gap repair with EP-PCR insert starting from 100 ng of template (4A) and 1000 ng (5A).


A
hUsb1
KpnI
SalI
LEU2
AMP
2μ
Mutagenesis
to introduce
AscI site
PCR + gel
purification
hUsb1
AscI
SalI
LEU2
AMP
2μ
hUsb1
PCR
EP-PCR
hUsb1
AscI
SalI
LEU2
AMP
2μ
hUsb1-EP
AscI
SalI
LEU2
AMP
2μ
Digestion with
AscI and SalI
Gap repair
Gap repair
hUsb1
LEU2
AMP
2μ
hUsb1
LEU2
AMP
2μ
Selection on
-LEU DO + 5-FOA
Plasmid isolation
Sequencing

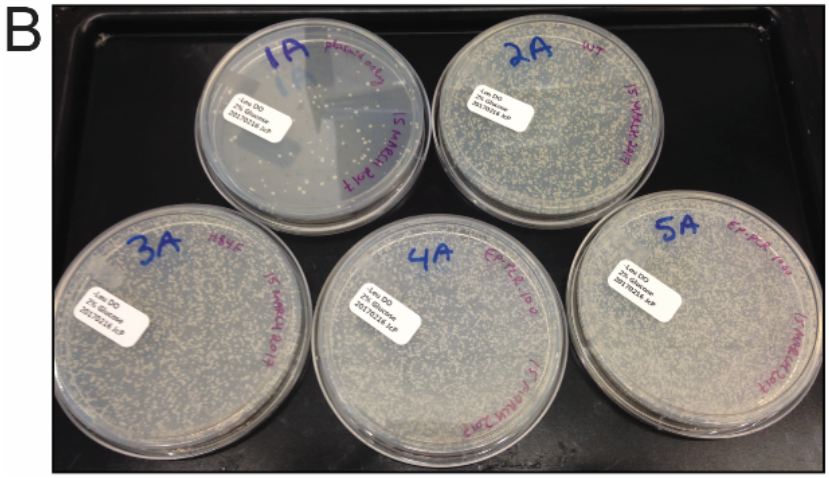

B
1A
2A
3A
4A
5A

Figure A3-2. Screen results.

A) After plasmid rescue of screen hits, isolated plasmids were re-transformed into yTJC0700 and assayed for growth on media lacking leucine and containing 5-FOA. Bold font indicates mutations that were selected and tested in (B).

B) Mutations within the crystallizable domain of hUsb1 were selected from the screen and cloned into p425-GPD-hUsb1, then assayed as in (A).

C) Screen mutations within the crystallizable domain of hUsb1 (PDB 5V1M) are shown in pink sticks and spheres. The active residues are shown in blue sticks and spheres.

D) Screen mutations were identified in Q125 and S126, which are adjacent to H84 and active site residues H120, S122, H208, and S210.

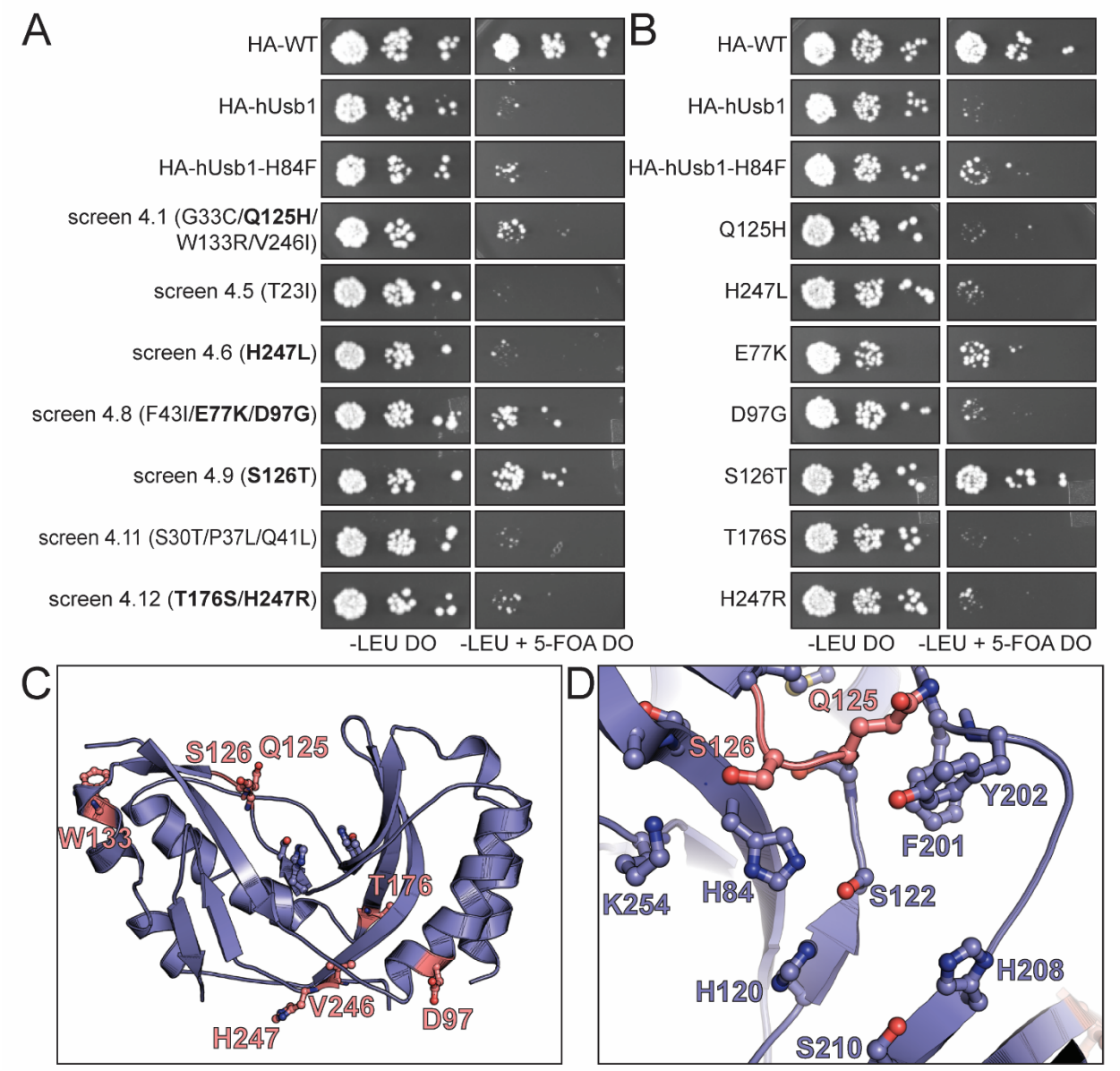

A
HA-WT
HA-hUsb1
HA-hUsb1-H84F
screen 4.1 (G33C/Q125H/ W133R/V246I)
screen 4.5 (T23I)
screen 4.6 (H247L)
screen 4.8 (F43I/E77K/D97G)
screen 4.9 (S126T)
screen 4.11 (S30T/P37L/Q41L)
screen 4.12 (T176S/H247R)
-LEU DO
-LEU + 5-FOA DO
B
HA-WT
HA-hUsb1
HA-hUsb1-H84F
Q125H
H247L
E77K
D97G
S126T
T176S
H247R
-LEU DO
-LEU + 5-FOA DO
C
S126
Q125
W133
T176
V246
D97
H247
D
Q125
S126
Y202
F201
K254
H84
S122
H120
H208
S210

Table A3-3. Sequencing results from isolated screen mutants.

Screen culture	Mutations (bold font indicates additional testing in Figure A3-2B)
hUsb1-WT gap repair control	
2.1	WT
hUsb1-H84F gap repair control	
3.1	H84F
3.2	H84F
hUsb1 EP-PCR 100 ng template	
4.1	G33C (G100>T), **Q125H** (G375>T), W133R (T397>C), V246I (G736>A)
4.3	WT
4.5	T23I (C71>T)
4.6	**H247L** (A740>T)
4.8	F43I (T127>A), **E77K** (G229>A), **D97G** (A290>G), silent mutation (T309>A), silent mutation (T513>A)
4.9	Silent mutation (T222>C), **S126T** (G377>C)
4.11	S30T (G89>C), P37L (C110>T), Q41L (A122>T)
4.12	Silent mutation (A177>C), **T176S** (A526>T), **H247R** (A740>G)
4.13	E57K (G169>A)
hUsb1 EP-PCR 1000 ng template	
5.1	WT
5.4	WT
5.5	WT

A3.4 References

Didychuk, A. L., E. J. Montemayor, T. J. Carrocci, A. T. DeLaitsch, S. E. Lucarelli *et al.*, 2017 Usb1 controls U6 snRNP assembly through evolutionarily divergent cyclic phosphodiesterase activities. Nat Commun 8**:** 497.

Giaever, G., A. M. Chu, L. Ni, C. Connelly, L. Riles *et al.*, 2002 Functional profiling of the Saccharomyces cerevisiae genome. Nature 418**:** 387-391.

Gietz, R. D., and R. A. Woods, 2002 Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. Methods Enzymol 350**:** 87-96.

Hilcenko, C., P. J. Simpson, A. J. Finch, F. R. Bowler, M. J. Churcher *et al.*, 2013 Aberrant 3' oligoadenylation of spliceosomal U6 small nuclear RNA in poikiloderma with neutropenia. Blood 121**:** 1028-1038.

Mroczek, S., J. Krwawicz, J. Kutner, M. Lazniewski, I. Kucinski *et al.*, 2012 C16orf57, a gene mutated in poikiloderma with neutropenia, encodes a putative phosphodiesterase responsible for the U6 snRNA 3' end modification. Genes Dev 26**:** 1911-1925.

Appendix 4: Genetic interactions between Usb1 and U6 pathway factors

A4.1 Overview

Genomic deletion of yeast *USB1* is lethal, but can be rescued by overexpression of U6 snRNA (HILCENKO *et al.* 2013). I hypothesized that other U6-associated proteins would have genetic interactions with Usb1. Deletion of the non-essential yeast gene *LHP1* does not rescue deletion of *USB1*, but is synthetically sick with mutations in yUsb1 that reduce activity. However, overexpression of both Prp24 and Lhp1 rescues deletion of yUsb1. Interestingly, Lhp1 overexpression increases the total amount of free U6, yet lowers the amount of U4 in U4/U6 and produces an unidentified slower migrating U4/U6 band on a native gel. These data are consistent with the hypothesis that Usb1 stabilizes U6 snRNA and that stability is dependent on protection of the 3' end, which can be compensated for by overexpression of other U6-binding factors.

A4.2 Materials and Methods

A4.2.1 Yeast strains and plasmids

The *lhp1Δ* strain was constructed by transforming BJ2168 (*MATa leu2 trp1 ura3-52 prb1-1122 pep4-2 prc1-407 gal2*) with linear DNA amplified from pUG6 (Euroscarf) containing the *kanMX* gene flanked by regions homologous to immediately up and downstream of the *LHP1* gene. The *usb1Δlhp1Δ, and usb1Δcus2Δ* strains were constructed by transforming yTJC0700 (*MATa leu2 trp1 ura3-52 prb1-1122 pep4-2 prc1-407 gal2 USB1::hph*) with linear DNA amplified using Herculase II Fusion DNA polymerase (Agilent Technologies) from pUG6 (Euroscarf) containing the *kanMX* gene flanked by regions homologous to immediately up and downstream of the *LHP1* or *CUS2* genes. Cells were grown on YPD for one day, then replica plated onto media containing 200 µg/mL G418. Individual colonies were screened and deletion of *LHP1* and *CUS2* was confirmed by PCR.

Overexpression plasmids containing the open reading frames (ORFs) for Prp24, Lhp1, and Lsm8 were generated by PCR amplification from yeast genomic DNA using primers that added an upstream BamHI and downstream SalI site, then cloned into BamHI/SalI-cut p425-GPD. All mutations, including addition of a single N-terminal HA tag, were generated using inverse PCR as previously described (DIDYCHUK *et al.* 2017).

A4.2.2 Yeast *USB1* complementation assays

yTJC0700 was transformed with pRS414 or p425-GPD plasmids using the lithium acetate method (GIETZ AND WOODS 2002). Growth phenotypes were assessed by spotting 10-fold serial dilutions (OD_{600} = 0.5, 0.05, 0.005) onto solid medium lacking leucine and containing 1 mg/mL 5-FOA (for p425-GPD plasmids) or media lacking tryptophan and containing 1 mg/mL 5-FOA (for pRS414 plasmids). Plates were incubated at 30°C for 3 days. For solution hybridization experiments, transformants were streaked from media lacking leucine to media lacking leucine and containing 1 mg/mL 5-FOA, then restreaked back onto media lacking leucine. The low copy

yUsb1 strain was transformed with p425-GPD empty vector and was not streaked onto media containing 5-FOA in order to retain the low copy pRS416-yUsb1 plasmid.

A4.2.3 Western blotting

Yeast were grown in selective media and total protein was isolated by trichloroacetic acid precipitation (CARROCCI *et al.* 2017). Protein concentration was normalized by A_{280} measurement and equivalent amounts were separated on a 4-20% Criterion TGX midi protein gel (200 V for 1 hour; Bio-Rad Laboratories) and subsequently transferred to a nitrocellulose membrane (30 minutes, 100 V, 4°C). The membrane was blocked using 5% (w/v) non-fat dry milk and probed using an HA antibody conjugated to horseradish peroxidase used at 1:5000 dilution (Sigma Aldrich 11667475001). Goat α-mouse HRP antibodies were used at 1:10,000 dilution (Bio-Rad Laboratories 1706515). Blots were developed using Clarity Western ECL substrate (Bio-Rad Laboratories) and imaged using an ImageQuant LAS 4000 Imager (GE Healthcare Life Sciences).

A4.2.4 Solution hybridization

Total RNA was prepared under non-denaturing conditions from 10 OD_{600} units of yeast grown in YPD media in logarithmic growth at 30°C as previously described (BURKE *et al.* 2015). Briefly, cells were pelleted, washed in RNA buffer (500 mM NaCl, 200 mM Tris pH 7.5, 100 mM EDTA), resuspended in 300 µL RNA buffer, 200 µL acid washed glass beads, and 300 µL 25:24:1 phenol/chloform/isoamyl-alcohol equilibrated against RNA buffer. Cells were vortexed in a Disruptor Genie for two minutes with rests on ice, then centrifuged at max speed to remove cell debris and glass beads. The aqueous phase was transferred to a new tube and extracted against 300 µL equilibrated phenol/chloroform/isoamyl-alcohol mixture, centrifuged, and ethanol precipitated.

DNA oligomers for solution hybridization are listed in **Table A4-1**. Oligomers were 5' end labeled with [γ-^{32}P] (Perkin Elmer) using OptiKinase (Affymetrix). Solution hybridization was performed as previously described (LI AND BROW 1993; BURKE *et al.* 2015). Total RNA was hybridized with radiolabeled U1-SH and U4-14B or U6-SH in 150 mM NaCl, 50 mM Tris pH 7.5,

1 mM EDTA for 15 minutes at 37°C, then incubated on ice for 5 minutes prior to separation on a 9% polyacrylamide gel (29:1 acrylamide:bisacrylamide, 89 mM Tris borate, 2 mM EDTA pH 8.0) for 7 hours at 175 V. Gels were dried under vacuum and exposed to a phosphorimager screen, imaged using a Typhoon FLA 9000 (GE Healthcare Life Sciences), and analyzed using ImageJ software.

Table A4-1. List of synthetic DNA oligonucleotides.

For strain construction and confirmation via PCR	
Lhp1_deletion_f	5'-GCTATGATGTATTTTTTTCTTTTTCCAAAAAATGGGTTCTATTTGGTTC TACTGGAACTAAAGTAGCATCTGCAAAGAAGTACCAGCTGAAGCTTCGTA CGCTGCAGG-3'
Lhp1_deletion_r	5'-GTACATATGTTAGTACATATATTCTTAATATGCTATGATAATGAGAT ACGAGAACCAGAAGAAACACAAGAACCGCGGCCGCATAGGCCACTA GTGGATCTGATATCACCTA-3'
Lhp1_delcon_f	5'-GCAATATCTCAAGATTTTTGAA-3'
Lhp1_delcon_r	5'-GGAGACAAATAATATAGTATTA-3'
Cus2_deletion_f	5'-GTAAAAGGCTGCAGACCCTTGAACGCACTACACTTTTCACTGTATTC CGTTACCAACATTACTGTTTGACCCAGCTGAAGCTTCGTACGCTGCAGG-3'
Cus2_deletion_r	5'-GTTTATTACATTGTGTTCACTTATTTTACTTAAGTACCAAAGATATGG GGGTAATACATTTTTTTACTGCACCGCGGCCGCATAGGCCACTAGTGGA TCTGATATCACCTA-3'
Cus2_delcon_f	5'-GGAAGAAACAGGACAACAGG-3'
Cus2_delcon_r	5'-TTCGATTCTCGTACCGCTC-3'
Primers to make p425-GPD-Prp24, Lhp1, Cus2, Lsm8, and Lsm8ΔC	
BamHI-Prp24-f	5'-ACGGGATCCATATGGAGTATGGACATCACGCTAGA-3'
SalI-Prp24-r	5'-GGCGTCGACCTACTCACCTAGAAACATCTTGCGAAAATC-3'
BamHI-Lhp1-f	5'-ACGGGATCCATATGTCTGAAAAACCACAACAAGAG-3'
SalI-Lhp1-r	5'-GGCGTCGACTCACTCCTTGTGCTCCTCATCGTCATCG-3'
BamHI-Cus2-f	5'-ACGGGATCCATATGGATGCTGATGAATTGGAATTGAAAGG-3'
SalI-Cus2-r	5'-GGCGTCGACCTATATAAGGTCATCTTCCACTTCGCTGT-3'
BamHI-Lsm8-f	5'-CGGGATCCATATGTCAGCCACCTTGAAAGACTAC-3'
SalI-Lsm8-r	5'-GGCGTCGACTTATTTTGTCTTTGATTCGTACAC-3'
SalI-Lsm8ΔC-r	5'-GGCGTCGACTTAGACCTTCTTTTCGTCTATAGGAGC-3'
p425-up-HA-r	5'-AGCGTAATCTGGTACGTCGTATGGGTACATATGGATCCACTAGTT CTAGAATCCGTCG-3'
Prp24-f	5'-ATGGAGTATGGACATCACGCTAGACCAGAT-3'
Lhp1-f	5'-ATGTCTGAAAAACCACAACAAGAGGAGCAAG-3'
Cus2-f	5'-ATGGATGCTGATGAATTGGAATTGAAAGG-3'

A4.3 Results and discussion

Based on the observation that Usb1 processing prevented Lhp1 binding, but Lsm2-8 could bind both before and after Usb1 processing (DIDYCHUK *et al.* 2017), I hypothesized that the need for Usb1 could be bypassed by removing Lhp1. Usb1 might only be required to ensure that Lhp1 is removed so that Lsm2-8 can bind and incorporate the snRNA into the spliceosome. To this end, I generated a strain in which the genomic copies of both *USB1* and *LHP1* are deleted. *LHP1* is non-essential in yeast and its deletion has no reported phenotype (YOO AND WOLIN 1994; LIN-MARQ AND CLARKSON 1995). However, deletion of *LHP1* does not rescue deletion of *USB1* (**Figure A4-1**). Deletion of *CUS2*, which interacts with Usb1 in a Y2H assay (**Appendix I**), also does not rescue deletion of *USB1* (data not shown).

I next wondered if a synthetic effect between *USB1* and *LHP1* existed, as previous work demonstrated that *LHP1* becomes essential when non-essential subunits of Lsm2-8 complex (Lsm6 and Lsm7) are deleted (PANNONE *et al.* 2001). I therefore wondered if deletion of *LHP1* would exacerbate mutations that impair the activity of yUsb1. To test this, I used two mutations that have slow growth phenotypes, yUsb1-F78A and Y55A (**Figure A4-1A**). Growth comparison of these strains relative to strains lacking *LHP1* show that deletion of *LHP1* has a synthetic growth (**Figure A4-1B**). Indeed, yeast expressing Usb1 F78A as their only copy are inviable in combination with *LHP1Δ* (**Figure A4-1B**). This result is consistent with previous results that suggest yeast minimally require either Lhp1 or Lsm2-8 to bind U6 and protect it from exonucleases within the cell. While either weakening Lsm2-8 binding (by reducing the level of 3' modified U6 by reducing Usb1 activity) or removing Lhp1 from the cell are not lethal by themselves, the combined effect is lethal in yeast.

Based on the previously published result that overexpression of U6 snRNA complements *USB1Δ* in yeast (HILCENKO *et al.* 2013), we tested if overexpression of U6 binding proteins could also rescue *USB1Δ*. We find that overexpression of HA-tagged Lhp1 and Prp24 rescues *USB1Δ* (**Figure A4-2A**). This result, along with the result that overexpression of U6 snRNA can

complement *USB1Δ* (HILCENKO et al. 2013), suggests that U6 processing by Usb1 can be bypassed either by increasing the levels of U6 RNA via overexpression or by stabilizing U6 RNA by increasing levels of U6 binding proteins. The mechanism by which Lhp1 overexpression rescues *USB1Δ* is presumably by protecting U6 from cellular exonucleases, thereby resulting in higher levels of intact U6 RNA. It is surprising that Prp24 overexpression also rescues *USB1Δ*, as Prp24 binds the asymmetric bulge of U6 RNA and there is no evidence to suggest that it interacts with the 3' tail (TEPLOVA *et al.* 2006; MONTEMAYOR *et al.* 2014). Prp24 may be protecting the "core" of U6 snRNA, which may be sufficient for yeast viability. Previous work has suggested that the 3' tail of U6 (nucleotides 94-112) increases splicing efficiency but is not essential for splicing *in vitro* (FABRIZIO et al. 1989; RYAN et al. 2002). However, further truncation of U6 is deleterious, as nucleotides 1-86 are insufficient to support yeast growth (BORDONNE AND GUTHRIE 1992). Alternatively, increased levels of Prp24 may compensate for *USB1Δ* by improving recruitment and co-binding of Lsm2-8 to U6 despite U6 lacking the native 3' phosphate end modification. Overexpression of full-length Lsm8, but not Lsm8 lacking the C-terminal residues 84-109 (ΔC), causes a slow growth phenotype (data not shown). Overexpression of Cus2 does not complement deletion of Usb1 (data not shown).

I next wondered how U4/U6 levels were affected in yeast lacking Usb1 and overexpressing Prp24, Lhp1, or hUsb1-H84F. I used solution hybridization to determine the levels of free U4 and U6 and of base paired U4/U6 (**Figure A4-2C**). When yUsb1 was on a low copy (pRS416) plasmid, the percentage of U4 in U4/U6 (60-69%) was similar to previous reports (**Figure A4-2D**) (BURKE *et al.* 2015). However, when yUsb1 was expressed from a high copy plasmid (p425-GPD), U4/U6 levels were lower despite only moderately reduced overall free U6 levels (**Figure A4-2D,E**). Yeast that lack yUsb1 and grow due to overexpression of Prp24 or hUsb1-H84F have significantly reduced levels of U4/U6 (5-7% U4 in U4/U6 for hUsb1-H84F, 10-16% U4 in U4/U6 for Prp24) and of free U6 (**Figure A4-2D,E**). Interestingly, cells that lack yUsb1 but have Lhp1 overexpressed have reduced U4 in U4/U6 (**Figure A4-2D**) yet have ~2.5-fold more

free U6 (**Figure A4-2E**). This could be a side effect of Lhp1's putative role in assisting RNA polymerase III with release and recycling (FRENCH *et al.* 2008).

Cells overexpressing Lhp1 also have an additional slower migrating U4/U6 band (**Figure A4-2C**, lanes 9-12). Since the RNA is isolated under conditions that should remove bound proteins (via phenol/chloroform/isoamyl alcohol extraction), it is unclear what this species could be. Since both it is present in lanes with both U4 and U6 probes, it could be U4/U6 traveling with a significantly different conformation or could be U4/U6 with Lhp1 that was not removed by phenol extraction.

It would be interesting to further examine U6 levels and effect on turnover in yeast lacking Usb1 and overexpressing U6 or other U6-binding proteins. Is the 3' tail of U6 protected from exonucleases in the absence of yUsb1, and how is the turnover rate of U6 affected? Is the 3' tail A-rich instead of U-rich, and is the 3' tail longer than in wild-type strains? Additionally, it would be interesting to examine snRNP composition in these yeast. Is Lsm2-8 recruited to U6 RNA despite it lacking the proper 3' end modification? Does Lhp1 remain bound to U6 in spliceosomal complexes where the 3' diol has not been processed to a 3' phosphate?

Figure A4-1. Genomic deletion of *LHP1* genetically interacts with mutations in *USB1*.

A) Mutations in yUsb1 (Y55A and F78A) cause slow growth phenotypes relative to wild-type Usb1. Loss of the wild-type yUsb1 plasmid is lethal.

B) Mutations as in (A) are exacerbated by genomic deletion of *LHP1*. Loss of the wild-type yUsb1 plasmid is lethal in a *usb1Δlhp1Δ* background.

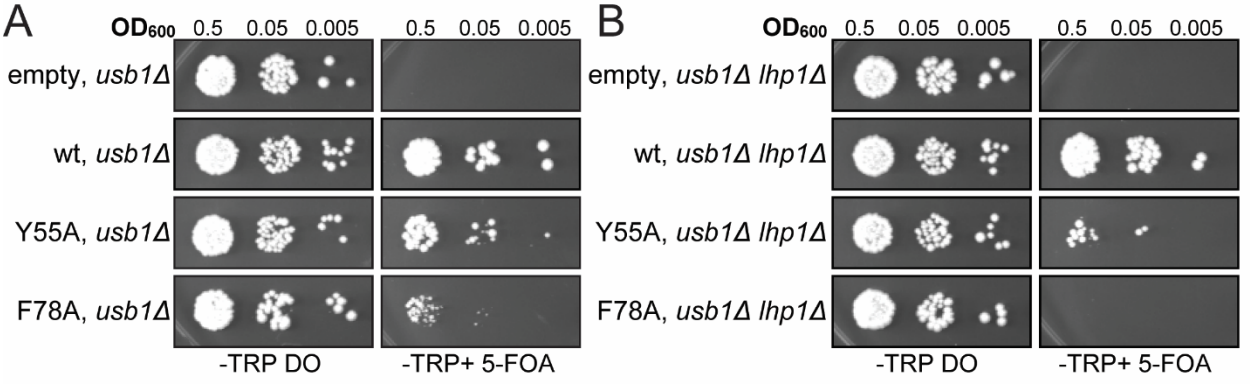


Figure A4-2. Usb1 genetically interacts with U6 binding proteins Lhp1 and Prp24.

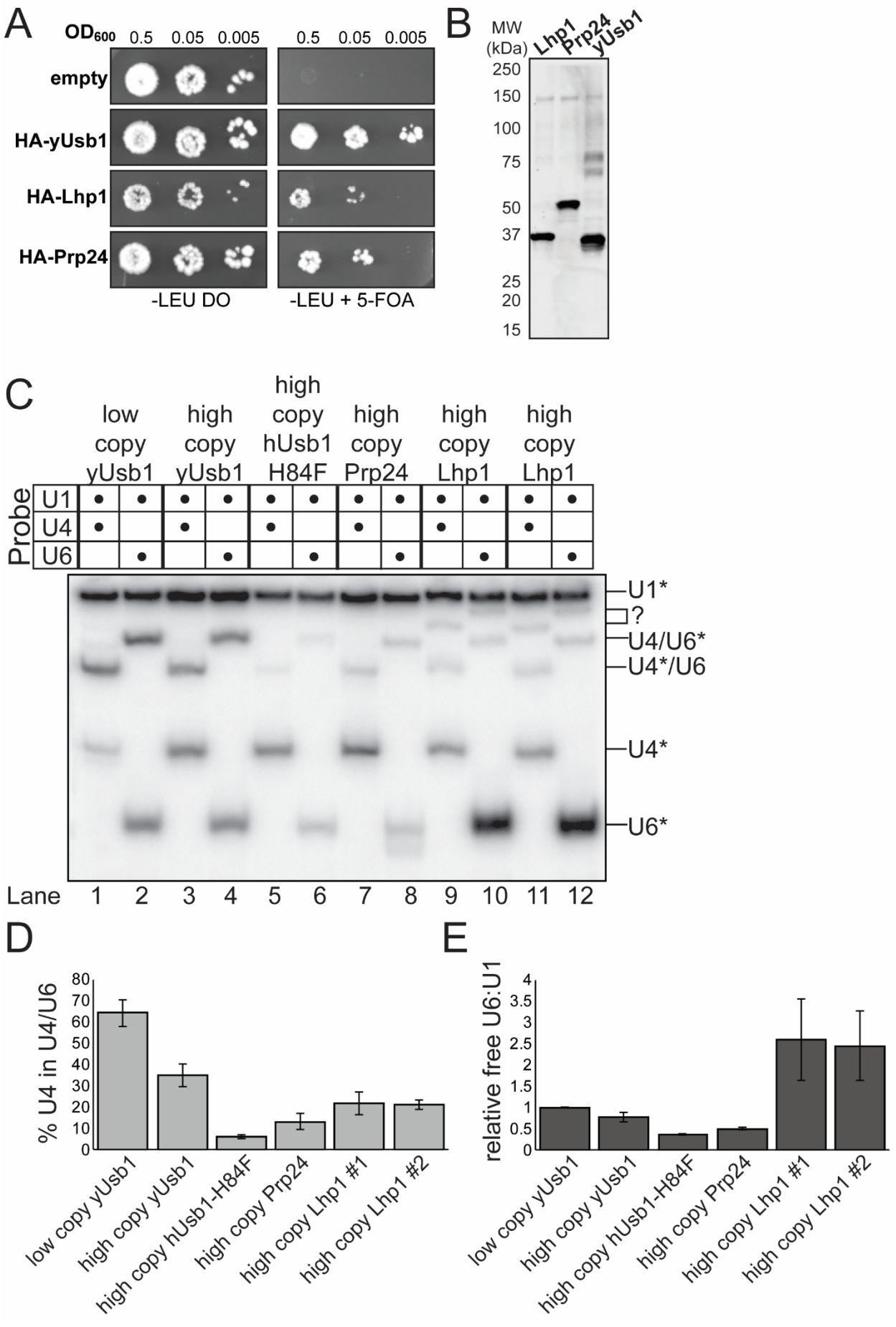
A) Overexpression of HA-tagged Lhp1 and Prp24 can complement deletion of Usb1.

B) SDS-PAGE and western blotting analysis of protein expression detected using a primary antibody to the genetically encoded N-terminal HA epitope tag for yeast strains in (A).

C) Solution hybridization of whole-cell RNA extracted under non-denaturing conditions to detect free U4, free U6, and base paired U4/U6. Lhp1 #1 and #2 are biological replicates.

D) Quantification of the relative amounts of percentage of U4 present in U4/U6 in two biological replicates. Error bars indicate range.

E) Total U6 levels normalized to the U1 in two biological replicates. Error bars indicate range.


A
OD600 0.5 0.05 0.005 0.5 0.05 0.005
empty
HA-yUsb1
HA-Lhp1
HA-Prp24
-LEU DO
-LEU + 5-FOA
B
MW (kDa)
Lhp1
Prp24
yUsb1
250
150
100
75
50
37
25
20
15
C
low copy yUsb1
high copy yUsb1
high copy hUsb1 H84F
high copy Prp24
high copy Lhp1
high copy Lhp1
Probe
U1
U4
U6
U1*
?
U4/U6*
U4*/U6
U4*
U6*
Lane 1 2 3 4 5 6 7 8 9 10 11 12
D
% U4 in U4/U6
low copy yUsb1
high copy yUsb1
high copy hUsb1-H84F
high copy Prp24
high copy Lhp1 #1
high copy Lhp1 #2
E
relative free U6:U1
low copy yUsb1
high copy yUsb1
high copy hUsb1-H84F
high copy Prp24
high copy Lhp1 #1
high copy Lhp1 #2

A4.4 References

Bordonne, R., and C. Guthrie, 1992 Human and human-yeast chimeric U6 snRNA genes identify structural elements required for expression in yeast. Nucleic Acids Res 20**:** 479-485.

Burke, J. E., S. E. Butcher and D. A. Brow, 2015 Spliceosome assembly in the absence of stable U4/U6 RNA pairing. RNA 21**:** 923-934.

Carrocci, T. J., D. M. Zoerner, J. C. Paulson and A. A. Hoskins, 2017 SF3b1 mutations associated with myelodysplastic syndromes alter the fidelity of branchsite selection in yeast. Nucleic Acids Res 45**:** 4837-4852.

Didychuk, A. L., E. J. Montemayor, T. J. Carrocci, A. T. DeLaitsch, S. E. Lucarelli *et al.*, 2017 Usb1 controls U6 snRNP assembly through evolutionarily divergent cyclic phosphodiesterase activities. Nat Commun 8**:** 497.

Fabrizio, P., D. S. McPheeters and J. Abelson, 1989 In vitro assembly of yeast U6 snRNP: a functional assay. Genes Dev 3**:** 2137-2150.

French, S. L., Y. N. Osheim, D. A. Schneider, M. L. Sikes, C. F. Fernandez *et al.*, 2008 Visual analysis of the yeast 5S rRNA gene transcriptome: regulation and role of La protein. Mol Cell Biol 28**:** 4576-4587.

Gietz, R. D., and R. A. Woods, 2002 Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. Methods Enzymol 350**:** 87-96.

Hilcenko, C., P. J. Simpson, A. J. Finch, F. R. Bowler, M. J. Churcher *et al.*, 2013 Aberrant 3' oligoadenylation of spliceosomal U6 small nuclear RNA in poikiloderma with neutropenia. Blood 121**:** 1028-1038.

Li, Z., and D. A. Brow, 1993 A rapid assay for quantitative detection of specific RNAs. Nucleic Acids Res 21**:** 4645-4646.

Lin-Marq, N., and S. G. Clarkson, 1995 A yeast RNA binding protein that resembles the human autoantigen La. J Mol Biol 245**:** 81-85.

Montemayor, E. J., E. C. Curran, H. H. Liao, K. L. Andrews, C. N. Treba *et al.*, 2014 Core structure of the U6 small nuclear ribonucleoprotein at 1.7-A resolution. Nat Struct Mol Biol 21**:** 544-551.

Pannone, B. K., S. D. Kim, D. A. Noe and S. L. Wolin, 2001 Multiple functional interactions between components of the Lsm2-Lsm8 complex, U6 snRNA, and the yeast La protein. Genetics 158**:** 187-196.

Ryan, D. E., S. W. Stevens and J. Abelson, 2002 The 5' and 3' domains of yeast U6 snRNA: Lsm proteins facilitate binding of Prp24 protein to the U6 telestem region. RNA 8**:** 1011-1033.

Teplova, M., Y. R. Yuan, A. T. Phan, L. Malinina, S. Ilin *et al.*, 2006 Structural basis for recognition and sequestration of UUU(OH) 3' temini of nascent RNA polymerase III transcripts by La, a rheumatic disease autoantigen. Mol Cell 21**:** 75-85.

Yoo, C. J., and S. L. Wolin, 1994 La proteins from Drosophila melanogaster and Saccharomyces cerevisiae: a yeast homolog of the La autoantigen is dispensable for growth. Mol Cell Biol 14**:** 5412-5424.

Appendix 5: NMR Spectroscopy of yUsb1

This work was performed with the help of Andy DeLaitsch, Ronnie Frederick, Marco Tonelli, and Gabriel Cornilescu, as well as helpful discussions with Katie Henzler-Wildman. Sam Butcher and Woonghee Lee assigned spectra.

A5.1 Overview

Comparison of the structures of yUsb1 and hUsb1 with a substrate analog revealed that structural changes in yUsb1 are likely to occur to accommodate the RNA substrate (**Appendix 2**). Determining the structure of yUsb1 with a substrate analog may improve our understanding of substrate recognition by Usb1. However, the crystallization conditions of yUsb1 contained 2 M $(NH_4)_2SO4$, preventing substrate RNA or analogs from easily being soaked into the crystal. The dynamics of loops flanking the active site of yUsb1 may be important for activity (**Appendix 2**). Thus, determining the structure and dynamic properties of these loops in solution is essential. Additionally, unambiguous determination of the catalytic mechanism and assignment of the catalytic acid and base is most straightforwardly done via NMR by calculating the p*K*a of the active site histidines. I demonstrate that yUsb1 is amenable to structure determination via NMR, that chemical shift perturbations occur upon addition of RNA substrate, and that resonances for the active site histidine imidazole nitrogens can be observed and display pH sensitive chemical shifts.

A5.2 Materials and Methods

A5.2.1 Cloning, protein production, and purification

Recombinant wild-type yUsb1 71-290 and mutants were produced as previously described (DIDYCHUK *et al.* 2017). Briefly, mutations were introduced into the pET28b-HT-yUsb1 1-290 via inverse PCR and ligation (**Table A5-1**). Protein was produced in *E. coli* BL21 STAR (DE3) pLysS (Invitrogen) at 37°C to an OD_{600} of ~0.6-0.8, induced by addition of 1 mM IPTG, and grown for 3-4 hours at 37°C. Cells were collected by centrifugation, resuspended in IMAC buffer (500 mM NaCl, 50 mM HEPES acid, 50 mM sodium HEPES base, 15 mM imidazole base, 10% glycerol, 1 mM TCEP-HCl, pH ~7.5) containing DNase I and lysozyme, sonicated, and centrifuged to remove insoluble material. Protein was purified by Ni-NTA agarose chromatography by step elution with IMAC buffer supplemented with 500 mM imidazole, dialyzed overnight at 4°C into IEX buffer (100 mM NaCl, 10 mM HEPES acid, 10 mM sodium HEPES base, 10% glycerol, 1 mM TCEP, pH ~ 6.2) with 1 mg TEV protease. Precipitated protein was removed by centrifugation, and Usb1 was further purified via cation-exchange chromatography (HiTrap S, GE Healthcare) in IEX buffer with gradient elution against IEX buffer containing 2 M NaCl.

^{15}N-labeled Usb1 71-290 was produced in *E. coli* BL21 STAR (DE3) pLysS (Invitrogen). A 3 mL LB culture containing 50 µg/mL kanamycin, 35 µg/mL chloramphenicol, and 1% glucose was grown at 37°C for 3 hours, then 1 mL of this culture was added to 50 mL of non-inducing MDAG media (2 mM $MgSO_4$, 0.2 mL of 5000x metals mix (BLOMMEL *et al.* 2007), 0.5% glucose, 0.25% aspartate, 1x NPS (50 mM NH_4Cl, 5 mM Na_2HPO_4, 25 mM KH_2PO_4, 5 mM Na_2SO_4), 0.1 mg/mL methionine, 0.1 mg/mL 17A (no C, Y, M) mix, 1 mL of 1000x B12 vitamin cocktail (BLOMMEL *et al.* 2007) containing 50 µg/mL kanamycin and 35 µg/mL chloramphenicol and grown at 25°C for 24 hours. The next day, 20 mL of this culture was diluted into 1 L of M9-^{15}N media (1x M9 salts (50 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.6 mM NaCl), 0.2 mL of 5000x metals mix, 1 mL of 1000x B12 vitamin cocktail, 2 mM $MgSO_4$, 30 µg/mL thiamine, 4 g/L glucose, 1 g/L $^{15}NH_4Cl$, 0.1 mM $CaCl_2$) resulting in an OD_{600} of ~ 0.15. The culture was grown to an OD_{600} of 0.6 then induced

with 1 mM IPTG and grown for 37°C for 4 hours and purified as above. Double labeled $^{13}C^{15}N$ samples were produced in the same way, with the exception that the M9-$^{15}N^{13}C$ contained 2 g/L ^{13}C-glucose. Triple labelled samples were grown in M9-$^{15}N^{13}C^{2}H$ media (MORRISON *et al.* 2015) and purified as described above. NMR samples were dialyzed and concentrated into 50 mM NaCl, 20 mM bis-tris, 10 mM HCl, 1 mM TCEP-HCl pH ~6.2.

A5.2.2 *In vitro* exonuclease activity assays

Exoribonuclease assays were performed as previously described (DIDYCHUK *et al.* 2017). Time course assays used 500 nM yUsb1 and 100 nM FAM-U6 95-112(UUU)-3'OH substrate in 50 mM NaCl, 50 mM KCl, 20 mM bis-tris, 10 mM HCl, 1 mM TCEP-HCl, 10% surcrose, 5% glycerol, 0.5 mM EDTA, 0.005% Triton X-100, 0.1 mg/mL BSA. Reactions were quenched by the addition of an equal volume of 100% formamide, resolved on a 20% 19:1 acrylamide:bis-acrylamide PAGE gel containing 8 M urea, 89 mM Tris borate, and 2 mM EDTA. Gels were imaged directly through low fluorescence glass plates (CBS Scientific) using a Typhoon FLA 9000 (GE Healthcare Life Sciences). For time course experiments, the percentage of the substrate processed was calculated using the ratio of product(s) to total signal at 0, 5, 15, 30, and 60 minute time points. Data were fit to a one-phase exponential association equation $Y=Y_0+(Y_{max} - Y_0)*(1-e^{-kx})$, where Y is the % of the substrate processed, x is the time, and *k* is the rate constant (GraphPad Prism 4).

A5.2.3 Titration with RNA

A short oligouridine substrate containing a non-cleavable internal deoxyuridine (5'-UdUU-3') was purchased from Dharmacon and resuspended to a final concentration of 10 mM in 50 mM NaCl, 20 mM bis-tris, 10 mM HCl, 1 mM TCEP-HCl, pH 6.4. This RNA was added to 633 µM ^{15}N-labeled Usb1 71-290 in 50 mM NaCl, 20 mM bis-tris, 10 mM HCl, 1 mM TCEP-HCl, pH 6.3, 50 µM DSS, 10% D_2O in 0.25x, 0.5x, 1x, 1.5x, 2x, and 3x molar ratios.

U6 96-112 with two additional 3' uridine residues and an additional 5' guanine (U6 G96-112UU) was synthesized via *in vitro* transcription from a double stranded synthetic DNA

oligonucleotide template (Integrated DNA Technologies) using recombinant His_6-tagged T7 RNA polymerase in 40 mM Tris-Cl pH 8.0, 1 mM spermidine, 0.01% Triton X-100, 38 mM $MgCl_2$, 5 mM dithiothreitol, 4.1 mM ATP, 1.4 mM CTP, 5.5 mM GTP, 15.1 mM UTP, and 1 µM dsDNA template. U6 G96-112U was purified via denaturing PAGE (20% 19:1 acrylamide:bis-acrylamide containing 8 M urea, 89 mM Tris borate, and 2 mM EDTA), extracted from the gel into 0.3 M NaOAc, 1 mM EDTA overnight, diluting 4-fold with water, then purified via anion exchange chromatography (HiTrap Q FF, GE Healthcare) using 100 mM NaCl, 20 mM bis-tris, 10 mM HCl, 1 mM EDTA pH~6.5, and step eluted with buffer containing 2 M NaCl. U6 G96-112U was concentrated and buffer exchanged into 50 mM NaCl, 20 mM bis-tris, 10 mM HCl, 1 mM TCEP-HCl pH~6.2 using a 3 kDa cutoff spin filter (Amicon) to a final concentration of 1280 µM. A spectrum for the apo ^{15}N-labeled Usb1 71-290 sample was collected, then U6 G96-112U was added for final concentrations of ~180 µM Usb1 and ~210 µM U6 G96-112UU.

A5.2.4 pH titration

^{15}N-labeled Usb1 71-290 was dialyzed into 100 mM NaCl, 20 mM bis-tris, 1 mM TCEP-HCl where the pH had been adjusted with HCl to 5.8 or 7.2. Samples containing ~14 mg/mL ^{15}N-Usb1 71-290, 92 mM NaCl, 18 mM bis-tris, 0.92 mM TCEP-HCl, 8% D_2O, 15 µM DSS were prepared in the low and high pH buffers. The pH of the sample at 35°C was measured directly using a Biotrode pH Sensor (Hamilton Robotics). After spectra for the low (final pH 5.65) and high (final pH 7.20) samples had been recorded, 70 µL of each sample was removed and added to the other sample to produce samples of intermediate pH (6.10 and 6.85). The process was repeated to produce samples at pH 6.25 and 6.55. The pH of the samples was measured before and after each experiment, and little drift in pH (< 0.03 pH units) was observed.

A5.2.5 NMR spectroscopy

All NMR data were collected at the National Magnetic Resonance Facility at Madison (NMRFAM). Spectra were collected on 600 or 900 MHz Bruker Avance spectrometers. Data collection at 35°C improved peak dispersion. Data was processed with NMRPipe and NMRFAM-

SPARKY (DELAGLIO *et al.* 1995; LEE *et al.* 2015) was used it visualize processed data. 2D ^{1}H-^{15}N TROSY-HSQC spectra were used for RNA titration experiments and 2D ^{1}H-^{15}N HSQC experiments were used to monitor histidine chemical shifts. 2D ^{1}H-^{15}N HSQC and 3D HNCACAB, CBCA(CO)NH, HNCA, HN(CO)CA, HNCO, and NOESY ^{15}N-HSQC experiments have been collected to be used to assign backbone resonances.

Table A5-1. List of synthetic DNA oligonucleotides.

Primer name	Primer sequence
H133N-f	5'-AACGTTTCCTTAACTCGATCGTTGTTGTTTGAAAC-3'
H133-r	5'-TAAAGGTTTGGGGGCTCCCAAGTGTGAGATGAAC-3'
H231N-f	5'-AACGTTTCAATTGCCATCGCTAGCAACCCTTCAAAAG-3'
H231-r	5'-TAAATTTTGCCTACTTACAATTAAATCCTGATAG-3'
H90N-f	5'-AATCGACAACTTCAGAAAATCATATGTAAATATAAAG-3'
H90-r	5'-AATCGCTGGAGTCGGACGCCACTCAAAATA-3'
H125N-f	5'-AATTTGGGAGCCCCCAAACCTTTACACGTTTCC-3'
H125-r	5'-TGAGATGAACAACGGGTCAAAATCCACCAG-3'
H149-f	5'-AATGTTTTCATTCAAGAAATGCGTAATGGTTTAC-3'
H149-r	5'-TCTTTGTTCTTCAGTTTCAAACAACAACGA-3'
H283N-f_pET	5'-AATTCCATTAGAATCCCTTTTAACTAACTCGAG-3'
H283-r	5'-ACGGTTTTCATCACAATAGATAGAGTTAAC-3'
U6 G96-112(UU) template	5'-AAAAAACGAAATAAATCTCCTATAGTGAGTCGTATTAGAA-3'
U6 G96-112(UU) complement	5'-TTCTAATACGACTCACTATAGGAGATTTATTTCGTTTTTT-3'
UdUU (Dharmacon)	5'-rUdUrU-3'

A5.3 Results and discussion

A5.3.1 Yeast Usb1 is amenable to structure determination via NMR

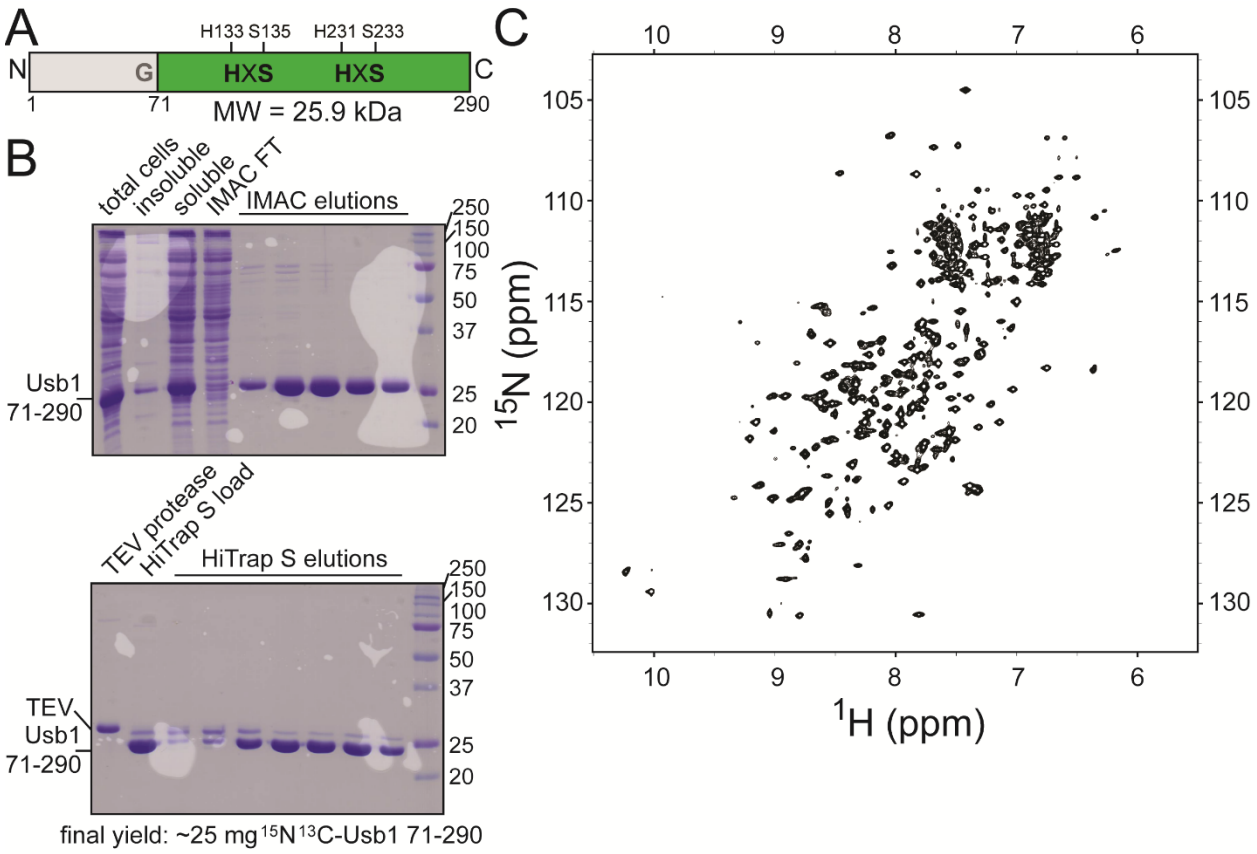
Yeast Usb1 is a 33.7 kDa protein that consists of a 25.9 kDa catalytic domain (residues 71-290) and an N-terminal domain that is predicted to be unstructured (DIDYCHUK *et al.* 2017) (**Figure A5-1A**). The structure of the catalytic domain has been determined by x-ray crystallography and exhibits a typical 2H phosphodiesterase fold (DIDYCHUK *et al.* 2017). However, structural changes likely occur in order to accommodate the RNA substrate (DIDYCHUK *et al.* 2017) (**Appendix 2**). Determining the solution structure of the catalytic domain of yUsb1 may yield useful information and can be used to determine the co-structure of yUsb1 with an RNA substrate. The catalytic domain of yUsb1 is amenable for NMR, as its size is below the ~35 kDa practical limit for structure determination and because the protein is easy to express and purify (**Figure A5-1B**). The 2D ^{1}H-^{15}N TROSY spectrum of yUsb1 71-290 shows that the peaks are well-dispersed, indicating that structure determination by NMR should be feasible (**Figure A5-1C**). Backbone assignments for yUsb1 are currently ongoing. Once the backbone is assigned, the 2D ^{1}H-^{15}N TROSY of full-length yUsb1 could be measured to determine if the N-terminal domain is unstructured as predicted.

Figure A5-1. yUsb1 is amenable to structure determination via NMR.

A) Primary structure of yUsb1 NMR construct, containing residues 71-290 and an N-terminal glycine left from TEV cleavage. The final product has a molecular weight of 25.9 kDa.

B) yUsb1 can be efficiently produced in labeled media. Top: initial expression and purification with IMAC. Bottom: second step IEX purification to remove TEV protease. Final yield from 1 L of M9-$^{15}N^{13}C$ media is ~25 mg of protein.

C) The 2D-^{1}H-^{15}N TROSY-HSQC spectrum at 900 MHz shows well-dispersed peaks, suggesting that the protein is folded and amenable to structure determination via NMR.



A5.3.2 Binding to RNA can be observed via NMR

Chemical shift perturbation can be used to characterize the affinity of enzyme-substrate interactions as well as map the interaction site on an enzyme (WILLIAMSON 2013). When a short non-cleavable RNA substrate analogue containing an internal deoxyuridine (UdUU) is titrated into yUsb1, several peaks shift (**Figure A5-2A**). Four of these peaks lie in a region that contains the tentative assignments for the active site serines S135 and S233 (**Figure A5-2B**). These peaks shift as more substrate is added (rather than appearing at a new chemical shift), suggesting that yUsb1 and the substrate analogue are in the fast exchange regime and k_{off} is fast, consistent with a weak affinity or high (micromolar range) K_d.

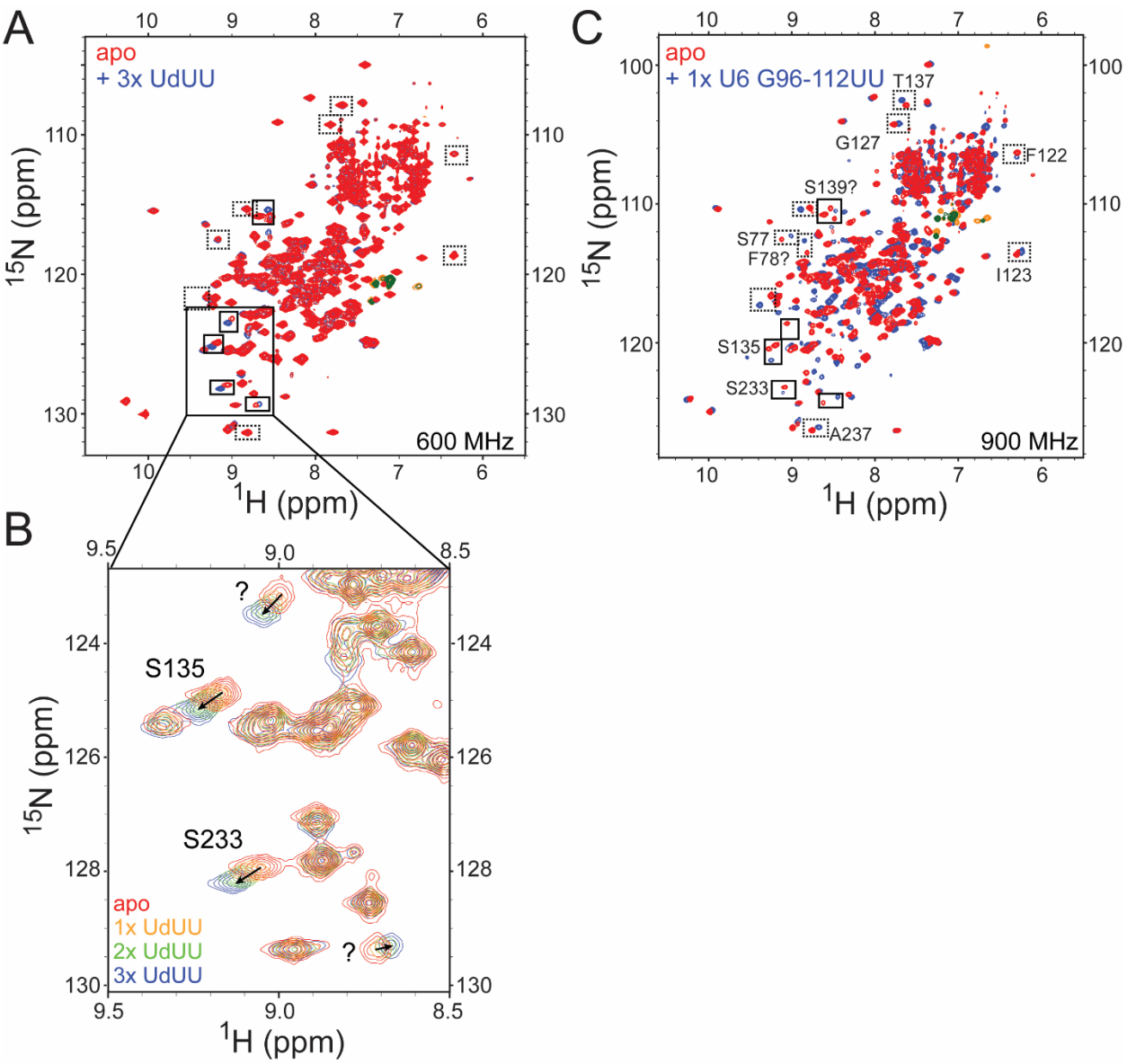
Interestingly, when a 20 nucleotide long RNA substrate is used, many additional peaks are shifted (**Figure A5-2C**). This indicates that additional contacts between yUsb1 and an RNA substrate occur, which may be important for substrate recognition. Many of these residues have been tentatively assigned, and F78 (which plays a role in catalytic activity; see **Appendix 2**) and F122, I123, and G127 (in Loop 2; see **Appendix 2**) have chemical shift perturbations, suggesting that they play a role in substrate recognition or undergo conformational changes upon substrate binding.

Figure A5-2. Titration with RNA reveals residues that interact with the RNA substrate.

A) 2D-^{1}H-^{15}N TROSY-HSQC spectra of yUsb1 before (red) and after addition of 3-fold molar excess of a short non-cleavable RNA oligonucleotide (UdUU). Residues boxed in solid lines shift upon addition of the short RNA oligonucleotide, while residues boxed in dotted lines only shift upon addition of a longer RNA (shown in (C)). Tentative assignments are indicated.

B) Zoom of the spectrum in (A) where several peaks undergo chemical shift perturbation. Tentative assignments are indicated.

C) 2D-^{1}H-^{15}N TROSY-HSQC spectra of yUsb1 before (red) and after addition of an equimolar amount of an RNA oligonucleotide containing U6 nucleotides G96-112 with two additional uridine nucleotides. Solid boxed lines indicate residues that move upon addition of either the UdUU RNA or the longer U6 RNA, while dotted lines indicate residues that shift only in the presence of the longer substrate. Tentative assignments are indicated.


A
apo
+ 3x UdUU
600 MHz
^{15}N (ppm)
^{1}H (ppm)
B
S135
S233
?
?
apo
1x UdUU
2x UdUU
3x UdUU
C
apo
+ 1x U6 G96-112UU
T137
G127
F122
S139?
S77
F78?
I123
S135
S233
A237
900 MHz

A5.3.3 The active site histidines can be observed and are pH sensitive

The reaction mechanism for Usb1 likely involves the active site histidines acting as the catalytic acid and base, similar to the mechanism of RNase A. In one (exonucleolytic) step, the catalytic base abstracts a proton from the 2' hydroxyl of the substrate, which then attacks the phosphate to form a 2',3'-cyclic phosphate and release the terminal nucleotide. Based on the structure of hUsb1 with 5'UMP (DIDYCHUK *et al.* 2017), H133 likely acts as the catalytic base and H231 acts as the catalytic acid in first step chemistry. These assignments can be confirmed by determining the p*K*a of each histidine, as the catalytic base should have a p*K*a ~1.5 pH units lower than catalytic acid and should have a p*K*a lower than the pH optimum of 6.5 (NIELSEN AND MCCAMMON 2003), although the mechanistically similar RNase A has active site histidine p*K*a's of 5.8 and 6.2 (THOMPSON AND RAINES 1994).

Histidine side chains have unique chemical shifts that can be measured using 2D ^{1}H-^{15}N HSQC experiments, with ^{15}N chemical shifts in the 160-220 ppm range (compared to the 100-140 ppm window for backbone amide groups). Histidine side chains can have 1-4 peaks (Nε2-Hε1, Nε2-Hδ2, Nδ1-Hε1, or Nδ1-Hδ2), depending upon its environment, p*K*a, and tautomerization state. yUsb1 has six histidine residues: the active site histidines (H133 and H135), three solvent-exposed histidines (H125, H149, and H283), and one buried histidine (H90). Five histidine side chains can be observed (**Figure A5-3A**). These peaks are highly sensitive to pH (**Figure A5-3B**). Unfortunately, many of the peaks are weak and only two side chains can be titrated over the pH range 5.65 - 7.20 (**Figure A5-3B**). The histidine side chain Nδ1 or Nε2 and nitrogen resonances have not yet been assigned.

In order to determine which peaks come from the active site histidines, mutants of the two active site histidines were generated (H133N and H231N; **Figure A5-3C,D**). Interpretation of these spectra was inconclusive, as both mutants lacked several of the original peaks. Additionally, mutating one active site histidine to asparagine may change the p*K*a of the other active site

histidine and therefore change its chemical shift(s). However, at 7.0 ppm, H133N has a peak while H231N lacks a peak, suggesting that this might be assigned to H231.

In order to simplify the spectra, the non-active site histidines were mutated to asparagines. These histidines lie on one face of yUsb1 and are not near the active site (**Figure A5-3E**). Mutating all four histidines to asparagine ("4HIS") has marginal effect on yUsb1 activity *in vitro* (**Figure A5-3F**). The 2D ^{1}H-^{15}N HSQC of the 4HIS mutant is greatly reduced in complexity and peaks from two side chains can be observed (**Figure A5-3G**). One of these sets of peaks may be the titratable histidine "3", while the other histidine may be the weak set of peaks corresponding to histidine "4" (**Figure A5-3A**). Based on the missing peak in **Figure A5-3D**, the second weak set of peaks may be assigned to histidine 231 and the first set of peaks assigned to histidine 133 (**Figure A5-3G**). The pH titration shown in **Figure A5-3B** could be repeated using the 4HIS mutant and extended over a larger pH range in order to calculate p*K*a's, and the two histidines could be assigned by measuring the spectra of 4HIS/H133N and 4HIS/H231N mutants.

The unambiguous assignment of catalytic acid and base will provide us insight into the reaction mechanism of yUsb1. Measuring the p*K*a of active site histidines in yUsb1 and hUsb1 will help us understand why the two homologs have different pH optimums and different propensity for second step cyclic phosphodiesterase activity. The p*K*a of the active site histidines could be measured in the context of a F78H or other mutants to determine how residues flanking the active site affect activity and p*K*a's.

Figure A5-3. Active site histidines are pH sensitive.

A) 2D-^{1}H-^{15}N HSQC spectrum showing the histidine side chains for yUsb1 at pH 5.65. Histidine side chains can have 1-4 peaks per side chains depending on protonation state and dynamic exchange. yUsb1 has six histidine residues (two in the active site, three solvent exposed, and one buried). Five side chains can be identified in the spectra at pH 5.65.

B) 2D-^{1}H-^{15}N HSQC spectra at pH 5.65, 6.10, 6.25, 6.55, 6.85, and 7.20. Residue 1 is only observable at pH 5.65. Residues 2 and 4 can be observed at pH 5.65, 6.10, and 6.25. The chemical shifts for residue 5 do not significantly change over the course of the titration. Residue 3 is highly titratable and can be observed over the course of the experiment.

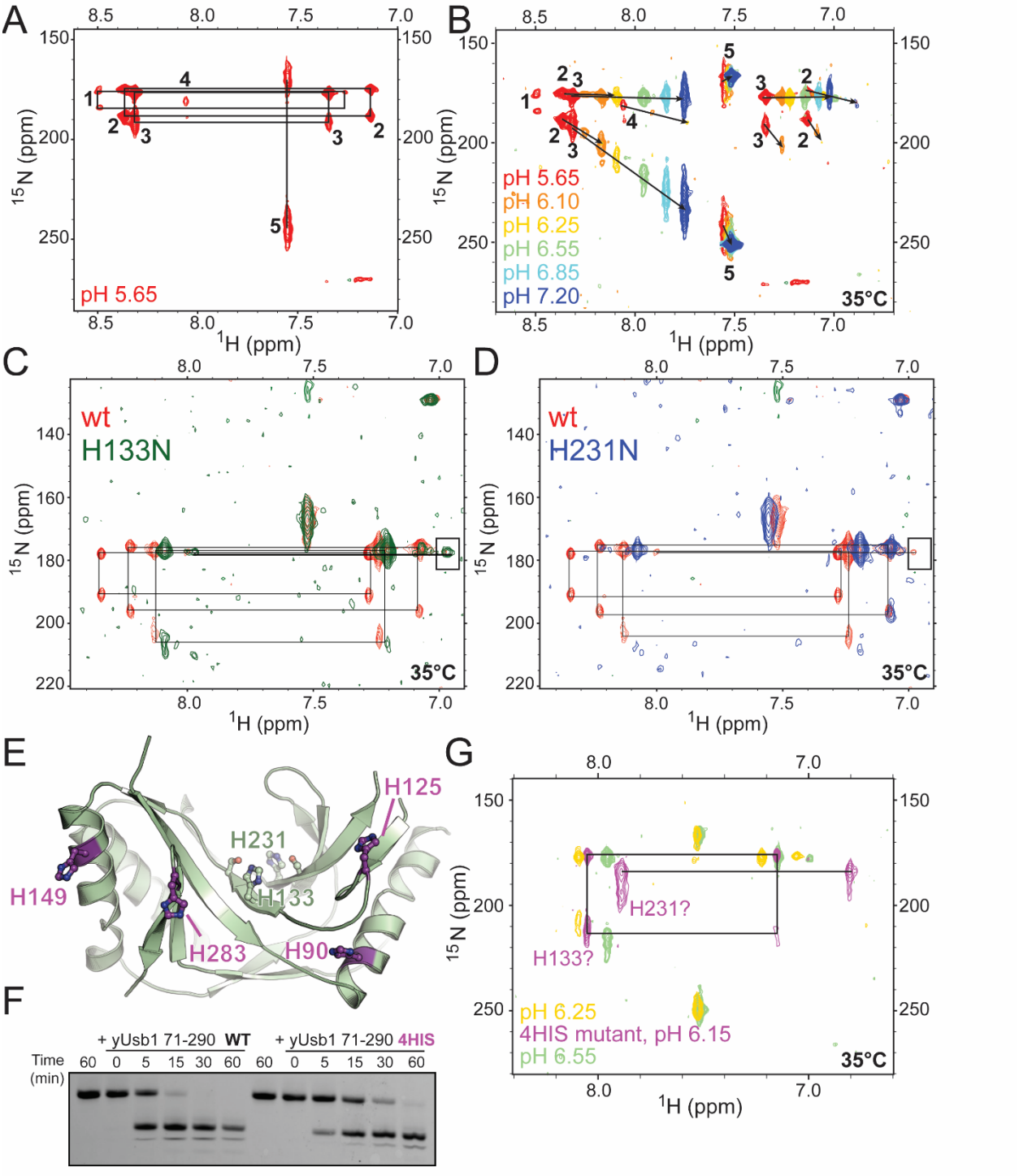
C) 2D-^{1}H-^{15}N HSQC spectrum of yUsb1-H133N (green) in comparison to wt yUsb1 (red).

D) 2D-^{1}H-^{15}N HSQC spectrum of yUsb1-H231N (green) in comparison to wt yUsb1 (red).

E) yUsb1 contains six histidine residues, shown in sticks and spheres. The active site histidines H133 and H231 and active site serines are shown in green. Of the remaining histidine residues (purple), three (H125, H149, and H283) are solvent exposed, while residue H90 is buried.

F) Mutation of the four non-active site histidines to asparagine (4HIS) has little effect on the catalytic activity of yUsb1 *in vitro*.

G) 2D-^{1}H-^{15}N HSQC spectrum of yUsb1-4HIS (purple) in comparison to wt yUsb1 at pH 6.25 (yellow) and pH 6.55 (green).


A
B
C
D
E
F
G
pH 5.65
pH 6.10
pH 6.25
pH 6.55
pH 6.85
pH 7.20
35°C
wt
H133N
H231N
H125
H231
H133
H149
H283
H90
+ yUsb1 71-290 WT
+ yUsb1 71-290 4HIS
Time (min)
60 0 5 15 30 60 60 0 5 15 30 60
H231?
H133?
pH 6.25
4HIS mutant, pH 6.15
pH 6.55
^{15}N (ppm)
^{1}H (ppm)

A5.4 References

Blommel, P. G., K. J. Becker, P. Duvnjak and B. G. Fox, 2007 Enhanced bacterial protein expression during auto-induction obtained by alteration of lac repressor dosage and medium composition. Biotechnol Prog 23**:** 585-598.

Delaglio, F., S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer *et al.*, 1995 NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6**:** 277-293.

Didychuk, A. L., E. J. Montemayor, T. J. Carrocci, A. T. DeLaitsch, S. E. Lucarelli *et al.*, 2017 Usb1 controls U6 snRNP assembly through evolutionarily divergent cyclic phosphodiesterase activities. Nat Commun 8**:** 497.

Lee, W., M. Tonelli and J. L. Markley, 2015 NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. Bioinformatics 31**:** 1325-1327.

Morrison, E. A., A. E. Robinson, Y. Liu and K. A. Henzler-Wildman, 2015 Asymmetric protonation of EmrE. J Gen Physiol 146**:** 445-461.

Nielsen, J. E., and J. A. McCammon, 2003 Calculating pKa values in enzyme active sites. Protein Sci 12**:** 1894-1901.

Thompson, J. E., and R. T. Raines, 1994 Value of general Acid-base catalysis to ribonuclease a. J Am Chem Soc 116**:** 5467-5468.

Williamson, M. P., 2013 Using chemical shift perturbation to characterise ligand binding. Prog Nucl Magn Reson Spectrosc 73**:** 1-16.

Appendix 6: U6 snRNP properties from the yeast *Kluveromyces lactis*

This work was started as a collaboration with Tucker Carrocci with the help of Andrew DeLaitsch.

A6.1 Overview

The U6 snRNP exhibits a unique interlocked topology between U6 snRNA and its binding partner Prp24. Understanding how this unprecedented topology is formed and resolved is critical to understanding how Prp24, an ATP-independent RNA chaperone, is able to recognize U6 snRNA and catalyze its annealing to U4 snRNA to allow assembly into the spliceosome. We show that a close homolog of Prp24 from *K. lactis* can functionally replace Prp24 from *S. cerevisiae*. The *K. lactis* Prp24 (KlPrp24) behaves similarly to *S. cerevisiae* (ScPrp24) *in vitro*. KlPrp24 lacking its native cysteines retains U6 binding *in vitro*. This cysteine-less construct could be used to install probes at pairs of locations in Prp24 to measure distance and dynamics within the protein as it binds U6 and anneals U6 to U4. Study of homologs of the U6 snRNP may reveal common determinants of high-affinity RNA binding and the interlocked topology, as well as give insight into the mechanism of U4/U6 annealing.

A6.2 Materials and Methods

A6.2.1 Cloning and protein expression

The coding sequence for Prp24 from *Kluveromyces lactis* was codon optimized for expression in *E. coli* and contained substitutions for its two native cysteines (C60L and C379S). The sequence of the synthetic gene is shown in **Table A6-1**. This gene was cloned into NdeI/BamHI-cut pET28b vectors containing an N-terminal hexahistidine and TEV protease cleavage site ("HT") or containing an N-terminal hexahistidine and TEV protease cleavage site followed by a biotin acceptor peptide sequence ("HT-AP"). Mutations were introduced via inverse PCR with Phusion DNA polymerase with GC buffer, Dpn1 treatment, self-ligation using T4 PNK and T4 DNA ligase, and transformation into *E. coli* 5α competent cells (all New England Biolabs). Resulting clones were transformed into *E. coli* Rosetta 2 (DE3) cells (Novagen) and grown in LB at 37°C to an OD_{600} of 0.6-0.8, placed on ice for 30 minutes, then induced by addition of 1 mM IPTG and growth at 16°C for 20 hours. Cells were collected by centrifugation, resuspended in IMAC buffer (500 mM NaCl, 50 mM HEPES acid, 50 mM sodium HEPES base, 15 mM imidazole base, 10% glycerol, 1 mM TCEP-HCl) supplemented with DNase I and lysozyme. Cells were lysed via sonication and insoluble material was removed by centrifugation. KlPrp24 was purified by Ni-NTA agarose chromatography with step elution using IMAC buffer supplemented with 500 mM imidazole. Eluate and 1 mg TEV protease were dialyzed at 4°C overnight into IEX buffer (100 mM NaCl, 10 mM HEPES acid, 10 mM HEPES base, 10% glycerol, 1 mM TCEP-HCl, pH ~7.0), precipitated protein was removed by centrifugation, then KlPrp24 was purified via cation-exchange chromatography (HiTrap S, GE Healthcare) in IEX buffer with gradient elution against IEX buffer containing 2 M NaCl. All protein samples were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis to assess purity.

A6.2.2 RNA production

Wild-type and A62G or A79G-stabilized U6 RNA containing nucleotides 30-99 with two additional cytidines (U6 30-99CC) (MONTEMAYOR *et al.* 2014; MONTEMAYOR *et al.* 2017) was *in*

vitro transcribed from synthetic DNA oligonucleotides (Integrated DNA Technologies) (**Table A6-2**) using recombinant 6HIS-tagged T7 RNA polymerase. *In vitro* transcription reactions contained 1 µM template dsDNA, stoichiometric NTPs (7.2 mM ATP, 4.2 mM CTP, 4.2 mM GTP, 4.4 mM UTP), 40 mM Tris-Cl pH 8.0, 1 mM spermidine, 0.01% Triton X-100, 38 mM $MgCl_2$, and 5 mM dithiothreitol. The transcription reaction was quenched by addition of 45 mM EDTA and RNAs were purified via denaturing 10% 19:1 acrylamide:bis-acrylamide gel electrophoresis containing 8 M urea, 89 mM Tris borate, 2 mM EDTA. RNA was visualized by UV shadowing and extracted from the gel via passive diffusion into 0.3 M sodium acetate, 1 mM EDTA, pH 5.2. RNA was further purified via anion-exchange (HiTrap Q, GE Healthcare) using IEX buffer (100 mM NaCl, 20 mM bis-tris, 10 mM HCl, 1 mM TCEP-HCl, pH ~6.2), with step elution using IEX buffer containing 2 M NaCl. RNA was concentrated using a 10 kDa spin filter and concentration was estimated using the A_{260} and calculated extinction coefficients.

Full-length U6 RNA binding assays used a 5'-Cy3 fluorescently labeled RNA containing U6 nucleotides 1-112 that was produced via splinted ligation using a 5'-Cy3 U6 1-12 RNA oligonucleotide (Integrated DNA Technologies) and U6 13-112 that was produced via *in vitro* transcription from a modified plasmid containing a 5' hammerhead ribozyme and a 3' HDV ribozyme. The U6 13-112 RNA was phosphorylated with ATP and T4 PNK, which also removed the 3' cyclic phosphate produced by 3' HDV cleavage. The Cy3-U6 1-12 and U6 13-112 RNAs were ligated using T4 RNA ligase 2 (New England Biolabs) at 37°C for 2 hours using a DNA splint that was complementary to U6 nucleotides 1-30, with a Cy3-U6:splint:U6 13-112 ratio of 2:1.5:1. The ligation product was purified by urea PAGE and ion exchange as above.

Full-length 3'-Cy5 fluorescently labeled U4 (nucleotides 1-160) was produced via splinted ligation as described above using a synthetic RNA oligonucleotide consisting of U4 145-160 with a 3'-Cy3 moiety, *in vitro* transcribed U4 1-145 that was produced from a pUC57 plasmid containing a 5' hammerhead and a 3' HDV ribozyme, and a DNA oligonucleotide complementary to nucleotides 120-160 of U4. U4 1-145 was treated with T4 PNK in the absence of ATP to remove

the 3' cyclic phosphate and the U4 145-160-Cy3 RNA was 5’ phosphorylated with T4 PNK in the presence of ATP as described above. After splinted ligation, the resulting U4 1-160-Cy5 was purified by urea PAGE and ion exchange as described above.

A6.2.3 Gel shift assays

RNA binding reactions with Prp24 were carried out using 1 nM Cy3-U6 1-112 or 5 nM U4 1-160-Cy5. RNAs were refolded by heating to 90°C for 2 min in 2x RNA binding buffer (100 mM KCl, 20% sucrose, 20 mM bis-tris, 10 mM HCl, 1 mM EDTA acid, 1 mM TCEP-HCl, 0.01% Triton X-100, pH~ 6.5, 0.2 mg/mL tRNA, 0.02 mg/mL sodium heparin) and snap cooled on ice. Proteins were prepared as 2x stocks in protein binding buffer (100 mM KCl, 20% sucrose, 20 mM bis-tris, 10 mM HCl, 1 mM EDTA acid, 1 mM TCEP-HCl, 0.01% Triton X-100, pH~ 7, 0.2 mg/mL BSA). Binding reactions contained equal volumes of 2x RNA and 2x protein stocks. Samples were incubated at 22°C for 20 min prior to loading onto 16.5 × 22 cm 6% polyacrylamide gels (29:1 acrylamide:bis-acrylamide, 89 mM Tris borate, 2 mM EDTA pH 8.0). Samples were electrophoresed for 1.5 h at 3 W at 4°C. Gels were imaged directly through low fluorescence glass plates (CBS Scientific) using a Typhoon FLA 9000 imager (GE Healthcare Life Sciences). Results were analyzed using ImageJ software and binding curves were fit using nonlinear regression in GraphPad Prism 4 to the Hill equation: % bound = (B_{max} * [Prp24]^H)/(K_d^H + [Prp24]^H). B_{max} was restrained to be between 0 and 100%, and H (Hill coefficient) and K_d were restrained to be >0. Binding affinities are reported for three technical replicates.

Annealing reactions were carried out at 22°C in binding buffer (as described above) containing 5 nM U4 1-160-Cy5, 25 nM U6 1-112 RNA and 200 nM Prp24 protein. RNAs were refolded by heating to 90°C for 2 min in 2x RNA binding buffer followed by snap cooling on ice. Reactions were stopped by the addition of 2 µl of proteinase K buffer (0.5% sodium dodecyl sulphate, 0.3 mg/ml tRNA, 5 mM $CaCl_2$, 30 mM HEPES pH 7.0, 0.2 mg/ml proteinase K) prior to separation on a 6% polyacrylamide gel (29:1 acrylamide:bis-acrylamide, 89 mM Tris borate, 2 mM EDTA pH 8.0). Samples were electrophoresed for 2 h at 3 W at 4°C and analyzed as above.

Annealing rates were calculated using the ratio of free U4 to U4/U6 in Proteinase K-treated lanes. Resulting data were then fit using non-linear regression in GraphPad Prism 4 to a one-phase exponential association equation: $Y = Y_{max}*(1-e\text{^}(-kt))$ where Y is the % U4 in U4/U6, *k* is the rate constant, and t is the time. Annealing rates are reported for three technical replicates.

A6.2.4 U6-Prp24 complex formation

The optimal stoichiometry of U6:Prp24 was determined empirically using non-denaturing polyacrylamide gel electrophoresis. U6-Prp24 complexes were reconstituted using ~5-10 µM U6 30-99CC (wt, A62G, A79G, or A62G/A79G) and ~1.75-2.5 molar excess of KlPrp24 ΔNΔC (29-395) in 100 mM NaCl, 20 mM Tris pH 7.5, 1 mM TCEP-HCl, 10% glycerol and incubated at 22°C for 20 minutes prior to purification via anion-exchange chromatography (HiTrap Q, GE Healthcare) in 100 mM KCl, 10 mM Tris base, 10 mM HEPES acid, 1 mM TCEP-HCl, 2 mM $MgCl_2$, 5% glycerol and eluted via gradient elution against buffer containing 2 M KCl. The snRNP complex and unbound RNA could be separated as the snRNP eluted at ~57 mS/cm and the free RNA eluted at ~72 mS/cm. Elution fractions were analyzed by non-denaturing polyacrylamide gel electrophoresis and concentrated to a final concentration of ~5 mg/mL using a 10 kDa spin filter.

A6.2.5 Site-directed labeling

KlPrp24 containing an N-terminal biotin acceptor peptide (AP), lacking its native cysteines (C60L/C379V), and harboring two reintroduced cysteine residues (N173C/D337C) was expressed and purified as above with the exception that IEX buffer lacked TCEP-HCl. Protein was labeled using a 20-fold molar excess of maleimide-Cy3 or Cy5 in 300 mM NaCl, 50 mM bis-tris pH 6.5, 10% glycerol in the dark at 4°C overnight. Free dye was removed via cation-exchange chromatography as described above.

A6.2.6 Structure visualization and generation of homology model

The model of KlPrp24 with U6 was generated using the Phyre2 server with one-to-one threading (KELLEY *et al.* 2015) using the coding region of KlPrp24 and the coordinates of ScPrp24 from PDB 4N0T (MONTEMAYOR *et al.* 2014). All figures were generated with PyMOL

(http://www.pymol.org). Electrostatic surfaces were generated with APBS as implemented in PyMOL (BAKER *et al.* 2001).

A6.2.7 Yeast *PRP24* complementation assay

The coding sequence for *Kluveromyces lactis* Prp24, codon optimized for expression in *Saccharomyces cerevisiae*, with flanking regions from the ScPrp24 gene, was cloned via BamHI and SalI into pUC19. The sequence for the coding and flanking regions used is listed in **Table A6-1**. Mutations in KlPrp24 were introduced via inverse PCR and self-ligation as described above and sequence verified (**Table A6-2**). PCR with primers complementary to the EcoRV and XhoI sites was used to subclone KlPrp24 with flanking regions into pRS313 backbone derived from pRS313-ScPrp24 that had been cut with EcoRV/XhoI.

Alleles of Prp24 in pRS313 were transformed into LL101 (*MAT***a** *his3 leu2 trp1 ura3 met2 can1 ade2 lys2 prp24-Δ1::ADE2* [pUN50-*PRP24*]) using the lithium acetate method (GIETZ AND WOODS 2002) and selected on solid media lacking histidine. Transformants were restreaked onto media lacking histidine and containing 1 mg/mL 5-fluoroorotic acid to select against the wild-type plasmid, then restreaked back onto media lacking histidine. Viable clones were grown in YPD liquid media overnight, then resuspended in 10% glycerol to an OD_{600} of 0.5. Serial eight-fold dilutions were plated on solid YPD media and incubated at 16, 23, 30, and 37°C.

Table A6-1. Nucleotide sequence of synthetic genes.

KlPrp24 Geneblock, optimized for *E. coli* expression, with C60L/C379S (underlined) mutations. Cloned into pET28b-HT and pET28b-HT-AP with NdeI/BamHI.	ACGGCCATATGAATGCTGAGAACAACAAGCGCCGTCCGGAAGA CGAGAACGTGAGTGTGGAGGAGGCAGCTACGTTAAAAAAAGTG AAGAACTTGCACACGCATAACCGCGAATATACCACTGTTCTGGT TAAAAATCTGCCGCCGAACTATAATCATCATAAGGTTCGTCGTT ATTTCAAGACTttgGGGTCTATCCTGCAAATCGATGTCACCGACT CAACGGACGGTGATTCTAAGTTAGCTCGCATCGAATTTTCAAGT TACGACCAAGCATTGACTGCCGTGTCGCGCACTTTAAAAAAAAT TGGGTTCCATCAGATCACTGTAGAGCAACTTACTGATTCCACTA TTTGGGTTACGAATTTCCCTCCAGGCTACGATGCGTCAAAACTG CGTAACTTGCTGAGTCAGTATATTGATAGTCCCATCCTGAGCAT CCGCTTACCTTCACTTGCTTTTAATAGTCGCCGCCGCTTCGCAT ACGTCGATCTTGTATCCCCTGAGGTGGCCAAAAAAGCCACCAA TCGTCTGAATGGAATTGAGATCAATAACTATAAATTAGTTGCTAA GATTAGCAATGTCAACGAGCGTACCGAACGCACGGATAACGCG ATCTCTGATGGCCGCGAGATTATCGTTAAAAATCTGCCCGATGA CATTACGATTGATGATCTGATTAAGATGTTCAATGAATTTGGAGA TACCGAGAAGGCACGCATTGTGACGGGCGACGAGACGAATCCT GGACGCCACTCACGTTACGCTTTTATCACTTTCAAGAACAAAGC CTCGGCTGATAACGCGCTGTCGTTGAACGGCGCGGTTATGAAT GGCAAGCCCTTGCATGTTTCAAAAGTCATGCGTAAGGCTTACCT TGAGCGTCAGGAGGTGAAACGCTTGTTAGCATCGCGCAAGGAG AACCCAAAGGTAGTTGGGATCTATCCCCTTAGCGACCAGATTAC TCCTGATCAGATCTTGGCCTTCATCACTGAGAAGGCTCAGATTC GCCCtCCCGACATCACGAAGGTGCTTTTGGTATCCGACCACGAA GGAGCGCTTATCATTGCTGATAAAGAGGCGATTACTGCTAAGGT ATCACTTGCAATCAATGGGTATAAATTCAAGAATCGCATCTTAAA GtctGTTTCCGCCCACGAGCTTAGCCTGCATTACCCAAACGAGCA TAAGACCACCCTTAACCAACCTAAACACGCCGTCGTAAAAGAGA TTAAGAGTCAACAAGCCTCTCAAAACGATAAGAAACTTTCAAAC GACGATTTTCGTAAATTATTTCTGTCTAAATAAGGATCCCTGA
KlPrp24 Geneblock, codon optimized for yeast expression, with flanking ScPrp24 up and downstream regions (used for *in vivo* complementation assays). Coding region is underlined. Cloned into pRS313-ScPrp24 cut with EcoRV/XhoI (for wild-type). Cloned into pUC19 cut with BamHI/SalI to introduce mutagenize, then cloned this into	ACGGCGATATCAGATATTCTTTGCTATCCATGTCCGATTTATTCT GATAATATTTCCGCTTGTATCATTCAATTTTCTATGTACATCATG ACAATGAATTGTAGAGTGCATCAATGTCTTTTTTAAAGTCATCGC TAAAAAATTTTCAGGCGGGAACTTATTCAAAATGGCAGATTTACT ACAACATTGAAAACCCTTTATATTTATTATATGAGAATCATTAGC GATAAACAAGCCCACATATTCACTTCATAATCTGTGATAATGAAC GCGGAAAACAATAAAAGGAGACCGGAGGACGAGAACGTTAGC GTCGAAGAAGCGGCCACGCTGAAGAAGGTTAAGAACCTGCACA CACACAACAGAGAGTACACCACCGTGCTGGTCAAAAACCTGCC GCCTAACTATAACCACCATAAAGTAAGGCGTTATTTCAAAACGT GTGGAAGTATATTGCAAATTGACGTCACTGACTCTACCGATGGA GACTCAAAACTGGCGAGAATAGAGTTTTCAAGTTATGATCAAGC ACTAACGGCAGTGAGTAGGACACTGAAGAAGATCGGATTTCAT CAAATAACCGTTGAACAGCTTACGGACTCCACCATCTGGGTGA CTAATTTTCCCCCCGGTTATGACGCCAGTAAGCTGAGAAATCTT CTAAGCCAGTATATAGACAGCCCAATTTTGTCCATAAGGTTGCC GTCTCTTGCGTTCAACTCAAGGAGACGTTTTGCCTACGTAGACT TAGTCAGTCCTGAAGTCGCCAAGAAGGCAACGAATAGACTGAA CGGTATAGAGATAAACAACTACAAACTAGTAGCCAAAATCTCTA

pRS313-ScPrp24 cut with EcoRV/XhoI.	ATGTAAATGAAAGAACGGAGCGTACAGATAACGCAATTAGCGA CGGGAGGGAAATAATAGTAAAGAACTTGCCGGATGATATTACTA TCGACGATCTTATCAAGATGTTTAACGAATTTGGGGATACTGAG AAGGCAAGAATCGTAACCGGTGATGAGACAAATCCTGGGAGAC ATAGCAGATACGCCTTCATCACCTTCAAAAATAAGGCCTCCGCG GATAACGCTCTATCACTAAATGGGGCCGTAATGAACGGTAAGC CTTTACATGTATCTAAGGTTATGAGAAAGGCCTACTTGGAGCGT CAGGAGGTGAAAAGGCTACTAGCAAGCAGGAAGGAGAACCCG AAGGTGGTGGGCATCTATCCACTGTCAGACCAAATCACACCTG ATCAAATACTAGCCTTTATCACTGAGAAAGCCCAAATCAGACCC CCGGATATAACTAAAGTCCTGCTAGTCTCTGACCACGAGGGAG CACTAATAATTGCTGACAAGGAAGCAATAACGGCGAAGGTTAGT CTGGCGATTAACGGATATAAATTTAAGAATAGAATACTAAAATGT GTGTCAGCTCACGAACTTTCTCTGCACTATCCCAACGAGCATAA AACTACCCTGAACCAGCCTAAACACGCAGTGGTTAAAGAGATCA AATCCCAACAAGCTTCCCAGAACGATAAGAAACTTAGCAATGAC GACTTCAGGAAATTATTTCTATCCAAGTAAACTAACCGAACCGG AGAAAGACGAACTTCTAATAGGATGTCATTTTAGCAGTGAATAT AAACCCAAGTTTCTTTCCAAACGTTCACATACATCAAGCAGCCA ACATTCGAGAGAGCAGAGATTACGACCTCGAGCTGA

Table A6-2. List of synthetic DNA oligonucleotides.

Oligos to make pRS313-KlPrp24 +/- flank and mutants	
Oligos to make pUC19-KlPrp24 +/- flank	
BamHI-KL-f	5'-ATGAGGATCCCCAACCATATGATATCAGATATTCTTTGC-3'
SalI-KL-r	5'-TCGTGTCGACGGGCCCCCCCTCGAGGTCGTAATCTCTG-3'
Oligos to introduce mutations into KlPrp24 +/- flank	
313-C60L-f	5'-TTGGGAAGTATATTGCAAATTGACGTCACTGAC-3'
313-C60L-r	5'-CGTTTTGAAATAACGCCTTACTTTATGGTG-3'
313-C379V-f	5'-GTTGTGTCAGCTCACGAACTTTCTCTGCACTATC-3'
313-C379V-r	5'-TTTTAGTATTCTATTCTTAAATTTATATCC-3'
313-N173C-f	5'-TGTAGACTGAACGGTATAGAGATAAACAACTAC-3'
313-N173C-r	5'-CGTTGCCTTCTTGGCGACTTCAGGACTGAC-3'
313-D337C-f	5'-TGTATAACTAAAGTCCTGCTAGTCTCTGACCAC-3'
313-D337C-r	5'-CGGGGGTCTGATTTGGGCTTTCTCAGTGAT-3'
Oligos to move EcoRV/XhoI fragment of KlPrp24 +/- flank into pRS313 plasmid	
313KL-PCR-f	5'-CTTTCCAACCATATGATATCAGATATTCTTTG-3'
313KL-PCR-r	5'-CCCCCCCTCGAGGTCGTAATCTCTGCTCTCTCG-3'
Oligos to sequence pRS313-KlPrp24 +/- flank and variants	
M13-forward(-20)	5'-GTAAAACGACGGCCAGT-3'
M13-reverse(-27)	5'-CAGGAAACAGCTATGAC-3'
313-Kl-484seq-for	5'-ACGTCACTGACTCTACCGATG-3'
313-Kl-906seq-rev	5'-GTCGATAGTAATATCATCCGGC-3'
Oligos to make pET28b-HT- and pET28b-HT-AP-KlPrp24 mutations and truncations	
Truncations	
pET28b-TEV-rev	5'-ATGGCCCTGGAAGTACAGGTTCTCGCTGCTG-3'
AP-rev	5'-ATGTTCATGCCATTCAATTTTCTGCGCTTC-3'
KlPrp24-22-f	5'-ACGTTAAAAAAAGTGAAGAACTTGCACACG-3'
KlPrp24-29-f	5'-TTGCACACGCATAACCGCGAATATACCACTG-3'
pET28b-stop-end-for	5'-TAAGGATCCGAATTCGAGCTCCGTCGACAAGC-3'
KlPrp24-395-rev	5'-GGTCTTATGCTCGTTTGGGTAATGCAGGCTAAG-3'
Mutations to return to wild-type	
KlPrp24-L60C-f	5'-TGCGGGTCTATCCTGCAAATCGATGTCACCGAC-3'
KlPrp24-60-r	5'-AGTCTTGAAATAACGACGAACCTTATGATG-3'
KlPrp24-S379C-f	5'-TGCGTTTCCGCCCACGAGCTTAGCCTGCATTAC-3'
KlPrp24-379-r	5'-CTTTAAGATGCGATTCTTGAATTTATACCC-3'
Mutations to replace native cysteines	
KlPrp24-C60E-f	5'-GAAGGGTCTATCCTGCAAATCGATGTCACCGAC-3'
KlPrp24-C60M-f	5'-ATGGGGTCTATCCTGCAAATCGATGTCACCGAC-3'
KlPrp24-C60K-f	5'-AAAGGGTCTATCCTGCAAATCGATGTCACCGAC-3'
KlPrp24-C379V-f	5'-GTTGTTTCCGCCCACGAGCTTAGCCTGCATTAC-3'
KlPrp24-C379M-f	5'-ATGGTTTCCGCCCACGAGCTTAGCCTGCATTAC-3'
Mutations to introduce new cysteines	
KlPrp24-D87C-f	5'- TGCCAAGCATTGACTGCCGTGTCGCGCACTTTAA -3'
KlPrp24-D87-r	5'- GTAACTTGAAAATTCGATGCGAGCTAACTTAG-3'
KlPrp24-S94C-f	5'-TGCCGCACTTTAAAAAAAATTGGGTTCCATCAG-3'
KlPrp24-S94-r	5'-CACGGCAGTCAATGCTTGGTCGTAACTTG-3'
KlPrp24-L97C-f	5'-TGCAAAAAAATTGGGTTCCATCAGATCACTGTAG-3'

KlPrp24-L97-r	5'-AGTGCGCGACACGGCAGTCAATGCTTGGTC-3'
KlPrp24-N173C-f	5'-TGCCGTCTGAATGGAATTGAGATCAATAACTAT-3'
KlPrp24-N173-r	5'-GGTGGCTTTTTTGGCCACCTCAGGGGATAC-3'
KlPrp24-K225C-f	5'-TGCATGTTCAATGAATTTGGAGATACCGAGAAG-3'
KlPrp24-K225-r	5'-AATCAGATCATCAATCGTAATGTCATCGGG-3'
KlPrp24-N265C-f	5'- TGCGCGCTGTCGTTGAACGGCGCGGTTATGAATG-3'
KlPrp24-N265-r	5'- ATCAGCCGAGGCTTTGTTCTTGAAAGTGAT-3'
KlPrp24-S268C-f	5'-TGCTTGAACGGCGCGGTTATGAATGGCAAGCCC-3'
KlPrp24-S268-r	5'-CAGCGCGTTATCAGCCGAGGCTTTGTTCTTG-3'
KlPrp24-D337C-f	5'-TGCATCACGAAGGTGCTTTTGGTATCCGACCAC-3'
KlPrp24-D337-r	5'-GGGAGGGCGAATCTGAGCCTTCTCAGTGATG-3'
Other mutations	
KlPrp24-N49D/H50A-f	5'-GATGCGCATAAGGTTCGTCGTTATTTCAAGACTTTG-3'
KlPrp24-N49-r	5'-ATAGTTCGGCGGCAGATTTTTAACCAGAAC-3'
KlPrp24-K77A-f	5'-GCGTTAGCTCGCATCGAATTTTCAAGTTACGAC-3'
KlPrp24-K77-r	5'-AGAATCACCGTCCGTTGAGTCGGTGACATC-3'
KlPrp24-M285D-f	5'-GATCGTAAGGCTTACCTTGAGCGTCAGGAGGTG-3'
KlPrp24-M285-r	5'-GACTTTTGAAACATGCAAGGGCTTGCCATTC-3'
Oligos to sequence pET28b-HT- and pET28b-HT-AP-KlPrp24 and variants	
T7-forward	5'-TAATACGACTCACTATAGGG-3'
pET28b-rev-seq	5'-CGCCAATCCGGATATAGTTCC-3'
DNA oligonucleotides used for RNA transcription	
U6-30-99CC-WT-temp	5'-TTCTAATACGACTCACTATAGGTCAATTTGAAACAATACAGA GATGATCAGCAGTTCCCCTGCATAAGGATGAACCGTTTTACAA AGAGACC-3'
U6-30-99CC-WT-comp	5'-GGTCTCTTTGTAAAACGGTTCATCCTTATGCAGGGGAACTG CTGATCATCTCTGTATTGTTTCAAATTGACCTATAGTGAGTCGT ATTAGAA-3'
U6-30-99CC-A62G-temp	5'-TTCTAATACGACTCACTATAGGTCAATTTGAAACAATACAGA GATGATCAGCGGTTCCCCTGCATAAGGATGAACCGTTTTACAA AGAGACC-3'
U6-30-99CC-A62G-comp	5'-GGTCTCTTTGTAAAACGGTTCATCCTTATGCAGGGGAACCG CTGATCATCTCTGTATTGTTTCAAATTGACCTATAGTGAGTCGT ATTAGAA-3'
U6-30-99CC-A62G/A79G-temp	5'-TTCTAATACGACTCACTATAGGTCAATTTGAAACAATACAGA GATGATCAGCGGTTCCCCTGCATAAGGGTGAACCGTTTTACAA AGAGACC-3'
U6-30-99CC-A62G/A79G-comp	5'-GGTCTCTTTGTAAAACGGTTCACCCTTATGCAGGGGAACCG CTGATCATCTCTGTATTGTTTCAAATTGACCTATAGTGAGTCGT ATTAGAA-3'

A6.3 Results and discussion

A6.3.1 *K. lactis* U6 snRNP components are conserved

Kluveromyces lactis is an ascomyceteous budding yeast that is closely related to *Saccharomyces cerevisiae* (SCHAFFRATH AND BREUNIG 2000). Prp24 is well-conserved between *K. lactis* and *S. cerevisiae*, with 41% amino acid identity and 62% similarity overall (**Figure A6-1A**). The domain structure is conserved and consists of four RNA recognition motifs (RRMs). RRMs 1 and 2 are the best conserved, while RRMs 3 and 4 are relatively poorly conserved. This is intriguing, considering that RRM1 does not contact RNA in the core snRNP structure and that RRM3 is responsible for binding the high affinity site on U6 that includes the ACAGAGA sequence (MONTEMAYOR *et al.* 2014). The SNFFL box, which is important for binding to the Lsm2-8 ring (RADER AND GUTHRIE 2002; MONTEMAYOR *et al.* in prep), is nearly perfectly conserved (**Figure A6-1A**). In contrast, a helix in RRM4 that contributes to the interaction with Lsm2-8 is poorly conserved (**Figure A6-1A**). While some residues that contact U6 in the crystal structure are conserved (**Figure A6-1B**), many contacts are not conserved. These include several amino acids that are in contact with U6 RNA in the crystal structure and are the location of suppressor mutations of U6-A62G cold sensitivity, including R153 (alanine in KlPrp24), P195 (valine in KlPrp24), M213 (isoleucine in KlPrp24), P243 (valine in KlPrp24), N253 (serine in KlPrp24), C254 (arginine in KlPrp24), S283 (histidine in KlPrp24), D288 (methionine in KlPrp24), and P291 (alanine in KlPrp24) (MONTEMAYOR *et al.* 2014). The most interesting of these is D288, which forms a unique “aspartate bridge” to coordinate U6-A42 and G55 to form the interlocked topology (MONTEMAYOR *et al.* 2014). Determining the crystal structure of the KlPrp24-U6 complex could aid in determining whether the interlocked topology is conserved and in determining which contacts and structural motifs are most essential for forming such a topology.

A model for *K. lactis* U6-Prp24 was generated using the Phyre2 server, which uses one-to-one threading based on a sequence and a structure (KELLEY *et al.* 2015). Although residues within the electropositive groove that were shown to be important for annealing *in vitro* (DIDYCHUK

et al. 2016) are poorly conserved (**Figure A6-1A**), the overall electropositive character of the groove is conserved in *K. lactis* Prp24, although it may be less positively charged in some areas (**Figure A6-1C**). Interestingly, of the tested electropositive groove residues, only positions N53, R81, and R131 are conserved in charge (**Figure A6-1A**). As discussed in **Appendix VI**, N53 and R81 may be more critical to yeast viability than other residues within the groove.

The *K. lactis* U6 snRNA is very well conserved, with perfect conservation between nucleotides 21-103 (**Figure A6-1C**). The 5' stem loop in KlU6 is less stable and is four base pairs shorter. The 3' tail of U6 is also shorter in *K. lactis*, suggesting that the U6 snRNP in *K. lactis* may be more compact. U4 snRNA is also well conserved, and U4/U6 Stems I and II are perfectly conserved (**Figure A6-1D**). The conservation of U6 and U4, along with the lower conservation in Prp24, makes *K. lactis* the perfect system to test hypotheses that could lead to a better understanding of how Prp24 binds U6 and anneals it to U4.

Figure A6-1. Prp24, U6, and U4 are conserved in *S. cerevisiae* and *K. lactis.*

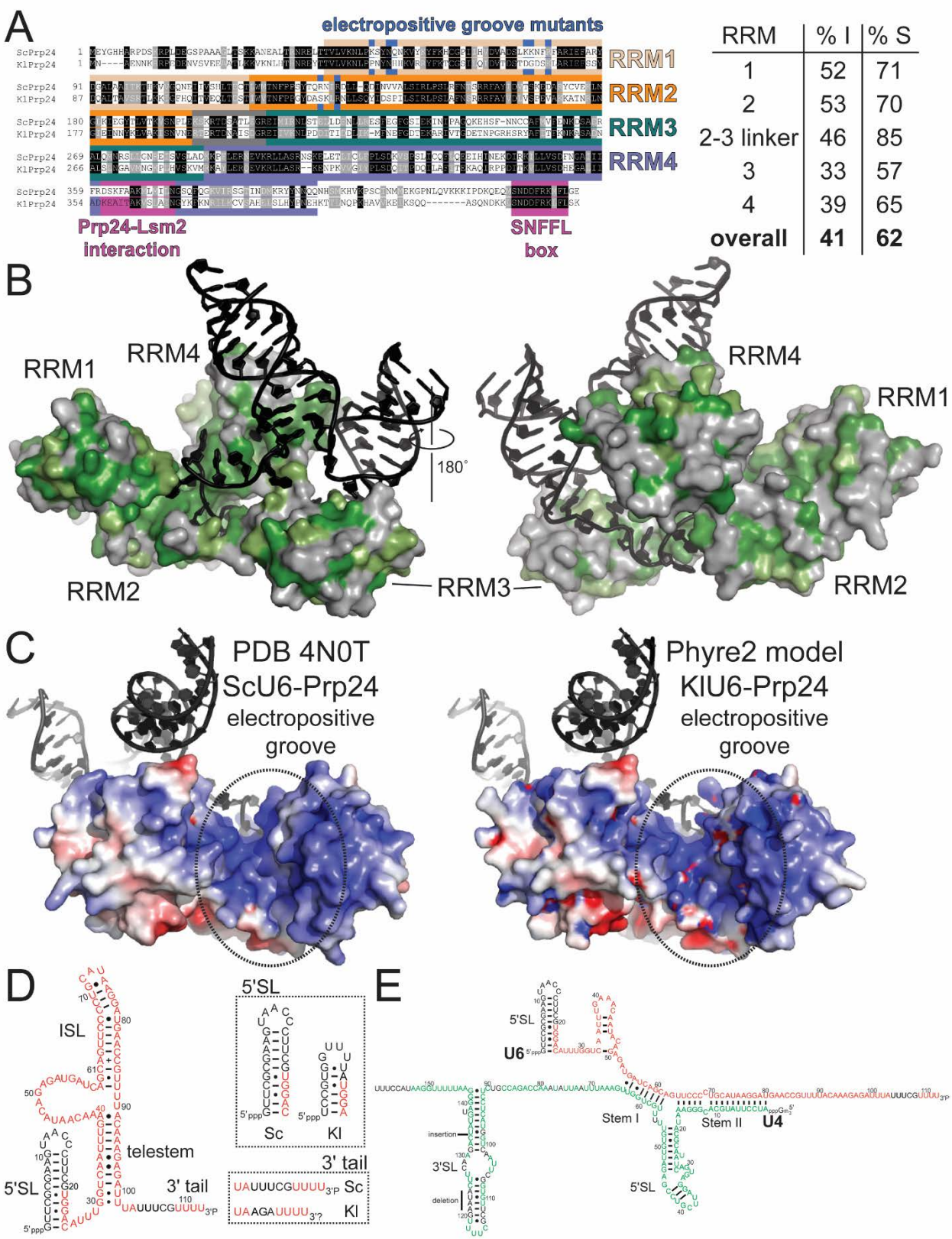
A) Alignment of Prp24 from *S. cerevisiae* (ScPrp24; top) and *K. lactis* (KlPrp24; bottom). RNA recognition motifs (RRMs) are highlighted in wheat, orange, green, and purple, as are known protein-protein interaction regions with the Lsm2-8 ring (magenta). Residues in the electropositive groove shown to affect annealing *in vitro* (DIDYCHUK et al. 2016) are highlighted in blue. The percentage of residues that are identical (% I) or similar (% S) in each domain or overall is shown to the right.

B) Many residues are identical (green) or similar (light green) in Prp24, with residues contacting U6 RNA generally more conserved.

C) Although the residues in the electropositive groove are poorly conserved, this region is predicted to be positively charged in KlPrp24.

D) U6 snRNA is highly conserved in *K. lactis* and *S. cerevisiae*. The sequence of ScU6 is shown, with nucleotides in red perfectly conserved in KlU6. The 5' SL is not conserved, with the KlU6 5' SL less stable than in ScU6. The 3' tail of U6 is also shorter in KlU6.

E) U4 and U6 are highly conserved in *K. lactis* and *S. cerevisiae*. The sequence of ScU6 and ScU4 is shown, with nucleotides in red and green conserved in KlU6 and KlU4. *K. lactis* U4/U6 likely adopts a similar three-helix junction containing U4/U6 Stem I, II, and the U4 5'SL.


A
electropositive groove mutants
RRM1
RRM2
RRM3
RRM4
Prp24-Lsm2 interaction
SNFFL box
RRM | % I | % S
1 | 52 | 71
2 | 53 | 70
2-3 linker | 46 | 85
3 | 33 | 57
4 | 39 | 65
overall | 41 | 62
B
RRM1
RRM4
RRM2
RRM3
180˚
C
PDB 4N0T
ScU6-Prp24
electropositive groove
Phyre2 model
KlU6-Prp24
electropositive groove
D
ISL
telestem
5'SL
3' tail
5'SL
Sc
Kl
3' tail
Sc
Kl
E
5'SL
U6
Stem I
Stem II
U4
insertion
deletion
3'SL
5'SL

A6.3.2 KlPrp24 binds U6 similarly to ScPrp24 *in vitro*

To establish if KlPrp24 has similar properties to ScPrp24 *in vitro*, we tested the affinity of KlPrp24 for fluorescently labeled ScU6 (**Figure A6-2A**). KlPrp24 behaves similarly to ScPrp24 and forms a single species when bound to U6 on a native gel which travels with similar mobility to ScPrp24 (**Figure A6-2A**). KlPrp24 binds ScU6 very tightly with 1.6 nM affinity, which is ~1.4 fold better than ScPrp24 (**Figure A6-2B**). As expected, this suggests that the regions which are not conserved in U6 (the 5'SL and 3' tail) do not contribute to Prp24 binding. Thus, ScU6 can be used for biochemical studies in lieu of KlU6 *in vitro*.

A6.3.3 KlPrp24 anneals U4 and U6 slower than ScPrp24 *in vitro* and has altered U4 binding properties

Next, we tested the ability of KlPrp24 to anneal ScU6 and ScU4 (**Figure A6-2C**). Interestingly, KlPrp24 anneals U4 and U6 ~2-fold slower than ScPrp24, despite improved U6 binding (**Figure A6-2D**). This small difference in annealing activity could be due to several possibilities: (1) that KlPrp24 has sequence preference for KlU4 and KlU6, (2) that tighter U6 binding by KlPrp24 prevents annealing by stabilizing the reactants or (3) that KlPrp24 is a worse annealer because of its electropositive groove is less positively charged, or because it lacks residues important for annealing by ScPrp24. The first possibility seems unlikely due to the high conservation of U4 and U6 (**Figure A6-1C-D**), while the second possibility seems unlikely because inclusion of the Lsm2-8 ring enhances annealing while increasing the fraction of U6 bound (DIDYCHUK *et al.* 2016; MONTEMAYOR *et al.* in prep). The third possibility could be tested by introducing new positive charges into the electropositive groove to make it more similar to the ScPrp24 groove.

Next, we tested the affinity of Prp24 for U4 snRNA. ScPrp24 weakly binds U4 with a K_d of ~340 nM (**Figure A6-2E**). At high concentrations of ScPrp24 it binds non-specifically, but it forms a specific U4-Prp24 band up to 800 nM (**Figure A6-2E**). Interestingly, KlPrp24 binds ~2.5 worse, with a high nanomolar K_d and Hill coefficient (**Figure A6-2E-G**). Specific binding of KlPrp24 to U4

is nearly undetectable (**Figure A6-2E**). In the conditions of the annealing assay (200 nM Prp24), 10-fold more U4 is bound by ScPrp24 than KlPrp24. These data suggest that the electropositive groove, which is more positively charged in ScPrp24, contributes to annealing by electrostatic recruitment of U4. The electropositive groove may promote annealing by increasing the local concentration of U4, which is then captured by a transiently unwound U6 ISL.

Figure A6-2. KlPrp24 can bind and anneal U4 and U6 *in vitro*.

A) ScPrp24 (top) and KlPrp24 (bottom) bind ScU6 RNA with low nanomolar affinity.

B) Quantification of Prp24 binding to U6 in (A). Plotted data points represent the average of three technical replicates ± s.d.

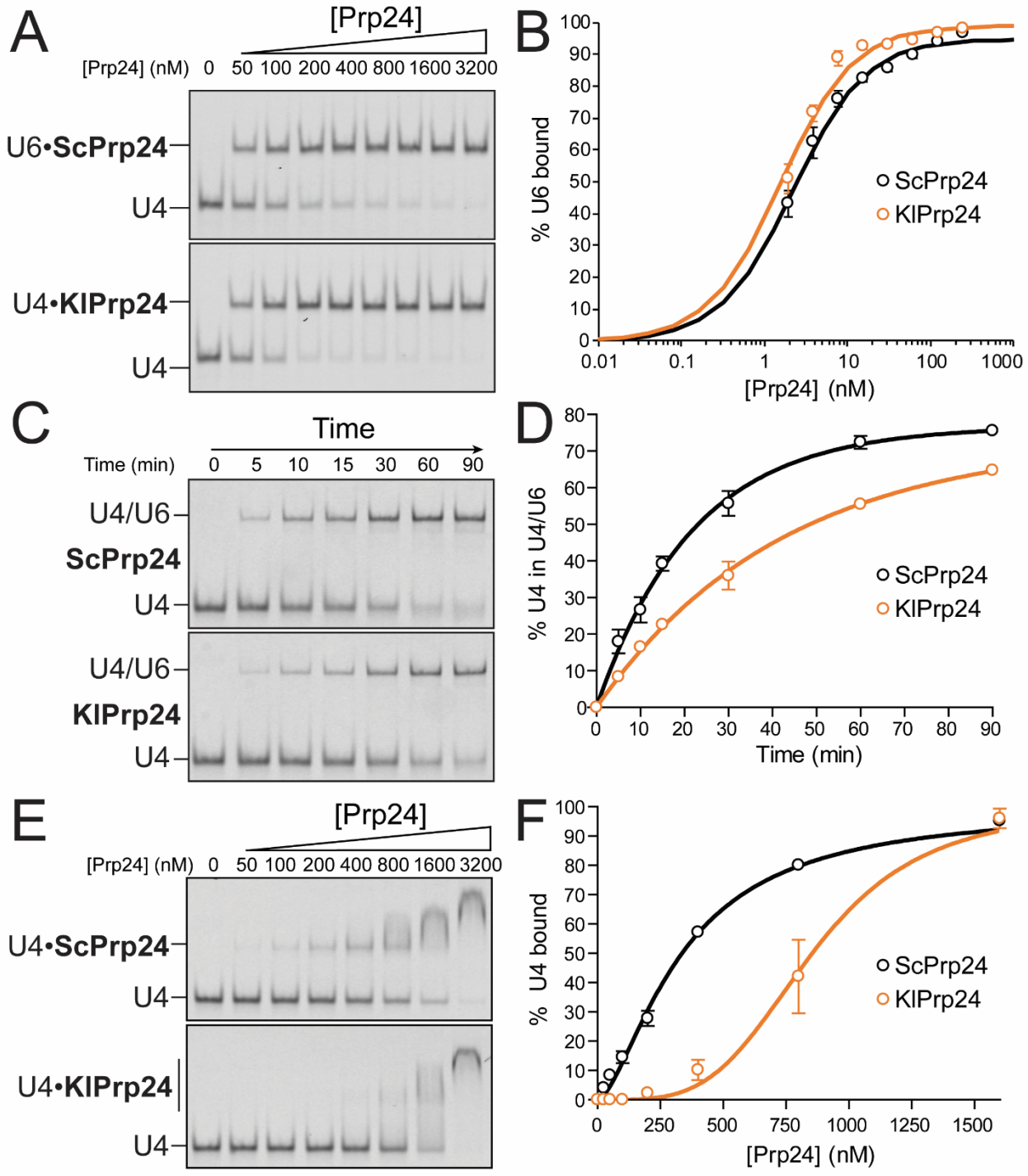
C) ScPrp24 (top) and KlPrp24 (bottom) can anneal ScU6 and ScU4.

D) Quantification of annealing in (C). Plotted data points represent the average of three technical replicates ± s.d.

E) ScPrp24 (top) binds ScU4 RNA with high nanomolar affinity and forms a specific complex. KlPrp24 (bottom) binds ScU4 RNA with higher nanomolar affinity, likely due to nonspecific binding at high Prp24 concentrations.

F) Quantification of Prp24 binding to U4 in (E). Plotted data points represent the average of three technical replicates ± s.d.

G) Values for fit parameters for binding and annealing by ScPrp24 and KlPrp24 in (B), (D), and (F).



G

	U6 binding		U4/U6 annealing		U4 binding		
	K_d (nM)	B_{max} (% U6)	k (min^{-1})	Y_{max}(% U4)	K_d (nM)	B_{max}(% U4)	Hill
ScPrp24	2.2 ± 0.1	95 ± 1	0.046 ± 0.003	77 ± 2	340 ± 20	100 ± 4	1.6 ± 0.1
KlPrp24	1.6 ± 0.1	99 ± 1	0.024 ± 0.002	73 ± 3	850 ± 90	100 ± 10	4 ± 1
	KlPrp24 ~1.4 fold better		ScPrp24 ~1.9 fold better		ScPrp24 ~2.5 fold better		

A6.3.4 The *K. lactis* U6 snRNP can be co-purified and is crystallizable

As discussed above, determining the co-structure of KlPrp24 with U6 will allow us to determine which protein-RNA contacts are essential for formation of the interlocked topology. First, we sought to determine if KlPrp24 behave similarly at high concentration *in vitro*. Using U6 30-99CC, which is crystallizable with ScPrp24 (MONTEMAYOR *et al.* 2014; MONTEMAYOR *et al.* 2017) and it perfectly conserved in *K. lactis*, we show that KlPrp24 binds this RNA stoichiometrically (**Figure A6-3A**) and that this complex can be co-purified via anion exchange (**Figure A6-3B**). Complexes formed with U6 containing stabilizing mutations (A62G, A79G, and A62G/A79G) are crystallizable (**Figure A6-3C**). Crystal screening is in progress and crystallization conditions must be optimized, but one crystal produced diffraction data that extends to ~5.5 Å (**Figure A6-3D**). With further optimization, the structure of the KlU6 snRNP core could be determined.

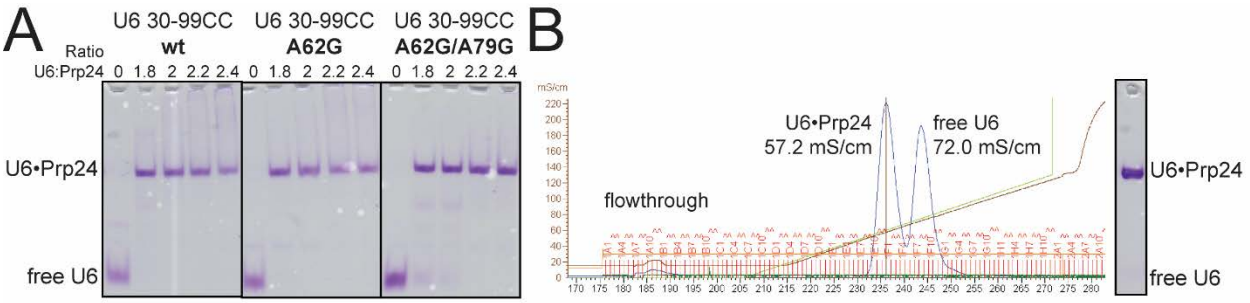
Figure A6-3. The core of the *K. lactis* snRNP can be purified and crystallized.

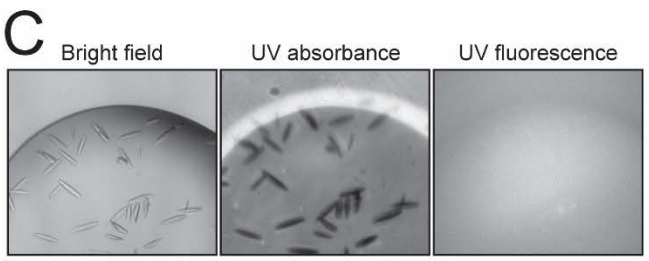
A) KlPrp24 lacking the N- and C-terminal domains (KlPrp24 29-395) binds U6 30-99CC RNA with the wild-type sequence or with stabilizing mutations A62G or A62G/A79G. Binding is stoichiometric and KlPrp24-U6 runs as a single complex.

B) The KlU6 snRNP core (KlPrp24 with U6 30-99CC) is salt-stable and can be copurified via anion exchange.

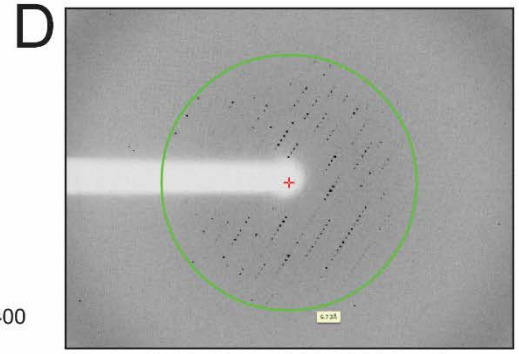
C) The KlU6 snRNP core is crystallizable. These crystals absorb UV well, suggesting that they contain nucleic acid. These crystals have poor UV fluorescence properties, consistent with KlPrp24 29-395 having only one tryptophan residues that is predicted to be buried.

D) One crystal form diffracts x-rays to medium resolution (~5.5 Å).





PegRx E6 KlPrp24 ΔNΔC•U6 30-99CC-A62G
1:3 dilution into 1 mM $MnCl_2$.
Reservoir: 10% 2-propanol, 0.1 M sodium citrate pH 5.0, 26% PEG-400
Diffracts to ~9 Å



KlPrp24 ΔNΔC•U6 30-99CC-A62G/A79G
Reservoir: 200 mM NH_4Cl, 30% PEG-MME 2000,
100 mM sodium citrate pH 5.5
Diffracts to ~5.5 Å

A6.3.5 KlPrp24 does not require its native cysteines for binding *in vitro*

KlPrp24 contains two native cysteine residues (C60 and C379), while ScPrp24 contains eight cysteines. Native cysteines must be removed prior to labeling with thiol-specific reagents. KlPrp24 C60 is conserved in ScPrp24 and is packed within an α-helix in RRM1 (**Figure A6-4A**). KlPrp24 C379 is a serine in ScPrp24 and is also predicted to be buried within the protein (**Figure A6-4A**). Based on its packing, C60 was first mutated to leucine while based on the presence of a serine in ScPrp24, KlPrp24-C379 was first mutated to serine. While KlPrp24-C60L had little effect on *in vitro* binding affinity or behavior, KlPrp24-C379S bound U6 weakly and resulted in heterogeneous bands on a native binding gel (**Figure A6-4B**). KlPrp24-C379V and C379M were tested to find a suitable replacement for C379 (**Figure A6-4B**). KlPrp24-C379V bound U6 tightly, while C379M displayed binding properties similar to C379S (**Figure A6-4B**). Upon closer inspection of the KlPrp24 model, we realized that KlPrp24 C379 is close to a serine that ScPrp24 lacks at that position (KlPrp24-S364/ScPrp24-L369) (**Figure A6-4A**). Perhaps having placing two serine residues (in the case of KlPrp24-C379S) or a serine and a methionine (in the case of KlPrp24-C379M) within 5 Å destabilizes the fold of RRM4. These data suggest that the two native cysteines in KlPrp24 can be replaced with leucine and valine to generate a cysteine-less Prp24 that retains binding properties that are similar to wild-type Prp24.

A6.3.6 KlPrp24 can be site-specifically labeled using reintroduced cysteines

Since the native cysteines in KlPrp24 can be replaced without affecting U6 binding affinity, the next step was to engineer a cysteine-less construct (C60L/C379V) and reintroduce non-native cysteines at locations of interest. Positions N173 and D337 were selected for cysteine reintroduction, as they were predicted to be surface-exposed and to be far from the U6 binding surface (**Figure A6-4C**). Additionally, they were far apart in secondary structure (in RRMs 2 and 4 respectively) yet should be ~44 Å apart based on the U6•Prp24 structure (**Figure A6-4C**), which would correspond to a high (~0.9) FRET efficiency if these positions were labeled with Cy3 and Cy5. KlPrp24 was engineered with an N-terminal biotin acceptor peptide (AP tag) for surface

immobilization, the native cysteines removed (C60L/C379V), and new cysteines reintroduced at positions N173 and D337 (N173C/D337C). This version of the protein was termed “AP-4mut” and binds U6 RNA with similar affinity to wild-type KlPrp24 (**Figure A6-4D**). This protein can be labeled with maleimide-conjugated Cy3, Cy5, or with both (**Figure A6-4E**). Versions of the protein containing only one cysteine residue (at position N173 or D337) can also be labeled (**Figure A6-4F**). These proteins bind similarly before and after maleimide labeling, suggesting that the conjugation of the fluorphore does not interfere with RNA binding. However, labeled protein non-specifically enters the gel (potentially due to the presence of tRNA and heparin in the assay conditions) (**Figure A6-4F**). The assay conditions and mutation locations (as well as the substitutions at C60 and C379) could be further optimized and co-opted for single molecule FRET experiments.

Figure A6-4. Native cysteines can be removed from KlPrp24 without drastically affecting binding activity.

A) KlPrp24 contains two cysteines: C60 (top) and C379 (bottom). C60, in RRM1 is conserved in ScPrp24 and is predicted to be buried. C379, in RRM4, is a serine in ScPrp24.

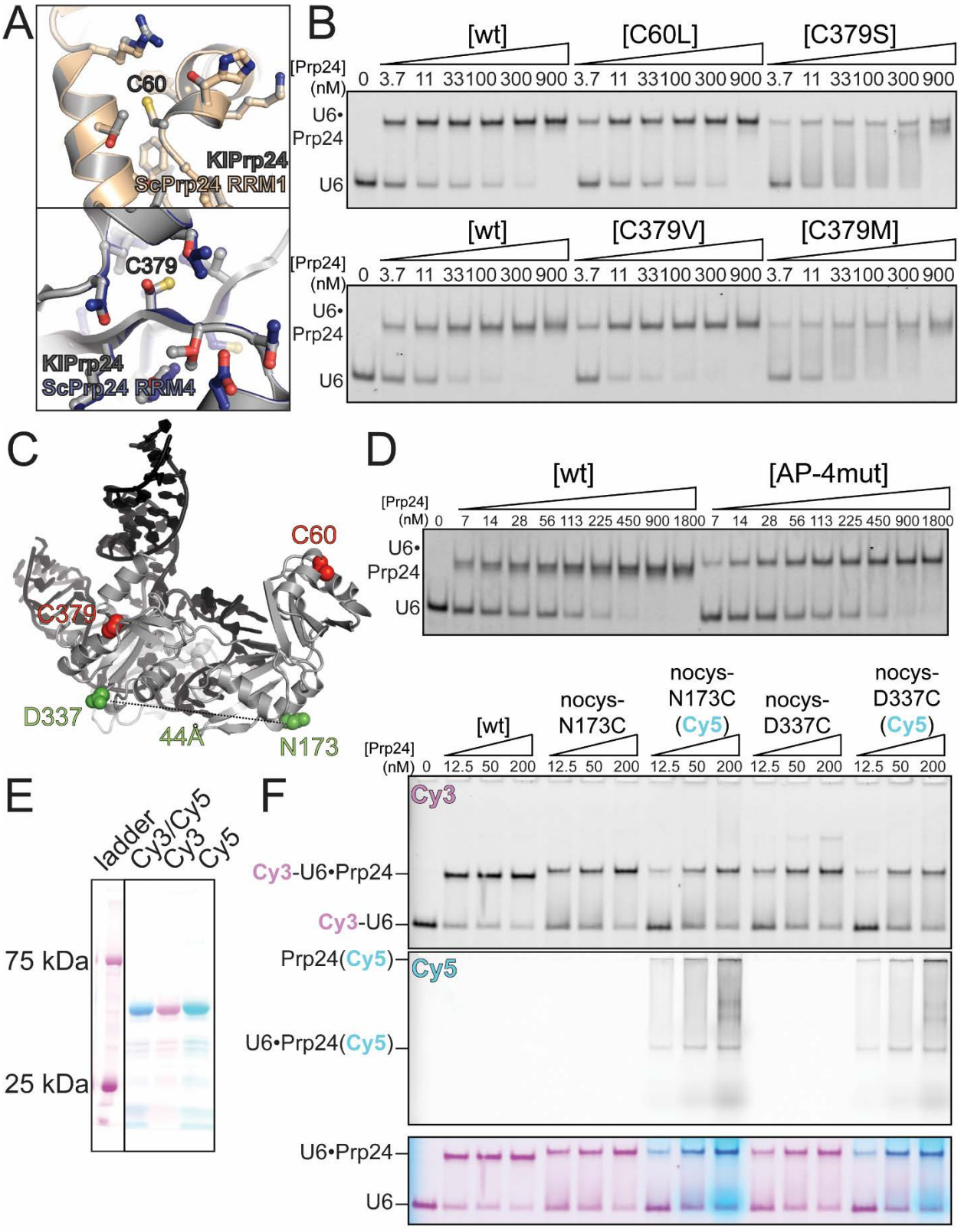
B) Substitution of C60 for a leucine in KlPrp24 has little effect on binding behavior, but substitution of C379 for serine resulted in lower affinity and altered gel mobility (top). Substitution of C379 for methionine (bottom) resulted in similar altered binding affinity, but substitution for leucine resulted in similar binding as wild-type.

C) The native cysteine C60 and C379 (red) are located in RRMs 1 and 4. Re-introduced cysteines at N173 and D337 (green) are in RRMs 2 and 4 and are predicted to be 44 Å apart in the U6 snRNP.

D) The AP-4mut protein binds to U6 RNA similarly to wild-type KlPrp24.

E) The AP-4mut protein can be fluorescently labeled with maleimide-conjugated Cy3 and/or Cy5.

F) Cy5-labeled KlPrp24 containing an AP tag and C60L/C379V/N173C or C60L/C379V/D337C mutations binds Cy3-labeled U6, but non-specifically enters the gel and is trapped in the gel wells.


A
C60
KlPrp24
ScPrp24 RRM1
C379
KlPrp24
ScPrp24 RRM4
B
[wt]
[C60L]
[C379S]
[Prp24] (nM)
0 3.7 11 33 100 300 900 3.7 11 33 100 300 900 3.7 11 33 100 300 900
U6•Prp24
U6
[wt]
[C379V]
[C379M]
[Prp24] (nM)
0 3.7 11 33 100 300 900 3.7 11 33 100 300 900 3.7 11 33 100 300 900
U6•Prp24
U6
C
C60
C379
D337
44Å
N173
D
[wt]
[AP-4mut]
[Prp24] (nM)
0 7 14 28 56 113 225 450 900 1800 7 14 28 56 113 225 450 900 1800
U6•Prp24
U6
[wt]
nocys-N173C
nocys-N173C (Cy5)
nocys-D337C
nocys-D337C (Cy5)
[Prp24] (nM)
0 12.5 50 200 12.5 50 200 12.5 50 200 12.5 50 200 12.5 50 200
E
ladder
Cy3/Cy5
Cy3
Cy5
75 kDa
25 kDa
F
Cy3
Cy3-U6•Prp24
Cy3-U6
Prp24(Cy5)
Cy5
U6•Prp24(Cy5)
U6•Prp24
U6

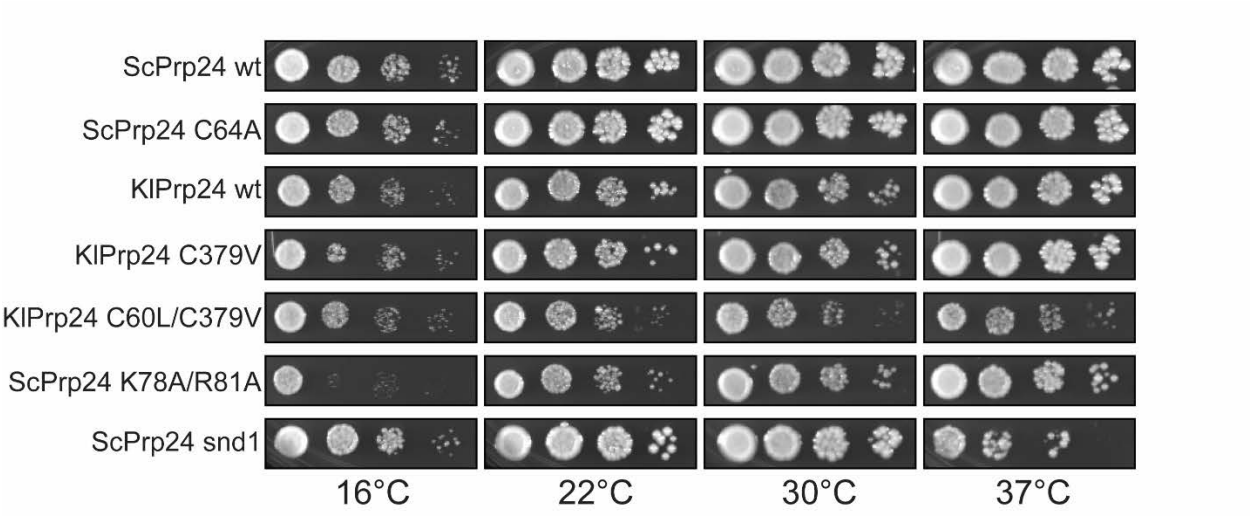
A6.3.7 KlPrp24 can complement ScPrp24 *in vivo*

Since KlPrp24 can bind and anneal ScU6 and U4 *in vitro*, we expected that KlPrp24 could complement *S. cerevisiae PRP24*. Indeed, KlPrp24 expressed from a shuffle plasmid containing the flanking ScPrp24 up and downstream regions can functionally replace ScPrp24 (**Figure A6-5**). Yeast with KlPrp24 as the only copy of Prp24 have a mild growth phenotype at temperatures below 37°C (**Figure A6-5**). However, this cold sensitivity is not as severe as that of an electropositive groove mutant of ScPrp24 (ScPrp24 K78A/R81A) that is known to affect annealing *in vitro* (DIDYCHUK et al. 2016). The mild cold sensitivity for KlPrp24 is consistent with its ~2-fold slower annealing kinetics (**Figure A6-2D**).

We then asked if the yeast could survive with KlPrp24 lacking its native cysteines. KlPrp24-C379V behaved similarly to wild-type KlPrp24, with a mild cold-sensitive growth phenotype (**Figure A6-5**). It is unsurprising that the C379V mutation has little effect as this mutation does not greatly reduce the affinity of KlPrp24 for U6 *in vitro* (**Figure A6-4B**). Interestingly, mutation of the first cysteine (C60L) in combination with C379V does not affect growth at low temperatures, but does introduce a temperature sensitive phenotype at 37°C similar to the phenotype seen in the ScPrp24-snd1 mutation, which destabilizes a protein-protein interaction between Prp24 and Lsm2 (MONTEMAYOR *et al.* in prep). This result is surprising, as this region of Prp24 is not expected to form a protein-protein contact with Lsm2-8 (MONTEMAYOR *et al.* in prep). Additionally, in ScPrp24, this cysteine (C64 in ScPrp24), while conserved, has no phenotype upon mutation to alanine (**Figure A6-5**) or serine (MONTEMAYOR *et al.* 2017). Intriguingly, in one crystal structure of the U6 snRNP core, this cysteine formed an intramolecular cross-link with H63 and K99 of ScPrp24 of unknown function (MONTEMAYOR *et al.* 2017). The function of ScPrp24-C64/KlPrp24-C60, and of RRM1 in general, remains to be determined.

Figure A6-5. KlPrp24 can complement ScPrp24 *in vivo.*

KlPrp24 can complement ScPrp24 but has a mild growth defect at 30°C and low temperatures. KlPrp24-C379V behaves similarly to KlPrp24 wt, but mutation of both cysteines in KlPrp24 (C60L/C379V) results in an additional temperature sensitive growth phenotype. For comparison, ScPrp24-C64A is included as an example of a mutation in Prp24 that lacks a growth phenotype (MONTEMAYOR *et al.* 2017), ScPrp24-K78A/R81A is included as an example of a mutation in Prp24 with a cold-sensitive growth phenotype (Appendix 8), and ScPrp24-snd1 is included as an example of a mutation in Prp24 with a temperature-sensitive phenotype (MONTEMAYOR *et al.* in prep).



A6.4 References

Baker, N. A., D. Sept, S. Joseph, M. J. Holst and J. A. McCammon, 2001 Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A 98**:** 10037-10041.

Didychuk, A. L., E. J. Montemayor, D. A. Brow and S. E. Butcher, 2016 Structural requirements for protein-catalyzed annealing of U4 and U6 RNAs during di-snRNP assembly. Nucleic Acids Res 44**:** 1398-1410.

Gietz, R. D., and R. A. Woods, 2002 Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. Methods Enzymol 350**:** 87-96.

Kelley, L. A., S. Mezulis, C. M. Yates, M. N. Wass and M. J. Sternberg, 2015 The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10**:** 845-858.

Montemayor, E. J., E. C. Curran, H. H. Liao, K. L. Andrews, C. N. Treba *et al.*, 2014 Core structure of the U6 small nuclear ribonucleoprotein at 1.7-A resolution. Nat Struct Mol Biol 21**:** 544-551.

Montemayor, E. J., A. L. Didychuk, H. Liao, P. Hu, D. A. Brow *et al.*, 2017 Structure and conformational plasticity of the U6 small nuclear ribonucleoprotein core. Acta Crystallogr D Struct Biol 73**:** 1-8.

Montemayor, E. J., A. L. Didychuk, G. K. Sidhu, A. D. Yake, D. A. Brow *et al.*, in prep.

Rader, S. D., and C. Guthrie, 2002 A conserved Lsm-interaction motif in Prp24 required for efficient U4/U6 di-snRNP formation. RNA 8**:** 1378-1392.

Schaffrath, R., and K. D. Breunig, 2000 Genetics and molecular physiology of the yeast Kluyveromyces lactis. Fungal Genet Biol 30**:** 173-190.

Appendix 7: Additional insights into Prp24-mediated annealing

This work was performed with undergraduate researchers Andrew DeLaitsch and Matthew Larson.

A7.1 Overview

Despite identification of important structural features in U6 and Prp24 for Prp24-mediated annealing (DIDYCHUK *et al.* 2016), the mechanism of annealing remains poorly understood. The mechanistic role of the electropositive groove is undefined, as is the mechanism by which Lsm2-8 enhances Prp24-mediated annealing. Furthermore, the order of U4/U6 stem formation and the initiation point for stem exchange are unknown. The contribution of U4 snRNP proteins to free U4 snRNP structure is unknown and it is unknown if these proteins contribute to Prp24-mediated annealing in yeast. U4/U6 annealing in humans is driven by RNA-RNA interactions between U4 and U6 and also between protein-protein interactions between hPrp24 and hPrp3 (MEDENBACH *et al.* 2004). Finally, the mechanism by which Prp24 dissociates from U4/U6 and how the interlocked U6-Prp24 topology is resolved is unknown. Prp24 remains tightly bound to the U4/U6 product *in vitro* (GHETTI *et al.* 1995; DIDYCHUK *et al.* 2016), yet is displaced after U4/U6 di-snRNP formation (SHANNON AND GUTHRIE 1991; JANDROSITZ AND GUTHRIE 1995). In this appendix, I describe the effect of destabilizing and stabilizing mutations in U4 RNA on annealing rate, the contribution of U4 snRNP proteins Snu13 and Prp31 to Prp24-mediated annealing, and experiments that narrow down the binding site of Prp24 on U4/U6. This work will form the basis for further understanding of the Prp24-mediated annealing mechanism.

A7.2 Materials and methods

A7.2.1 Protein production

Saccharomyces cerevisiae Prp24 protein was expressed and purified as previously described (DIDYCHUK *et al.* 2016). Briefly, Prp24 was expressed in *E. coli* Star (DE3) pLysS cells (Invitrogen) from a modified pET15b plasmid encoding an N-terminal decahistidine tag, TEV cleavage site, and Prp24 residues 1-444. Cells were grown in LB at 37°C to an OD_{600} of 0.6-0.8, and induced by addition of 1 mM IPTG and growth at 20°C for 20 hours. Cells were collected by centrifugation, resuspended in IMAC buffer (500 mM NaCl, 50 mM HEPES acid, 50 mM sodium HEPES base, 15 mM imidazole base, 10% glycerol, 1 mM TCEP-HCl) supplemented with DNase I and lysozyme. Cells were lysed via sonication and insoluble material was removed by centrifugation. Prp24 was purified by Ni-NTA agarose chromatography with step elution using IMAC buffer supplemented with 500 mM imidazole. Eluate and 1 mg TEV protease were dialyzed at 4°C overnight into IEX buffer (100 mM NaCl, 10 mM HEPES acid, 10 mM HEPES base, 10% glycerol, 1 mM TCEP-HCl, pH ~7.0), precipitated protein was removed by centrifugation, then Prp24 was purified via cation-exchange chromatography (HiTrap S, GE Healthcare) in IEX buffer with gradient elution against IEX buffer containing 2 M NaCl.

The fSNAP-Prp24 fusion protein was expressed from a modified pET-Duet1 plasmid that included an N-terminal hexahistidine tag, TEV protease site, and fSNAP tag. Briefly, the fSNAP open reading frame with N-terminal hexahistidine tag and TEV protease site was PCR amplified from p11-fSNAP with primers to add NcoI and BamHI restriction sites and subcloned into a pET-Duet1 vector (EMD Biosciences). The Prp24 open reading frame was amplified from *S. cerevisiae* genomic DNA with primers to add BamHI and SalI restriction sites and cloned into the pET-Duet1-fSNAP vector. The fSNAP-Prp24 fusion protein was expressed in *E. coli* Star (DE3) pLysS cells and purified as described above. Protein was labeled with a two-fold molar excess of SNAP-Surface 649 (New England Biolabs) in 150 mM NaCl, 50 mM HEPES pH 7.5, 1 mM DTT at 4°C

overnight. Excess dye was removed using a PD Minitrap G-25 column (GE Healthcare) in labeling buffer.

Full-length Prp31 (FL; 1-494) and C-terminally truncated Prp31 (ΔC; 1-343) were expressed from a modified pGEX-6P-1 plasmid containing an N-terminal GST tag, hexahistidine tag, and TEV protease cleavage site (a kind gift from Kiyoshi Nagai) (HARDIN *et al.* 2015). Snu13 was expressed from a modified pET15b plasmid encoding an N-terminal decahistidine tag, TEV cleavage site, and the Snu13 residues 1-126. Prp31 FL, Prp31ΔC, and Snu13 were expressed and purified as above with the exception that cells were grown after induction at 16°C for 24 hours. All protein samples were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis to assess purity.

A7.2.2 RNA production

U4, U6 and U4/U6 RNAs were *in vitro* transcribed using recombinant His_6-tagged T7 RNA polymerase. U4 RNA was transcribed from a modified pUC118 plasmid template containing the T7 polymerase promoter sequence, two additional G nucleotides at the transcription start site, and a BsaI restriction site at the end of the template to allow for run-off transcription (**Figure A7-1A, 6**), or from a pUC57 plasmid containing the T7 polymerase promoter sequence, a 5' hammerhead ribozyme, U4, a 3' HDV ribozyme, and a BsaI restriction site (**Figure A7-1B, 2, 3**). U6 RNA was transcribed from a modified pUC57 plasmid containing the T7 polymerase promoter sequence, an additional G nucleotide at the beginning of the transcript, a 3' HDV ribozyme, and a BamHI restriction site. Linked U4/U6 constructs were transcribed from a modified pUC57 plasmid containing the T7 polymerase promoter sequence, and varying sequences corresponding to U4 and U6 followed by a 3' HDV ribozyme (**Figure A7-7, 8**). Mutations were introduced via inverse PCR with Phusion DNA polymerase in HF buffer, followed by Dpn1 treatment and self-ligation using T4 PNK and T4 DNA ligase, and transformation into *E. coli* 5α competent cells (all New England Biolabs).

In vitro transcription reactions contained 0.1-0.2 ng/µL template dsDNA, stoichiometric NTPs (7.2 mM ATP, 4.2 mM CTP, 4.2 mM GTP, 4.4 mM UTP), 40 mM Tris-Cl pH 8.0, 1 mM spermidine, 0.01% Triton X-100, 38 mM $MgCl_2$, and 5 mM dithiothreitol. The transcription reaction was quenched by addition of 45 mM EDTA and RNAs were purified via denaturing 10% 19:1 acrylamide:bis-acrylamide gel electrophoresis containing 8 M urea, 89 mM Tris borate, 2 mM EDTA. RNA was visualized by UV shadowing and extracted from the gel via passive diffusion into 0.3 M sodium acetate, 1 mM EDTA, pH 5.2. RNA was further purified via anion-exchange (HiTrap Q, GE Healthcare) using IEX buffer (100 mM NaCl, 20 mM bis-tris, 10 mM HCl, 1 mM TCEP-HCl, pH ~6.2), with step elution using IEX buffer containing 2 M NaCl. RNA was concentrated using a 10 kDa spin filter and concentration was determined using the A_{260} and calculated extinction coefficients.

Full-length 3'-Cy5-labeled U4 (nucleotides 1-160) used in **Figure A7-4** was produced via splinted ligation using a synthetic RNA oligonucleotide consisting of U4 145-160 with a 3'-Cy3 moiety and *in vitro* transcribed U4 1-145 that was produced from a pUC57 plasmid containing a 5' hammerhead and a 3' HDV ribozyme. A DNA oligonucleotide complementary to nucleotides 120-160 of U4 was used as a splint. U4 1-145 was treated with T4 PNK in the absence of ATP to remove the 3' cyclic phosphate and the U4 145-160-Cy3 RNA was 5' phosphorylated with T4 PNK in the presence of ATP. After splinted ligation, the resulting U4 1-160-Cy5 was purified by urea PAGE and ion exchange as described above.

U6 RNA binding assays in **Figure A7-5** used a 5'-Cy3-labeled RNA containing U6 nucleotides 1-112 that was produced via splinted ligation using a 5'-Cy3 U6 1-12 RNA oligonucleotide (Integrated DNA Technologies) and U6 13-112 that was produced via *in vitro* transcription from a modified plasmid containing a 5' hammerhead ribozyme and a 3' HDV ribozyme. The U6 13-112 RNA was phosphorylated with ATP and T4 PNK, which also removed the 3' cyclic phosphate produced by 3' HDV cleavage. The Cy3-U6 1-12 and U6 13-112 RNAs were ligated using T4 RNA ligase 2 (New England Biolabs) at 37°C for 2 hours using a DNA splint

that was complementary to U6 nucleotides 1-30, with a Cy3-U6:splint:U6 13-112 ratio of 2:1.5:1. The ligation product was purified by urea PAGE and ion exchange as above.

A7.2.3 Gel shift assays

RNAs were radiolabeled as previously described (DIDYCHUK *et al.* 2016). Briefly, RNAs were dephosphorylated with 10 units of CIP (New England Biolabs), extracted with phenol:chloroform:isoamyl alcohol (25:24:1), and ethanol precipitated. RNAs were pelleted by centrifugation and resuspended in water prior to labeling with [γ-^{32}P] ATP (3000 Ci/mmol). RNA was gel purified as described above, extracted by passive diffusion into 0.3 M sodium acetate, 1 mM EDTA, ethanol precipitated, and resuspended in water.

Binding of RNAs with Prp24 were carried out using trace (<1 nM) radiolabeled RNA or 0.5 - 2 nM fluorescently labeled RNA. RNAs were refolded by heating to 90°C for 2 min in 2x RNA glycerol binding buffer (100 mM KCl, 20% glycerol, 10 mM HEPES acid, 10 mM sodium HEPES base, 1 mM EDTA acid, 1 mM TCEP-HCl, 0.01% Triton X-100, pH~ 7.0, 0.2 mg/mL tRNA, 0.02 mg/mL sodium heparin; **Figure A7-1,2,3,6,7,8**) or 2x RNA sucrose binding buffer (100 mM KCl, 20% sucrose, 20 mM bis-tris, 10 mM HCl, 1 mM EDTA acid, 1 mM TCEP-HCl, 0.01% Triton X-100, pH~ 6.5, 0.2 mg/mL tRNA, 0.02 mg/mL sodium heparin; **Figure A7-4,5**) and snap cooled on ice. Proteins were prepared as 2x stocks in protein binding buffer (RNA buffer lacking tRNA and sodium heparin and containing 0.2 mg/mL BSA). Binding reactions contained equal volumes of 2x RNA and 2x protein stocks. Samples were incubated at 22°C for 20 min prior to loading onto 16.5 × 22 cm 6% polyacrylamide gels (29:1 acrylamide:bis-acrylamide, 89 mM Tris borate, 2 mM EDTA pH 8.0). Gels were run for 1.5-2 h at 3 W or for 2-3 hours at 150V at 4°C.

Annealing reactions were carried out at 22°C-30°C in binding buffer (as described above) containing <5 nM radiolabeled U4 RNA, 25 nM U6 1-112 RNA and 250 nM Prp24 protein. Annealing reactions shown in **Figure A7-4,5** contained 2 nM U4 1-160-Cy5, 10 nM U6 G1-112, 200 nM Prp24, and if indicated 100 nM Snu13 or 100 nM Snu13 and 100 nM Prp31. RNAs were refolded by heating to 90°C for 2 min in 2x RNA binding buffer followed by snap cooling on ice.

Reactions were quenched by the addition of 2 µl of proteinase K buffer (0.5% sodium dodecyl sulphate, 0.3 mg/ml tRNA, 5 mM $CaCl_2$, 30 mM HEPES pH 7.0, 0.2 mg/ml proteinase K) prior to separation on a 6% polyacrylamide gel (29:1 acrylamide:bis-acrylamide, 89 mM Tris borate, 2 mM EDTA pH 8.0). Gels were run 1.5-2 h at 3 W or for 2-3 hours at 150V at 4°C.

Radioactive gels were dried on Biorad filter paper, exposed to a PhosphorImager screen, and imaged on a Typhoon FLA 9000 imager (GE Healthcare Life Sciences). Fluorescent gels were imaged directly through low fluorescence glass plates (CBS Scientific) using a Typhoon FLA 9000 imager (GE Healthcare Life Sciences). Annealing rates were calculated using the ratio of free U4 to U4/U6 in Proteinase K-treated lanes. Resulting data were then fit using non-linear regression in GraphPad Prism 4 to a one-phase exponential association equation: $Y = Y_{max}*(1-e\text{^}(-k t))$ where Y is the % U4 in U4/U6, *k* is the rate constant, and t is the time. Unless otherwise indicated, annealing rates are reported for three technical replicates.

A7.3 Results and discussion

A7.3.1 Changing the 5' sequence of U4 RNA affects annealing *in vitro*

Previous annealing experiments used a U4 construct that contained two additional 5' guanines to facilitate *in vitro* transcription by T7 RNA polymerase (**Figure A7-1A**) (DIDYCHUK *et al.* 2016). To test whether these additional nucleotides (or the potential non-templated nucleotides at the 3' end from runoff transcription) affected annealing *in vitro*, I generated a new plasmid that contained a 5' hammerhead that undergoes self-cleavage to produce a 5'OH followed by the native sequence (**Figure A7-1B**). The 5'OH was a fortunate product as it allowed direct radiolabeling via T4 PNK (bypassing the CIP step). This construct also had a 3' HDV ribozyme, which self cleaves to produce a 2',3'-cyclic phosphate at the 3' end of U4, which is the lowest affinity 3' end modification for yeast Lsm2-8 binding (DIDYCHUK *et al.* 2017), and therefore helps to prevent non-specific Lsm2-8 binding to U4. To my surprise, the new construct lacking the additional guanines annealed to a greater extent and with faster kinetics than the previous construct (**Figure A7-1C**). Fortunately, the observations made for *in vitro* annealing reported in Didychuk *et al.* 2016 are also true for the new construct. Mutations within the electropositive groove of Prp24 (Prp24-R131A/R134A) reduce annealing (**Figure A7-1D**) and mutation of U6 to stabilize the base of the telestem greatly increases the annealing rate (**Figure A7-1E**). I propose that this new construct more closely resembles the *in vivo* U4 snRNA, as the old construct is predicted by Mfold (ZUKER 2003) to adopt a fold that does not contain the U4 5'SL (**Figure A7-1A**) that is bound by Snu13 and Prp31 (MARMIER-GOURRIER *et al.* 2003; DOBBYN AND O'KEEFE 2004; LIU *et al.* 2007; NGUYEN *et al.* 2016). In contrast, the new construct is predicted to contain the U4 5'SL (**Figure A7-1B**). This serves as a lesson of how we must make every attempt to design *in vitro* experiments to mimic *in vivo* reality as closely as possible, and is a reminder that non-templated nucleotides in RNA (much like cloning scars in proteins) can affect *in vitro* results.

Figure A7-1. Additional 5' guanosines in U4 affect annealing *in vitro.*

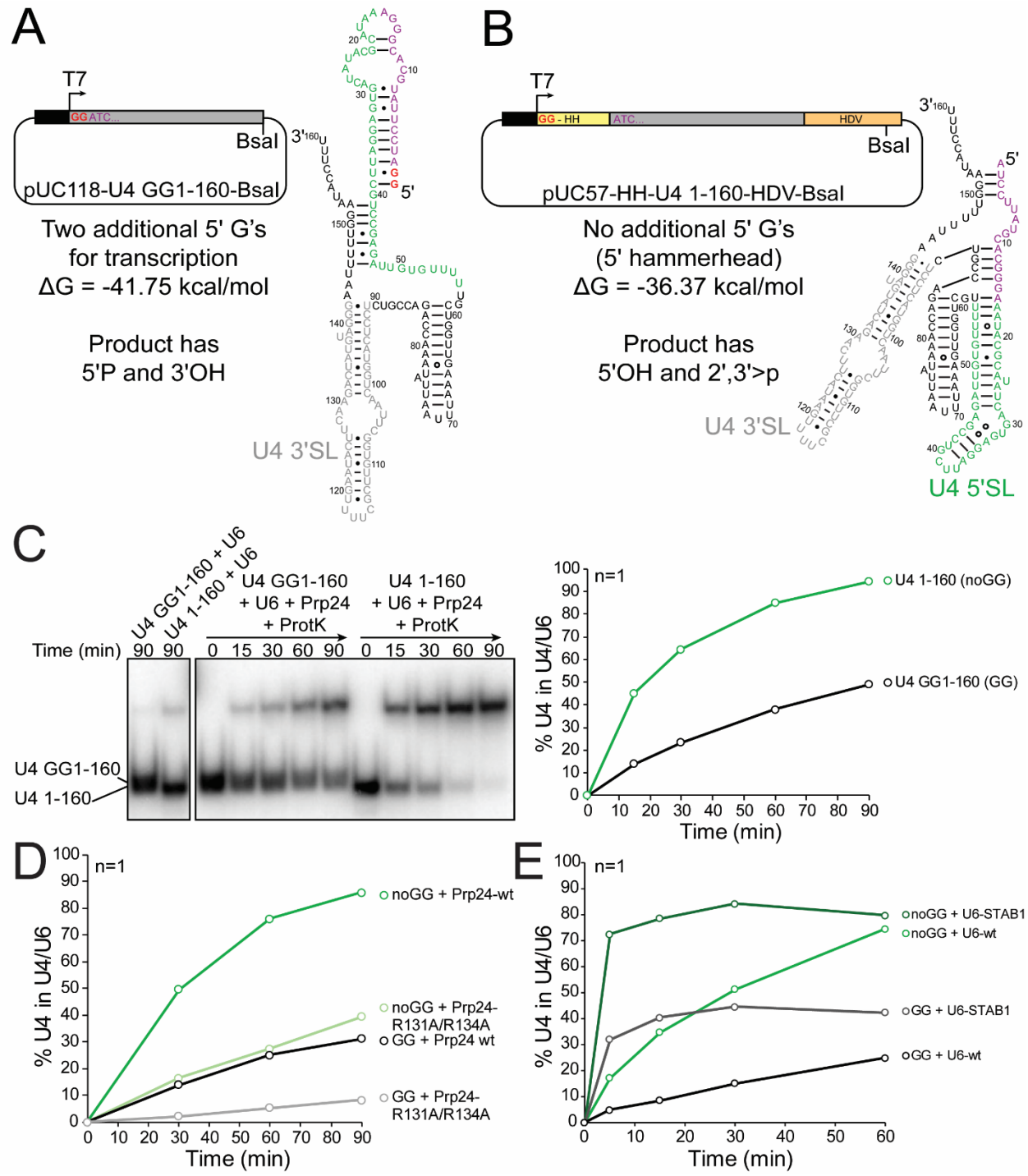
A) The U4 construct used in Didychuk *et al.* 2016 was transcribed *in vitro* from pUC118-U4 GG1-160-BsaI, which has two additional 5' guanosines for efficient transcription. The product contains a 5' phosphate and 3'OH. The predicted fold for this RNA (ZUKER 2003) contains the U4 3'SL (grey) but does not contain the U4 5'SL (green). Additionally, this fold sequesters the nucleotides that form U4/U6 Stem II (purple) in a stem. The predicted stability of this fold is -41.75 kcal/mol.

B) A new U4 construct lacking the additional 5' guanosines was transcribed *in vitro* from pUC57-HH-U4 1-160-HDV-BsaI, which contains a 5' hammerhead ribozyme and 3' HDV ribozyme that undergo self-cleavage to product U4 1-160 with a 5'OH and 3' cyclic phosphate. The Mfold-predicted fold contains the U4 3'SL (grey) and 5'SL (green), with the U4/U6 Stem II-forming nucleotides (purple) less stably paired than in (A).

C) Left: U4 1-160 lacking 5' additional guanines is annealed to U6 RNA faster than U4 containing additional guanines. Protein-free annealing is also more significant for U4 lacking the 5' additional guanines. Right: Quantification of annealing assay.

D) An electropositive groove mutant, Prp24-R131A/R134A (DIDYCHUK *et al.* 2016) reduces U4/U6 annealing for U4 with and without 5' additional guanines.

E) Stabilization of the U6 telestem (U6-STAB1) (DIDYCHUK *et al.* 2016) improves Prp24-mediated annealing for U4 with and without 5' additional guanines.


A
T7
GG ATC...
BsaI
pUC118-U4 GG1-160-BsaI
Two additional 5' G's
for transcription
ΔG = -41.75 kcal/mol
Product has
5'P and 3'OH
U4 3'SL
B
T7
GG - HH
ATC...
HDV
BsaI
pUC57-HH-U4 1-160-HDV-BsaI
No additional 5' G's
(5' hammerhead)
ΔG = -36.37 kcal/mol
Product has
5'OH and 2',3'>p
U4 3'SL
U4 5'SL
C
U4 GG1-160 + U6
U4 1-160 + U6
U4 GG1-160
+ U6 + Prp24
+ ProtK
U4 1-160
+ U6 + Prp24
+ ProtK
Time (min)
90 90
0 15 30 60 90
0 15 30 60 90
U4 GG1-160
U4 1-160
n=1
% U4 in U4/U6
Time (min)
U4 1-160 (noGG)
U4 GG1-160 (GG)
D
n=1
% U4 in U4/U6
Time (min)
noGG + Prp24-wt
noGG + Prp24-R131A/R134A
GG + Prp24 wt
GG + Prp24-R131A/R134A
E
n=1
% U4 in U4/U6
Time (min)
noGG + U6-STAB1
noGG + U6-wt
GG + U6-STAB1
GG + U6-wt

A7.3.2 Prp24-mediated annealing is efficient at lower temperatures

Previous annealing reactions were carried out at 30°C to mimic the optimum growth temperature for yeast (DIDYCHUK *et al.* 2016). To test the effect of temperature on annealing, I measured Prp24-mediated annealing rates at 10, 22, 30, and 37°C (**Figure A7-2A,B,D**). The annealing rates and extents at 10, 22, and 30°C were similar, while the annealing rate at 37°C was slightly slower (**Figure A7-2A,B,D**). As expected, the efficiency of protein-free annealing increased with increasing temperature (**Figure A7-2C,D**). Although the annealing rates at 10, 22, and 30°C are similar, the increase in protein-free annealing means that Prp24-mediated annealing is likely most efficient at lower temperatures. The reduction in Prp24-mediated annealing rate at high temperatures (37°C) could be due to disruption of U6 RNA secondary structure that is important for tight Prp24 binding or due to destabilization of the U4/U6 interaction. For simplicity, subsequent experiments were carried out at 22°C.

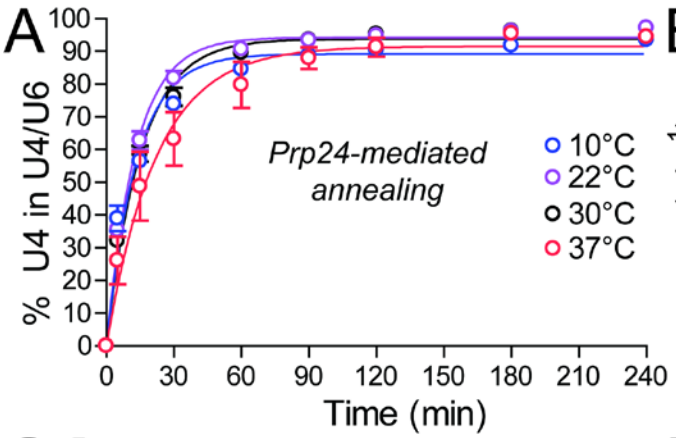
Figure A7-2. Prp24 anneals U4 and U6 *in vitro* efficiently over a range of temperatures.

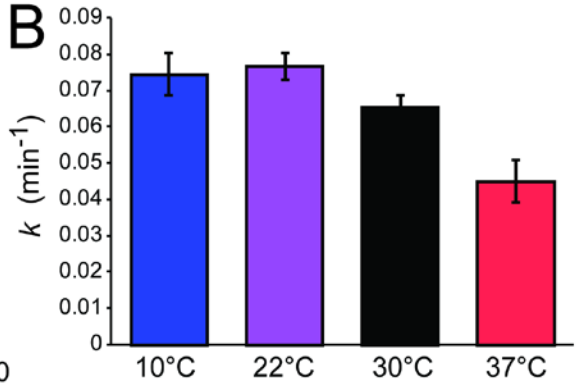
A) Prp24-mediated annealing is efficient over the range of 10°C - 37°C.

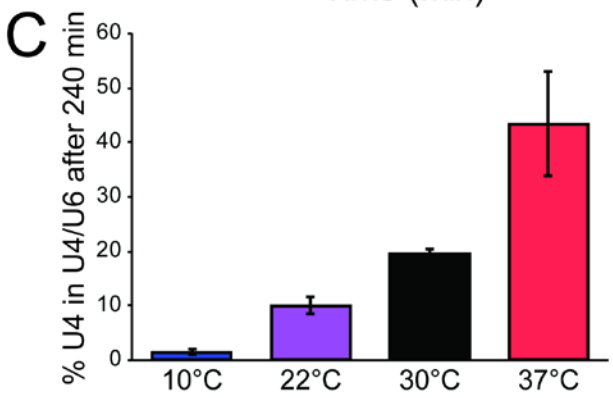
B) Graph of Prp24-mediated annealing rates for 10°C, 22°C, 30°C, and 37°C.

B) Protein free-annealing increases with increasing temperatures as measured by the amount of U4 in U4/U6 after 240 minutes in the absence of Prp24.

D) Annealing rates and extents for Prp24-mediated annealing at different temperatures and the observed % U4 in U4/U6 after 240 minutes that is a product of protein-free annealing.







D

	k (min^{-1})	Y_{max}(% U4)	% U4 in U4/U6 after 240 min
10°C	0.075 ± 0.006	89 ± 1	1.5 ± 0.5
22°C	0.077 ± 0.004	94 ± 1	10 ± 2
30°C	0.065 ± 0.003	94 ± 1	20 ± 1
37°C	0.045 ± 0.006	91 ± 3	40 ± 10

A7.3.3 Effect of mutations in U4 RNA on annealing *in vitro*

To further understand the annealing mechanism, I sought to test the effect of mutations in U4 RNA. The annealing rate with different U4 mutations in U4/U6 Stem I, II, and III were measured (**Figure A7-3A**). Disrupting a potential base pair in U4/U6 Stem III (U4-G84A) has no effect on annealing rate *in vitro* (**Figure A7-3B**). U4/U6 Stem III does not form in the U4/U6.U5 tri-snRNP and this region of U4 is bound by Brr2 (NGUYEN *et al.* 2016). While the stem could form in the U4/U6 di-snRNP (where Brr2 is not present), the similar annealing rate suggests that this stem does not significantly contribute to U4/U6 annealing or U4/U6 stability *in vivo*.

In contrast, mutations in U4/U6 Stem I (**Figure A7-3C**) and Stem II (**Figure A7-3D**) affect Prp24-mediated annealing *in vitro.* Interestingly, mutations that destabilize G-C base pairs have different effects based on their position in the stem. In general, destabilizing mutations that are further from the effective ends of helices or base mismatches have a larger effect on annealing (either amplitudes, rates, or both). Note that it is likely that U4/U6 stems I and II are either partially or fully coaxially stacked after annealing (CORNILESCU *et al.* 2016; NGUYEN *et al.* 2016). Effects on the amplitudes of annealing may result from thermodynamic destabilization of the annealed U4/U6 product. U4-G58C is worse than U4-G61C (**Figure A7-3C**), which may indicate that the more central U4-G58 nucleotide plays a greater thermodynamic role in annealing than U4-G61. Similarly, the more central U4-C4A mutation is more deleterious than U4-C3A (**Figure A7-3D**). U4-G9C is worse than U4-C10A (**Figure A7-3D**), perhaps because U4-C10 is closer to the C-C mismatch in Stem II. Interestingly, U4-G13C has little effect, potentially because it is directly adjacent to the C-C mismatch in Stem II (**Figure A7-3D**). U4-G14C and U4-G15C have increasingly detrimental effects (**Figure A7-3D**). U4-G15C is by far the most deleterious mutation, with drastically reduced rates and amplitudes of annealing. This is consistent with U4-G15 playing a critical role in annealing, perhaps by playing a role in initiation of duplex formation during annealing. Interestingly, despite the fact that U4-C10, like U4-G13, is adjacent to the C-C mismatch, the U4-C10A mutation is more deleterious than U4-G13C (**Figure A7-3D**). This could

suggest that U4-C10 plays a critical role in annealing and could support a kissing loop model of duplex formation initiation.

The observation that mutations in U4 affect both the annealing kinetics and the maximum extent of annealing (presumably by destabilizing the U4/U6 product) makes unambiguous interpretation of these results difficult at this stage. Additionally, mutations in U4 likely affect free U4 structure, which could in turn affect annealing (see **A7.3.1**). To tease out the different contributing factors, the stability and structure of free U4 and U4/U6 could be measured using melting curves and chemical probing. Determining which region(s) of U4/U6 are most important for annealing rate independent of stability could aid in determining which regions form first. This could help differentiate between potential models of annealing (i.e. kissing loop formation between U6 nucleotides 71-75 and U4 nucleotides 6-10 versus formation of Stem I followed by unwinding of the U6 ISL and Stem II formation). A kissing loop model would predict the most important U4 nucleotides to be C10, G9, U8, A7 and U6 because these positions pair with the U6 ISL loop nucleotides G71, C72, A73, U74, and A75. As discussed above, a strong effect of mutations at U4-G9 and C10 was observed. The kissing loop hypothesis could be further tested by investigating the effects of mutations at U4-U8 and U4-U6. Positions that have a larger effect on annealing rate than would be predicted based on their destabilization of U4/U6 may be important for initial recognition and strand exchange.

Destabilizing mutations in U4 and U6 can be rescued by combining them in a compensatory fashion. Disrupting the U4-G9/U6-C72 base pair via mutation in either U4 (U4-G9C) or U6 (U6-C72G) reduces the annealing rate significantly (**Figure A7-3E**). However, combining the two variants rescues the annealing rate and extent to wild-type levels (**Figure A7-3E**). This pair of mutations was selected because U6-C72 is in a loop at the top of the U6-ISL, which allows mutation without affecting free U6 structure or ablating U6-Prp24 contacts. This result parallels experiments performed in yeast, where U4-G9C causes a cold sensitive

phenotype that is rescued by U6-C72G (SHANNON AND GUTHRIE 1991). Interestingly, U6-C72G does not appear to cause a cold sensitive phenotype by itself (SHANNON AND GUTHRIE 1991).

Next, I sought to test whether mutants expected to stabilize U4/U6 improved annealing. U4-C12G is expected to replace the C-C mismatch with a G-C base pair in U4/U6 Stem II (**Figure A7-3F**). U4-A16G is expected to shift base pairing to produce a new G-C base pair (**Figure A7-3F**). Both mutations are expected to accelerate annealing via stabilization of the product U4/U6. Indeed, both U4-C12G and A16G improve Prp24-mediated annealing rate by ~30% (**Figure A7-3G**). Additionally, these stabilizing mutations can act as cis suppressors of the destabilizing mutation U4-G14C (**Figure A7-3H**). The double mutant U4-C12G/G14C has slower annealing kinetics than wild-type yet has a similar maximum extent annealed (**Figure A7-3H**). In contrast, the U4-G14C/A16G mutant has faster kinetics than wild-type, yet has a lower maximum extent annealed (**Figure A7-3H**). These differential effects could be caused by effects on free U4 stability or conformation. Finally, U4-A16G can also act as a trans suppressor of U6-A62G slow annealing (**Figure A7-3I**). U4-A16G also suppresses the cold-sensitivity of U6-A62G *in vivo* (VIDAVER *et al.* 1999). Thus, mutations in U4 RNA that stabilize U4/U6 Stem II can suppress annealing defects caused by destabilizing mutations in U4 or U6 RNA. This in turn indicates that there are no strict sequence-specific requirements for Prp24-mediated annealing at these positions. Continued work to tease out the relative contributions of free U4 structure and U4/U6 stability will provide new insight into the annealing mechanism.

Figure A7-3. Mutations in U4 RNA have complex effects on U4/U6 annealing *in vitro.*

A) Secondary structure of U4/U6 with mutation locations used in B-I highlighted.

B) Disruption of potential secondary structure in U4/U6 Stem III region via the U4-G84A substitution does not affect annealing kinetics or maximum % U4 in U4/U6.

C) Disruption of secondary structure in U4/U6 Stem I via the mutations U4-G58C and G61C have different effects, with U4-G61C more severely affecting maximum % U4 in U4/U6.

D) Disruption of secondary structure in U4/U6 Stem II has different effects, with U4-G13C having the least effect and G15C the most.

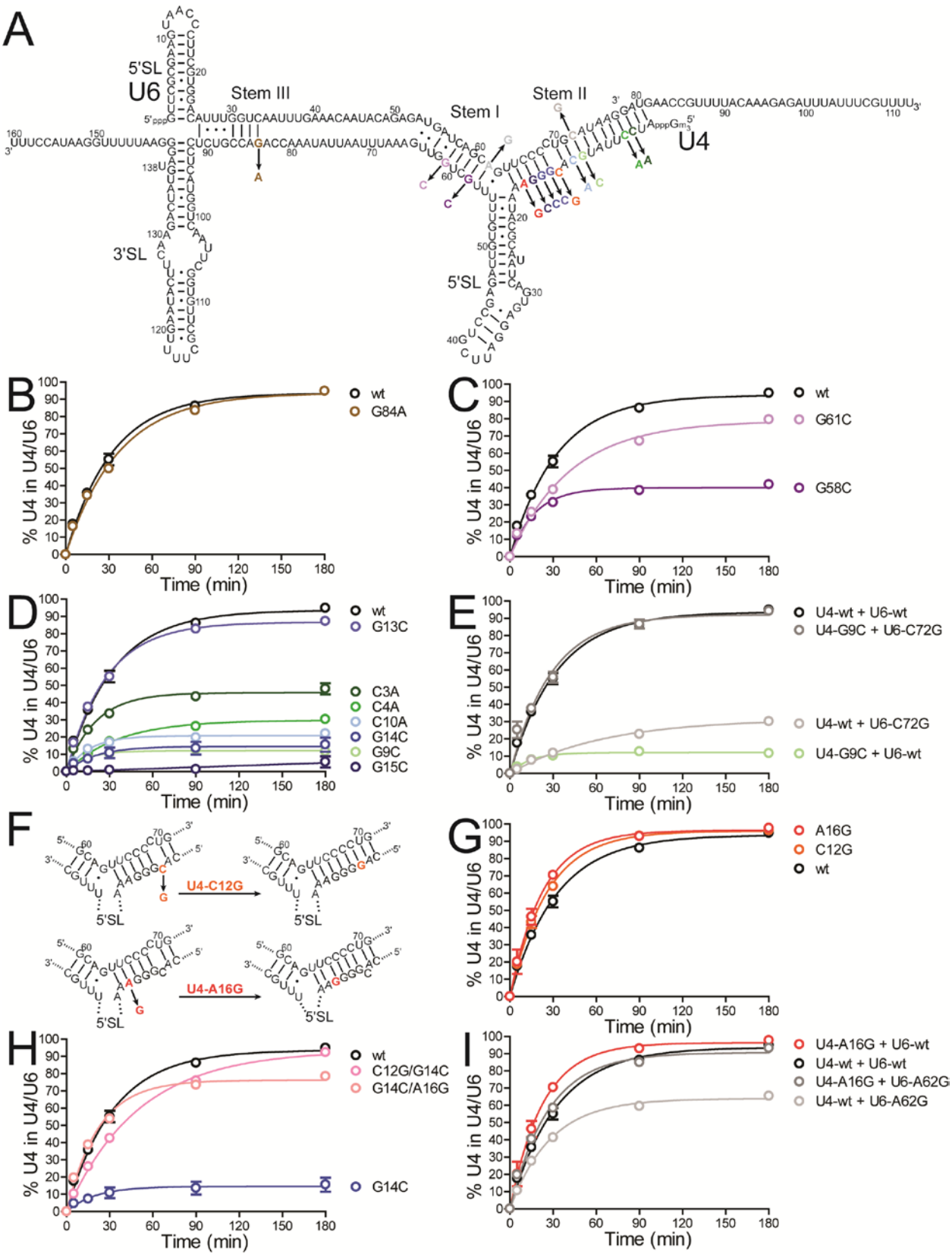
E) The U6-C72G mutation decreases annealing rate and maximum % U4 in U4/U6 similar to U4-G9C (its base pairing partner). Including both mutations restores annealing to wild-type rate and extent.

F) U4-C12G and U4-A16G mutations are expected to stabilize U4/U6 Stem II. U4-C12G seals the C-C mismatch at, while U4-A16G likely causes a rearrangement of basepairing to add a G-C base pair.

G) Stabilization of U4/U6 Stem II via U4-A16G and U4-C12G mutations improve annealing rates.

H) Stabilization of U4/U6 Stem II via U4-A16G and U4-C12G mutations rescues the destabilizing mutant U4-G14C.

I) Stabilization of U4/U6 Stem II via U4-A16G and U4-C12G mutations rescues the U6-A62G mutation, which reduces U4/U6 annealing by stabilizing the free U6 ISL.


A
5'SL
U6
Stem III
Stem I
Stem II
U4
3'SL
5'SL
B
% U4 in U4/U6
Time (min)
wt
G84A
C
wt
G61C
G58C
D
wt
G13C
C3A
C4A
C10A
G14C
G9C
G15C
E
U4-wt + U6-wt
U4-G9C + U6-C72G
U4-wt + U6-C72G
U4-G9C + U6-wt
F
U4-C12G
U4-A16G
G
A16G
C12G
wt
H
wt
C12G/G14C
G14C/A16G
G14C
I
U4-A16G + U6-wt
U4-wt + U6-wt
U4-A16G + U6-A62G
U4-wt + U6-A62G

A7.3.4 U4 snRNP proteins Snu13 and Prp31 do not contribute to annealing *in vitro*

We hypothesized that protein-protein interactions between Prp24 and U4 snRNP or U4/U6 di-snRNP proteins could contribute to annealing. Therefore, we sought to determine if the U4 snRNP proteins Snu13 and Prp31 affect Prp24-mediated annealing. First, we tested if Snu13 or Prp31 catalyze annealing in the absence of Prp24. After 90 minutes, <5% of U4 is incorporated into U4/U6 in the absence of Prp24 (**Figure A7-4A**; lane 1). Snu13 binds U4 tightly and remains bound to the U4/U6 product (**Figure A7-4A**; lane 2). Upon treatment with proteinase K, this resolves to free U4 and U4/U6, with similar U4/U6 levels to the protein-free control (**Figure A7-4A**; lane 3). Prp31 lacking its unstructured C-terminal domain (ΔC) binds U4 and U4/U6 in the presence of Snu13 (**Figure A7-4A**; lane 5). Upon proteinase K treatment, this resolves to free U4 and U4/U6 (**Figure A7-4A**; lane 6). Although more U4 is incorporated into U4/U6 in the presence of Snu13/Prp31ΔC, this difference is small (**Figure A7-4A**; lanes 4 and 6). We conclude that Snu13 and Prp31 bind U4 tightly and do not significantly contribute to annealing in the absence of Prp24.

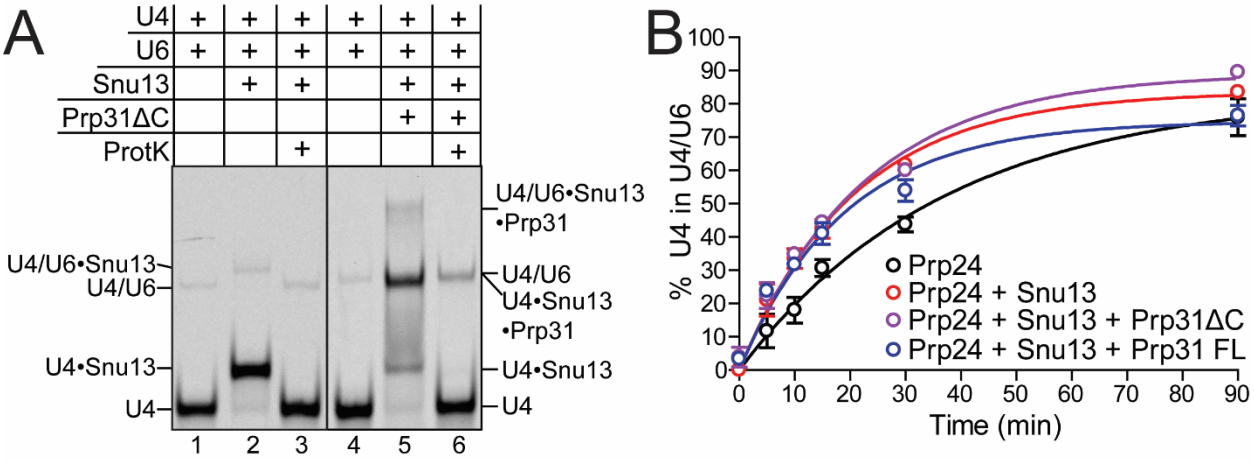
Next, Prp24-mediated annealing rates were measured in the presence of U4 pre-bound by Snu13, Snu13 with Prp31ΔC, and Snu13 with full-length Prp31 (FL) (**Figure A7-4B**). Inclusion of Snu13 increased the Prp24-mediated annealing rate ~1.7 fold (**Figure A7-4B, C**). Including Prp31ΔC or full-length Prp31 did not additionally stimulate annealing (**Figure A7-4B, C**). This subtle Snu13-mediated increase in annealing rate could be the result of stabilizing a conformation of U4 that contains the U4 5'SL (see **A7.3.1**). As Snu13 and Prp31 primarily bind the U4 5'SL and have minimal contact with U4/U6 Stem I and II (NGUYEN *et al.* 2016), they should stabilize the reactant (U4) and product (U4/U6) similarly and not contribute to annealing. In contrast, the U4/U6 di-snRNP specific protein Prp3 makes extensive contacts with U4/U6 Stem II and U6 80-88, which could stabilize U4/U6 and promote annealing. The effect of Prp3/Prp4 on Prp24-free and -mediated annealing should be tested *in vitro*.

Figure A7-4. U4 snRNP proteins do not significantly contribute to U4/U6 annealing *in vitro.*

A) Snu13 and Prp31ΔC do not significantly contribute to annealing in the absence of Prp24. U4 and U6 alone, with Snu13 (100 nM), or with Snu13 and Prp31ΔC (100 nM each) were incubated for 90 minutes. U4 is efficiently bound by Snu13 and Prp31ΔC, but after treatment with proteinase K minimal U4 has been incorporated into U4/U6.

B) The addition of Snu13 alone or Snu13 with Prp31 FL or Prp31ΔC moderately improves Prp24-mediated annealing rate.

C) Quantification of annealing rates and extent shown in (B).



C

U4/U6 annealing

	k (min^{-1})	Y_{max}(% U4)	fold change rel. to wt
Prp24	0.027 ± 0.004	83 ± 6	1
Prp24 + Snu13	0.048 ± 0.004	84 ± 3	1.8
Prp24 + Snu13 + Prp31ΔC	0.045 ± 0.004	89 ± 3	1.7
Prp24 + Snu13 + Prp31 FL	0.053 ± 0.005	75 ± 3	2.0

A7.3.5 An N-terminal fSNAP tag does not affect Prp24-mediated annealing

Fluorescently-labeled Prp24 could be used for a number of useful and informative experiments, including single-molecule colocalization studies involving U6, Prp24, other U6 binding proteins (including Usb1, Lsm2-8, and Lhp1) or experiments involving U4 RNA and its associated proteins. Additionally, a fluorescently-labeled Prp24 construct could be used in the *in vitro* annealing assay to monitor Prp24 release from U4/U6 in the presence of U4 snRNP proteins. The U4/U6 product with Snu13 and Prp31 bound is ~155 kDa, and unambiguous determination of whether Prp24 is still bound via gel mobility is difficult due to its large size. Inclusion of other relevant proteins, including Lsm2-8 (~87 kDa) or Prp3/Prp4 (108 kDa) increase the size and complexity of such an experiment. Thus, finding a fluorescent labeling solution for Prp24 that retains its binding and annealing characteristics is essential. *S. cerevisiae* Prp24 contains eight native cysteines, making site-specific labeling via thiol-specific conjugated dyes difficult (see **Appendix 6**).

Another method of fluorescent labeling is via a protein fusion. The fSNAP protein is relatively small (20 kDa) and forms a covalent bond with a benzylguanine fluorophores (JUILLERAT *et al.* 2003). Because the C-terminal domain of Prp24 is involved in binding to Lsm2-8 (FROMONT-RACINE *et al.* 2000; RADER AND GUTHRIE 2002), we used an N-terminal fSNAP fusion of Prp24 (**Figure A7-5A**). The fSNAP-Prp24 fusion binds U6 tightly, while the fSNAP domain alone does not bind U6 RNA (**Figure A7-5B**). Additionally, fSNAP-Prp24 anneals U4 and U6 RNAs *in vitro* with similar kinetics and maximum extent annealed to wild-type Prp24 (**Figure A7-5C**). Thus, the N-terminal fSNAP fusion of Prp24 behaves similarly to wild-type Prp24.

The fSNAP-Prp24 fusion protein can be labeled with a Cy5-conjugated benzylguanine dye and visualized in a U6 gel-shift assay (**Figure A7-5D**). Some fluorescently-labeled Prp24 does non-specifically enter the gel, but this could potentially be avoided by additional purification steps. Both wild-type Prp24 and fSNAP-Prp24 also tightly bind U6 that has been heat-annealed to a biotinylated DNA complementary to the U6 5'SL (**Figure A7-5D**). These materials (Cy3-labeled

U6 bound to a biotinylated DNA oligo and Cy5-labeled fSNAP-Prp24) could be used for a colocalization single molecule spectroscopy (CoSMoS) experiment wherein U6 is attached to the slide surface via the biotinylated DNA and fSNAP-Prp24 is in solution (**Figure A7-5E**). This could be used to monitor binding properties (k_{off} and K_d) for Prp24 and U6 and could be used to test the effect of mutations in Prp24 and U6. Additionally, they could be used for a three-color experiment where another U6-binding protein (i.e. Lhp1 or Lsm2-8) is fluorescently labeled. This could be used to test the effect of protein-protein interactions between Prp24 and Lsm2-8 that promote co-binding, or between Prp24 and Lhp1 that results in anti-cooperative binding as observed in gel shift assays (DIDYCHUK *et al.* 2017). It would be interesting to see if Prp24 actively displaces Lhp1 from U6 or if Prp24 binding merely prevents re-capture by Lhp1.

Figure A7-5. An N-terminal fSNAP fusion of Prp24 behaves like Prp24 *in vitro.*

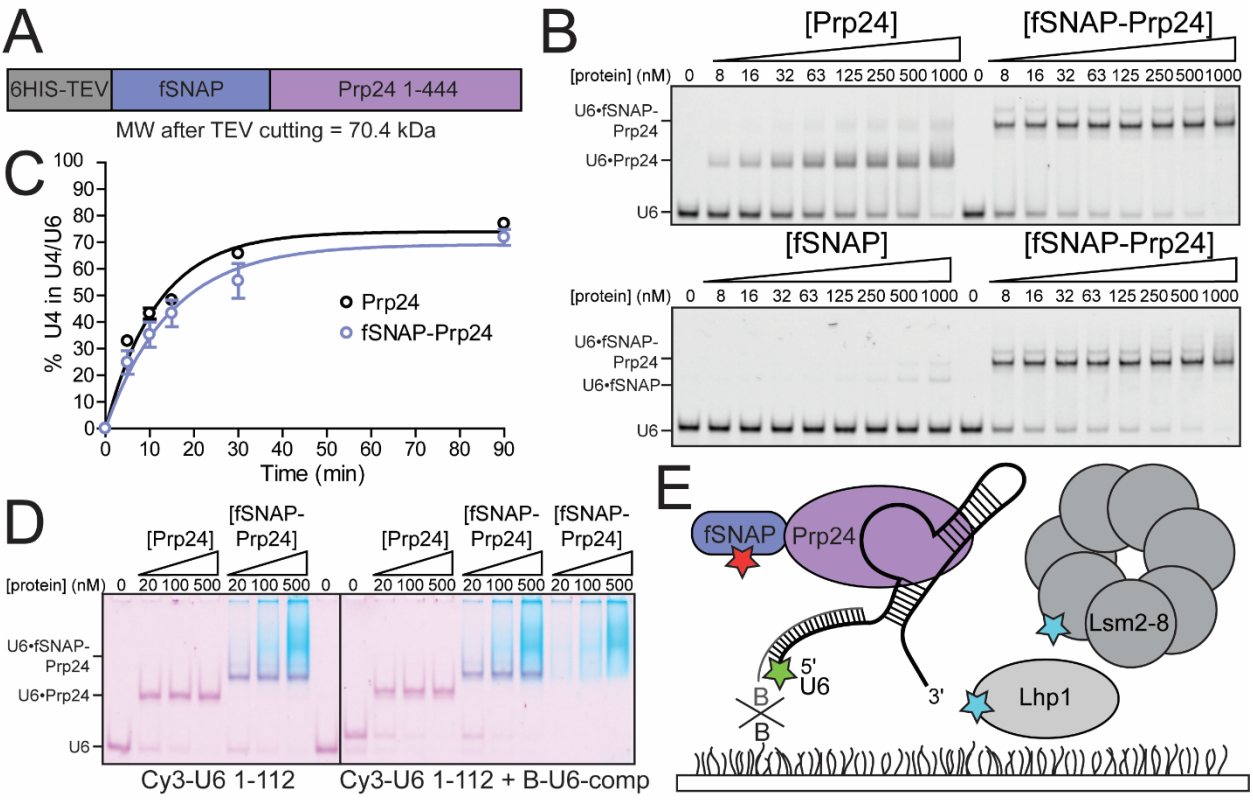
A) Secondary structure of the N-terminal fSNAP fusion of Prp24.

B) Top: fSNAP-Prp24 binds with similar affinity to full-length Prp24. Bottom: fSNAP protein alone does not bind U6 RNA.

C) fSNAP-Prp24 anneals U4 and U6 with similar kinetics and to a similar extent compared to full-length Prp24.

D) Both Prp24 and fSNAP-Prp24 can bind U6 and U6 that has been heat annealed to a biotinylated DNA that is complementary to the U6 5'SL.

E) The biotinylated complementary DNA could be used to tether fluorescently labeled U6 to a slide surface for CoSMoS experiments with the fSNAP-Prp24 fusion, which can be fluorescently labeled. Binding kinetics could be measured in the presence and absence of U6 binding proteins Lhp1 or Lsm2-8 labeled with a third fluorophore color.



A7.3.6 A composite U6-Prp24 binding surface does not improve Prp24 affinity for U4

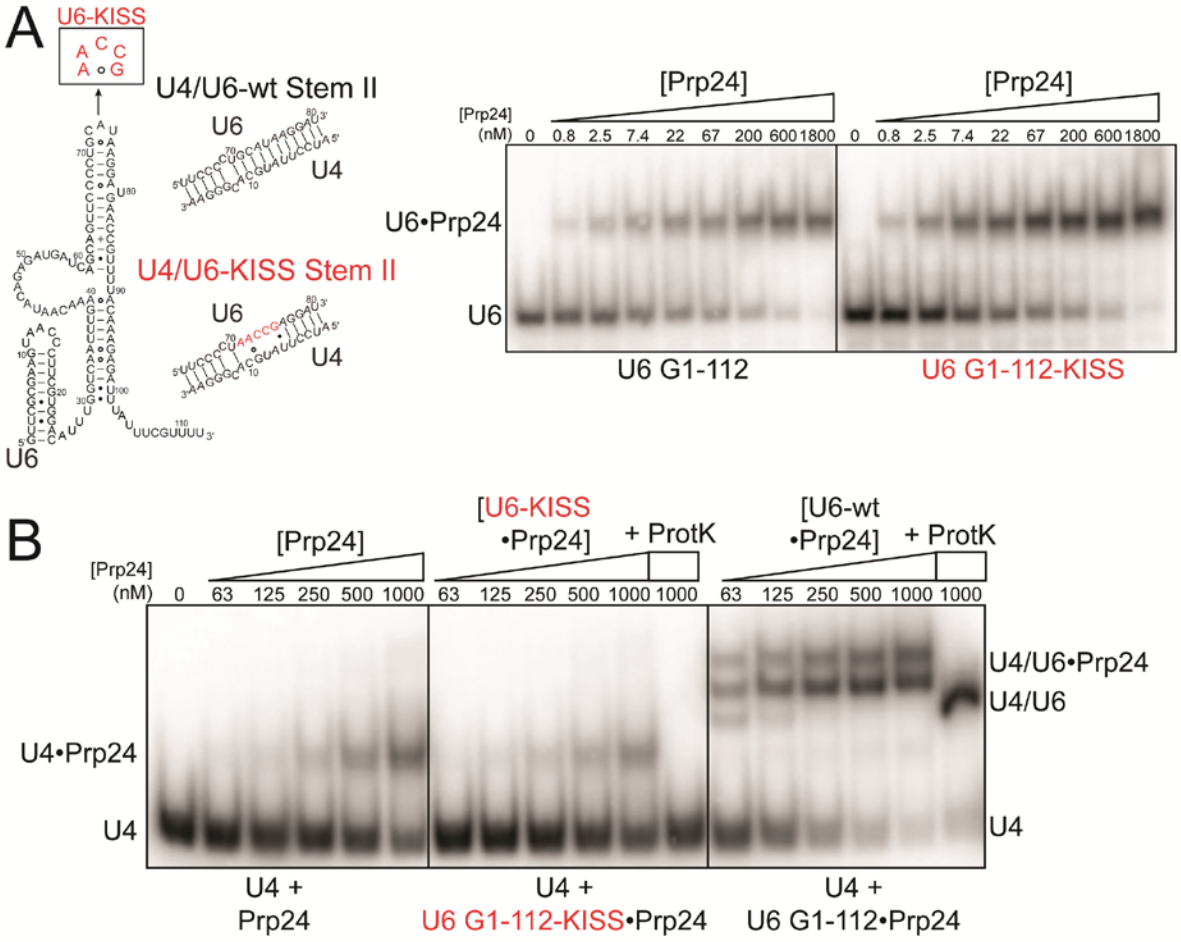
The mechanistic role of the electropositive groove is unclear. One possibility is that it directly binds U4 RNA. Prp24 alone has weak affinity for U4 (DIDYCHUK *et al.* 2016). However, this could because the electropositive groove does not form in the absence of U6 RNA. To test the hypothesis that U6 binding to Prp24 results in a high affinity U4 binding site in the electropositive groove, I used a mutant of U6 that is unable to form U4/U6 yet retains high affinity Prp24 binding. This mutant, "U6-KISS", contains five mutations at the top of the U6 ISL that disrupt U4/U6 Stem II (**Figure A7-6A**). U6-KISS is bound by Prp24 with similar affinity to wild-type U6 (**Figure A7-6B**), yet cannot anneal to U4 (data not shown). The affinity for U4 of Prp24 alone and Prp24 pre-incubated with an equimolar amount of U6-KISS was compared (**Figure A7-6C**). No ternary complex containing U6-KISS, Prp24, and U4 was observed. In contrast, when Prp24 pre-incubated with an equimolar amount of wild-type U6 is titrated, U4/U6•Prp24 ternary complex is observed, which resolves to U4/U6 upon treatment with proteinase K (**Figure A7-6C**). Thus, ternary complex is dependent upon having a U6 loop sequence that is complementary to U4, and stabilization of the electropositive groove in Prp24 by binding of U6 RNA is insufficient for stable recruitment of U4 into a ternary complex.

Figure A7-6. Prp24 affinity for U4 is not improved by the presence of U6.

A) The U6-KISS mutant (G71A/C72A/A73C/U74C/A75G) does not disrupt free U6 snRNA structure or protein binding, but will prevent stable formation of U4/U6 Stem II.

B) Prp24 can bind the U6-KISS mutant similar to wild-type U6 RNA.

C) Free Prp24 and Prp24 pre-incubated with U6-KISS bind U4 RNA poorly. Prp24 pre-incubated with wild-type U6 RNA efficiently forms U4/U6 and U4/U6•Prp24, which is resolved to U4/U6 after treatment with proteinase K.



A7.3.7 Prp24 binds U4/U6 with high affinity

Prp24 binds the product U4/U6 tightly *in vitro*, with affinities comparable or better than Prp24 for U6 alone (GHETTI *et al.* 1995; DIDYCHUK *et al.* 2016), yet is displaced after annealing is complete *in vivo* (SHANNON AND GUTHRIE 1991; JANDROSITZ AND GUTHRIE 1995). Identifying the high-affinity binding sites of Prp24 on U4/U6 could help us understand the annealing mechanism and the mechanism of Prp24 displacement. If the structure of U4/U6•Prp24 was known, one could use molecular dynamics simulations to model the annealing reaction.

As previously reported, even in the presence of non-specific inhibitors, Prp24 binds a non-specific RNA (human U1) with weak affinity (DIDYCHUK *et al.* 2016) (**Figure A7-7A**), yet has no affinity for a minimal linked U4/U6 RNA (CORNILESCU *et al.* 2016) (**Figure A7-7A**). However, when a single-stranded region corresponding to U6 nucleotides 39-55 is appended to this construct, Prp24 affinity improves (**Figure A7-7A**). Instead of building out from this construct, I made a linked construct containing full-length U4 and U6 RNA (U4/U6-link; **Figure A7-7B**). Prp24 binds this construct tightly with ~1 nM affinity (**Figure A7-7B**), a significant improvement from the minimal construct (U4/U6-112). However, significant laddering at high concentration of Prp24 was observed, suggestive of non-specific binding. Thus, a smaller construct lacking the U6 5'SL and U4 3'SL (U4/U6-small) was tested (**Figure A7-7B**). This construct displayed ~20 nM affinity on par with the affinity of Prp24 for U6 RNA (DIDYCHUK *et al.* 2016) and did not ladder at high concentrations of Prp24 (**Figure A7-7B**). Since this construct could conceivably reform U6 30-101, which includes the high affinity portion of U6 and could form the U6 telestem, I tested the effect of removing U6 nucleotides 81-112 (U4/U6-smallΔ81-112; **Figure A7-7C**). Prp24 bound this construct with similar ~20 nM affinity (**Figure A7-7C**). Finally, the U4/U6-smallΔ81-112 construct was further modified to remove part of U4/U6 Stem III (U4/U6-v1) to generate a minimal construct that binds with similar affinity to U6 RNA (**Figure A7-7D**).

To identify the high affinity site for Prp24, I further truncated U4/U6-v1. Interestingly, further truncation near the putative high affinity site in U6 strongly inhibits Prp24 binding (**Figure**

A7-8A). This pattern also exists when the termini are inverted (**Figure A7-8B**). Prp24 binds U4/U6-v6 worse than U4/U6-v1, even though they contain very similar sequences. This suggests that Prp24 prefers to bind the U4/U6 Stem III/I region when it contains base pairs enforced by the UUCG tetraloop in U4/U6-v1. Truncating U4/U6 Stem II or the U4 5'SL (Figure A7-8C) has no effect on Prp24 affinity for U4/U6. These experiments suggest that Prp24 binds to U4/U6 in the region between U4/U6 Stem III and Stem I that contains the high affinity site in U6 as observed in the crystal structure of U6-Prp24 (nucleotides U6 37-58) (MONTEMAYOR *et al.* 2014). However, secondary structure (base pairing with U4 RNA) also influences the affinity of Prp24 for U4/U6. These data suggest that the U6-Prp24 interlocked topology may persist in U4/U6. It would be interesting to test this by investigating the salt dependence of Prp24 binding, and to structurally characterize the U4/U6•Prp24 annealing intermediate and determine how the interlocked topology is resolved.

The minimal U4/U6 constructs described here could be used for structural studies. Linking U4 and U6 reduces the degrees of freedom and difficult in generating U4/U6•Prp24 complexes for structural studies. The minimal U4/U6-v1 RNA (~45 kDa) could be used with Prp24 (~50 kDa) and Snu13/Prp31 (~70 kDa) to generate a ~165 kDa complex suitable for cryo-EM. The U4/U6-small RNA (~58 kDa) could also be used and the Lsm2-8 complex included to make a ~265 kDa complex consisting of U4/U6-small•Prp24•Lsm2-8•Snu13•Prp31. The U4 5'SL can be replaced with a tetraloop-receptor (data not shown) which could alternatively be used to increase the size of the U4/U6 complex for cryo-EM studies.

Figure A7-7. Prp24 binds linked U4/U6 RNAs.

A) Prp24 non-specifically and weakly binds human U1 RNA and does not bind a linked U4/U6 construct (CORNILESCU *et al.* 2016). However, inverting the termini and adding nucleotides corresponding to the Prp24 binding site in U6-Prp24 (U6 nucleotides 39-58; green) improves binding.

B) Prp24 very tightly binds a linked U4/U6 construct containing full-length U4 and U6. Removing the U4 3'SL and U6 5'SL (U4/U6-small) does not reduce binding affinity and reduces laddering at high concentrations of Prp24. Nucleotides 37-58 of U6, which Prp24 contacts in the U6-Prp24 crystal structure (MONTEMAYOR *et al.* 2014) are highlighted in green.

C) Removing nucleotides in U6 that can form the U6 telestem (nucleotides 81-112) does not reduce affinity of Prp24 for U4/U6-small.

D) Prp24 binds U6 G1-112 and a linked U4/U6 that lacks the putative U4/U6 Stem III region (U4/U6-v1) with similar affinity and mobility.

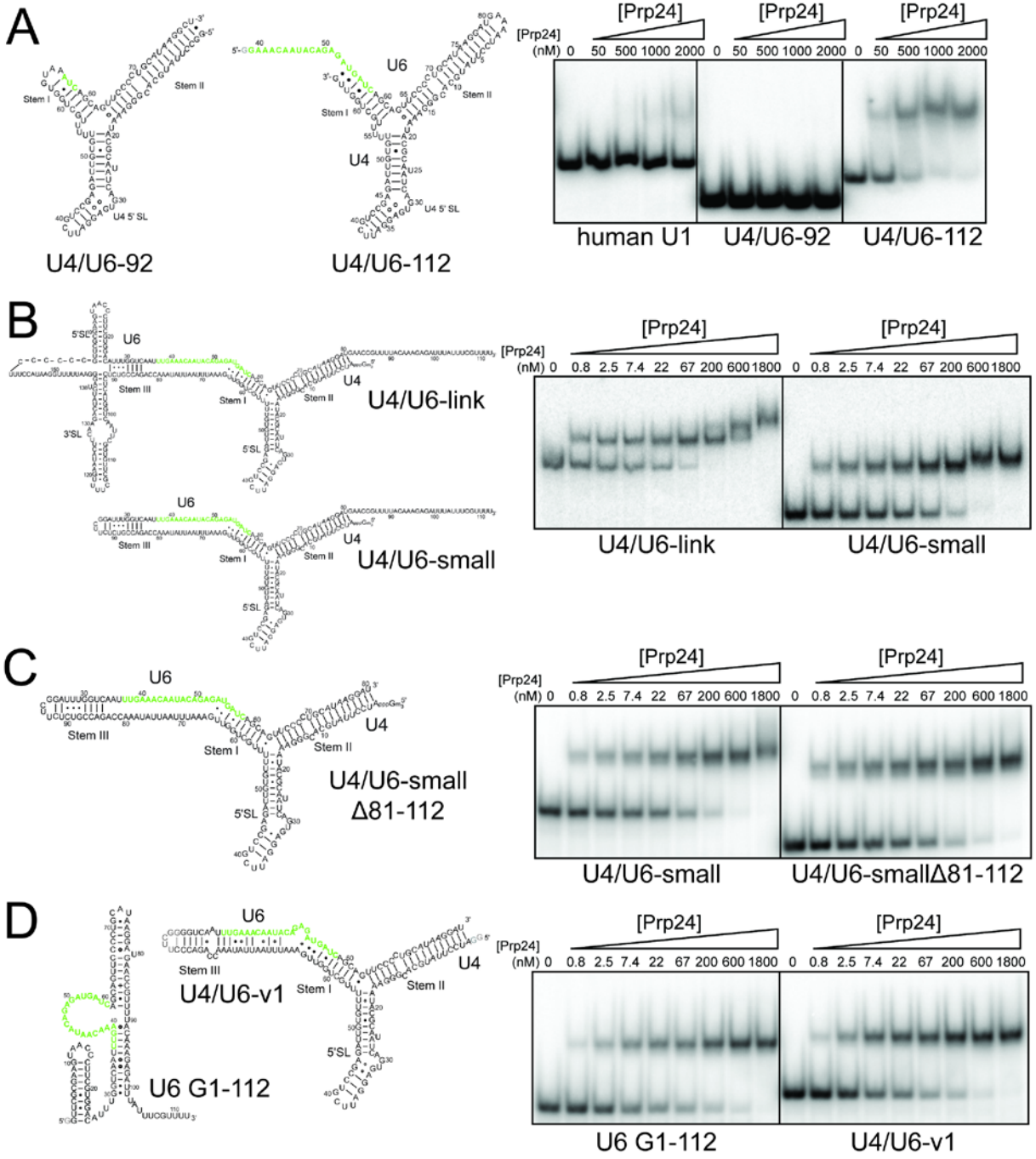
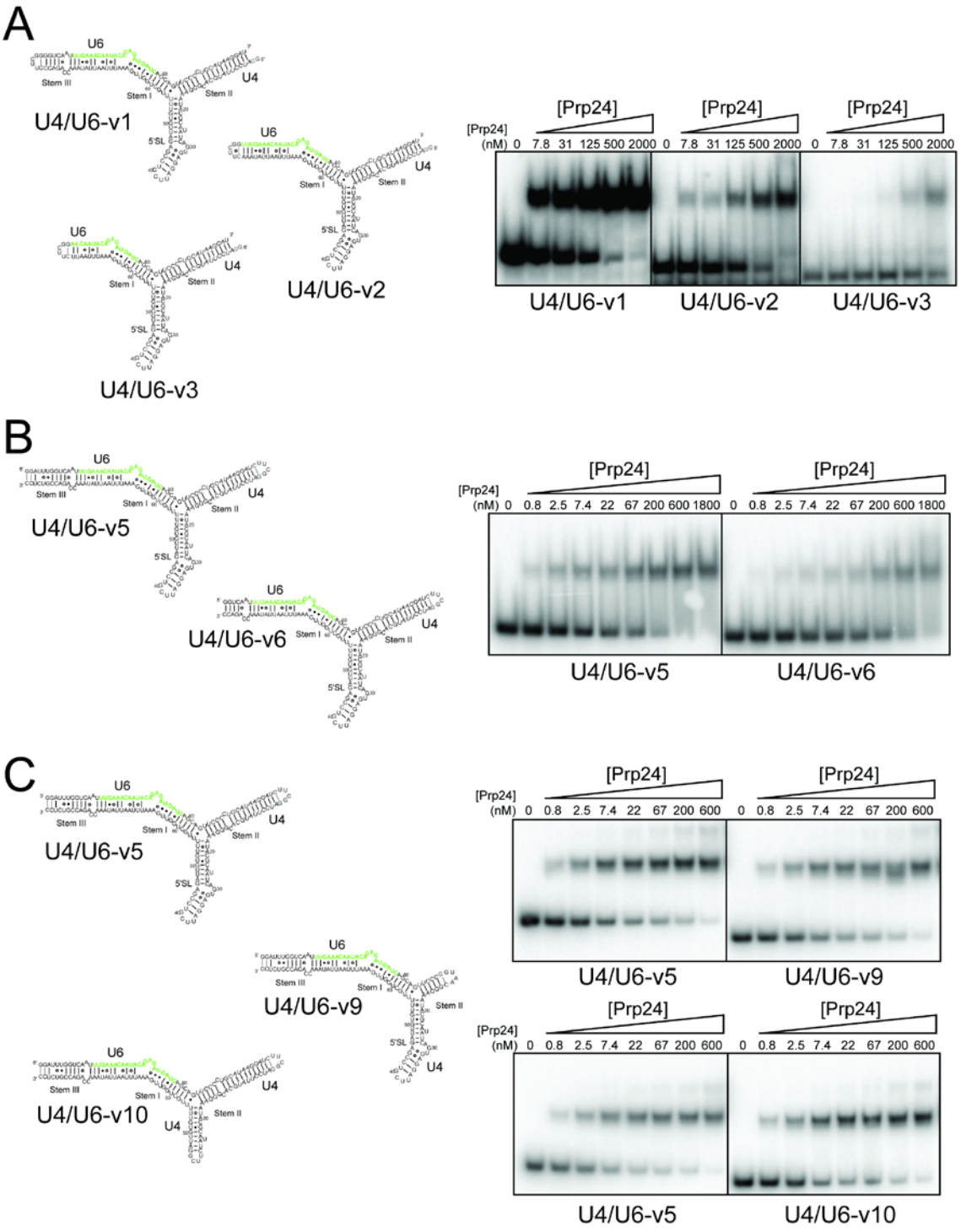

A
U4/U6-92
U4/U6-112
[Prp24]
(nM) 0 50 500 1000 2000
human U1
U4/U6-92
U4/U6-112
B
U4/U6-link
U4/U6-small
[Prp24]
(nM) 0 0.8 2.5 7.4 22 67 200 600 1800
U4/U6-link
U4/U6-small
C
U4/U6-small
Δ81-112
U4/U6-small
U4/U6-smallΔ81-112
D
U4/U6-v1
U6 G1-112
U6 G1-112
U4/U6-v1

Figure A7-8. Prp24 binds truncated linked U4/U6 RNAs.

A) Further reduction of linked U4/U6 into the putative high-affinity binding sequence (green) drastically reduces the affinity of Prp24 for U4/U6.

B) Inverting the termini of the linked U4/U6 construct (U4/U6-v5) does not significantly affect the affinity of Prp24 for U4/U6. Truncation into Stem III (U4/U6-v6) near the putative high-affinity binding sequence does reduce affinity.

C) Removing the majority of U4/U6 Stem II (U4/U6-v9) or the U4 5'SL (U4/U6-v10) does significantly affect Prp24 affinity for U4/U6.


A
U6
U4
Stem III
Stem I
Stem II
5'SL
U4/U6-v1
U4/U6-v2
U4/U6-v3
[Prp24]
[Prp24] (nM) 0 7.8 31 125 500 2000
U4/U6-v1
U4/U6-v2
U4/U6-v3
B
U4/U6-v5
U4/U6-v6
[Prp24] (nM) 0 0.8 2.5 7.4 22 67 200 600 1800
C
U4/U6-v5
U4/U6-v9
U4/U6-v10
[Prp24] (nM) 0 0.8 2.5 7.4 22 67 200 600

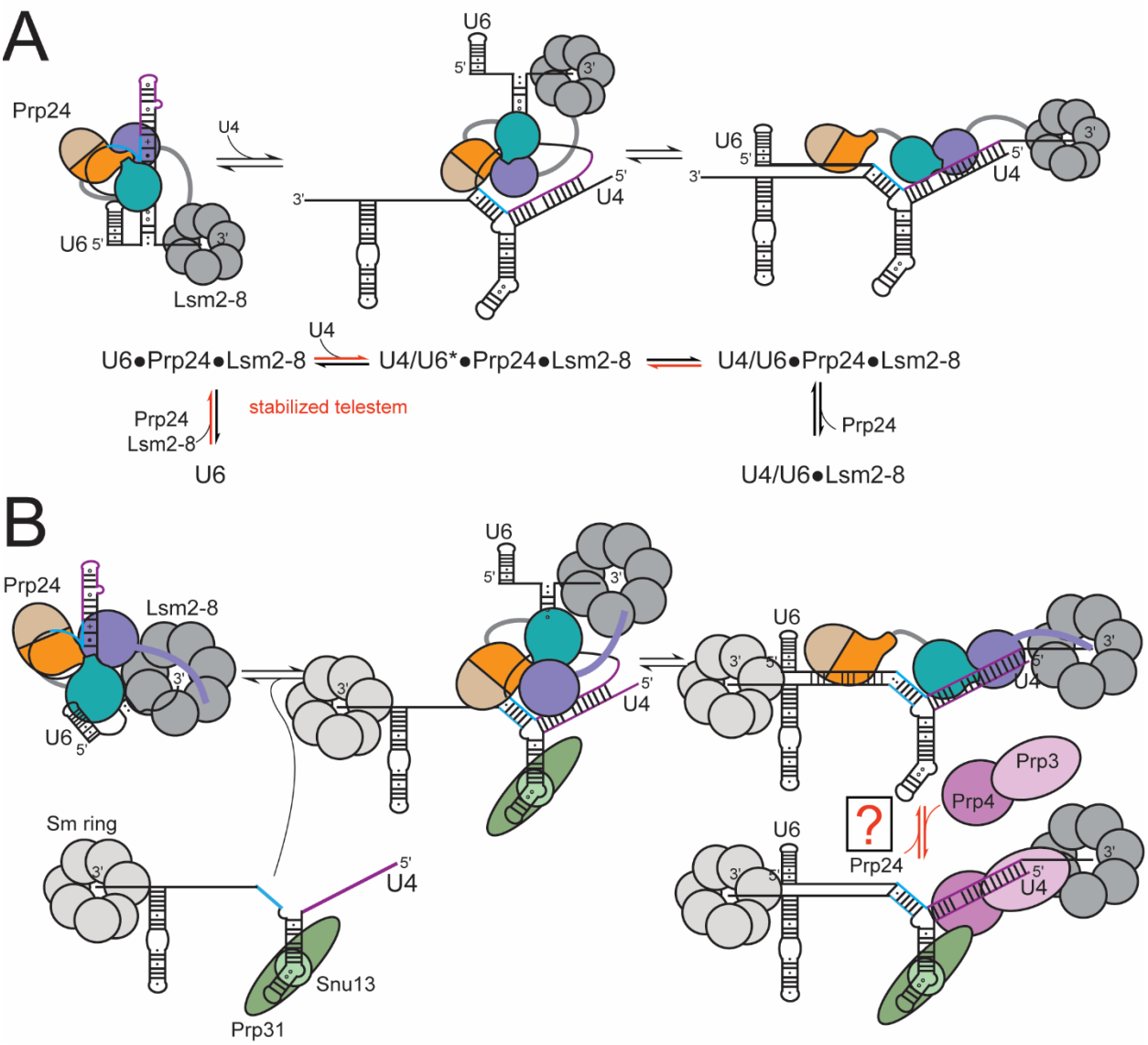
A7.3.8 Conclusions and future directions

This appendix extends the data presented in Didychuk *et al.* 2016 (**Chapter 4**) and describes several interesting future directions and tools to study the mechanism of Prp24-mediated annealing. This appendix expands our knowledge of Prp24-mediated annealing (**Figure A7-9**). I previously demonstrated that annealing depends on telestem formation, the electropositive groove, and Lsm2-8 binding (**Figure A7-9A**) (DIDYCHUK *et al.* 2016). In this Appendix, I show that Snu13 and Prp31 have subtle effects on annealing, likely by promoting an annealing-accessible conformation of U4 RNA (**Figure A7-9B**). Additionally, I characterized the binding site of Prp24 on U4/U6 and showed that its high affinity site lies between the U6 5'SL and U4/U6 Stem I. However, many interesting questions remain. How does the free U4 structure influence annealing? Which region of U4/U6 forms first? How do the electropositive groove, telestem, and Lsm2-8 ring enhance annealing? How is the interlocked topology of Prp24 resolved after completion of annealing? How does Prp3/Prp4 influence annealing and Prp24 dissociation? All of these questions could be answered using the *in vitro* annealing assay alongside other biophysical techniques.

Figure A7-9. Extended model for annealing pathway.

A) Simplified annealing model proposed in Didychuk *et al.* 2016.

B) U4 snRNP proteins Snu13 and Prp31 likely do not contribute to annealing. The contribution of U4/U6 di-snRNP proteins Prp3 and Prp4 should be measured to determine if they contribute to Prp24-mediated annealing or dissociation of Prp24. Prp24 binds U4/U6, likely in the region between Stem III and Stem I. Prp24 dissociates from the U4/U6 product via an undetermined mechanism.



A7.4 References

Cornilescu, G., A. L. Didychuk, M. L. Rodgers, L. A. Michael, J. E. Burke *et al.*, 2016 Structural Analysis of Multi-Helical RNAs by NMR-SAXS/WAXS: Application to the U4/U6 di-snRNA. J Mol Biol 428**:** 777-789.

Didychuk, A. L., E. J. Montemayor, D. A. Brow and S. E. Butcher, 2016 Structural requirements for protein-catalyzed annealing of U4 and U6 RNAs during di-snRNP assembly. Nucleic Acids Res 44**:** 1398-1410.

Didychuk, A. L., E. J. Montemayor, T. J. Carrocci, A. T. DeLaitsch, S. E. Lucarelli *et al.*, 2017 Usb1 controls U6 snRNP assembly through evolutionarily divergent cyclic phosphodiesterase activities. Nat Commun 8**:** 497.

Dobbyn, H. C., and R. T. O'Keefe, 2004 Analysis of Snu13p mutations reveals differential interactions with the U4 snRNA and U3 snoRNA. RNA 10**:** 308-320.

Fromont-Racine, M., A. E. Mayes, A. Brunet-Simon, J. C. Rain, A. Colley *et al.*, 2000 Genome-wide protein interaction screens reveal functional networks involving Sm-like proteins. Yeast 17**:** 95-110.

Ghetti, A., M. Company and J. Abelson, 1995 Specificity of Prp24 binding to RNA: a role for Prp24 in the dynamic interaction of U4 and U6 snRNAs. RNA 1**:** 132-145.

Hardin, J. W., C. Warnasooriya, Y. Kondo, K. Nagai and D. Rueda, 2015 Assembly and dynamics of the U4/U6 di-snRNP by single-molecule FRET. Nucleic Acids Res 43**:** 10963-10974.

Jandrositz, A., and C. Guthrie, 1995 Evidence for a Prp24 binding site in U6 snRNA and in a putative intermediate in the annealing of U6 and U4 snRNAs. EMBO J 14**:** 820-832.

Juillerat, A., T. Gronemeyer, A. Keppler, S. Gendreizig, H. Pick *et al.*, 2003 Directed evolution of O6-alkylguanine-DNA alkyltransferase for efficient labeling of fusion proteins with small molecules in vivo. Chem Biol 10**:** 313-317.

Liu, S., P. Li, O. Dybkov, S. Nottrott, K. Hartmuth *et al.*, 2007 Binding of the human Prp31 Nop domain to a composite RNA-protein platform in U4 snRNP. Science 316**:** 115-120.

Marmier-Gourrier, N., A. Clery, V. Senty-Segault, B. Charpentier, F. Schlotter *et al.*, 2003 A structural, phylogenetic, and functional study of 15.5-kD/Snu13 protein binding on U3 small nucleolar RNA. RNA 9**:** 821-838.

Medenbach, J., S. Schreiner, S. Liu, R. Luhrmann and A. Bindereif, 2004 Human U4/U6 snRNP recycling factor p110: mutational analysis reveals the function of the tetratricopeptide repeat domain in recycling. Mol Cell Biol 24**:** 7392-7401.

Montemayor, E. J., E. C. Curran, H. H. Liao, K. L. Andrews, C. N. Treba *et al.*, 2014 Core structure of the U6 small nuclear ribonucleoprotein at 1.7-A resolution. Nat Struct Mol Biol 21**:** 544-551.

Nguyen, T. H. D., W. P. Galej, X. C. Bai, C. Oubridge, A. J. Newman *et al.*, 2016 Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 A resolution. Nature 530**:** 298-302.

Rader, S. D., and C. Guthrie, 2002 A conserved Lsm-interaction motif in Prp24 required for efficient U4/U6 di-snRNP formation. RNA 8**:** 1378-1392.

Shannon, K. W., and C. Guthrie, 1991 Suppressors of a U4 snRNA mutation define a novel U6 snRNP protein with RNA-binding motifs. Genes Dev 5**:** 773-785.

Vidaver, R. M., D. M. Fortner, L. S. Loos-Austin and D. A. Brow, 1999 Multiple functions of Saccharomyces cerevisiae splicing protein Prp24 in U6 RNA structural rearrangements. Genetics 153**:** 1205-1218.

Zuker, M., 2003 Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31**:** 3406-3415.

Appendix 8: Electropositive groove mutants in Prp24 affect yeast viability

A8.1 Overview

We previously reported that mutations in Prp24 that reduce positive charge or introduce negative charge within the electropositive groove nearly eliminate Prp24-mediated annealing *in vitro* (DIDYCHUK *et al.* 2016). Importantly, these mutations do not affect Prp24-U6 binding, suggesting that the electropositive groove plays a specific role in recruiting U4 for annealing or stabilizing a transiently unwound intermediate in the annealing pathway. To test the role of the electropositive groove *in vivo*, I introduced mutant alleles of Prp24 into yeast and assayed for viability and growth phenotype. Interestingly, single mutations within the electropositive groove are viable and have little effect on growth. The double mutants N53D/Q54E and K78A/R81A have slow growth phenotypes. K78A/R81A is viable, but sick at the permissive temperature. These results suggest that the electropositive groove is involved in annealing *in vivo* and is important for growth. Additionally, it suggests that the effects of removing charge in the electropositive groove are cumulative and that yeast can tolerate less efficient U4/U6 annealing to a point.

A8.2 Materials and Methods

A8.2.1 Yeast plasmids and *PRP24* complementation assays

Mutations in pRS313-ScPrp24 were introduced via inverse PCR and ligation as previously described (DIDYCHUK *et al.* 2016). Mutant alleles of *PRP24* in pRS313 were transformed into LL101 (*MAT***a** *his3 leu2 trp1 ura3 met2 can1 ade2 lys2 prp24-Δ1::ADE2* [pUN50-*PRP24*]) using the lithium acetate method (GIETZ AND WOODS 2002). Growth phenotypes were assessed by pipetting 2.5 µL of 10-fold serial dilutions (OD_{600} = 0.5, 0.05, 0.005) onto solid medium lacking histidine and containing 1 mg/mL 5-FOA. Plates were incubated at 30°C for 3 days.

Temperature sensitivity for mutant alleles of *PRP24* was assessed after 5-FOA shuffle to remove the pUN50-*PRP24* wild type plasmid. Yeast were assayed as above on solid YPD medium and incubated at 16°C, 22°C, 30°C or 37°C.

A8.3 Results and discussion

Yeast containing single substitutions in Prp24 to remove positive charge or introduce negative charge within the electropositive groove were viable (**Figure A8-1A**). These included mutations that were tested *in vitro* (K50A, N53D, Q54E, K77A, K78A, R81A) as well as additional mutations K56A, K59A, H69A, and K380A. Mutants N53D and R81A appeared to give a mild slow-growth phenotype (**Figure A8-1A**). Mutation of R81 to lysine did not have an effect, suggesting that positive charge, rather than any specific contact, is important for its function. When pairs of electropositive groove residues are mutated, the effect on viability is much more striking (**Figure A8-1B**). Mutation of N53D in combination with Q54E, or R81A with K78A, yielded very few colonies that had a slow growth phenotype. Interestingly, the double mutant R131A/R134A had no apparent phenotype, despite reducing U4/U6 annealing rates similarly to the N53D/Q54E and K78A/R81A mutants *in vitro* (DIDYCHUK *et al.* 2016). The K78A/R81A mutant is viable after 5-FOA selection to remove wild-type Prp24, but possesses a cold sensitive phenotype (**Figure A8-1C**).

When the K78A/R81A mutant was tested in the U4/U6/Prp24 triple-disruption strain JEB100 (*MAT***a** *snr14Δ::trp1::ADE2 snr6Δ::LEU2 prp24-ΔORF::KANMX4 trp1 ura3 lys2 his3 ade2* [pRS316-U4/U6/Prp24]), where both Prp24 and U4+U6 are plasmid-borne, no phenotype was observed with the K78A/R81A mutation, likely due to increased U4 and U6 levels as measured by solution hybridization (data not shown). This suggests that Prp24-mediated annealing can be partially bypassed by increasing the concentration of U4 and U6 snRNAs. Indeed, overexpression of U4 and U6 snRNAs can rescue deletion of *PRP24 in vivo* (D. Brow, personal communication).

The mutations that appear to most severely affect yeast growth- N53, Q54, and R81- are localized to a ~9 Å wide patch within the electropositive groove (**Figure A8-1D**). This suggests that the “active site” of annealing within the electropositive groove of Prp24 can be narrowed down. Interestingly, this region (and particularly N53, Q54, and R81) contacts the loop of the ISL

(nucleotides 72-76) from a neighboring complex in the crystal lattice of the U6 snRNP (E. Montemayor, personal communication).

It would be interesting to test additional combinations of groove mutations to see if the effect of charge ablation is cumulative, or if there is indeed a region of the electropositive groove that is more important for annealing *in vivo*. The cold sensitive K78A/R81A mutation could be used to select for suppressor mutations in U6, U4, or other U4 or U6 snRNP components that improve protein-free annealing or bypass U4/U6 annealing entirely. Single mutations within the electropositive groove could also be tested for annealing activity *in vitro* to test if they ablate annealing or they have an intermediate effect. Determining if annealing activity is “buffered” against loss of positive charge or if loss of positive charge has a cumulative effect on annealing rate could be important for understanding the mechanism of annealing.

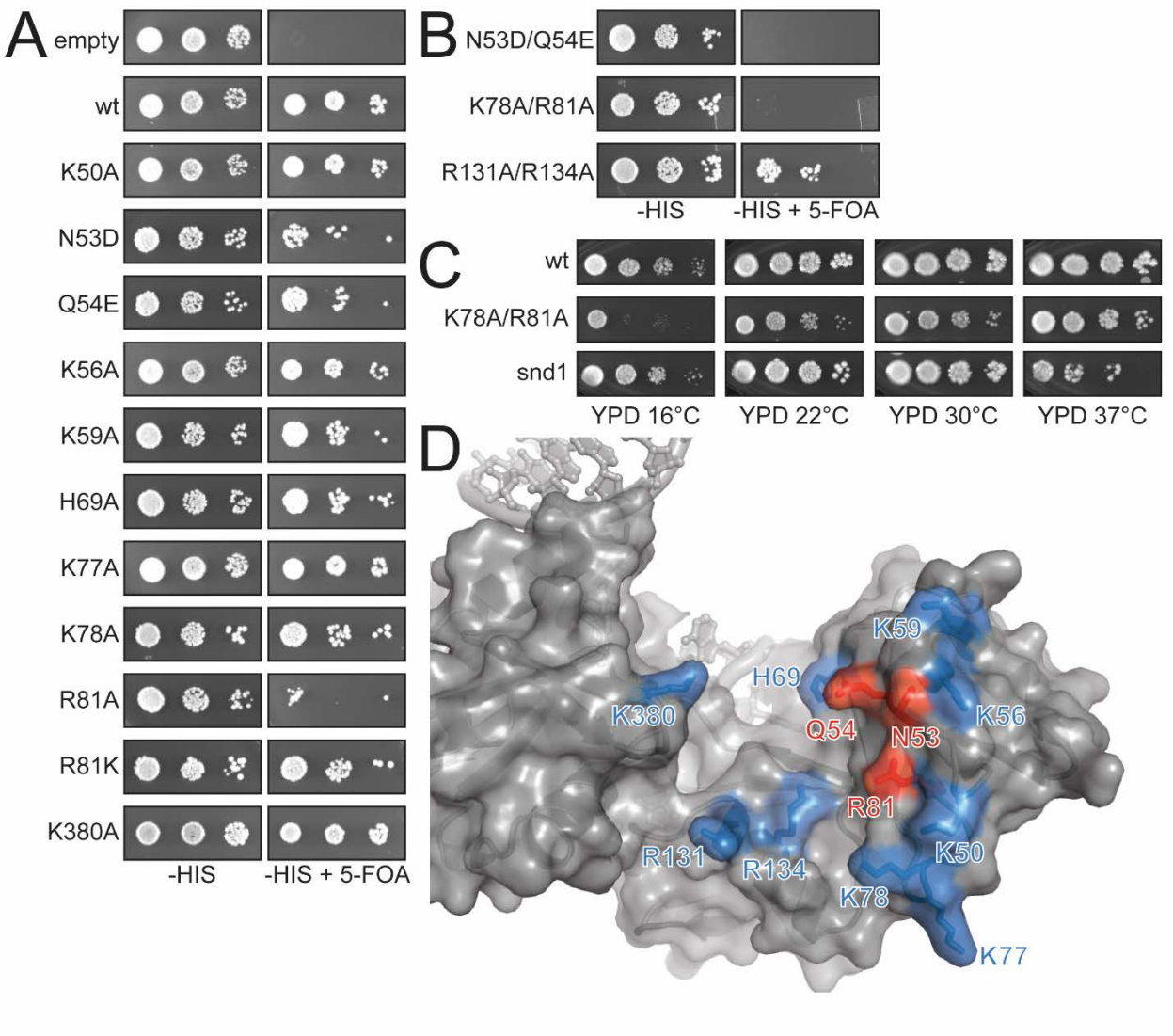
Figure A8-1. Electropositive groove mutants in Prp24 affect yeast growth.

A) Single mutations to reduce positive charge or introduce negative charge into residues within the electropositive groove are viable.

B) Double mutations within the electropositive groove tend to have a more significant growth phenotype. The double mutant R131A/R134A, which ablates annealing *in vitro*, has little effect *in vivo*.

C) Mutations within Prp24 can have opposite effects on temperature sensitivity. The snd1 mutation (D361A/S262A/K363A/L369A/M370A/I371A/N373A), which disrupts a protein-protein interface between Prp24 and Lsm2, is heat sensitive at 37°C. The K78A/R81A double mutation is cold sensitive, as it grows similar to wild-type at 37°C but worse at lower temperatures.

D) Mutations that more severely affect yeast growth (N53, Q54, and R81; red) are surface-exposed and close together in the structure of U6-Prp24 (PDB 4N0T). Positively charged residues within the electropositive groove that have little effect individually are highlighted in blue.


A
empty
wt
K50A
N53D
Q54E
K56A
K59A
H69A
K77A
K78A
R81A
R81K
K380A
-HIS
-HIS + 5-FOA
B
N53D/Q54E
K78A/R81A
R131A/R134A
-HIS
-HIS + 5-FOA
C
wt
K78A/R81A
snd1
YPD 16°C
YPD 22°C
YPD 30°C
YPD 37°C
D
K59
H69
K380
Q54
N53
K56
R81
R131
R134
K50
K78
K77

A8.4. References

Didychuk, A. L., E. J. Montemayor, D. A. Brow and S. E. Butcher, 2016 Structural requirements for protein-catalyzed annealing of U4 and U6 RNAs during di-snRNP assembly. Nucleic Acids Res 44**:** 1398-1410.

Gietz, R. D., and R. A. Woods, 2002 Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. Methods Enzymol 350**:** 87-96.

Allison Louise Didychuk
Department of Biochemistry
University of Wisconsin-Madison

Phone: 608-263-4081
E-mail: didychuk@wisc.edu

Education

2012 - Present	Ph.D. Candidate in Biophysics, expected graduation December 2017 University of Wisconsin, *Madison, WI*
2008 - 2012	B.S. Biology, Chemistry, Mathematics, summa cum laude University of Jamestown, *Jamestown, ND*

Research Experience

2012 - Present	University of Wisconsin, *Madison, WI* Ph.D. Advisor: Dr. Samuel E. Butcher *Research Area*: Structure and molecular mechanism of assembly of the spliceosomal U4/U6 di-snRNP complex; U6 snRNA 3' end processing
Summer 2011	NRC-CNRC Plant Biotechnology Institute, *Saskatoon, SK, Canada* Advisor: Dr. Michele Loewen *Research Area:* Structure of RuBisCo from *Pisum sativum*
Summer 2010	NRC-CNRC Plant Biotechnology Institute, *Saskatoon, SK, Canada* Advisor: Dr. Michele Loewen *Research Area:* Structural characterization of the N-terminal domain of the fungal G-protein-coupled receptor Ste2p

Publications

Didychuk, A.L., Butcher, S.E., Brow, D.A. The life of U6 small nuclear RNA, from cradle to grave. (*In preparation)*

Montemayor, E.J., **Didychuk, A.L.**, Sidhu, G.K., Yake, A.M., Brow, D.A., Butcher, S.E. Architecture of the U6 snRNP and recognition of 3'-end processed U6 RNA. (*In preparation)*

Didychuk A.L., Montemayor E.J., Carrocci T.J., DeLaitsch A.T., Lucarelli S., Westler, W.M., Brow D.A., Hoskins A.A., Butcher S.E. (2017) Usb1 controls U6 snRNP assembly through evolutionarily divergent phosphodiesterase activities. *Nature Communications* 8(1):497.

Montemayor, E.J., **Didychuk, A.L.**, Liao, H., Hu, P., Brow, D.A., Butcher, S.E. (2016) Structure and conformational plasticity of the U6 small nuclear ribonucleoprotein core. *Acta Cryst. D.* D73, 1-8.

Rodgers, M.L., **Didychuk, A.L.,** Butcher, S.E., Brow, D.A., Hoskins, A.A. (2016) A multi-step model for facilitated unwinding of the yeast U4/U6 RNA duplex. *Nucleic Acids Research* 44(22):10912-10928.

Cornilescu, G.*****, **Didychuk, A.L.***, Rodgers, M.L., Michael, L.A., Tonelli, M., Burke, J.E., Montemayor, E.J., Hoskins, A.A., Butcher, S.E. (2016) Structural analysis of multi-helical RNAs by NMR-SAXS/WAXS: Application to the U4/U6 di-snRNA. *Journal of Molecular Biology* 428(5 Pt A):777-89. **[*co-first authors]**

Didychuk, A.L., Montemayor, E.J., Brow, D.A., Butcher, S.E. (2016) Structural requirements for protein-catalyzed annealing of U4 and U6 RNAs during di-snRNP assembly. *Nucleic Acids Research* 44(3):1398-410.

Loewen, P.C., **Didychuk, A.L.**, Switala, J., Perez-Luque, R., Fita, I., Loewen, M.C. (2012) Structure of *Pisum sativum* Rubisco with bound ribulose 1,5-bisphosphate. *Acta. Cryst. D* 69: 10-14.

Oral and Poster Presentations

2017 "*Structures and evolutionarily divergent phosphodiesterase activities of Usb1 from yeast and human*" 11th Cold Spring Harbor Meeting on Eukaryotic mRNA Processing, Cold Spring Harbor, NY (**Poster**)

2017 "*Structures and evolutionarily divergent phosphodiesterase activities of Usb1 from yeast and human*" Graduate Women in Science National Conference, Sioux Falls, SD (**Poster**)

2017 "*Structures and evolutionarily divergent phosphodiesterase activities of Usb1 from yeast and human*" 22nd Annual Meeting of the RNA Society, Prague, Czech Republic (**Talk**)

2016 *"Usb1: A Means to a (3') End"* 2016 IPiB Retreat Elected Student Speaker, Madison, WI (**Talk**)

2016 *"Biochemical and Structural Characterization of the Yeast U6 Biogenesis Protein Usb1 Reveals a Mechanism for Coupling snRNA Processing with snRNP Assembly"* 21st Annual Meeting of the RNA Society, Kyoto, Japan (**Talk**)

2015 *"Efficient annealing of spliceosomal U4 and U6 RNAs by Prp24 requires an electropositive groove, the U6 telestem, and the Lsm ring"*, 10th Cold Spring Harbor Meeting on Eukaryotic mRNA Processing, Cold Spring Harbor, NY (**Talk**)

2015 *"Efficient annealing of spliceosomal U4 and U6 RNAs by Prp24 requires an electropositive groove, the U6 telestem, and the Lsm ring"*, 20th Annual Meeting of the RNA Society, Madison, WI (**Talk**)

2014 *"Mechanistic and structural studies of U4/U6 annealing by Prp24"*, 19th Annual Meeting of the RNA Society, Quebec City, Canada (**Poster**)

Professional Activities

2016 - Present *Graduate Women in Science (GWIS) National Treasurer*
Elected National Treasurer for GWIS, a 501(c)(3) non-profit organization. Responsible for general financial oversight, banking, and record keeping. Serve on the Executive Board and Finance Committee.

2016 - 2017 *Biophysics Program Steering Committee Student Representative*
Assist in organizing Biophysics Program annual recruiting weekend events; represent Biophysics students at Steering Committee meetings.

2014 - 2017 *RNA Society Junior Scientist Committee*
Organized junior scientist events (career development workshop, pre-meeting networking activity, social) at annual RNA Society meetings (2015 in Madison, WI; 2016 in Kyoto, Japan; 2017 in Prague, Czech Republic). Served as a Session Chair for Career Development Workshops at RNA 2015 and 2016 by organizing and moderating panels.

2016 *Student Research Grants Competition Peer Reviewer*
Reviewed applications for the SRGC Research Travel Awards for the May awards cycle.

2013 – 2016 *Graduate Women in Science (GWIS) Madison Chapter*
Served on the Outreach Committee from 2013-2014 and as the Outreach Committee Chair since 2014. Organized science outreach events on and off campus, attended biweekly officer meetings. Organized a booth at the USA Science & Engineering Festival (April 2016) in Washington, D.C.

Fellowships & Awards

2017	2017 Biophysics Colloquium "Best Student Talk"
2017	2017 IPiB Retreat Poster Award
2016 - 2017	Louis and Elsa Thomsen Distinguished Graduate Student Fellowship
2015	Graduate Women in Science Madison Chapter Ruth Dickie Grant-in-aid
2011 - 2012	College Fellow in Chemistry, University of Jamestown
2011 - 2012	College Fellow in Biology, University of Jamestown
2010 - 2012	College Fellow in Mathematics, University of Jamestown
2008 - 2012	Dean's List, University of Jamestown

2009	Freshman Mathematics Award, University of Jamestown
2008 - 2012	Wilson Scholarship, University of Jamestown (Full Tuition Scholarship)
2008	Governor General's Academic Medal, Vincent Massey High School

Undergraduate Mentoring

2016 - Present	*Stefani Lucarelli* Biochemistry Major 2016 Sophomore Undergraduate Research Award 2017 Biochemistry Undergraduate Summer Research Scholarship
2015 - 2017	*Matthew Larson* Biochemistry Major 2017 Biochemistry Undergraduate Summer Research Scholarship
2014 - 2017	*Andrew DeLaitsch* (graduated May 2017) Biochemistry & Economics Major 2016 Hilldale Award Biochemistry Department Super-G Program 2016
2016	Research Mentor Training Course, Delta Program, UW-Madison

Teaching Experience

Instructor, Biochemistry 551: Biochemical Methods (2017)
Developed syllabus and led a journal-club style seminar course for senior Biochemistry undergraduates
University of Wisconsin-Madison
Teaching Assistant, Biology 305: Cell Biology (2010)
University of Jamestown
Teaching Assistant, Chemistry 343 & 344: Organic Chemistry I & II (2010-2012)
University of Jamestown
Teaching Assistant, Chemistry 133 & 134: General Chemistry I & II (2009-2010)
University of Jamestown
Tutor, Mathematics Department (2009-2012)
University of Jamestown

References

Prof. Samuel E. Butcher
University of Wisconsin-Madison
Department of Biochemistry
Phone: 608-263-3890
Email: sebutcher@wisc.edu

Prof. David A. Brow
University of Wisconsin-Madison
Department of Biomolecular Chemistry
Phone: 608-262-1475
Email: dabrow@wisc.edu

Prof. Aaron Hoskins
University of Wisconsin-Madison
Department of Biochemistry
Phone: 608-890-3101
Email: ahoskins@wisc.edu