THE UNIVERSITY OF CHICAGO

PROOFREADING AND PROGRESSION AT THE CATALYTIC STAGES OF SPLICING: NOVEL INSIGHTS INTO THE ROLES AND REGULATION OF DEAH-BOX ATPASES

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN CELL AND MOLECULAR BIOLOGY

BY

CHRISTOPHER ANTHONY JASON CRADDOCK

CHICAGO, ILLINOIS

MARCH 2022

In dedication to Jennifer Frances Farley,

without whom none of this would be possible

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank Emina Stojkovic, Lucia Rothman-Denes, and Jon Staley for seeing the potential in me and putting me in the position to perform the research contained within this dissertation. I would especially like to thank Jon Staley for his mentorship, as he took a chance on me as an undergraduate from a small state university as part of an NSF Research Experiences for Undergraduates program in the Summer of 2012. In his lab, he has given me every opportunity to fully explore research directions and ideas I came up with, for better or for worse. I would also like to thank Deepti Bellur, my first mentor in the lab, for being an exemplary example of efficiency that I could never quite match. I would also like to thank my classmates for being such inspirations, especially fellow science dads Bill Richter and Will Yee, and members of my cohort for the good times and support. Thank you also to members of the various communities I've been a part of: the Chicago RNA Club, UC-GRIT, the HHMI Gilliam Fellows, and the eighth floor of Cummings Life Sciences Center. I would also like to thank my committee: Ben Glick, Alex Ruthenburg, and Joe Piccirilli for their mentorship. Last but not least, thank you to Thomas Campbell for writing the last-minute letter of recommendation that got my foot into the door at University of Chicago.

Most importantly, none of this would be possible without my partner, Jenne Farley. Thank you for lifting me up during the many lows I have experienced during this journey. Thank you for being a great partner, for your unwavering support, and your clear-headed guidance in times of stress. Thank you for making me a better person, and for being a great mother to our children Simone, Lionel, and Noemie. You are truly a model for the kind of person I hope to someday become for you and our family.

ABSTRACT

Pre-mRNA splicing is an essential step in eukaryotic gene expression, wherein interrupting sequences ("introns") are excised from newly transcribed precursor messenger RNA (pre-mRNA). At the same time, expressed sequences ("exons") are ligated together to form mature mRNA that codes for a given polypeptide product. As errors in splicing may lead to insertion of aberrant intronic sequence or deletion of exonic sequence, fidelity in splicing is necessary to ensure proper gene expression. This fidelity is maintained by a set of ATPases belonging to the SF2 family of nucleic acid helicases. Here, I focus on two of these helicases: the DEAH-box ATPases Prp16 and Prp43. Through characterization of disease-correlated mutations in *DHX38*, the gene encoding human Prp16, I present evidence for a conserved motif that mediates an autoinhibitory interaction between the N-terminal and helicase domains of Prp16. I also present evidence for a novel role for Prp43 in discarding splicing intermediates that are deficient in transition between the first and second catalytic steps of pre-mRNA splicing. These data deepen our knowledge of how DEAH-box ATPases ensure faithful gene expression.

CHAPTER 1

INTRODUCTION TO PRE-mRNA SPLICING AND

SPLICING REGULATION BY DEAH-BOX ATPases

The central dogma: how RNA plays an essential role in gene expression

Gene expression is the process of generating biologically active molecules from instructions encoded in the DNA. The "central dogma" of gene expression posits information flows from genes in DNA to proteins, with RNA being an intermediary molecule (Crick 1970). DNA carries heritable genetic information, which is used to code for proteins that carry out cellular function. In the classical view of the central dogma of gene expression, RNA is an intermediary molecule, copied ("transcribed") from DNA by RNA polymerases to be used as a template ("messenger RNA" or "mRNA") by ribosomes to create polypeptide chains ("translation") of the correct sequence to generate active protein.

However, advancements in the field of molecular biology since the inception of the central dogma have illuminated myriad roles for RNA molecules in catalyzing and regulating each step of gene expression beyond just acting as a messenger. Ribosomal RNAs catalyze the phosphodiester bond linkage necessary for polypeptide chain synthesis, and transfer RNAs carry amino acids to the ribosome to extend growing polypeptide chains (reviewed in Ramakrishnan 2002; Schmeing and Ramakrishnan 2009). RNAs that remain bound to RNA polymerase II can regulate the transcription of nearby genes (Werner et al. 2017), and small nuclear RNAs ("snRNA") act in processing the transcription products from RNA polymerase II to create mature mRNA from precursor mRNA ("pre-mRNA") in a process known as pre-mRNA splicing (reviewed in Staley and Guthrie 1998; Wahl, Will, and Lührmann 2009; Wilkinson, Charenton, and Nagai 2020).

Pre-mRNA splicing is an essential step in eukaryotic gene expression

Gene expression in eukaryotic cells is complicated by several factors. One factor is the presence of a nuclear membrane surrounding DNA, which resides in the nucleus, and the ribosomes, which reside in the cytoplasm, uncoupling the physical locations of transcription and translation. Because of this, mRNA generated in the nucleus must be exported through the nuclear membrane to the cytoplasm for translation (**Fig. 1.1**). Facilitating mRNA export are several factors that bind to the mRNA molecule that target it to the nuclear pore complex (reviewed in Niño et al. 2013). These export factors bind to adaptor proteins, which in turn bind to key features of an mRNA molecule that play a role in efficient and faithful translation once the mRNA is exported to the cytoplasm. These include adaptors that bind to the 5' 7-methylguanosine cap, which functions in translation initiation (Merrick and Pavitt 2018); the poly-A tail, which confers mRNA stability and promotes translation efficiency (Jackson, Hellen, and Pestova 2010); and, in higher eukaryotes, the THO/TREX and exon-junction complexes, which are deposited during pre-mRNA splicing (Moore and Proudfoot 2009)**.**

Pre-mRNA splicing is the process in which interrupting sequences ("introns") are excised and flanking expressed sequences ("exons") are ligated together to produce a mature mRNA that then goes on to code for the required polypeptide sequence (Staley and Guthrie 1998; Wahl, Will, and Lührmann 2009; Wilkinson, Charenton, and Nagai 2020). The number of introns within a eukaryotic genome can range from a relatively small number in lower eukaryotes such as budding yeast (Spingola et al. 1999) to hundreds of thousands in higher eukaryotes such as humans (Hong, Scofield, and Lynch 2006). These introns may contain sequences of completely different genes, or may

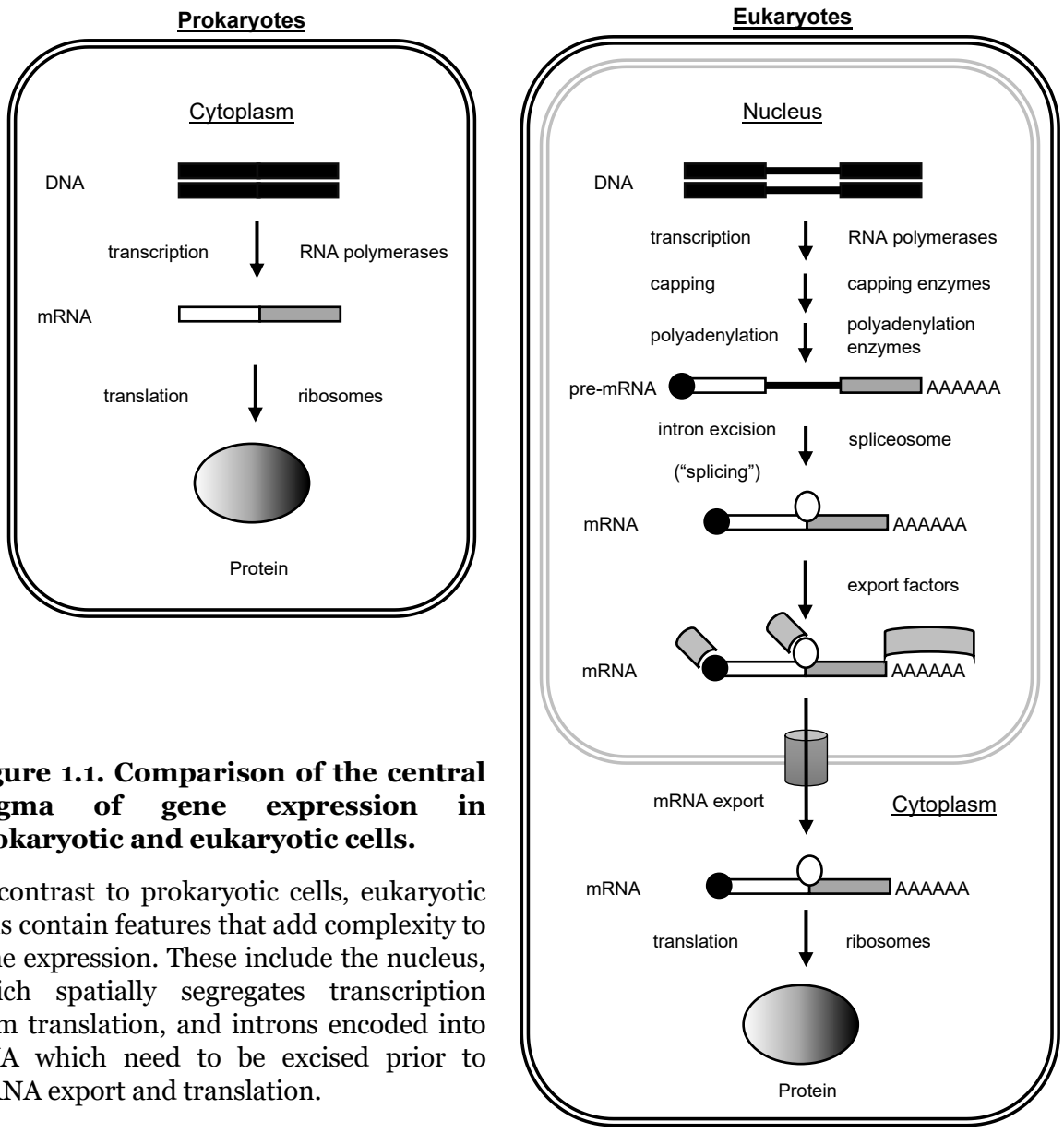

Figure 1.1. Comparison of the central dogma of gene expression in prokaryotic and eukaryotic cells.

In contrast to prokaryotic cells, eukaryotic cells contain features that add complexity to gene expression. These include the nucleus, which spatially segregates transcription from translation, and introns encoded into DNA which need to be excised prior to mRNA export and translation.

function in promoting or delaying mRNA accumulation (reviewed in Carmel and Chorev 2012; Rose 2019), but since they do not function in the downstream steps of translation or mRNA decay their excision is necessary for faithful gene expression to proceed.

During splicing, exons can also be selectively included or excluded from the resultant mRNA. This process, called alternative splicing, occurs in 95% of the genes in the human genome (Pan et al. 2008). Alternative splicing can be regulated on a tissue-

specific level to generate multiple isoforms from a single gene (E. T. Wang et al. 2008), affording higher eukaryotes an increase in proteome diversity that can be regulated at a granular level. The importance of pre-mRNA splicing, and of the regulation of pre-mRNA splicing, is underscored by the fact that 35% of disease-causing mutations within genes affect splicing (Manning and Cooper 2017), and mutations that affect components of the splicing machinery are also causal for disease (G.-S. Wang and Cooper 2007). In this thesis, I will detail how mutations within a splicing factor (Chapter 2), as well as a substitution within an intron (Chapter 3) dysregulates splicing alter gene expression.

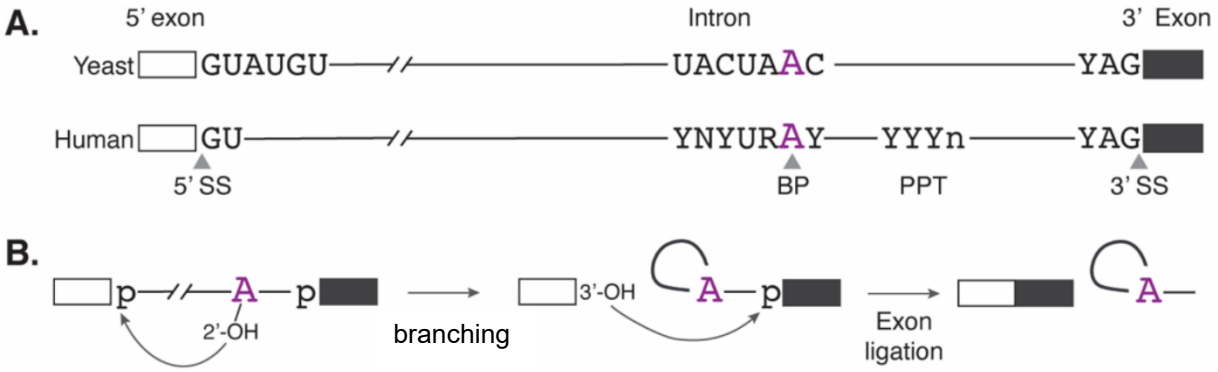

Figure 1.2. Pre-mRNA splicing is catalyzed between three reactive intronic elements defined by short nucleotide consensus sequences.

A. The 5' splice site, branch point, and 3' splice site are defined by the 5' splice site consensus sequence, branch site consensus sequence, and 3' splice site consensus sequence. While rigidly defined in budding yeast introns, these splice sites can deviate across introns in human genes.

B. Pre-mRNA splicing involves two transesterification reactions. During branching, a 2' hydroxyl group on the branch point adenosine attacks the 5' splice site phosphate. During exon ligation, the 3' hydroxyl on the free 5'exon attacks the 3' splice site phosphate, fully excising the intron while at the same time ligating the flanking exons. (Figure courtesy of Yi Zeng.)

Intron structure across evolutionary taxa

Introns contain three reactive elements that are used to catalyze the pre-mRNA splicing reaction: the 5' splice site phosphate, the 3' splice site phosphate, and the branch

site adenosine ("brA"). These reactive regions lie in close proximity to short nucleotide sequences known as the 5' splice site consensus sequence ("5'SS"), 3' splice site consensus sequence ("3'SS"), and the branch site ("BS"). While in budding yeast, these consensus sequences are rigidly defined, in higher eukaryotes such as human these sequences are more degenerate (**Fig 1.2A**) (Wahl, Will, and Lührmann 2009). Nevertheless, common intronic features within these consensus sequences are found across evolutionary taxa. The first two residues of the intron, which makes up part of the 5'SS, are GT (encoding "GU" after being transcribed to RNA), while the ultimate and penultimate residues of the intron that make up the 3'SS are AG. The brA of the BS is invariant, and the residue that lies two nucleotides upstream is a T (encoding "U"). These nucleotides not only define intronic reactive elements, but they also play a role in forming a catalytic core to properly position these reactive elements for pre-mRNA splicing catalysis.

Pre-mRNA splicing chemistry and spliceosome assembly

Pre-mRNA splicing (herein referred to as "splicing") occurs through two transesterification reactions (**Fig. 1.2B**): branching, in which the 2' hydroxyl group on the backbone of the brA attacks the 5' phosphate at the 5' splice site to form a branched lariat intermediate and free 5' exon; and exon ligation, in which the 3' hydroxyl group of the free 5' exon attacks the 5' phosphate of the 3' splice site to fully excise the intron and, concomitantly, ligate the flanking 5' and 3' exons together.

Despite the simple two-step chemistry of splicing, this process is catalyzed by a massive molecular machine known as the spliceosome. The spliceosome is a ribonucleoprotein machine, consisting of five snRNA components and over 80 protein

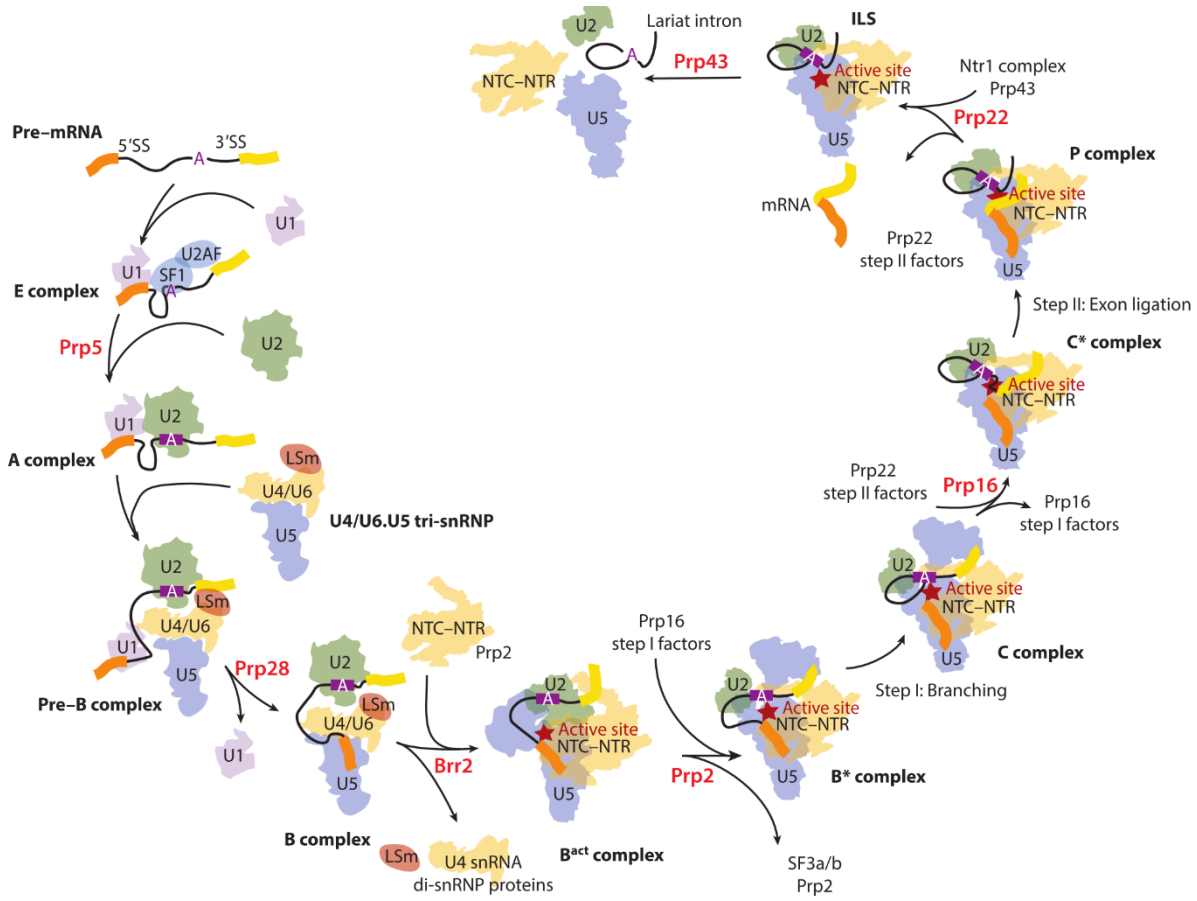

Figure 1.3. Overview of spliceosome assembly & dynamics during pre-mRNA splicing.

The spliceosome consists of several small nuclear ribonucleoprotein subunits (snRNPs), in addition to a large protein complex (NTC) and several accessory factors belonging to the DEAH-box ATPase family of RNA helicases. These components assemble stepwise *de novo* onto each introns that is to be excised, with several wholescale rearrangements of these components occurring at each step of assembly and catalysis. (Figure adapted from Wilkinson et al., 2019)

components that are conserved from yeast to humans. These snRNAs and protein components form several modular small nuclear ribonucleoprotein complexes ("snRNPs"), where one of the five snRNAs (U1, U2, U4, U5, and U6) are associated with several protein factors (Staley and Guthrie 1998). Additionally, a protein-protein complex associated with the protein factor Prp19p ("NineTeen Complex", or "NTC") is also necessary at the catalytic stages of splicing (Wahl, Will, and Lührmann 2009).

These snRNPs, the NTC, and several additional soluble factors assemble stepwise *de novo* onto each intron that is to be excised (**Fig. 1.3**) (Staley and Guthrie 1998; Wahl, Will, and Lührmann 2009; Wilkinson, Charenton, and Nagai 2020). Intron recognition and splicing catalysis rely largely on RNA-RNA interactions between splice site consensus sequences within the intron and the snRNA components of the spliceosome. Initial recognition of an intron occurs through binding of the U1 snRNP to the 5'SS ("E complex"), followed by binding of the U2 snRNP to the BS ("A complex"). A pre-formed complex known as the U4/U6●U5 triple-snRNP ("tri-snRNP") then adds to the intron ("pre-B complex"). Within the tri-snRNP are U6 snRNA, which catalyzes both transesterification reactions (Fica, Tuttle et al. 2013), base-paired to U4 snRNA which prevents U6 snRNA from adopting a catalytic conformation. The 5'SS is then handed off to U6 snRNA (Staley and Guthrie 1999) to form the "B complex"), and U4 snRNA is subsequently released by the tri-snRNP helicase Brr2p (Raghunathan and Guthrie 1998) to allow for splicing catalysis to begin ("B-activated" or "B^{act} complex").

Splicing progression relies on the activity of RNA helicases

The spliceosome is an exquisitely dynamic ribonucleoprotein machine, with wholescale rearrangements and binding or shedding of several snRNP and protein factors during the assembly, activation, catalysis, and disassembly steps required for each splicing event (**Fig 1.3**) (reviewed in Staley and Guthrie 1998; Wahl, Will, and Lührmann 2009; Wilkinson, Charenton, and Nagai 2020). This dynamicity of the spliceosome is attributed to conserved DExH/D-box helicases that act on it (Staley and Guthrie 1998). During initial spliceosome assembly, the ATPase activities of the DEAD-box helicases Prp5p and Prp28p are required for U2 snRNP binding to the BS and U6 snRNP binding

to the 5'SS, respectively (Ruby, Chang, and Abelson 1993; Staley and Guthrie 1999; Xu et al. 2004). Brr2p, mentioned above, is a DEIH-box ATPase required to release the U4 snRNP (Raghunathan and Guthrie 1998) which allows U6 snRNA to adopt a catalytically active conformation and also allows recruitment of the NTC to complete formation of the catalytically active spliceosome (Wahl, Will, and Lührmann 2009; Wilkinson, Charenton, and Nagai 2020)

Downstream of Brr2-mediated spliceosome activation, four DEAH-box ATPases (Prp2p, Prp16p, Prp22p, Prp43p) act to rearrange the spliceosome for catalysis of branching and exon ligation. Prp2p and Prp16p largely act to reposition a helix formed between the BS and U2 snRNA ("BS:U2 duplex") to either juxtapose the brA next to the 5' splice site phosphate in the catalytic core of the spliceosome (Haselbach et al. 2018) or shift the brA to allow for juxtaposition of the 3' splice site phosphate with 3' hydroxyl of the free 5' phosphate (Fica et al. 2017; S. Liu et al. 2017a; Wilkinson et al. 2017). Thus, Prp2p ATPase activity is required for the spliceosome to rearrange from the B^{act} conformation to one competent for catalysis of branching ("B* complex" or "branching conformation"), while Prp16p ATPase activity is required to rearrange the spliceosome from a post-branching conformation ("C complex") to one competent for catalysis of exon ligation ("C* complex" or "exon ligation conformation") (B. Schwer and Guthrie 1992). After exon ligation has been catalyzed, Prp22p acts on the post-catalytic spliceosome ("P complex") to release the exon-exon junction from the spliceosome (Company, Arenas, and Abelson 1991). Prp43p acts to disassemble the resultant intron-lariat spliceosome ("ILS complex") (Arenas and Abelson 1997; Martin, Schneider, and Schwer 2002), releasing the excised lariat intron for degradation and allowing for recycling of the snRNP

and NTC components of the spliceosome (Wahl, Will, and Lührmann 2009; Wilkinson, Charenton, and Nagai 2020).

In the context of acting on RNA-RNA duplexes in free solution, these DEAH-box ATPases bind to 3' ends of RNA, and translocate along RNA in a 3' to 5' direction to disrupt RNA-RNA structure (He, Andersen, and Nielsen 2010; He et al. 2017). However, in the context of the spliceosome, many of the RNA-RNA interactions these helicases are thought to disrupt are embedded deep in the spliceosome's center. Cryo-EM structures of the spliceosome also show density corresponding to Prp16p and Prp22p are located at the periphery of the spliceosome (S. Liu et al. 2017a; Zhan et al. 2018). DNA blocks, which prevent DEAH-box binding and translocation, placed just downstream of the RNA-RNA interactions disrupted by Prp16p, Prp22p, or Prp43p fail to block the rearrangements catalyzed by these helicases (Semlow et al. 2016; Toroney, Nielsen, and Staley 2019). Taken together, these data support a model in which DEAH-box ATPases act as "pullases," applying force to the interactions they remodel from a distance.

In addition to their ATP-dependent roles in initiating spliceosome rearrangements between catalytic steps, DEAH-box ATPases also have ATP-independent roles in promoting catalysis. Disabling the ATPase activity of Prp16p allows branching of an A-to-C mutation at brA ("brC") that would normally impair branching catalysis (Tseng, Liu, and Cheng 2011). Binding of Prp22p to the spliceosome both protects the 3'SS and allows exon ligation of 3'SS that are distal to the BS (Beate Schwer and Gross 1998). These data suggest Prp16p and Prp22p act to stabilize the branching and exon ligation catalytic conformations, respectively. Thus, these DEAH-box helicases do not only promote spliceosome rearrangement after catalysis of the forward reaction, but recruitment of

these factors promote the rate of the forward reaction as well. The requirement for these DEAH-box ATPases to bind prior to the catalysis of branching or exon ligation also positions these factors to enforce splicing fidelity by through kinetic proofreading.

RNA helicases ensure splicing fidelity through kinetic competition

In addition to their roles in splicing progression, DEAD- and DEAH-box ATPases also mediate splicing fidelity (reviewed in Semlow and Staley 2012). Splicing fidelity is postulated to occur through kinetic competition between the forward rate of the splicing reaction and the rate of the ATPase activity of the helicase. Two non-mutually exclusive mechanisms of kinetic proofreading may act to ensure splicing fidelity: the "sensor" and "timer" models (**Fig. 1.4**). In the sensor model, a suboptimal splice site triggers faster ATPase activity, leading to rejection of the suboptimal site prior to forward reaction. In the timer model, a suboptimal substrate slows the forward reaction rate, allowing time for the ATPase to reject the suboptimal substrate prior to catalysis or completion of the forward reaction.

In support of a kinetic proofreading model for splicing fidelity, reduced splicing of several consensus splice site mutations can be alleviated ("suppressed") by mutations that affect the RNA binding, helicase activity, or ATPase rate of DexD/H-box ATPases. The first of these "fidelity mutants" to be identified was *prp16-1*, which suppresses a branching defect conferred by brC (Couto et al. 1987). The *prp16-1* allele contains a mutation in the "A site" of Prp16p (S. Burgess, Couto, and Guthrie 1990), which lies in a conserved motif responsible for ATP binding (Fairman-Williams, Guenther, and Jankowsky 2010). Subsequent analysis of this and other alleles that suppress BS mutations found the

ATPase activity of Prp16p fidelity mutants is strongly reduced (S. M. Burgess and Guthrie 1993). Notably, these BS mutations specifically antagonize the branching step of splicing, and Prp16p activity is required after branching to remodel the spliceosome from the C complex to the exon ligation competent C* complex. Given the observation that Prp16p generally acts to move the U2:BS duplex out of the catalytic core both after successful branching (Fica et al. 2017; C. Yan et al. 2017) and when branching catalysis is inhibited by chemical modification of U6 snRNA or removing the reactive hydroxyl group from brA (Koodathingal and Staley 2013; Semlow et al. 2016), these results suggest branching is "proofread" through kinetic competition between branching catalysis and Prp16p ATPase activity.

In addition to Prp16p mediating branching fidelity, other DEAD- and DEAH-box ATPases act at other steps, likely through kinetic competition as well. Mutations in the helicase domain of Prp5p suppress the splicing defect of a mutation of the invariant U within the BS (Xu and Query 2007). Prp28p helicase mutations suppress the splicing

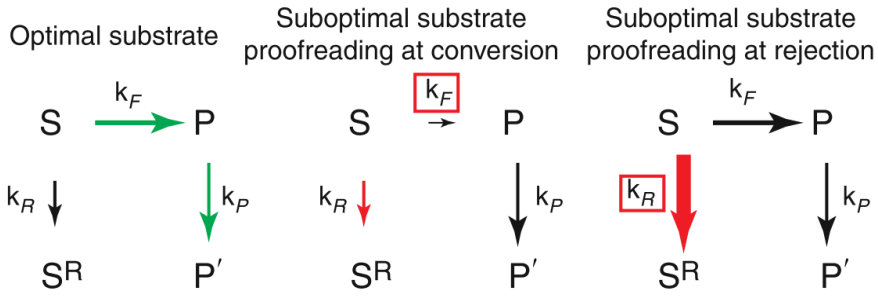

Figure 1.4. Comparison of kinetic proofreading schemes proposed to act in the spliceosome.

Kinetic proofreading consists of a competition between the rates of a forward reaction and a branched rejection pathway. Two non-mutually exclusive models can exist to ensure proofreading of a suboptimal substrate. In the "timer" model (center), the rate of the forward reaction is decreased relative to a fixed rejection rate. In the "sensor" model (right), a suboptimal substrate increases the rate of rejection relative to the forward reaction rate. (Figure adapted from Semlow and Staley, 2012).

defect of mutations in the 5'SS that weaken RNA-RNA interactions between the 5'SS and U6 snRNA (Yang et al. 2013). Cold-sensitive mutations of Prp2p, which indicate a deficiency in helicase activity, can suppress mutations that interfere with formation of the catalytic core (Wlodaver and Staley 2014). Prp22p helicase mutations suppress exon ligation defects conferred by point mutations within the 3'SS (Rabiah M Mayas, Maita, and Staley 2006).

The disassembly factor Prp43p acts after Prp16p-mediated (Koodathingal et al. 2010), Prp22p-mediated (Rabiah M Mayas, Maita, and Staley 2006), and likely after Prp2p-mediated proofreading (Jennifer Moore and Rebecca Toroney, unpublished data) to disassemble spliceosomes (Toroney, Nielsen, and Staley 2019). This disassembly is likely necessary to prevent re-sampling of the intron after proofreading activity (Semlow et al. 2016). However, recent studies suggest Prp43p ATPase activity can disassemble spliceosomes stalled at the step of U2 snRNP addition (Maul-Newby et al. 2021), can antagonize exon ligation and trigger splicing intermediate release in select budding and fission yeast transcripts (R. Kannan et al. 2013; Ram Kannan et al. 2015; Burke et al. 2018). These data, as well as the data I will detail in Chapter 3 of this thesis, suggest that Prp43p is, too, a *bona fide* fidelity factor that also proofreads spliceosomes deficient in splicing progression.

Mechanisms of DEAH-box ATPase activity and regulation

DEAH-box ATPases are members of the SF2 family of nucleic acid helicases (Fairman-Williams, Guenther, and Jankowsky 2010). These helicases are defined by the Aspartate-Glutamate-Alanine-Histidine residues in motif II of the helicase core, which

are responsible for ATP binding and hydrolysis (Jankowsky 2011). DEAH-box helicases can unwind RNA-RNA and RNA-DNA duplexes with 3' to 5' polarity *in vitro* in an ATP-dependent manner through an "inchworming" of the RecA2 and RecA1 RNA-binding domains to translocate along RNA (He et al. 2017). Although DEAH-box ATPases can translocate along RNA to disrupt RNA-RNA and RNA-DNA helices, these helicases acts as "molecular winches" to initiate spliceosomal rearrangements by pulling on the 3' end of their substrates while being anchored to the periphery of the spliceosome (Semlow et al. 2016; Toroney, Nielsen, and Staley 2019).

How the activities of DEAH-box ATPases are restricted to the spliceosome spatiotemporally remains an open question. Some DEAH-box ATPases, such as Prp2 and Prp43, require association with G-patch proteins to stimulate activity (Rong-Tzong Tsai et al. 2005) . These G-patch proteins stimulate DEAH-box ATPase activity by constricting movement of the RecA2 domain relative to RecA1, ensuring the helicase remains bound to and moves RNA processively (Bai et al. 2021). Fusing Prp43 with the G-patch of its cofactor Ntr1 leads to mistargeting of Prp43 activity (Fourmann et al. 2016) from its cognate substrate (Toroney, Nielsen, and Staley 2019). Notably, Ntr1 recruitment to the spliceosome is co-incident with Ntr2, and Ntr2 binding is incompatible with binding of factors associated with the catalytic stages of the spliceosome (Chen et al. 2013). Thus, for these spliceosomal DEAH-box ATPases, activity can be limited spatiotemporally by recruitment of G-patch co-factors.

Similarly, cryo-EM structures have shown recruitment of Prp16 and Prp22 can be promoted by the presence of "first step factors" such as Cwc25 and Yju2 for Prp16 or "second step factors" such as Slu7 and Prp18 for Prp22 (Galej et al. 2016; Wilkinson et al.

2017; S. Liu et al. 2017a). However, Prp16 and Prp22 do not require co-factors for stimulation of helicase activity *in vitro* (Y. Wang, Wagner, and Guthrie 1998; Tanaka and Schwer 2005), leading to the question of how these helicases are regulated. Notably, in contrast to Prp2 and Prp43, Prp16 and Prp22 have large N-terminal domains. Emerging evidence suggests that the N-terminal domains of SF2 family helicases can act to inhibit RNA binding, ATPase, and helicase activity (Absmeier et al. 2015; Absmeier, Santos, and Wahl 2019; Singh et al. 2021). In this thesis, I provide evidence for a role of the N-terminal domain of Prp16 in inhibiting spliceosomal activity (see Chapter 3), providing strong evidence that a subset of DEAH-box ATPases are regulated by autoinhibition.

CHAPTER 2

CHARACTERIZATION OF DISEASE-CORRELATED BIALLELIC MUTATIONS IN *DHX38* REVEAL A *cis*- INTERACTION BETWEEN THE N-TERMINAL AND HELICASE DOMAINS OF PRP16

This chapter is part of a collaborative project with Rafal Ploski of The Medical University of Warsaw. All experiments herein were designed and performed by Christopher A. J. Craddock except for Figure 2.1, which was performed by the Ploski lab.

Abstract

Errors in pre-mRNA splicing underlie a number of diseases, with many diseases stemming from dysregulation in branch site or 3' splice site choice. An increasing number of causal mutations are being found in factors associated with the spliceosome, a macromolecular machine that catalyzes pre-mRNA splicing. Here, we characterize biallelic mutations in *DHX38* that are correlated with a broad set of clinical phenotypes in a single patient. *DHX38* encodes Prp16, a DEAH-box ATPase implicated in both proofreading the branch site to maintain fidelity for the first catalytic step of pre-mRNA splicing as well as promoting rearrangements in the spliceosome that are necessary for the second catalytic step. Through characterizing these mutations, we found a previously undescribed N-terminal motif where the majority of disease-correlated mutations in Prp16 lie. We also find this motif, as well as the conserved motif VI in the helicase region of Prp16, mediate a *cis*- interaction between the N-terminal and helicase domains of Prp16. The residues of the N-terminal motif are predicted to bind to the RecA2 domain of the helicase core, suggesting this interaction is important for inhibiting the RNA binding activity of Prp16. We also present evidence that the N-terminal domain interacts with the N-terminal ratchet helix Brr2, and that a Brr2 mutation that weakens branch site fidelity abrogates this interaction. These data indicate the N-terminal domain autoinhibits Prp16 activity until it is recruited to the spliceosome and suggests that mutations either in the N-terminal motif or motif VI that disrupt the interaction between the N-terminal and helicase domains of Prp16 may be causal for disease by dysregulating Prp16 activity.

Introduction

Pre-mRNA splicing is an important step in eukaryotic gene expression wherein interrupting sequences ("introns") are excised and expressed regions ("exons") are ligated together to generate a mature mRNA molecule that can code for a specific polypeptide product. Regulation of pre-mRNA splicing through alternative inclusion or exclusion of exons through a process known as alternative splicing ("AS") allows for diversification of metazoan proteomes (Blencowe 2006). Upwards of 95% of pre-mRNA generated from human genes undergoes AS (Pan et al. 2008), and AS occurs on a tissue-specific level to mediate the production of alternative mRNA isoforms (E. T. Wang et al. 2008). In addition to inclusion or exclusion of sequences that code for a given polypeptide sequence, AS isoforms can also differ in the inclusion or exclusion of nucleotide sequence in regulatory untranslated regions ("UTRs") that lie upstream ("5'") or downstream ("3'") of the coding sequence ("CDS") (Pan et al. 2008).

In human neuronal tissue, AS patterns change with age, with upwards of 70% of this variation occurring during the first 14 years of development (Mazin et al. 2013). AS in developing neuronal progenitor cells ("NPCs") can impact cell physiology outcomes in brain development by altering subcellular protein localization through the inclusion or exclusion of exons to disrupt membrane targeting domains or protein localization signals (Xiaochang Zhang et al. 2016). AS of neuronal transcripts can also alter protein production at the molecular level through inclusion of poison exons that encode premature stop codons or altering mRNA lifetime through differential 3' UTR inclusion (Dillman et al. 2013; Xiaochang Zhang et al. 2016). AS is also necessary for neuronal

migration during brain development, as well as synapse formation and maintenance after neuronal maturation (Xiaochang Zhang et al. 2016; Vuong, Black, and Zheng 2016).

Dysregulation of AS, and splicing in general, underlies many diseases. Disruption of canonical AS patterns through loss of *trans*-acting splicing factors that modulate AS, such as NOVA2 and SRRM4, are observed in paraneoplastic opsoclonus-myoclonus ataxia and autism spectrum disorders, respectively. (Vuong, Black, and Zheng 2016; Irimia et al. 2014). Deletion mutants of the core spliceosome factor PRPF31 disrupts AS patterns in retinal organoids derived from patients with retinitis pigmentosa ("RP") (Buskin et al. 2018). Mutations in other core spliceosome factors are causal for disease, such as Brr2 also being attributed to RP (Zhao et al. 2009), or SF3b1 being linked to myleoplastic disorders (Tang et al. 2016; Carrocci et al. 2017).

The spliceosome is ribonucleic protein machine that catalyzes pre-mRNA splicing. Splicing catalysis is initiated through the stepwise assembly of the spliceosome onto the intron that is intended to be spliced out (Wahl, Will, and Lührmann 2009; Wilkinson, Charenton, and Nagai 2020). Assembly consists of the addition of ribonucleoprotein modules called snRNPs, each consisting of a small nuclear RNA ("snRNA") in complex with several conserved proteins. After initial assembly, the spliceosome undergoes several rearrangements of RNA-RNA and RNA-protein interactions to allow for catalysis of the branching and exon ligation transesterification steps that are necessary for full intron excision. These rearrangements are initiated by a set of eight RNA helicases, conserved from yeast to humans, belonging to the DExD/H-box family of ATPases (Staley and Guthrie 1998). In addition to their roles in promoting splicing progression, these RNA

helicases also act to maintain splicing fidelity by antagonizing splicing at incorrect nucleotide sequences (Semlow and Staley 2012).

Dysregulation of DExD/H-box ATPases is increasingly implicated in disease. Mutations in SF3b1 alter the allowable nucleotide sequences at the branch site, an intronic region that defines a reactant in the branching reaction, through altering binding of the DEAD-box ATPase Prp5 (Tang et al. 2016; Carrocci et al. 2017). Several mutations in the DEIH-box ATPase Brr2 that are causal for autosomal recessive RP lie in the ratchet helix of Brr2 and alter the ability for Brr2 to unwind RNA-RNA helices *in vitro* (Zhao et al. 2009; Ledoux and Guthrie 2016). Mutations associated with RP in the C-terminal fragment of Prp8, which binds to and modulates the activity of Brr2, also inhibit the unwindase activity of Brr2 (Maeder, Kutach, and Guthrie 2009). Notably, part of this C-terminal fragment forms a tail that binds into Brr2's helicase domain (Mozaffari-Jovin et al. 2013). Mutations that lie in *DHX38,* the gene encoding the DEAH-box ATPase Prp16 humans, have also recently been shown to be causal for autosomal recessive RP (Ajmal et al. 2014, 38; Latif et al. 2018).

Prp16 is a factor that binds transiently to the spliceosome and initiates spliceosome rearrangements necessary for the transition between the branching and exon ligation conformations (Beate Schwer and Guthrie 1991; B. Schwer and Guthrie 1992). Prp16 initiates these rearrangements by binding to the periphery of the spliceosome and pulling on the 3' end of the splicing substrate (Semlow et al. 2016; Galej et al. 2016; Zhan et al. 2018). Prp16 also acts to proofread the branch site of the pre-mRNA substrate to antagonize suboptimal nucleotide sequences in an ATPase-dependent manner (Couto et al. 1987; S. Burgess, Couto, and Guthrie 1990; S. M. Burgess and Guthrie 1993). Mutations

that reduce the ATPase activity of Prp16 allow usage of suboptimal sites (S. M. Burgess and Guthrie 1993), while perturbations in spliceosomal U6 snRNA that disallow branching chemistry are rendered susceptible to Prp16-dependent discard (Koodathingal et al. 2010), suggesting Prp16 enforces branch site fidelity through a kinetic proofreading model where the rate of ATPase activity competes with the rate of branching catalysis. Prp16 ATPase activity has also been shown to influence branch site usage to undock suboptimal substrates and allow for selection of alternative branch sites (Semlow et al. 2016). These data suggest the timing of Prp16 ATPase must be regulated to allow for splicing at the correct branch site sequence. As branch site selection influences 3' splice site ("3'SS") usage (see Chapter 3), and alternative 3'SS selection is a large driver of AS isoforms (E. T. Wang et al. 2008), mutations that disrupt regulation of Prp16 ATPase activity can have adverse effects on AS, which in turn can dysregulate neuronal development.

The N-terminal domain of Prp16 is necessary for regulating its function in the splicing pathway. This regulation is thought to be mediated spatiotemporally through the N-terminal domain mediating the recruitment of Prp16 to the spliceosome at the appropriate step in the splicing cycle (Y. Wang and Guthrie 1998). The N-terminal domain of Prp16 has also been shown to interact with the C-terminal helicase cassette of Brr2 via yeast-2-hybrid ("Y2H") assays, with this interaction leading to a decrease in Prp16 ATPase rate *in vitro* (Cordin et al. 2014), suggesting a regulatory function for the N-terminal domain. Despite these insights into the functional importance of the N-terminal domain in Prp16, this domain is thought to be poorly conserved across eukaryotic species (Y. Wang and Guthrie 1998).

Here, we characterize biallelic mutations in *DHX38* that are correlated with a clinical developmental phenotype. These mutations lie in separate alleles of *DHX38* and would cause missense mutations in separate regions of the Prp16 protein. We show through characterization of orthologous mutations in budding yeast *PRP16* that the maternal allele, which causes a G->R mutation in motif VI of Prp16, confers a cold-sensitive growth defect, consistent with a defect in unwindase activity. We also show the paternal allele, which causes an E->K mutation in the N-terminal domain, lies in close proximity to N-terminal mutations that are causal for RP. We discovered through sequence analysis that all disease-correlated mutations in the N-terminal domain ("NTD") lie in a highly conserved acidic motif ("N-terminal motif," or "NTM"). We further show through yeast-2-hybrid assays that the NTD interacts with the helicase-C-terminal-tail domain ("HCT") of Prp16, and mutations in motif VI and a charge-switch mutation of the NTM both disfavor this interaction. Finally, we show interaction between NTD fragments of Prp16 and Brr2 is abrogated by a mutation in the N-terminal ratchet helix of Brr2 associated with decreased branch site fidelity, suggesting Brr2 may sequester the NTD to activate Prp16 helicase function in the spliceosome. These data suggest a role for the NTM in mediating an autoinhibitory interaction between the NTD and HCT of Prp16, and further suggest the molecular phenotype underlying *DHX38*-associated disease is dysregulation of Prp16 autoinhibition.

Materials and Methods

Table 2.1: Yeast strains for *DHX38*/Prp16 study

YS107	*MATa ade2-101 his3Δ trp1 ura3-52 lys2-801 leu2 cup1Δ::ura3-52 prp16Δ::LYS2*, pSB2 (*PRP16 URA3 CEN4 ARS1*)	(S. M. Burgess and Guthrie 1993)
PJ69-4A	*MATa his3-200 leu2-3,112 trp1-901 ura3-52 gal4Δ LYS2::GAL1-HIS3 GAL2-ADE2 met2::GAL7-lacZ*	(P. James, Halladay, and Craig 1996)

Table 2.2: Plasmids for *DHX38*/Prp16 study

pSB58	*PRP16 TRP1 CEN4 ARS1*	(S. M. Burgess and Guthrie 1993)
pPR130	*BRR2 HIS3 CEN4*	(Raghunathan and Guthrie 1998)
pJPS2383	pPR130 *brr2-N1104A*	(Zhao et al. 2009)
pJPS2433	pPR130 *brr2-R1107L*	(Zhao et al. 2009)
pNRB271	*2μ LEU2* P_{ADH}*-GAL4-AD::DMC1-AD*	(Reitz, Grubb, and Bishop 2019)
pNRB272	*2μ TRP1* P_{ADH}*-GAL4-BD::DMC1-DBD*	(Gift from D. Bishop)
pJPS1971	*GPD-ACT1-CUP1 LEU2 CEN*	(Hilliker, Mefford, and Staley 2007)
pJPS1916	*GPD-act1-cup1-A259C* [brC] *LEU2 CEN*	(Hilliker, Mefford, and Staley 2007)
pJPS1922	*GPD-act1-cup1-U301G-A304G* [gAG] *LEU2 CEN*	(Hilliker, Mefford, and Staley 2007)

Table 2.3: Oligonucleotides for *DHX38*/Prp16 study

ACT1-CUP1_PE3	5'-ATTCGCTGAACCCGGTACC-3'
oCAJC-C28	5'-GCGGGAAGAACTGCGCCAAGAACTGCATATCGATTATAC-3'
oCAJC-C29	5'-GTATAATCGATATGCAGTTCTTGGCGCAGTTCTTCCCGC-3'
oCAJC-C30	5'-GAGAATGGTATGATAATGATAAGGATTATGGAAATTTGGTGCC-3'
oCAJC-C31	5'-GGCACCAAATTTCCATAATCCTTATCATTATCATACCATTCTC-3'
oCAJC-C71	5'-CGTCCAGCCAAGTATCCGCTTTATC-3'
oCAJC-C72	5'-CTAGCTCATCTTCATTATCGTCAAGGTCTAACTG-3'
oCAJC-D30	5'-ATCCATCGATGTCGACAAATAAAGAACTAACGCCTGGGTT-3'
oCAJC-D31	5'-AAAAAAGTATACAAAAACTAGTCTCGAGATTGTTGAATATGTTCCA-3'
oCAJC-D32	5'-ATCCATCGATGTCGACAAATCTTCTGGAGCTCCTCAAGAATAAATC-3'
oCAJC-D33	5'- AAAAAAGTATACAAAAACTAATGACCATTCTGTGTTTGCTGT-3'
oCAJC-D34	5'-ATCCATCGATGTCGACAAATATAAATAAGAGAAAGAAATTTAAGATCCAGTTAGACC-3'
oCAJC-D35	5'- AAAAAAGTATACAAAAACTAAAACTTTGATAAAAAAGGAGGGATCCAT-3'
oCAJC-D38	5'-TAGTTTTTGTATACTTTTTTAATGAAGATGACATTG-3'
oCAJC-D39	5'-ATTTGTCGACATCGATGGATC-3'
oCAJC-D40	5'-ATCCATCGATGTCGACAAATATGACTGAGCATGAAACGAAGG-3'
oCAJC-D41	5'-AAAAAAGTATACAAAAACTATTATTTCACATTTATTTCAAAGGACAACT-3'
oCAJC-E8	5'- ATCCATCGATGTCGACAAATTCAACATCAGAGCTTGGCAATG-3'
oCAJC-E10	5'-ATCCATCGATGTCGACAAATATGGGTCATTCGGGGCG-3'
oCAJC-E11	5'- AAAAAAGTATACAAAAACTATTTATTGTTATTCTCCAAGCCTAGTACTTC-3'
oCAJC-E13	5'- AAAAAAGTATACAAAAACTACTAAAAAAAAAGGCTTCCTTCTTTTGAAAG-3'
oCAJC-E15	5'-ATCCATCGATGTCGACAAATAAAGCAAAAGATAAGAGCAACCAAAAGATTTG-3'
oCAJC-E16	5'- AAAAAAGTATACAAAAACTAACTAAAAAAAAAGGCTTCCTTCTTTTGAAAG-3'
oCAJC-E38	5'-GATTTAAAAGACATTCCAGAAAAGGTTGTTCC-3'
oCAJC-E39	5'-GGAACAACCTTTTCTGGAATGTCTTTTAAATC-3'
oCAJC-E40	5'-CTAACGCCGACCAACAATCCGGAAGAGCGG-3'
oCAJC-E41	5'-CCGCTCTTCCGGATTGTTGGTCGGCGTTAG-3'
oCAJC-E42	5'-CTAACGCCGACCAAATTTCCGGAAGAGCGGG-3'
oCAJC-E43	5'-CCCGCTCTTCCGGAAATTTGGTCGGCGTTAG-3'
oCAJC-E44	5'-CTAACGCCGACCATAGATCCGGAAGAG-3'
oCAJC-E45	5'-CTCTTCCGGATCTATGGTCGGCGTTAG-3'

Prp16 allele generation: Plasmid pSB58 subjected to site-directed mutagenesis using oligo pairs oCAJC-C28/29 or oCAJC-C30/31 to generate *prp16-G696R* and *prp16-D179K*, respectively. To generate charge switch and alanine shave mutations of the NTM, pSB58 was amplified with oligo pair oCAJC-C71/72 to exclude the NTM region of Prp16. Gibson assembly was used to insert gBlocks containing codon-optimized sequence corresponding to the NTM with either the charge switch or alanine shave mutations.

Y2H plasmid generation: To generate bait plasmids for Y2H, plasmid pNRB272 was linearized with oligo pair oCAJC-D38/39. Fragments corresponding to Brr2 or Brr2-CC were generated by amplifying plasmid pPR130 with oligo pairs oCAJC-D40/41 or oCAJC-E8/D41, respectively, and inserted into linearized pNRB272 via Gibson assembly. Fragments corresponding to *brr2-N1104A* or *brr2-R1107L* were generated by amplifying plasmids pJPS2383 and pJPS2433, respectively, with oligo pair oCAJC-D40/41 and inserted into linearized pNRB272. A fragment corresponding to Prp16-HCT was generated with oligo pair oCAJC-E15/16 and inserted into linearized pNRB272. Mutations corresponding to *prp16-R686Q*, *prp16-R686I*, *prp16-Q685H*, or *brr2-L1951P* were generated via site-directed mutagenesis of bait plasmids using primer pairs oCAJC-E40/41, oCAJC-E42/43, oCAJC-E44/45, or oCAJC-E38/59, respectively.

To generate prey plasmids for Y2H, plasmid pNRB271 was linearized with oligo pair oCAJC-D38/39. Plasmid pS58 was amplified with oligo pairs oCAJC-D30/31, oCAJC-D32/33, oCAJC-D34/35, oCAJC-E10/11 or oCAJC-E10/13 to generate St1, St2, St3, NTD, or full-length *PRP16* sequence, respectively. *PRP16* fragments were inserted

into linearized pNRB271 via Gibson assembly. Mutations corresponding to *prp16-R686Q*, *prp16-R686I*, *prp16-Q685H* were generated via site-directed mutagenesis using primer pairs oCAJC-E40/41, oCAJC-E42/43, or oCAJC-E44/45, respectively.

Yeast transformation: Yeast transformations were performed using the rapid lithium acetate method (Gietz and Woods 2002) using 33 µl of yeast cell culture, 3.3 µl of 10 mg/mL salmon sperm DNA as a carrier, and 198 µl of 50% w/v PEG 3350, with a 30 minute incubation at 30 °C prior to a 15 minute heat shock at 42 °C. Transformed cells were plated on selective media, grown for 2 days at 30 °C, re-streaked onto fresh selective media, and grown for 3-5 days at 30 °C prior to subsequent analysis. For *PRP16* alleles, variants of pSB58 were transformed into yeast strain YS107. Transformants were plated onto 5-FOA media to select against retention of the wildtype pSB2 plasmid. For primer extension experiments, four independent strains from each *PRP16* allele transformation generated were re-streaked onto YPDA and transformed with *ACT1-CUP1* plasmids. For Y2H experiments, pNRB271 and pNRB272 variants were co-transformed into yeast strain PJ69-4A and plated on media to select for uptake of both plasmids.

Total RNA extraction: Yeast colonies were inoculated into 3 mL of complete synthetic media lacking uracil and grown overnight until saturation. Saturated cultures were backdiluted to an OD_{600} of 0.1-0.15 and grown at 30 °C for 2-3 doublings. Cultures were shifted to 16 °C for an additional 2 hours where indicated. A volume of 1.7 to 2 mL of these cultures was spun down, aspirated of supernatant, snap frozen in liquid nitrogen, and stored at -80 °C until ready for RNA extraction.

Yeast pellets were resuspended in 500 µl TRIzol, transferred to a 2 mL screw-cap tube containing 0.5 mm acid-washed glass beads, and lysed in a FastPrep-24 homogenizer for five one-minute cycles. An additional 500 µl TRIzol was added before proceeding with the manufacturer's instructions for RNA extraction. An additional chloroform extraction was performed with 400 µl choloroform prior to isopropanol precipitation. RNA pellets were resuspended in TE buffer pre-heated to 65 °C for at least ten minutes and adjusted to the same concentration prior to downstream experiments.

Primer Extension: Primer extension was performed with 4 to 12 µg of yeast total RNA using AMV reverse transcriptase. All experiments used used Quasar570 fluorophore-labeled ACT1-CUP1_PE3 primers (Biosearch, LLC). Total yeast RNA was added to an annealing buffer and primers were added to total yeast RNA to yield a final concentration of 1 µM each in 8 µL total volume (50 mM Tris-HCl pH 8.3, 60 mM NaCl, 10 mM DTT). The primer and total RNA solution was heated to 65 °C for 10 minutes and snap cooled on ice for at least 5 minutes. AMV RT and a dNTP mixture was added with 12 µL of a supplementary RT buffer containing spermidine and $Mg(OAc)_2$ for a final primer extension reaction mixture of 6U AMV RT, 1 µM each dNTP, 50 µM Tris-HCl (pH 8.3 at 42 °C), 60 µM NaCl, 30 µM $MgCl_2$, 10 µM DTT, and 1 µM spermidine. Primer extension was carried out for 60 minutes at 42 °C. Primer extension products were ethanol precipitates and resuspended in 5 µL of RNA loading buffer (95% formamide, 2x TBE, 0.5 µM EDTA). Extension products were visualized by running on a 12% denaturing polyacrylamide gel and visualized on a Typhoon scanner (Cytiva). Quantitation was performed using ImageQuant TL (Cytiva).

Results

DNA sequencing reveals biallelic DHX38 *mutations in a patient with a clinical developmental phenotype*

A four-year-old patient residing in Poland presented with several clinical phenotypes, to include blindness, retinal scarring, hypomyelination, epilepsy, stunted growth, and developmental retardation. Upon whole exome sequencing, biallelic mutations were found in *DHX38*, predicted to confer missense mutations in the resultant Prp16 protein (**Fig 2.1**). Notably, these mutations were inherited from both parents, who are heterozygous wildtype and present no clinical phenotype. Given the presence of retinal scarring, and previous identification of *DHX38* alleles that are causal for RP, we hypothesized the clinical phenotypes exhibited in our patient are caused by these biallelic mutations altering the function of Prp16p in our patient.

The maternal and paternal alleles affect different regions in the CDS of *DHX38*. The maternal cDNA sequence (**Fig. 2.1**, "matka") shows a G>C change at residue 2587, predicted to introduce a glycine to arginine mutation at position 863 in the Prp16 polypeptide sequence (herein referred to as "*dhx38-G683R*"). The paternal cDNA sequence (**Fig. 2.1**, "oiciec") shows a G>A change at residue 991, predicted to change a glutamine to a lysine at position 331 in the Prp16 protein (herein referred to as "*dhx38-E331K*"). The glycine residue at position 863 is conserved from *Saccharomyces cerevisiae* to humans, and lies in a motif that is strictly conserved across eukaryotic DEAH-box ATPases (Zhou and Reed 1998; Fairman-Williams, Guenther, and Jankowsky 2010). Although not strictly conserved, the charge of position E331 is conserved from *S.*

cerevisiae to humans; this residue is an aspartate in budding yeast (Zhou and Reed 1998). Given the possible conservation of these positions, we decided to characterize the functional impacts of these mutations using the well-characterized and genetically tractable model organism *S. cerevisiae* (herein referred to as "budding yeast").

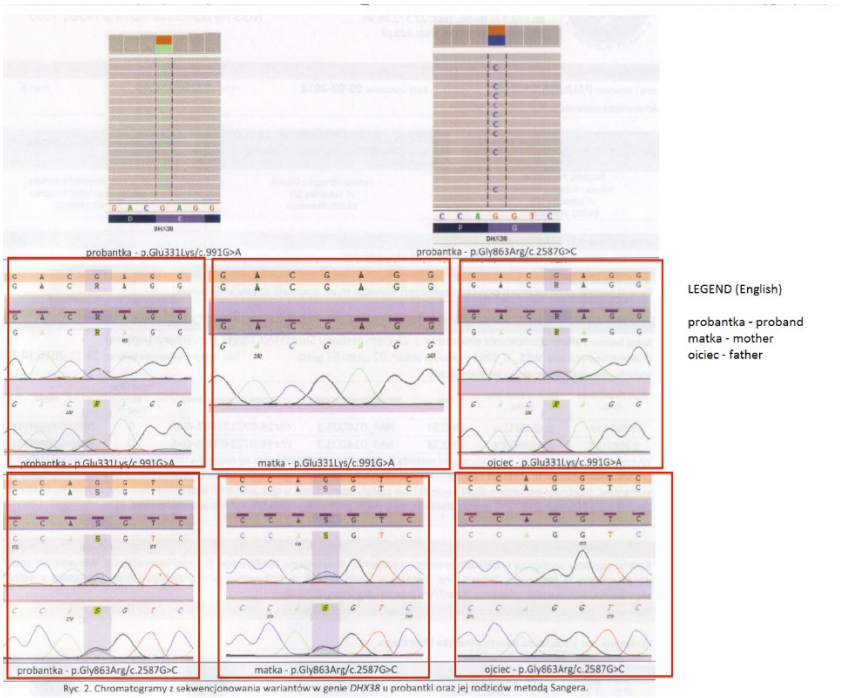

Figure 2.1. Genetic sequencing of the patient, mother, and father reveal inheritance of mutations in separate alleles of *DHX38*

Sequencing reveals missense mutations in either copy of *DHX38*, with a c.991G>A mutation in the paternal allele (oiciec) and a c.2587G>C mutation in the maternal allele (matka). These mutations are predicted to confer E331K and G863R mutations, respectively. Figure courtesy of Rafal Ploski.

The motif VI mutation in the maternal allele confers a cold-sensitive growth phenotype when introduced into yeast PRP16

The majority of spliceosome factors in budding yeast are conserved in humans (Wahl, Will, and Lührmann 2009; Wilkinson, Charenton, and Nagai 2020), and mutations causal for human disease have been successfully characterized using budding

yeast. In particular, mutations that causal for autosomal dominant RP in the *SNRNP200* gene in humans, which encodes human Brr2, have been shown to confer cold-sensitive growth defects when these mutations are introduced into budding yeast (Zhao et al. 2009; Ledoux and Guthrie 2016). As both the paternal allele in our patient and autosomal dominant RP mutations lie in the highly conserved helicase domain in Prp16 or Brr2, respectively, we expected to see similar effect when introducing our patient mutations in budding yeast.

The paternal allele harbors a mutation motif VI of the helicase domain (**Fig. 2.2A**), which is a motif conserved across DEAH-box ATPases and is responsible for regulating ATP binding and hydrolysis (Fairman-Williams, Guenther, and Jankowsky 2010). This mutation would change an uncharged glycine residue ("G") to a comparatively larger, polar arginine residue ("R") that restricts the allowable conformation that can be adopted by the polypeptide backbone. An orthologous residue in the structure of the model DEAH-box helicase Prp43 (He et al. 2017) shows this residue lies at the interface of a beta-sheet and a turn that connects to an alpha-helix that lies adjacent to the ATP binding site. When we mapped a G-to-R mutation at this position, all backbone-allowable rotamers caused steric hindrance with the structures around it, suggesting this mutation destabilizes this region in Prp16 (**Fig 2.2B**). Unfortunately, because the NTD of Prp16 is not conserved in Prp43, we could not perform this same analysis on the paternal *dhx38-E331K* mutation.

To determine whether these mutations impacted Prp16 function, we introduced mutations either orthologous to the maternal (*prp16-G696R*) or paternal alleles (*prp16-D179K*) into plasmid-borne copies of *PRP16* via site-directed mutagenesis. We then transformed these plasmids into yeast strains with the chromosomal copy of *PRP16* knocked out that contained a wildtype copy of *PRP16* on a plasmid that contained the

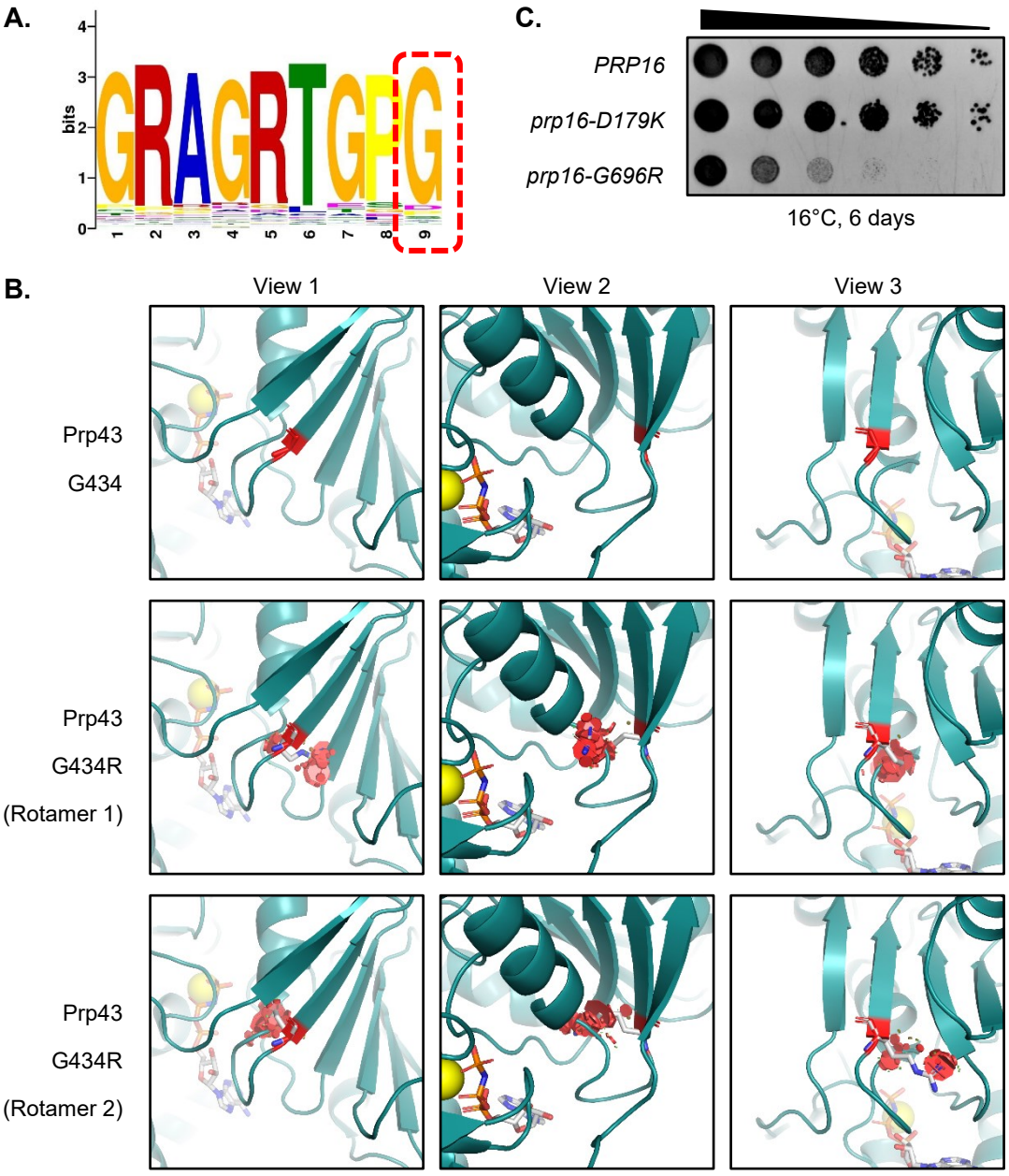

Figure 2.2. The maternal *dhx38-G863R* mutation lies in motif VI and confers a cold sensitive growth defect.

A. The *dhx38-G863R* mutation changes the last residue in the highly conserved motif VI.

B. Modeling in a *G>R* mutation in the crystal structure of the closely related DEAH-box ATPase Prp43 from *S. cerevisiae* (PDB accession number 5I8Q) suggests this mutation causes various steric clashes that would disrupt structural elements in Prp16.

C. A mutation orthologous to the maternal *dhx38-G863R* in *S. cerevisiae* (*prp16-G696R*) confers a cold-sensitive growth defect as a single copy, whereas a mutation orthologous to the paternal *dhx38-E331K* mutation (*prp16-D179K*), does not confer a growth defect at any temperature.

URA3 auxotrophic marker. We plated these transformants on 5-FOA-containing media

to select against retention of the wildtype *PRP16* plasmid, re-streaked colonies onto fresh YPDA media, grew colonies to saturation, spotted 1/10 serial dilutions of saturated culture onto YPDA plates, and incubated these strains at 16, 20, 25, 30, and 37 °C.

When comparing growth at several temperatures we found that *prp16-G696R*, but not *prp16-D179K*, conferred a noticeable cold-sensitive growth defect at 16 °C (**Fig 2.2C**). This result is consistent with mutations that affect helicase domains in spliceosomal helicases, and suggest a defect in ATP hydrolysis and/or helicase activity (Schneider, Hotz, and Schwer 2002; S. M. Burgess and Guthrie 1993). These data suggest that the paternal allele confers a defect in Prp16 helicase activity *in vivo*.

Disease-correlated mutations in DHX38 *that affect the N-terminus of Prp16 cluster in a conserved acidic motif*

To determine the whether the *prp16-D179K* mutation is truly orthologous to the *dhx38-E331K* mutation in the maternal allele of our patient, and whether this charge-switch mutation at this position had a consequence on function, we decided to look at the conservation of this residue across eukaryotic species. As a first step, we fetched protein sequences for Prp16 from fungal species in the Broad Institute's Fungal Orthogroups database (https://www.broadinstitute.org/fungal-genome-initiative), compared the aligned sequences, and found a 13-residue sequence that consisted largely of acidic amino acid residues (**Fig 2.3**). We did the same analysis using animal protein sequences fetched from the TreeFam database (Li et al. 2006), limiting our analysis to mammalian species due to poor sequence coverage of DHX38 protein sequences in other clades, and again found a highly similar motif (**Fig 2.4**). We then generated mammalian and fungal

consensus sequences using EMBOSS Cons and verified conservation by aligning the resultant consensus sequences using EMBOSS Stretcher (**Fig 2.5**).

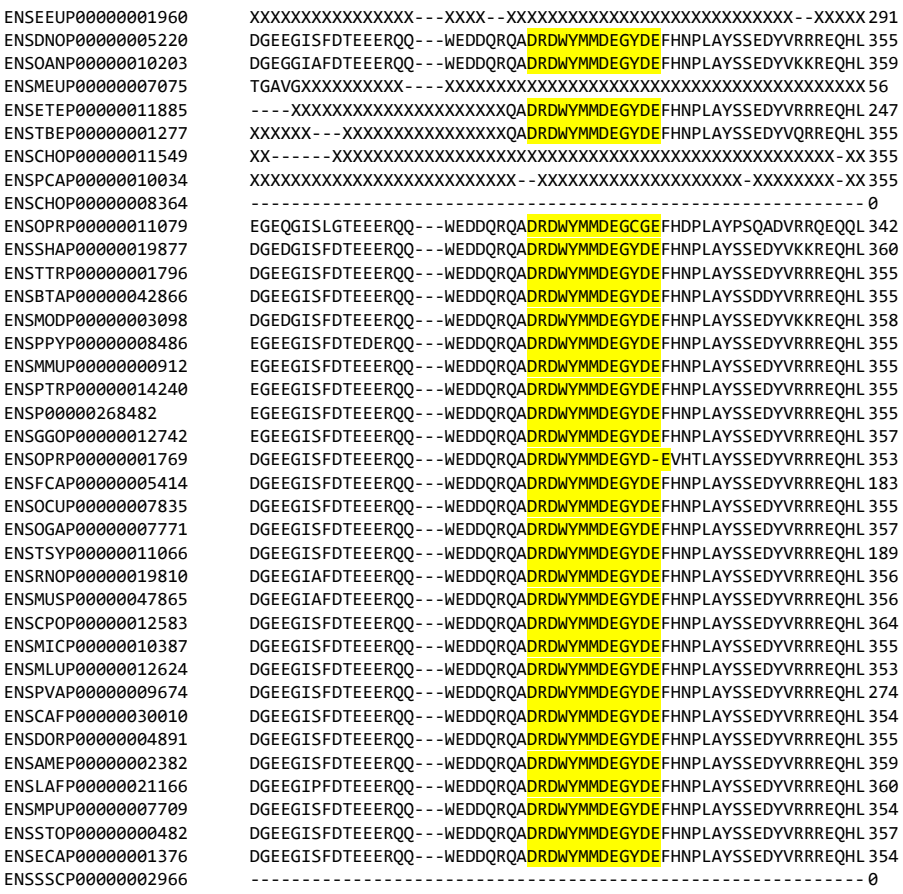

Figure 2.3. Alignment of mammalian DHX38 protein sequences reveal a strongly conserved acidic patch in the N-terminal domain.

A highly conserved stretch of residues is present in the N-terminus of DHX38. The putative 13-residue N-terminal motif is highlighted. DHX38 protein sequences pulled from the TreeFam database (http://www.treefam.org)

To further verify the presence of this conserved motif, we used MEME motif discovery (Bailey et al. 2015) using either Prp16 protein sequences from Fungal Orthogroups or DHX38 protein sequences from TreeFam. This analysis produced a conserved, 15-residue motif for fungal Prp16 (**Fig 2.6A**), of which 13 residues overlapped with the motif we discovered through sequence gazing and alignment. The TreeFam alignment pulled out a 12-residue motif that again overlapped with the 13-residue motif

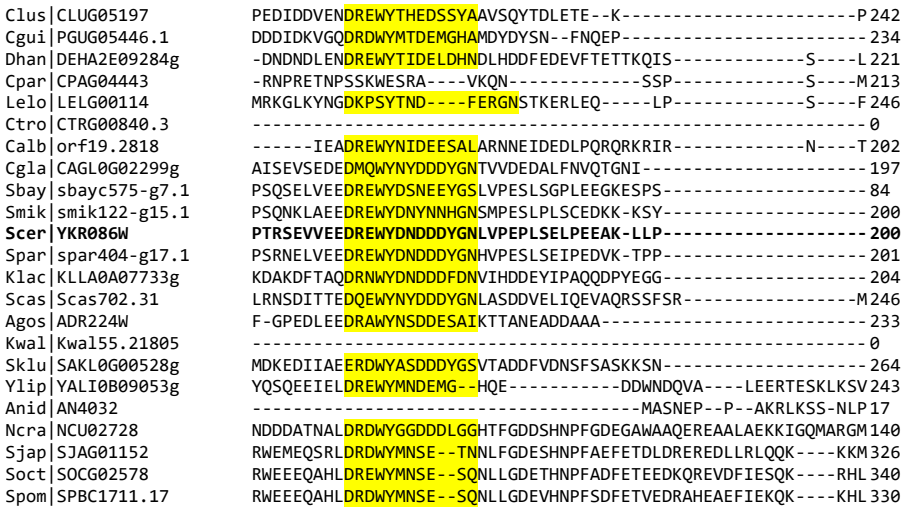

Figure 2.4. Alignment of fungal Prp16 protein sequences reveal a strongly conserved acidic patch in the N-terminal domain.

A highly conserved stretch of residues is present in the N-terminus of Prp16. The putative 13-residue N-terminal motif is highlighted. Prp16 protein sequences pulled from the Fungal Orthogroups database (https://www.broadinstitute.org/fungal-genome-initiative).

we discovered previously (**Fig 2.6B**). Given 12 of these residues are conserved between yeast and human (Zhou and Reed 1998), with the one residue downstream also being a charged residue (glutamate, or "E" in humans; asparagine, or "N" in yeast), we decided to define this N-terminal motif ("NTM") as 13 residues with the sequence DRDWYMMDEGYDE or DREWYDNDDYGN in humans and budding yeast, respectively.

Notably, two other mutations in *DHX38* that are causal for autosomal recessive RP also lie within this motif. These mutations either convert a polar, positively charged arginine residue to a polar uncharged glutamine residue (R324Q) or convert a glycine to a negatively charged aspartate residue (G332D, **Fig. 2.7A**). Aside from changing the charge identities of these residues in the NTM, the R324Q mutation removes a guanidinium group that can stabilize protein-protein contacts via π-π stacking interactions (Ferrari et al. 2020) and the G332D mutation would add a bulky sidechain which may prevent possible π-π interactions with its peptide backbone (Vernon et al.

```
DHX38_Mammal_cons    701 XXXGXXXXXXXEEXXXXGISFDTEEXXXXXERQQWXXXEDDQRQADRDWYMMDEGXYXXXDEFHNPLAYSSEDYVRRREQHLHKXQ 785
                         .       ..    ....:...     ...:.   ..:....||:||..|:. .    |.................|..... ...
Prp16_Fungal_cons    284 XXXXXXXXXXXXXXXXXXXXXXEXXXXXXXXXXXXEXXXXXEXXXXDREWYXNDDXXX---DYXXXXXXXXXXXXNXXXEDXXXXXX 365
```

Figure 2.5. Alignment of consensus sequences generated from mammalian DHX38 and fungal Prp16 protein sequences reveal the N-terminal motif is conserved from fungi to mammals.

Consensus sequences generated from mammalian DHX38 protein sequences (top) and fungal Prp16 sequences (bottom) show conservation of an acidic patch of residues in the N-terminal domain across these two clades.

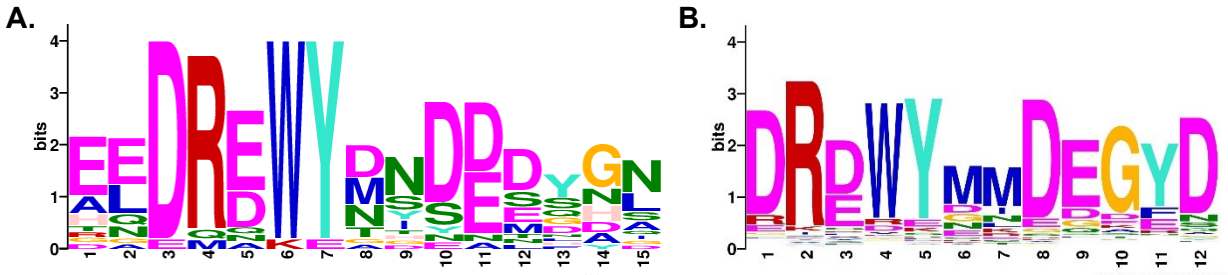

Figure 2.6. MEME motif discovery confirms the presence of the highly-conserved, acidic N-terminal motif in both animals and fungi.

A. Meme motif analysis performed on all DHX38 protein sequences from TreeFam pulls out an acidic motif which contains the 13-residue motif we identified through manual alignments.

B. Meme motif analysis of Prp16 protein sequences from the Funal Orthogroups database also pulls out a 13-residue motif containing the same residues we identified through alignments.

2018). As glycine residues lack a side chain, they can adopt a greater range of backbone conformations and thus are often found in regions of proteins where tight turns of the polypeptide backbone are necessary for proper structural conformation (see **Fig 2.2B**). Taken together, these observations suggest mutations in the NTM may confer disease phenotypes by disrupting Prp16 protein structure.

Mutations in the NTM do not confer growth defects in budding yeast

Given the conservation of the residues NTM, we hypothesized that more drastic mutations of the NTM could disrupt the function of Prp16 enough to confer a growth defect in yeast. Since *prp16-D179K* is a charge switch mutation, and that the NTM is

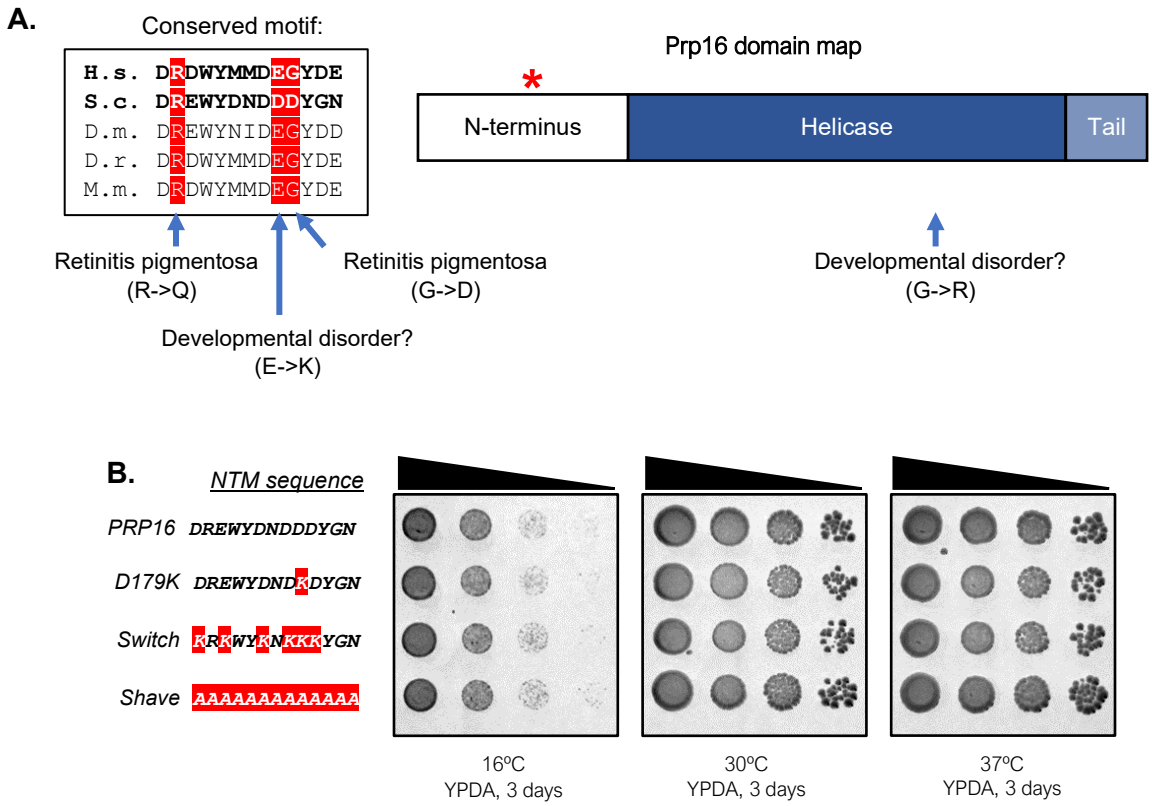

Figure 2.7. The N-terminal motif is a hot-spot for disease-correlated mutations in humans, but mutations in the NTM do not confer a growth defect in *S. cerevisiae*.

A. Domain map showing the location of mutations disease-correlated mutations. Three of four mutations lie in the conserved N-terminal motif.

B. Neither a charge-switch mutation nor an alanine shave of the N-terminal motif confer a growth defect in yeast.

highly negatively charged, we made a charge switch mutation in which all negatively charged residues in the NTM were changed to lysines ("Switch"). Given the identities of the residues would also play a role in side-chain-specific structural stabilization (see above), we also decided to make an "alanine shave" mutation in which all 13 residues were changed to alanine residues ("Shave"). We introduced these NTM mutations in a plasmid-borne copy of *PRP16*, shuffled these plasmids into yeast strains as described above, and again grew these strains at several different temperatures. Surprisingly, none of these drastic NTM mutations conferred a growth defect in yeast (**Fig. 2.7B**), suggesting these

mutations disrupt Prp16 function in a manner has a negligible effect in budding yeast as compared to humans (see Discussion).

The NTM may bind to the HCT domain of Prp16 to inhibit RNA binding or translocation

The NTD of Prp16 has been shown to interact with the HCT domain *in vitro* through immunoprecipitation assays (Y. Wang and Guthrie 1998). In addition, the sequence of the NTM is highly negatively charged, and bears a strong similarity to the sequence of the C-terminal tail of human Prp8 that has been shown to bind to the ratchet helix of Brr2 which binds RNA prior to spliceosome activation (Mozaffari-Jovin et al. 2013). Given these observations, we hypothesized the NTD of Prp16 can bind to the HCT, and that the NTM plays an important role in mediating this interaction.

To examine this, we downloaded the predicted yeast Prp16 structure from AlphaFold (Jumper et al. 2021), and superposed it with the structure of the model DEAH-box ATPase Prp43 bound to RNA (He et al. 2017). When we overlaid the NTM of Prp16 (**Fig. 2.8A**, yellow sticks) with Prp43 (teal), we saw that the positioning of the NTM would be sterically incompatible with where RNA (white) is bound to Prp43. Further, when we overlaid the helicase domains of the predicted Prp16 structure with the crystal structure of Prp43, we found the D179K mutation (red) is positioned near the 5' hairpin of the RecA2 domain (**Fig. 2.8B**, as denoted by residue E313 in Prp43 and residue E574 in Prp16, shown in sticks), which is important for Prp43 translocation along RNA (He et al. 2017). We also noticed the NTM does not traverse the entire channel in the way single-stranded RNA is positioned in the Prp43p crystal structure. Instead, AlphaFold predicts the NTM forms a loop to obscure only part of the helicase domain, akin to the way the tail of Prp8 binds to the RecA2 domain of the N-terminal Brr2 cassette to inhibit Brr2 helicase

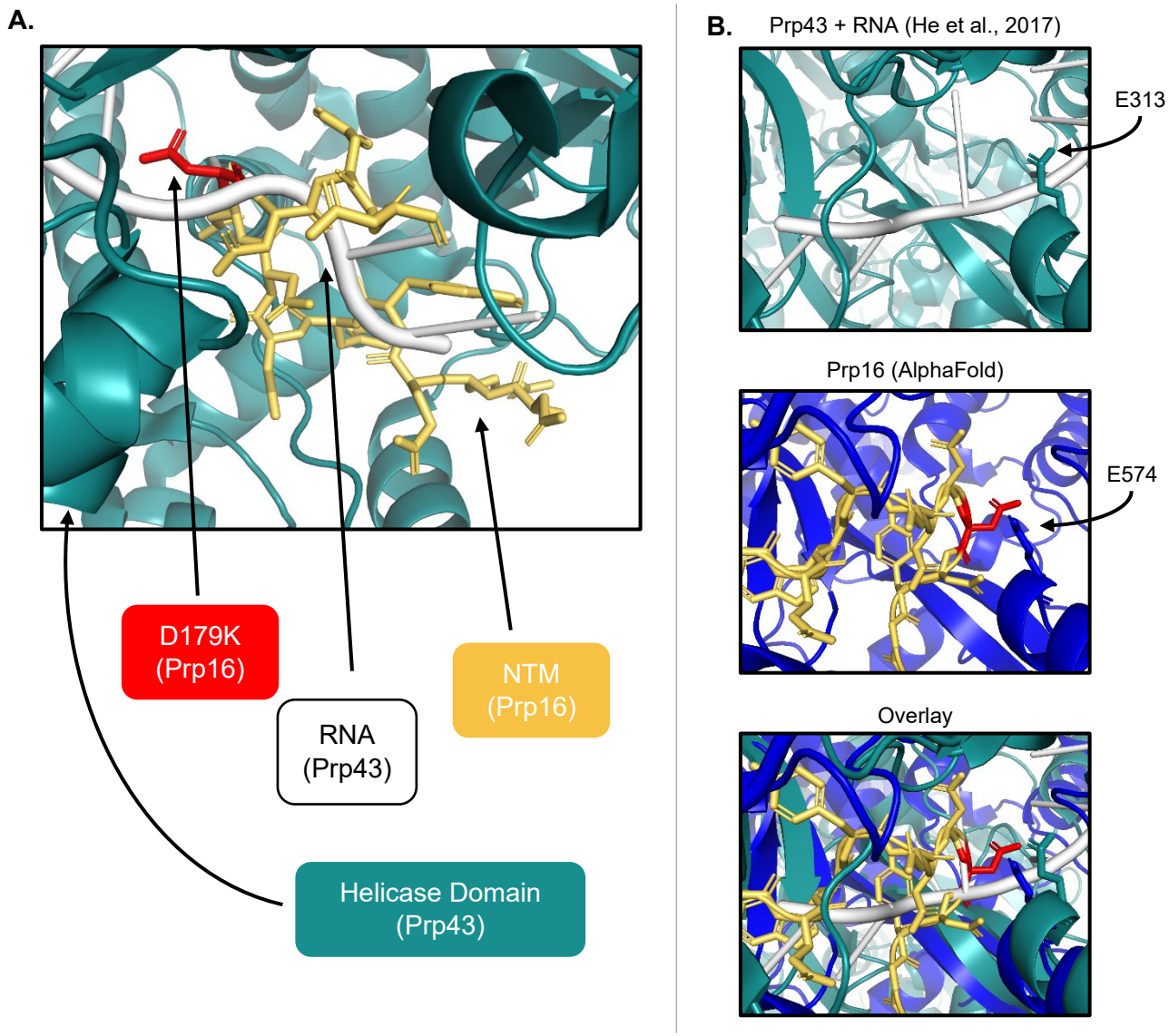

Figure 2.8. Comparison of the predicted Prp16 structure from AlphaFold with the crystal structure of Prp43 bound to RNA suggest the NTM may bind to the RecA2 domain of Prp16

A. Overlay of the N-terminal motif (gold) of Prp16 from AlphaFold (accession number P15938) overlaid onto the crystal structure of Prp43 (teal) bound to RNA (white) (PDB accession number 5I8Q) suggests the N-terminal motif would be sterically incompatible with RNA binding to the RecA2 domain of Prp16.

B. Alternate view of Prp43 bound to RNA (teal, white) and Prp16 (blue, gold). RecA2 is denoted by glutamate residues shown in sticks, bottom left.

activity (Mozaffari-Jovin et al. 2013). These observations suggest that the NTM mediates an interaction between the NTD and HCT domains of Prp16, possibly to autoinhibit Prp16

helicase activity.

Mutations in the NTM impede an interaction between the NTD and HCT domains of Prp16

To directly test whether the NTM can mediate an interaction between the NTD and HCT domains of Prp16, we tested for interaction between these domains using yeast-2-hybrid. We placed the HCT domain of Prp16 (defined as residues 321-1082) onto a bait vector containing a *TRP1* auxotrophic marker and placed the NTD of Prp16 (defined as residues 1-320) onto a prey vector containing the *LEU2* auxotrophic marker. We introduced NTM mutations into the NTD fragments as well as the *prp16-G696R* into the HCT domain. As a controls, we made a bait vector using Brr2 that is known to interact with N-terminal fragments of Prp16 via yeast-2-hybrid (van Nues and Beggs 2001) and included prey vectors that had either full-length *PRP16* or full-length *prp16-G696R* alleles. We co-transformed bait and prey plasmids into the PJ69-4A yeast strain that lacks *TRP1* and *LEU2* and contains a promoter that drives *HIS3* transcription only when an interaction between the bait and prey vectors occurs (Fromont-Racine, Rain, and Legrain 1997). We then spotted saturated liquid cultures of the resultant strains onto media lacking tryptophan, leucine, and histidine and incubated at a range of temperatures.

All four NTD prey vectors displayed a strong interaction with the HCT domain of Prp16, as determined by a strong growth phenotype on media lacking histidine when incubated at 30 °C (**Fig 2.9**). Growth of the strains containing NTD prey vectors in combination with HCT bait vectors displayed stronger growth when compared to strains containing NTD prey vectors with a Brr2 bait vector control, suggesting the NTD binds and preferentially to the HCT. Interestingly, this strong growth phenotype was reduced

when NTD prey vectors were combined with HCT bait vectors containing the *prp16-G696R* mutation, indicating this mutation in motif VI inhibits the interaction between the NTD and HCT domains in *trans* via our Y2H assay.

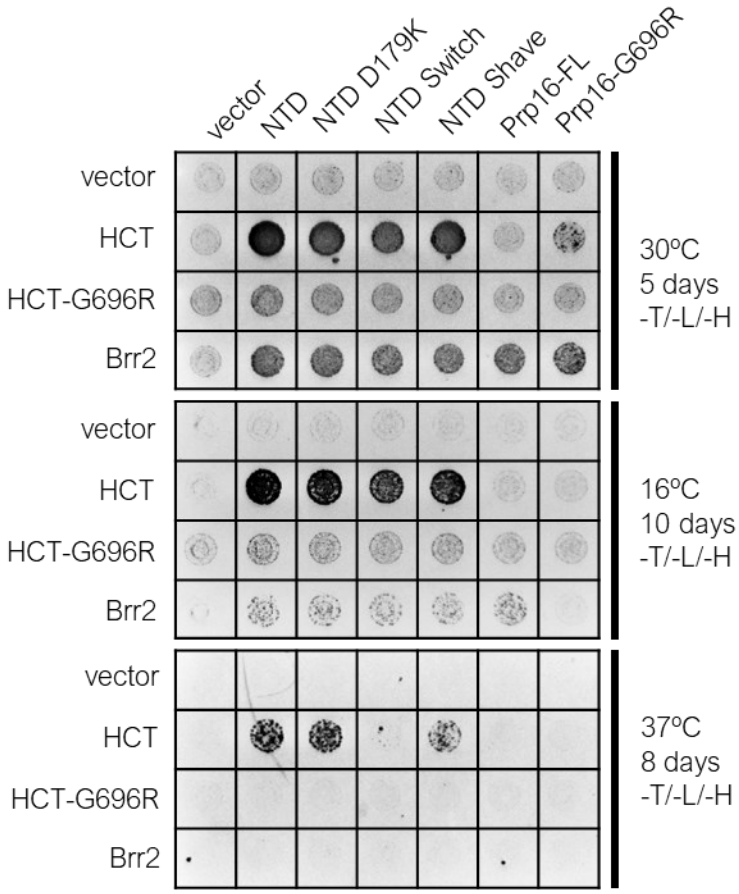

Figure 2.9. Both a charge switch mutation in the NTM and *prp16-G696R* mutation impairs an interaction between the N-terminal and helicase domains of Prp16.

Y2H shows the N-terminal domain (NTD) can interact with the helicase domain (HCT) of Prp16 in *trans*. The charge switch mutation in the NTD impairs this interaction at 37 °C, and a *prp16-G696R* mutation reduces this interaction at all temperatures. In the contest of full-length Prp16, the *prp16-G696R* mutation promotes interaction with the HCT domain in *trans* at 30 °C, but antagonizes interactions with the HCT of Prp16 and Brr2 at 16 and 37 °C.

Growth of strains containing NTD prey vectors and wildtype HCT bait vector was robust at 16 °C, with levels of growth above that seen with our Brr2 control. For both wildtype and *prp16-D179K* NTD prey vectors, growth was also robust at 37 °C while most other combinations failed to grow over a course of 8 days. However, the prey vectors

containing drastic alanine shave and charge displayed a reduction in interaction with the HCT bait vector, with the alanine shave mutation exhibiting a modest decrease in growth and the charge switch mutation failing to grow at all. These data are consistent with our hypothesis that the NTM mediates an interaction between the NTD and HCT.

We also found that a full-length *PRP16* prey vector failed to interact with a wildtype HCT bait vector at all temperatures, but a full-length *prp16-G696R* prey vector could form a weak interaction with a wildtype HCT bait vector. These data indicate the HCT domain can interact with the NTD of Prp16 in *cis*, and that the *prp16-G696R* motif VI mutation can inhibit interactions between the NTD and HCT in both *trans* and *cis*. These data also suggest that mutations in motif VI that interfere with the NTD and HCT interaction in *cis* can free up the NTD to promote interactions of the NTD in *trans*. Curiously, the weak interaction between full-length *prp16-G696R* and the HCT were lost at 37 °C and 16 °C, as was the relatively strong interaction between this prey construct and the Brr2 bait vector, suggesting that this motif VI mutation may impart cold sensitivity due to sequestration of the NTD at low temperatures.

Motif VI mutations reduce the interaction between the NTD and HCT domains of Prp16

Our results showing the *prp16-G696R* mutation also reduces the interaction between the NTD and HCT of Prp16 led us to the hypothesis that motif VI mutations in general may act in the same way. Notably, mutations in motif VI of Prp16 can suppress an L1951P mutation in the H2-Sec63-2 domain of Brr2 that impedes a Y2H interaction with full-length Prp16 (Cordin et al. 2014). Given our data above, we hypothesized that motif VI mutations generally inhibit interaction between the NTD and HCT of Prp16, allowing for increased Brr2 interaction. To test this, we performed a Y2H assay with either

prey vectors containing the Prp16-NTD or full-length *PRP16* with and without these motif VI mutations (R686I and R686Q). As a control, we also created a full-length *prp16-Q685H* prey vector, which contains a mutation that failed to increase interaction with Brr2 via Y2H (Cordin et al. 2014). We also created bait vectors containing these same mutations in the context of only the HCT of Prp16. We also generated bait vectors corresponding to the Brr2 C-terminal cassette (Brr2-CC), which contains the H2-Sec63-2 domain, either with or without the L1951P mutation. We co-transformed the bait and prey vectors and assayed for whether motif VI mutations interfered with the interaction between the NTD and HCT by looking for changes in growth at 30 °C.

We found that the NTD of Prp16 could interact with both wildtype and Q685H HCT domains via Y2H (**Fig 2.10A**). Likewise, we found that full-length wildtype and *prp16-Q685H* failed to interact with wildtype or Q685H HCT domains in *trans*, suggesting these HCT alleles maintain strong intramolecular interaction with the NTD in *cis*. Conversely, both R686I and R686Q mutations. in the context of the HCT domain only, strongly abrogated the interaction with the NTD in *trans*. The full-length constructs *prp16-R686I* and *prp16-R686Q* both allowed weak but noticeable Y2H interaction with the HCT domain in *trans* when compared to wildtype full-length *PRP16*, suggesting these mutations also disrupt a *cis* intramolecular interaction between the HCT and the NTD. Notably, in our Y2H assay the R686Q mutation in the HCT bait construct led to the strongest loss of Y2H efficiency with the NTD prey construct, whereas the *prp16-R686Q* full-length prey construct conferred the strongest growth phenotype when paired with the Q685H HCT and *brr2-CC-L1951P* bait constructs. The R686I mutation, on the other hand, led to an intermediate phenotype, suggesting an R-to-I mutation has less of an impact on disrupting the HCT and NTD interaction. These data indicate that motif VI

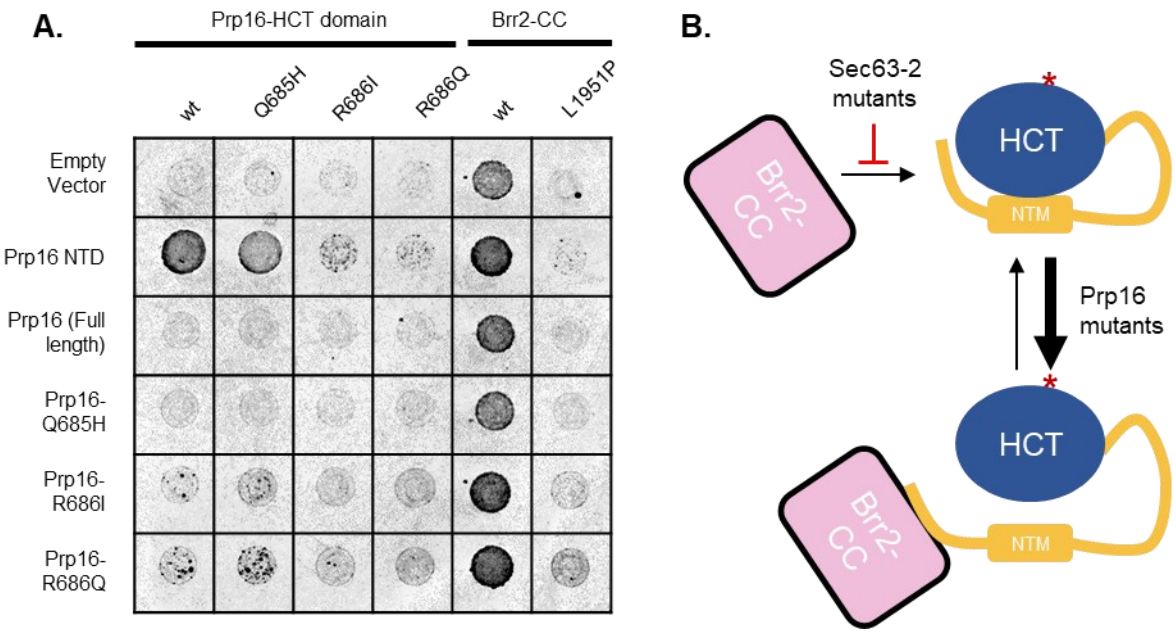

Figure 2.10. Mutations in motif VI destabilize an interaction between the NTD and HCT, allowing the NTD to more readily interact with other factors in *trans*

A. Y2H shows motif VI mutations R686I and R686Q decrease the interaction between the NTD and HCT in *trans*. These mutations were previously seen to increase interaction of full-length Prp16 with the C-terminal H2-Sec63-2 domain in Brr2. Q685H, which does not increase an interaction between Prp16 and Brr2, maintains an interaction with the NTD when placed in the HCT. When placed in full-length Prp16, R686I and R686Q also allow for an interaction with Prp16-HCT and Brr2-CC containing an L1951P mutation, suggesting these mutations at position R686 in full-length Prp16 interfere with antagonize an interaction between the HCT and the NTD in *cis*.

B. Model for how mutations in the C-terminal cassette of Brr2 and mutations in motif VI of Prp16 modulate interactions by interfering with interactions with the NTD of Prp16.

mutations inhibit an interaction between the NTD and HCT domains in both *trans* and *cis*, and that inhibiting this interaction frees the NTD to allow for greater interactions between Prp16 and other factors in *trans* (**Fig 2.10B**).

A mutation in the ratchet helix of Brr2 abolishes interaction with N-terminal fragments of Prp16

Our data suggest that the NTD plays an autoinhibitory function to repress the RNA binding activity of Prp16 prior to recruitment to the spliceosome. This led us to

hypothesize that Prp16 binding to the spliceosome relieves this autoinhibition by dislodging the NTD from the HCT domain of Prp16. Given prior Y2H data showing polypeptide sequence corresponding to fragments of the Prp16 NTD can bind to Brr2 (van Nues and Beggs 2001), we hypothesized that Brr2 relieves Prp16 autoinhibition after Prp16 is recruited to the spliceosome.

Cryo-EM structures of the spliceosome immediately after branching catalysis suggest Prp16 binds in close proximity to the N-terminal cassette of Brr2 (**Fig. 2.11A**) (Galej et al. 2016; Zhan et al. 2018). While the NTD of Prp16 is not present in these structures, the orientation of Prp16 shows its N-terminal region is positioned close to the ratchet helix of the N-terminal Brr2 cassette (**Fig. 2.11B**). Interestingly, the ratchet helix of the N-terminal cassette of Brr2 has been shown to bind the tail region of Prp8 which bears striking similarity to the identity of the first five residues of the NTM (DREDLY and DRDWYD, respectively) (Mozaffari-Jovin et al. 2013). Additionally, an N1104A mutation in the ratchet helix of the N-terminal cassette, which corresponds to the S1087A RP mutation in human Brr2, also leads to relaxation of BS fidelity in yeast (Mayerle and Guthrie 2018). Because the BS is proofread by Pp16, this observation suggests *brr2-N1104A* confers a dysregulation in Prp16 helicase activity. Given these data, we hypothesized the N1104A mutation in Brr2 led to relaxed BS fidelity by impeding an interaction between Brr2 and the NTD of Prp16.

To test this hypothesis, we made bait plasmids containing either wildtype, or *brr2-N1104A* alleles. As an additional control, we included the *brr2-R1107L* mutation, which confers decreased growth of yeast strains containing reporters constructs with aberrant BS sequences, suggesting this allele imparts a defect in splicing unrelated to inhibition of Prp16 activity. We co-transformed these bait plasmids with prey plasmids containing

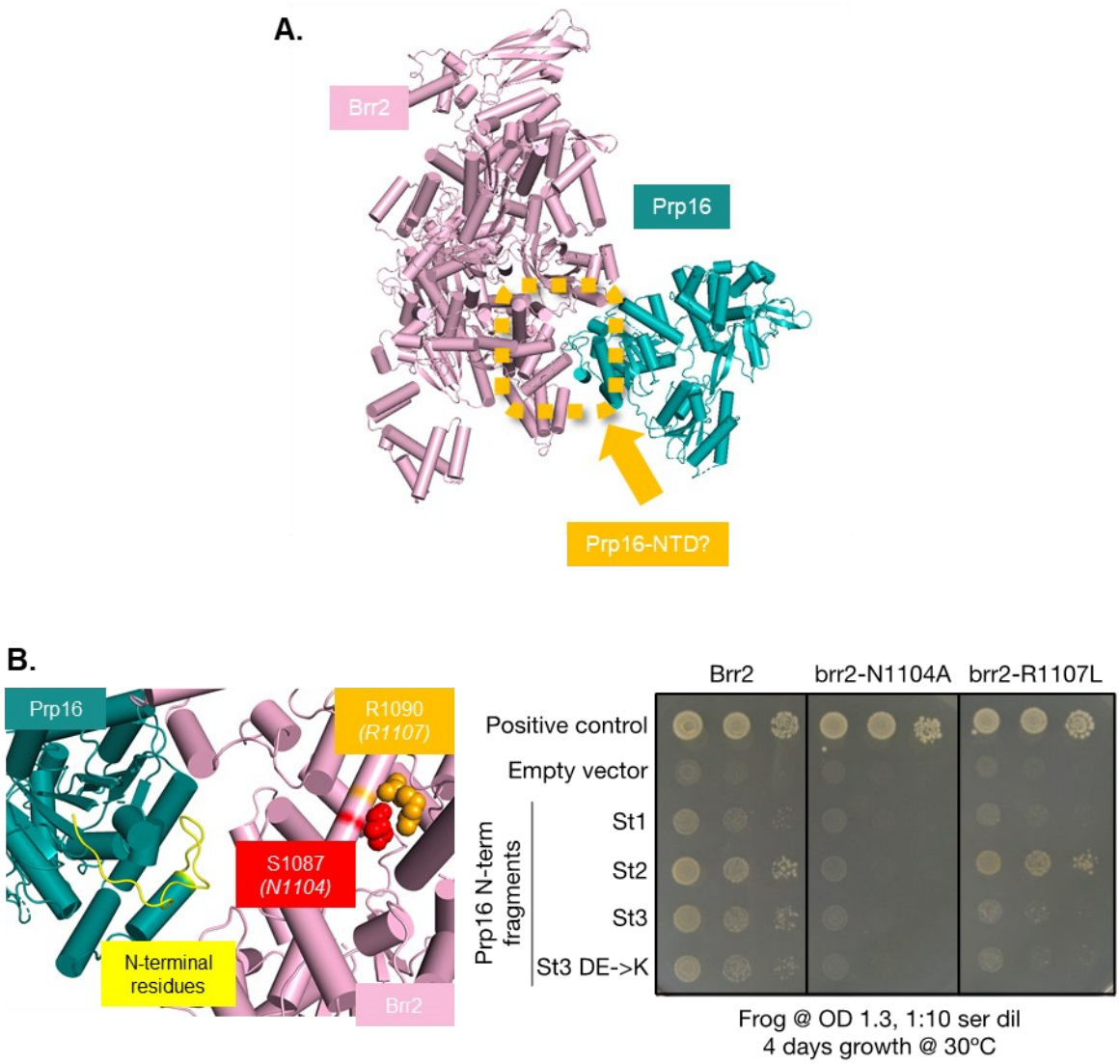

Figure 2.11. A mutation in the N-terminal ratchet helix of Brr2 abolishes an interaction between Brr2 and the NTD of Prp16.

A. A cryo-EM structure of Prp16 bound to the spliceosome (PDB accession number 5LJ5) shows Prp16 in close proximity to the N-terminal cassette of Brr2.

B. Close-up of the Brr2 N-terminal ratchet helix, with RP alleles highlighted and residues just downstream of the NTD of Prp16 highlighted. The *brr2-N1104A* mutation, which leads to decreased branch site fidelity in budding yeast, ablates interaction between Brr2 and the NTD of Prp16.

polypeptide sequences corresponding to NTD fragments shown to interact with full-length Brr2 previously (St1, St2, St3) (van Nues and Beggs 2001), and included a charge switch mutation in the St3 fragment. We then spotted saturated liquid cultures in 1/10

serial dilutions onto media lacking histidine and assayed for interaction by observing growth over four days at 30 °C. For a positive control, we used bait and prey plasmids corresponding to the *DMC1* binding and activation domains. We found that the *brr2-N1104A*, which conferred relaxed BS fidelity in budding yeast (Mayerle and Guthrie 2018), abrogated the Y2H interaction with all Prp16 NTD fragments tested (**Fig. 2.11C**). This is consistent with our hypothesis that *brr2-N1104A* interferes with binding of the NTD of Prp16 and suggests the fidelity defect of *brr2-N1104A* is due to an inability of Brr2 to derepress Prp16. Curiously, the *brr2-R1107L* mutation also abrogated the interaction of Brr2 with the St3 fragments of Prp16. Importantly, the St2 fragment of Prp16 does not include the NTM and instead includes 22 residues upstream of the St3 fragment (residues 65-154 for St2 vs 87-236 for St3), and the charge switch mutation of the NTM in St3 confers only a mild growth phenotype with a wildtype *BRR2* bait plasmid. These data suggest that derepression of Prp16 may rely on Brr2 binding the more N-terminal residues the Prp16 NTD, and that the NTM may instead be more important for the Prp16 autoinhibitory interaction. In this sense, Brr2 binding to more N-terminal regions of the NTD would pull the NTM out of proximity of the Prp16 RecA2 domain to derepress RNA binding activity.

Discussion

Disease-correlated DHX38 *alleles affect regions that mediate a* cis*-interaction between the NTD and HCT domains of Prp16*

In this study, we examined the impact of biallelic mutations in *DHX38* that are correlated with a clinical developmental phenotype. These mutations occur in separate alleles, with the paternal allele coding for a missense G->R mutation in motif VI of the

HCT domain and the maternal allele coding for a missense E->K mutation in the NTD domain of Prp16. Using Y2H, we determined both motif VI and a conserved NTM are both important for mediating an interaction between the NTD and HCT domains of Prp16. These data suggest that the mechanism underlying the developmental disorder in our patient may be destabilizing this interaction.

Numerous lines of evidence are emerging showing that RNA binding proteins and helicases involved in RNA metabolism employ an autoinhibition mechanism, wherein a *cis* interaction of a distal protein domain sterically blocks the RNA binding domain of a given protein. This has recently been shown in the crystal structure of the human processosome factor DHX37. Much like the spliceosome, the processosome is another macromolecular ribonucleoprotein machine that is involved in ribosomal RNA processing and relies on the activity of DHX37 for progression through these processing steps. In the crystal structure of DHX37, the NTD of this helicase forms a loop that obscures the RecA2 domain (Singh et al. 2021), similar to the predicted structure of Prp16 generated by Alphafold (**Fig. 2.8**) and akin to an N-terminal "plug" domain binding to RecA2 in Brr2 (Absmeier et al. 2015). Importantly, all the above helicases are proposed to translocate RNA in a 3' to 5' manner, and RecA2 binds the 5' end of RNA transcripts during translocation (He et al. 2017). Thus, NTD binding to the RecA2 domain could interfere with initial RNA binding to autoinhibit helicase activity.

Notably, motif VI lies within the RecA2 domain of DEAH-box ATPases (Fairman-Williams, Guenther, and Jankowsky 2010). Binding of RNA alters the conformation of the RecA2 domain in the model DEAH-box ATPase Prp43, leading to repositioning of several residues in motif VI (He et al. 2017). Part of this movement involves shifting of an alpha-helix within motif VI by 4 Å to allow for tighter interaction with ATP. The motif VI

mutation in our patient, *dhx38-G863R*, may introduce steric hindrance (**Fig. 2.2B**) to prevent the RecA2 domain from adopting a conformation that is competent for binding to the NTD of Prp16. Similarly, the mutations we tested at position R686 change this residue from a large, flexible arginine residue to shorter, more rigid isoleucine or glutamine residues, which may also interfere with a conformational change that allows for proper binding of RecA2 to the NTD. In this sense, both mutations in motif VI and the NTM would phenocopy each other in conferring an inability for RecA2 to bind the NTD.

Implications of Prp16 dysregulation for the human intronome

In comparison to the limited intronome of budding yeast, the vast majority of the human genome encodes intronic sequence. While only several hundred genes in budding yeast contain an intron (Spingola et al. 1999), every human gene contains an average of seven introns, each spanning several kilobases long (Hong, Scofield, and Lynch 2006). In addition to an increased number of introns, the splice site consensus sequences within each human intron are more degenerate than the strictly-conserved sequences in budding yeast (Wahl, Will, and Lührmann 2009). Notably, the BS sequences in humans can deviate markedly from the yeast consensus sequence of UACUAAC (Mercer et al. 2015; Zeng et al. 2021). The increased splicing load in addition to the higher number of potential splice sites present a challenge in maintaining quick and accurate splicing in human cells.

Prp16 is one of several factors that ensure BS fidelity through a kinetic proofreading mechanism. Prp16 activity is temporally limited to the branching catalytic step, where it disengages the BS from the catalytic core during proofreading (Semlow et al. 2016). Upstream of the catalytic stages of splicing, the DEAD-box ATPase Prp5 acts to proofread U2 snRNA binding to the BS of the intron (Xu and Query 2007). Recently,

mutations in the factor SF3b1 have been shown to modulate the recruitment of Prp5, which has the downstream affect of decreasing or increasing the likelihood of rejection of BS sequences that deviate from the consensus sequence in budding yeast (Carrocci et al. 2017; Tang et al. 2016). These SF3b1 mutations are causal for myelodysplastic syndromes, and lead to altered usage of BS sequences in humans.

As noted earlier, alternative BS selection impacts 3'SS selection (Carrocci et al. 2017; Tang et al. 2016; Zeng et al. 2021). A switch in alternative 3'SS selection has been shown to occur during development in *C. elegans* (Ragle et al. 2015), dysregulation of BS selection has a downstream impact on 3'SS selection in myelodysplastic disorders associated with SF3b1 (Darman et al. 2015), and optimal exon ligation decreases with increasing distance between the BS and 3'SS (Beate Schwer and Gross 1998 and see Chapter 3). Therefore, altered usage of BS genome-wide through Prp16 dysregulation can also lead to defects in downstream splicing events, altering a number of developmental and neuronal pathways, leading to the broadened phenotype we see in our patient.

A model for autoinhibition of Prp16 by the NTD

These data suggest that Prp16 activity is regulated spatiotemporally via its NTD, consistent with a previous study showing the NTD is important for Prp16 recruitment and activity (Y. Wang and Guthrie 1998). Our study shows that the NTD binds to the HCT domain of Prp16 (**Fig. 2.8, 2.9**) through a conserved NTM, which is predicted to bind to the RecA2 domain to sterically occlude RNA binding. We propose a model in which Prp16 not bound to the spliceosome is autoinhibited by its NTD (**Fig. 2.12**), and upon binding to the spliceosome the NTD is disengaged from the HCT to allow for RNA binding and ATPase activity. Our data showing a mutation in the ratchet helix of the N-terminal

cassette of Brr2 can abrogate the interaction with fragments of the NTD of Prp16 (**Fig. 2.11**) strongly suggest that Brr2 relieves the autoinhibition of Prp16, allowing for spatiotemporal regulation of Prp16 activity through recruitment to the spliceosome.

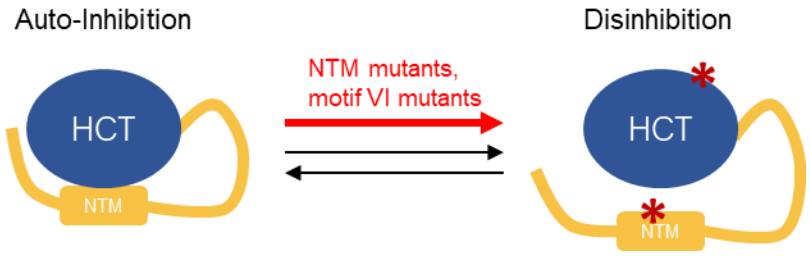

Figure 2.12. A model of Prp16 autoinhibition by a *cis* interaction between the NTD and HCT domains.

Prp16 autoinhibition would be mediated by the N-terminal motif binding to the RecA2 domain, preventing RNA binding and helicase activity. Mutations in the NTM and motif VI interfere with this autoinhibitory interaction, leading to a loss in spatiotemporal regulation of Prp16 activity to dysregulate branch site usage during pre-mRNA splicing.

CHAPTER 3

PRP43 DISCARDS SPLICING INTERMEDIATES DEFICIENT IN 3' SPLICE SITE BINDING

This chapter is in preparation for publication and adapted from the unpublished manuscript "Prp43p discards splicing intermediates deficient in 3' splice site binding." All experiments described herein were designed and performed by Christopher A. J. Craddock.

Abstract

Pre-mRNA splicing, an essential step in eukaryotic gene expression catalyzed by the spliceosome, must proceed with high fidelity to ensure faithful gene expression. Splicing fidelity is maintained by accessory factors belonging to the DEAH-box family of RNA helicases that transiently bind to and act on the spliceosome. Here, we instead reveal a novel role for the DEAH-box helicase Prp43p in proofreading 3' splice site binding in *Saccharomyces cerevisiae*. We found evidence that a mutation in the 5' splice site consensus sequence impairs entry into the exon ligation conformation, narrowing the optimal branch site to 3' splice site distance and deforms the 3' splice site binding pocket, all of which imply a defect for the 5' splice site mutation in 3' splice site docking. Further, we found that the exon ligation defect of this 5' splice site mutation is alleviated by a mutation in the ATPase center of Prp43p; similarly, in the context of a wildtype 5' splice site sequence, this Prp43p mutation allows increased exon ligation of substrates with longer branch site to 3' splice site distances. These data uncover a novel role for Prp43p in discarding splicing intermediates that inefficiently bind the 3' splice site.

Introduction

Pre-mRNA splicing is an essential step in eukaryotic gene expression. Splicing involves the excision of interrupting sequences (introns) out of precursor messenger RNA (pre-mRNA) and ligation of flanking expressed sequences (exons) to form mature mRNA that can be translated by the ribosome to a functional polypeptide product. Splicing is catalyzed in two successive transesterification steps, known as branching and exon ligation, by a multicomponent ribonucleoprotein machine, known as the spliceosome (Staley and Guthrie 1998). The spliceosome consists of five small nuclear

ribonucleoprotein complexes (snRNPs), each consisting of several protein factors in complex with a small nuclear RNA (snRNA) component. These snRNPs are added in a stepwise manner during spliceosome assembly, which occurs *de novo* on each intron that needs to be excised (Wahl, Will, and Lührmann 2009). After spliceosome assembly, several remodeling events significantly alter the conformation and composition of the protein and RNA components of the spliceosome to allow for catalysis of each transesterification reaction. Further rearrangements allow for release of the mRNA product and disassembly of the spliceosome to allow for recycling of snRNPs for subsequent splicing events (Hoskins and Moore 2012; Wan, Bai, and Shi 2019; Wilkinson, Charenton, and Nagai 2020). While rearrangements drive optimal substrates along the canonical splicing pathway, the same rearrangements also discriminate against suboptimal substrates to augment substrate specificity. Specificity in splicing in important to ensure the correct intronic sequence is excised from pre-mRNA in order to minimize deleterious effects in downstream steps of gene expression. Although dysregulation of gene expression through errors in splicing has been noted to underlie many diseases (Manning and Cooper 2017), how these errors are minimized through spliceosomal rearrangements remains to be fully investigated.

Spliceosomal rearrangements are initiated by accessory factors belonging to the SF2 superfamily of helicases (Fairman-Williams, Guenther, and Jankowsky 2010). Three DEAH-box RNA helicases act before and after catalysis of each transesterification reaction: Prp16p, Prp22p, and Prp43p. Prp16p initiates rearrangements after branching that allow for exon ligation (B. Schwer and Guthrie 1992; Semlow et al. 2016), and Prp22p releases mRNA after exon ligation (Company, Arenas, and Abelson 1991). After a successful round of splicing, Prp43p disassembles the spliceosome, releasing the excised

intron for degradation and freeing snRNP components for another round of splicing (Arenas and Abelson 1997; Martin, Schneider, and Schwer 2002).

In addition to their roles in splicing progression, the ATPase activities of Prp16p, Prp22p, and Prp43p play roles in maintaining splicing fidelity (Semlow and Staley 2012). Both Prp16p and Prp22p play a general role in disengaging splicing reactants from the catalytic core during splicing progression as well as during proofreading of suboptimal substrates (Semlow et al. 2016). These helicases likely proofread by acting either as "sensors" to disengage suboptimal substrates at a rate faster than optimal substrates (Semlow and Staley 2012) or as "molecular clocks" (Koodathingal and Staley 2013) with set rates that antagonize suboptimal reactants with relatively slow rates of branching (S. M. Burgess and Guthrie 1993) or exon ligation (Mayas, Maita, and Staley 2006). After rejection by Prp16p or Prp22p, suboptimal substrates are released from the spliceosome by Prp43p (Koodathingal *et al.* 2010; Mayas *et al.* 2010), likely to prevent potential usage of these suboptimal sites through re-sampling (Semlow et al. 2016; Toroney, Nielsen, and Staley 2019). Interestingly, in some cases discard of suboptimal 3'SS by Prp43p has been shown to act independently of Prp22p activity (Rabiah M Mayas, Maita, and Staley 2006; R. M. Mayas et al. 2010). Prp43p in humans has also been implicated in preventing accumulation of spliceosomes that are slowed in U2 snRNP addition (Maul-Newby *et al.* 2021). These observations suggest Prp43p not only disassembles spliceosomes bound to suboptimal substrates but also acts to ensure on-pathway spliceosomes progress through the splicing cycle at an optimal rate.

Prp22p strongly discriminates against point mutations within the 3' splice site consensus sequence (3'SS), comprised of the final three residues of the intron (Mayas *et al.* 2006). The 3'SS binds to the spliceosome through a network of several RNA-RNA

interactions (**Fig. 3.1D**). The ultimate guanosine residue of the intron (G-1) forms a non-Watson-Crick hydrogen bonding interaction with the first guanosine residue of the intron in the 5' splice site consensus sequence (G+1) (Liu *et al.* 2017). The G-1 residue also stacks onto residue A51 of U6 snRNA (U6-A51) to properly position the 2' hydroxyl of G-1 for efficient catalysis of exon ligation (Wilkinson et al. 2017). Binding of the 3'SS is further stabilized by hydrogen bonding between the Hoogsteen faces of the penultimate residue of the 3'SS (A-2) and the branch site adenosine (brA) (Liu *et al.* 2017; Wilkinson *et al.* 2017). Importantly, brA is covalently linked to G+1 after the branching reaction, creating the binding site for the 3'SS, making successful branching a prerequisite for efficient exon ligation. Single point mutations within the 3'SS at G-1 and A-2, as well as mutations at the G+1 and the brA residues that constitute the 3'SS binding site, are strongly antagonized *in vitro* by Prp22p ATPase activity (Mayas *et al.* 2006). Prp22p also proofreads a U-to-G transversion mutation at the antepenultimate position of the 3'SS (gAG), which weakens 3'SS binding through a steric clash with a conserved glutamine residue in the spliceosomal scaffolding protein Prp8 (Wilkinson et al. 2017). However, double mutations that change the identity at both G-1 and A-2 of the 3'SS not only block exon ligation but also are insensitive to Prp22p-mediated proofreading (Mayas *et al.* 2006), suggesting that substrates that are severely impaired in 3' splice site binding are proofread independent of Prp22p, perhaps by discard mediated by Prp43p.

Increasing evidence also indicates 5' splice site consensus sequence (5'SS) identity can mediate 3'SS choice (B. Schwer and Guthrie 1992; Hilliker, Mefford, and Staley 2007; Semlow et al. 2016), possibly by negatively impacting the efficiency of exon ligation. Previous characterization of an A-to-C substitution at position +3 of the 5'SS (A3c) in

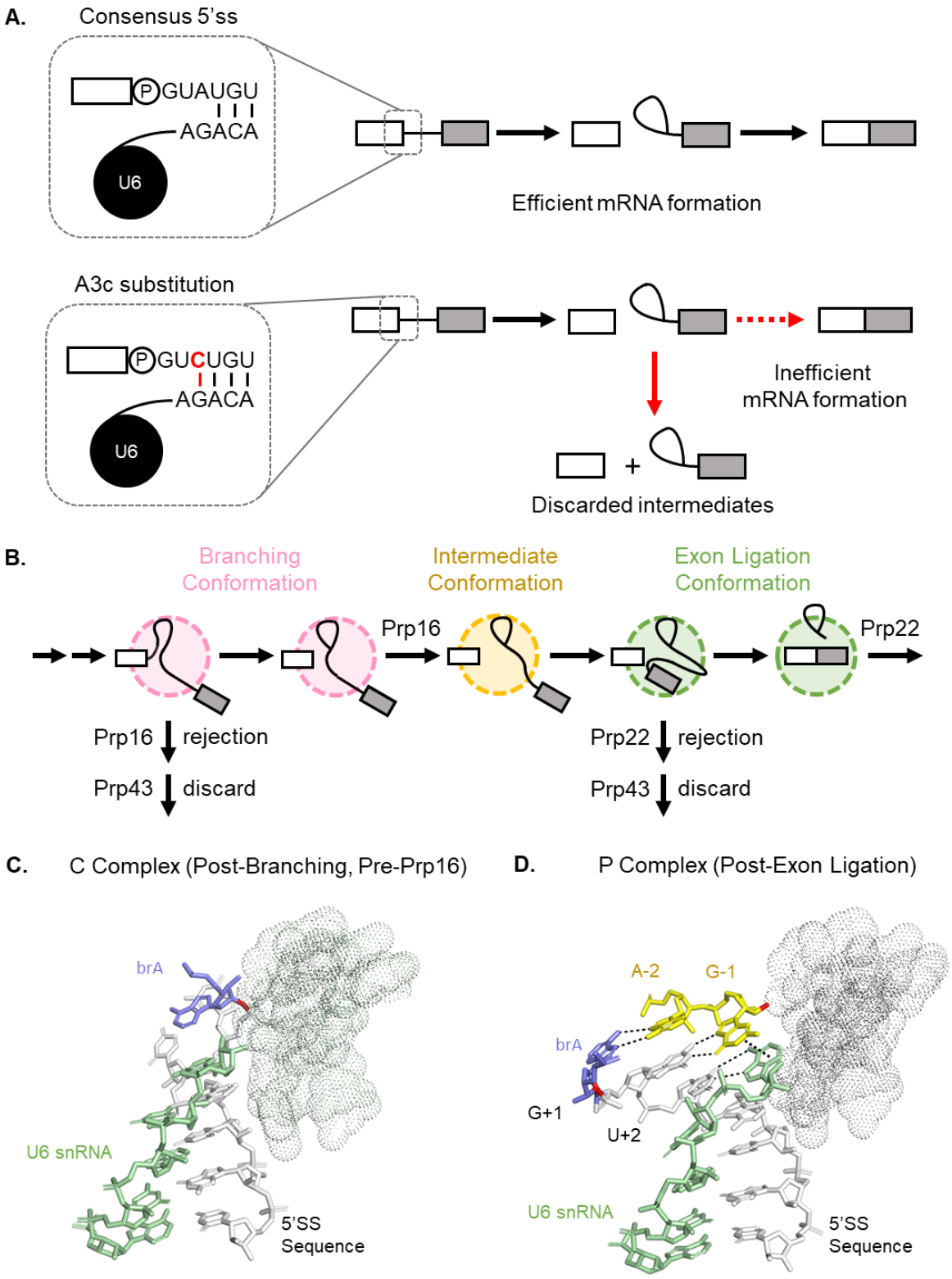

Figure 3.1. A3c blocks exon ligation at one of several steps along the splicing pathway.

Figure 3.1: A3c blocks exon ligation at one of several steps along the splicing pathway (continued).

A. The A3c substitution introduces an additional base pair between the 5'SS and U6 snRNA, leading to inefficient exon ligation and triggering discard of splicing intermediates.

B. Overview of the proofreading and progression steps at the catalytic stages of pre-mRNA splicing.

C. Cryo-EM structure of the catalytic core immediately after branching and before Prp16p ATPase activity (C* complex, PDB accession number 5LJ3) showing juxtaposition of brA (blue) and insertion of the reactive 2' hydroxyl (red) with the catalytic triplex (mesh).

D. Cryo-EM structure of the catalytic core immediately after exon ligation (P complex, PDB 6BK8) showing juxtaposition of 3'SS-G3 (yellow) and insertion of the reactive 2' hydroxyl (red) into the catalytic triplex. Stabilization of 3'SS-G3 binding is achieved by several interactions between the 3'SS and brA, U6-A51, and 5'SS residues.

budding yeast indicates A3c antagonizes exon ligation through the introduction a Watson-Crick base pair between the 5' splice site consensus sequence (**Fig. 3.1A**) (Konarska, Vilardell, and Query 2006). This A3c substitution is also found in the *TER1* intron in some fission yeasts introns, which use the spliceosome as 3' end processing machinery to generate telomerase RNA (Ram Kannan et al. 2015). As telomerase RNA is only encoded by the 5' exon of *TER1*, the free 5' exon intermediate must be released from the spliceosome prior to the exon ligation step of splicing to allow for incorporation into the telomere reverse transcriptase (TERT) complex (R. Kannan et al. 2013). These observations suggest A3c both blocks exon ligation and triggers spliceosome disassembly to free intermediates generated by the branching step of splicing. As Prp43p catalyzes spliceosome disassembly to release splicing intermediates, we hypothesized that Prp43p may also discard A3c intermediates that block splicing after branching catalysis.

After branching catalysis, the spliceosome undergoes at least two rearrangements to allow for exon ligation. Prp16p catalyzes rearrangement out of the branching

conformation in an ATP-dependent manner, removing the brA from the catalytic core and driving the spliceosome into a biochemically distinct, catalytically inert intermediate state. (B. Schwer and Guthrie 1992; Hilliker, Mefford, and Staley 2007; Ohrt et al. 2013; Semlow et al. 2016; Chung et al. 2019) The spliceosome then undergoes a separate, Prp16-independent arrangement that is accompanied by docking of the 3'SS and two "second-step factors" known at Slu7p and Prp18p, which in turn recruit Prp22p to stabilize the exon ligation conformation (Frank and Guthrie 1992; Beate Schwer and Gross 1998; S.-A. James, Turner, and Schwer 2002; Hilliker, Mefford, and Staley 2007; Ohrt et al. 2013; Chung et al. 2019). Given two rearrangements are necessary to allow for exon ligation, the step along the splicing pathway where A3c blocks exon ligation and triggers discard remains unknown.

Here, we present evidence that the A3c substitution antagonizes exon ligation by inhibiting 3'SS binding, and that Prp43p acts to discard intermediates that are generally deficient in 3'SS binding. Our data show A3c does not affect the Prp16p-dependent rearrangement in budding yeast, as we had anticipated, but instead affects a downstream rearrangement in U2 snRNA. Structural analyses *in silico* suggests A3c antagonizes 3'SS docking, as the implicated Watson-Crick base pair between A3c and U6 snRNA would displace key residues required for stable 3'SS binding. A3c narrows the range of BS to 3'SS distances that allow for efficient exon ligation when compared to introns with a wildtype 5'SS consensus sequence, indicative of a 3'SS binding defect. Mutations in both Prp22p and Prp43p alleviate the exon ligation defect of A3c, further implying 3'SS binding stability is compromised. Finally, we show Prp43p antagonizes exon ligation of introns containing a consensus 5'SS but long BS to 3'SS distances that compromise juxtaposition of the 3'SS with the catalytic core. These data establish evidence for a novel role for

Prp43p in discarding spliceosomes that inefficiently bind the 3'SS and further suggest Prp43p enforces an upper limit on BS to 3'SS distance.

Materials and Methods

Table 3.1: Yeast strains for Prp43p study

YHM118	*MATa ade2 his3 leu2 lys2 trp1 snr6::LEU2 snr20::LYS2 ura3* pU2U6U (*SNR6 SNR20 URA3 CEN*)	(Madhani and Guthrie 1994a)
YHM187	*MATa ade2 his3 leu2 lys2 trp1 prp16-302 snr6::LEU2 snr20::LYS2 ura3* pU2U6U (*SNR6 SNR20 URA3 CEN*))	(Madhani and Guthrie 1994b)
yJPS568	*MATa his3 leu2 lys2 met15 prp43 ::KanMX4 ura3* pJPS298 (pRS315 *LEU2 PRP43*)	(Leeds et al. 2006)
yJPS575	*MATa his3 leu2 lys2 met15 prp43 ::KanMX4 ura3* pJPS643 (pRS315 *LEU2 prp43-Q423N*)	(Leeds et al. 2006)
yJPS969	*MATa ade2-101 his3Δ200 leu2Δ1 lys2-801 prp22 ::LEU2 trp1-63 ura3-52* pJPS1493 (p358-PRP22 *TRP1 PRP22 CEN AMP*)	(Mayas *et al.* 2006)
yJPS972	*MATa ade2-101 his3Δ200 leu2Δ1 lys2-801 prp22 ::LEU2 trp1-63 ura3-52* pJPS1502 (p358-prp22-R805A *TRP1 prp22-R805A CEN AMP*)	(Mayas *et al.* 2006)

Table 3.2: Plasmids for Prp43p study

pCG92	GPD-*ACT1-CUP1 URA3 CEN*	(Lesser and Guthrie 1993)
pJPS1453	pCG92 *ACT1-CUP1-gAG*	(Mayas *et al.* 2006)
pJPS1462	pCG92 *ACT1-CUP1-A3c*	This study
pSX6	pSE358 *SNR6 TRP1 CEN*	(Madhani and Guthrie 1992)
pJPS796	*SNR20 LEU2 CEN*	(Hilliker *et al.* 2007)
pJPS870	pJPS796 *snr20-C59g*	(Hilliker *et al.* 2007)
pJPS871	pJPS796 *snr20-G100c*	(Hilliker *et al.* 2007)
pJPS895	pJPS796 *snr20-C59g-G100c*	(Hilliker *et al.* 2007)
pJPS216	pSE362 *SNR20 HIS3 CEN*	(Hilliker *et al.* 2007)
pJPS1695	pJPS216 *snr20-U98a*	(Hilliker *et al.* 2007)
pJPS1295	pJPS216 *snr20-U99a*	(Hilliker *et al.* 2007)
pJPS1546	pJPS216 *snr20-G100a*	(Hilliker *et al.* 2007)
pJPS1299	pJPS216 *snr20-U101a*	(Hilliker *et al.* 2007)
pJPS1702	pJPS216 *snr20-U102a*	(Hilliker *et al.* 2007)
pJPS1300	pJPS216 *snr20-A103u*	(Hilliker *et al.* 2007)
pJPS1542	pJPS216 *snr20-A103g*	(Hilliker *et al.* 2007)
pJPS1706	pJPS216 *snr20-C104a*	(Hilliker *et al.* 2007)
pJPS1374	pJPS216 *snr20-C104u*	(Hilliker *et al.* 2007)
pJPS1709	pJPS216 *snr20-A105u*	(Hilliker *et al.* 2007)
pJPS1708	pJPS216 *snr20-A105g*	(Hilliker *et al.* 2007)

Table 3.3: Oligonucleotides for Prp43p study

ACT1-CUP1_PE3	5'-ATTCGCTGAACCCGGTACC-3'
oCAJC-D64	5'-CTGTGTCTCATGTACTAACTGCTTCATTCTTTTTGTTGC-3'
oCAJC-D65	5'- GCAACAAAAAGAATGAAGCAGTTAGTACATGAGACACAG-3'
oCAJC-D70	5'- GTGTCTCATGTACTAACATATTATATGTTTAGAGGTTGG-3'
oCAJC-D71	5'-CCAACCTCTAAACATATAATATGTTAGTACATGAGACAC-3'
oCAJC-E74	5'-GTTAGTACATGAGACACAGTAGCAAATAAAAG-3'
oCAJC-F7	5'- TAGAGGTTGGTACCGGGTTCA-3'

Branch site to 3' splice site ACT1-CUP1 reporter constructs: pCG2 and pJPS1462 were subjected to two-stage site directed mutagenesis (W. Wang and Malcolm 1999) with oligonucleotide pairs oCAJC-D64/oCAJC-D65 and oCAJC-D70/oCAJC-D71 to generate Δ7 and Δ28 reporter variants, respectively. To replace the branch site to 3' splice site sequence with polyadenosine tracts, the *ACT1-CUP1* reporter plasmids were linearized with oCAJC-E74 and oCAJC-F7 and re-circularized using NEB HiFi assembly with a single-stranded DNA oligo containing 20 to 24 nucleotide homology on either end and the appropriate number of adenosine residues in between.

Yeast transformation: Yeast transformations were performed using the rapid lithium acetate method (Gietz and Woods 2002) using 33 µl of yeast cell culture, 3.3 µl of 10 mg/mL salmon sperm DNA as a carrier, and 198 µl of 50% w/v PEG 3350, with a 30 minute incubation at 30 °C prior to a 15 minute heat shock at 42 °C. Transformed cells were plated on selective media, grown for 2 days at 30 °C, re-streaked onto fresh selective media, and grown for 3-5 days at 30 °C prior to subsequent analysis.

For experiments involving U2 snRNA alleles, plasmids containing wildtype U6 snRNA and U2 snRNA derivatives were co-transformed as described above. pU2U6-URA containing cells were selected against by re-streaking on 5-FOA media and grown for 2 days at 30 °C. Colonies were then re-streaked onto YPDA media and then transformed with appropriate *ACT1-CUP1* constructs as described above.

Copper growth experiments: Media containing varying concentrations of copper sulfate were made using complete synthetic media lacking uracil. Bacto-agar was replaced with PHYTAGAR (Alfa Aesar) as the gelling agent to ensure no contaminating divalent metals were present. Yeast strains were transformed with *ACT1-CUP1* reporters, grown overnight to saturation, spotted onto plates with varying concentrations of copper sulfate in 1/1, 1/10, and 1/100 serial dilutions, and incubated at 30 °C for two to seven days. Images were captured using a Canon compact digital camera.

Total RNA extraction: Yeast colonies were inoculated into 3 mL of complete synthetic media lacking uracil and grown overnight until saturation. Saturated cultures were backdiluted to an OD_{600} of 0.1-0.15 and grown at 30 °C for 2-3 doublings. Cultures were shifted to 16 °C for an additional 2 hours where indicated. A volume of 1.7 to 2 mL of these cultures was spun down, aspirated of supernatant, snap frozen in liquid nitrogen, and stored at -80 °C until ready for RNA extraction.

Yeast pellets were resuspended in 500 µl TRIzol, transferred to a 2 mL screw-cap tube containing 0.5 mm acid-washed glass beads, and lysed in a FastPrep-24 homogenizer for five one-minute cycles. An additional 500 µl TRIzol was added before proceeding with the manufacturer's instructions for RNA extraction. An additional extraction was performed with 400 µl chloroform prior to isopropanol precipitation. RNA pellets were resuspended in 1.25x annealing buffer (62.5 µM Tris-HCl, pH 8.3 at 42 °C; 62.5 µM KCl, 12.5 µM DTT, 1.25 µM EDTA) pre-heated to 65 °C for at least ten minutes and adjusted to the same concentration prior to downstream experiments.

Primer Extension: Primer extension was performed with 4.4 to 12 μg of yeast total RNA using AMV reverse transcriptase. All experiments used radiolabeled primers, except for the screen involving U2 stem IIc mutations (**Fig. 3.3**), which used Quasar570 fluorophore-labeled ACT1-CUP1_PE3 primers (Biosearch, LLC). Primers were added to 4.8 μL of total yeast RNA to yield a final concentration of 1 μM each. The primer and total RNA solution was heated to 90 °C for 3 minutes and cooled to 48 °C for at least 5 minutes. AMV RT and a dNTP mixture was added to annealed RNA with a supplementary RT buffer containing spermidine and $MgCl_2$ yielding a final primer extension reaction mixture of 6U AMV RT, 1 μM each dNTP, 50 μM Tris-HCl (pH 8.3 at 42 °C), 50 μM KCl, 10 μM $MgCl_2$, 10 μM DTT, and 1 μM spermidine. RT reactions were supplemented with 4 mM NaPPi to inhibit RNase H activity to formation of aberrant products to reduce background and allow for better visualization of lariat intermediate (Figs. 2, 4, 5, 6). 50 ng of actinomycin D and additional KCl to a final concentration of 135 μM were added to reactions using poly-adenosine tracts (Figs. 5, 6) to further minimize RT slippage. Primer extension was carried out for 5 minutes at 37 °C, followed by 45 minutes at 42 °C. An equal volume of RNA loading buffer (95% formamide, 2x TBE, 0.5 μM EDTA, bromophenol blue) was added to the reaction mixture. Extension products were visualized by running on a 15% denaturing polyacrylamide gel and imaging with a phosphorimager. Quantitation was performed using FIJI version 1.53. In most cases, exon ligation efficiency was calculated by taking the ratio of mRNA/lariat intermediate. For reactions using fluorophore-labeled primers, exon ligation efficiency was calculated using a ratio of mRNA/(lariat intermediate + mRNA), due to low detection of lariat intermediate in wildtype samples.

In silico *Structural Comparison:* An *in silico* structure of a five nucleotide model RNA-RNA helix of the 5'SS:U6 interaction was generated in UCSF Chimera that includes additional base pairs between position +3 of the 5'SS and position +50 of U6 snRNA as well as and position +2 of the 5'SS and position +51 of U6 snRNA. This structure was superimposed onto cryo-EM structures of the C complex and P complex structures (PDB accession number 5LJ3 and 6BK8, respectively) (Galej et al. 2016; S. Liu et al. 2017a) using PyMol version 2.3.3 (Schrodinger, LLC).

Intramolecular RNA structure analysis: The predicted intramolecular RNA-RNA structure of the region between the BS and 3'SS of the *ACT1-CUP1* intron, as well as the consequences of mutations within this region, were determined using the online RNAfold server (Lorenz et al. 2011). The first eight nucleotides downstream of the BS all sequence downstream of the 3'SS were omitted from the analysis, as done in previous assessments of structures between the BS and 3'SS of yeast introns (Meyer et al. 2011).

Mathematical Modeling: Formulas for levels of RNA species at steady state were derived by hand, and data points were generated using MATLAB R2020a (The Mathworks, Inc.). Plots were generated using Microsoft Excel.

Results

A3c inhibits exon ligation downstream of the Prp16p-mediated rearrangement

Because A3c impairs exon ligation, perturbations that interfere with the transition between branching and exon ligation should exacerbate this phenotype. A3c hyperstabilizes the 5'SS and U6 snRNA interaction (5'SS:U6) in the branching

conformation (**Fig. 3.1A**) (Konarska *et al.* 2006), the G+1 of the 5'SS is covalently linked to the brA residue of the BS after branching catalysis, and Prp16p can detach the branch site consensus sequence (BS) from U2 snRNA to allow for alternative BS selection by pulling on the substrate (Semlow et al. 2016). Given these observations, we initially hypothesized Prp16p remodels the 5'SS:U6 interaction, and thus A3c would impair the Prp16p-dependent rearrangement step. To determine whether A3c blocks exon ligation at Prp16p-dependent rearrangement step, we first tested whether *prp16-302*, a *PRP16* mutant deficient in transitioning out of the branching conformation, would exacerbate a growth defect conferred by the exon ligation defect of A3c. As a control, we also tested whether the exon ligation defect of A3c would be affected by a *prp16-101* mutation, which does not exacerbate growth defects of 3'SS mutants (Query and Konarska 2004) but instead has a phenotype at the branching step of splicing by relaxing proofreading of branch site mutations (S. M. Burgess and Guthrie 1993). To test the affects of *prp16* alleles on the splicing of A3c, we used an *ACT1-CUP1* reporter gene, which consists of essentially the *ACT1* 5' exon and intron fused to a copy of the *CUP1* gene. Increased efficiency of *ACT1-CUP1* splicing leads to increased levels of Cup1p protein, allowing growth of *cup1Δ* strains on media containing increased levels of toxic copper ions. Changes in splicing efficiency due to perturbations in the *ACT1-CUP1* intron or mutations in spliceosome factors can therefore be indirectly determined through changes in the ability for yeast to grow on media with increased levels of copper sulfate. (Lesser and Guthrie 1993).

We transformed wild-type and mutated *ACT1-CUP1* reporters into wildtype or *prp16* mutant strains and assayed for growth at 30 °C on solid media containing varying concentrations of copper sulfate (**Fig. 3.2**). Wildtype *PRP16*, *prp16-101*, *and prp16-302* strains all exhibited strong growth with wildtype *ACT1-CUP1* reporters, with slight growth

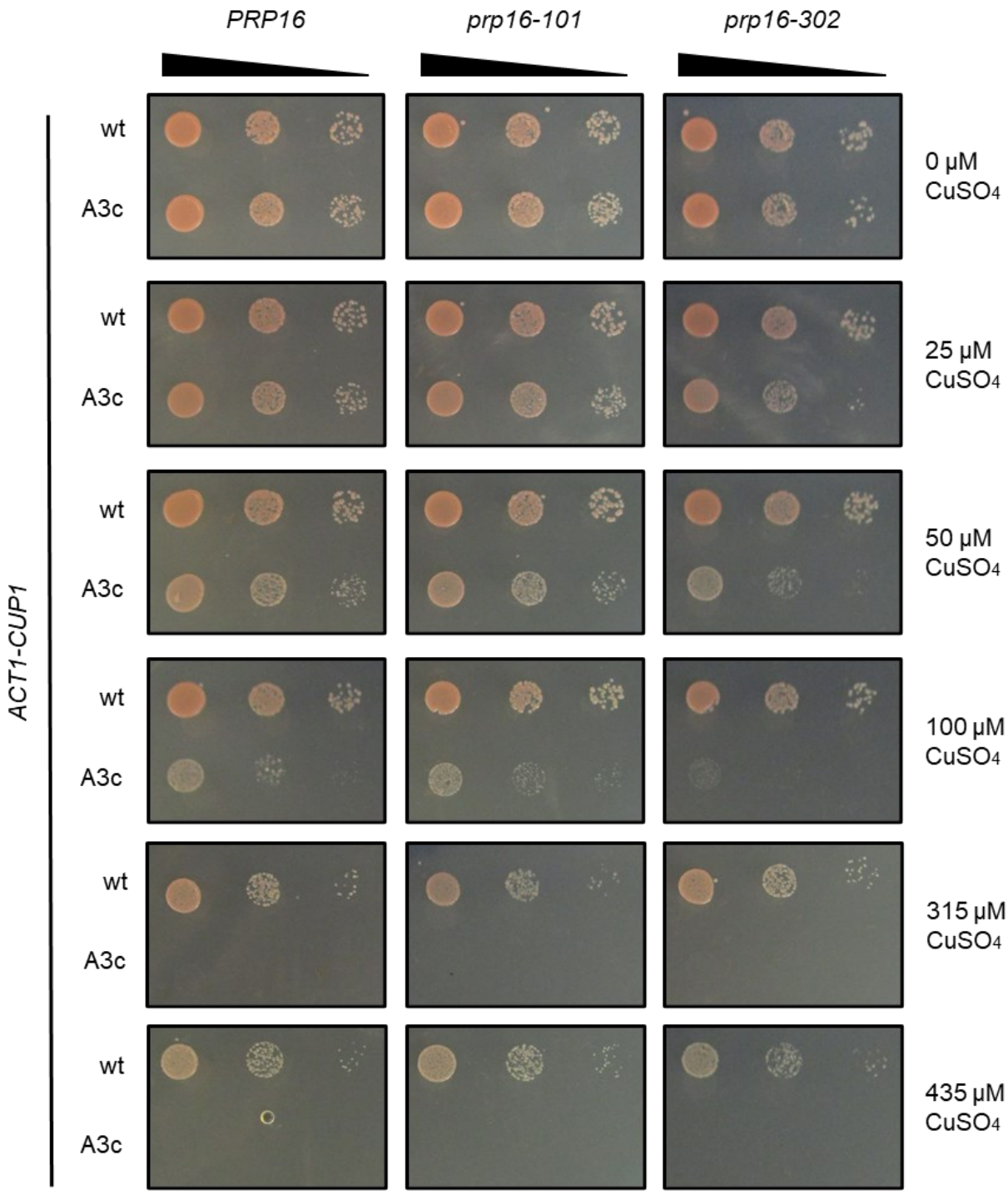

Figure 3.2 A Prp16p ATPase mutant exacerbates the growth defect of A3c.

Copper growth assay of yeast containing the branching fidelity mutation *prp16-101* or the exon ligation mutation *prp16-302* in combination with wildtype or A3c *ACT1-CUP1* reporters. The growth defect of A3c at higher copper concentrations is exacerbated by *prp16-302* but not *prp16-101*.

deficiencies beginning to show at 435 μM copper sulfate. This growth defect was slightly

exacerbated with *prp16-302* in the context of a wildtype *ACT1-CUP1* reporter, as reported previously (Query and Konarska 2004). In combination with an A3c *ACT1-CUP1* substrate, all yeast strains exhibited an increased sensitivity to copper. Wildtype *PRP16* began to show decreased growth when copper concentrations reached 100 μM. This growth phenotype was exacerbated when the A3c *ACT1-CUP1* substrate was combined with *prp16-302*, with growth defects beginning to show at 50 μM copper sulfate. By contrast, the growth defect of A3c was not affected when combined with *prp16-101*. These data indicate A3c is sensitized to mutations in *PRP16* that specifically impair transition out of the branching conformation.

However, given there are two rearrangements required to enter the exon ligaiton conformation, and mutations that impair the Prp16p-dependent transition can also exacerbate exon ligation defects of perturbations that act downstream of Prp16p activity (Query and Konarska 2004; L. Liu, Query, and Konarska 2007), we could not be sure A3c specifically antagonizes the Prp16p-dependent step. To this end, we leveraged a previously-described RNA-RNA remodeling event in U2 snRNA to determine where exactly A3c antagonizes exon ligation. Genetic evidence suggests U2 stem IIc, an intramolecular helix in U2 snRNA, is disrupted during the Prp16p-dependent transition out of the branching conformation (C complex) (**Fig. 3.4B**) (Hilliker, Mefford, and Staley 2007; Perriman and Ares 2007). U2 stem IIc disruption leads to formation of a mutually exclusive intramolecular structure, known as U2 stem-loop IIa, that stabilizes an intermediate conformation between the branching and exon ligation states. U2 stem-loop IIa is then disrupted in a Prp16p-independent manner during transition into the exon ligation conformation (C* complex), allowing re-formation of U2 stem IIc (Hilliker *et al.* 2007; Perriman and Ares 2007).

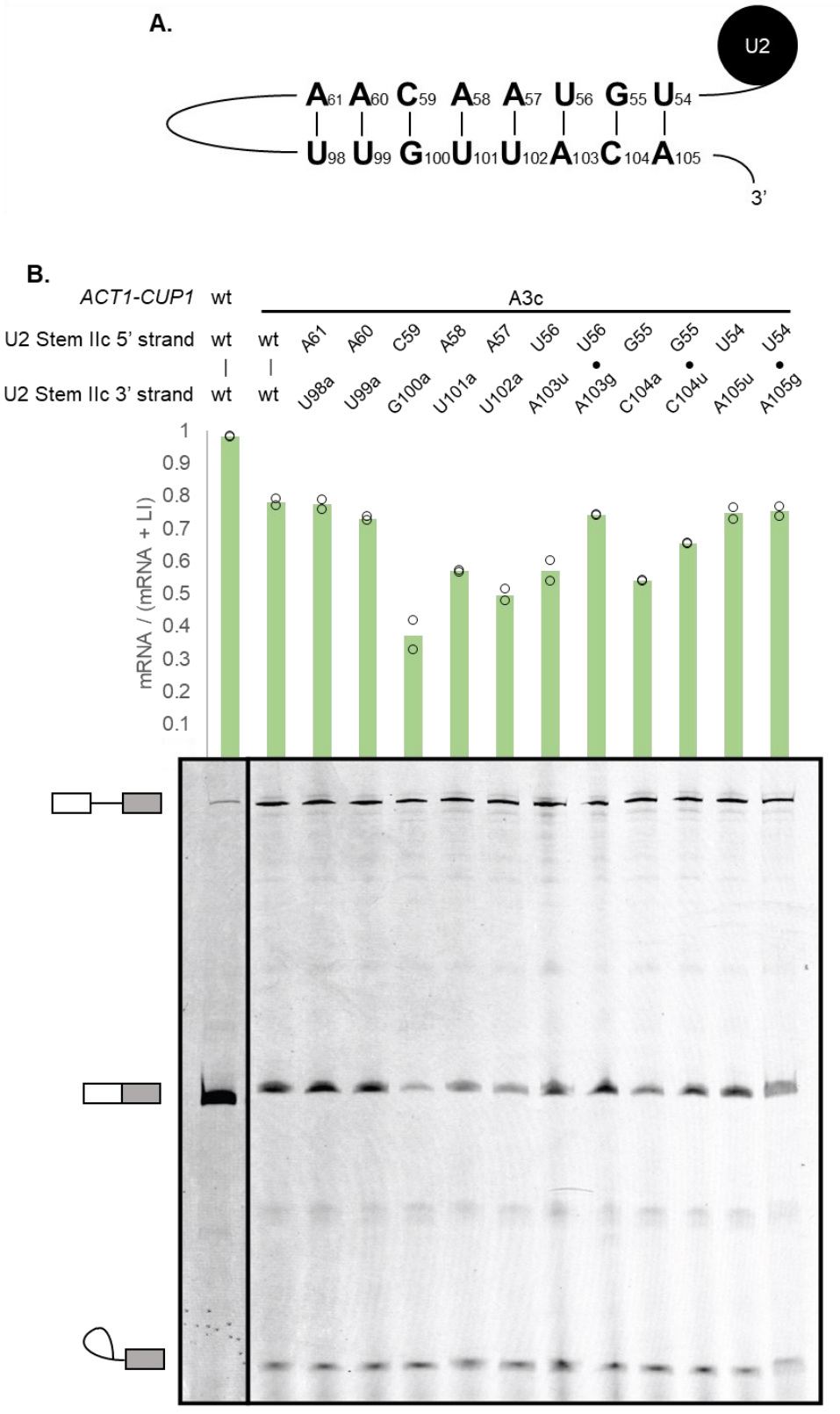

Figure 3.3: Destabilizing U2 stem IIc exacerbates the exon ligation defect of A3c.

Figure 3.3: Destabilizing U2 stem IIc exacerbates the exon ligation defect of A3c (continued).

A. Structure of the U2 stem IIc intramolecular RNA-RNA helix.

B. Point mutations along the 3' strand of U2 stem IIc were co-transformed with A3c *ACT1-CUP1* substrates and exon ligation efficiency was probed using primer extension. The exon ligation defect of A3c is exacerbated by mutations that abolish Watson-Crick base pairing in stem IIc. Exacerbation of the A3c exon ligation defect was more modest when changing a Watson-Crick base pair to a G●U wobble and was minimal when changing an A-U Watson-Crick base pair to a G●U wobble. Exacerbation was also minimal at the ends of the U2 stem IIc helix.

To determine whether A3c blocks exon ligation at or downstream of the Prp16p-dependent rearrangement, we destabilized U2 stem IIc and assayed for alleviation of the A3c exon ligation defect via primer extension. If A3c antagonizes exon ligation at the Prp16p-dependent rearrangement step, destabilization of U2 stem IIc would alleviate the exon ligation defect of A3c. Conversely, if A3c antagonizes rearrangements downstream of Prp16p activity, destabilization of U2 stem IIc would fail to alleviate and, instead, exacerbate the exon ligation defect of A3c. To destabilize U2 stem IIc, we first made mutations to the 3' strand of stem IIc (**Fig. 3.3A**) to prevent inadvertently destabilizing the intermediate state, because the 5' strand of stem IIc forms the loop of stem-loop IIa, and mutations within this loop promote transition into the exon ligation formation (Hilliker *et al.* 2007). Combining point mutations in the 3' strand of U2 stem IIc with A3c mutations did not alleviate the exon ligation defect of A3c. Instead, we found mutations that destabilized U2 stem IIc reduced the exon ligation efficiency of A3c by as much as two-fold, as determined by primer extension analysis (**Fig. 3.3B**). Mutations that abolished base-pairing between the 5' and 3' strands of U2 stem IIc exacerbated the exon ligation defect of A3c to a greater extent than those that changed the identity of the base-pair from a G-C Watson-Crick base pair to a G●U wobble. Additionally, abolishing more

strictly-conserved base pairs within the interior of U2 stem IIc exacerbated the exon ligation defect of A3c more than less conserved residues at the ends of the U2 stem IIc helix. These data strongly suggest that mutations in the 3' strand of U2 stem IIc exacerbate the exon ligation defect of A3c due to weakening U2 snRNA intramolecular structure.

To exclude the possibility that exacerbation of A3c could be due to these mutations having an effect outside of destabilizing U2 stem IIc, we chose the base pair that conferred the strongest exacerbation when abolished (U2-G100:U2-C59) and performed compensatory mutational analysis to determine whether restoring Watson-Crick pairing could rescue this exacerbation. We tested U2 stem IIc mutations at this base pair position against several perturbations that confer an exon ligation defect: A3c, which has an exon ligation defect that is exacerbated by a U2-G100a mutation (**Fig. 3.3B**) ; *prp16-302*, which has a cold sensitive growth defect that is suppressed or exacerbated by mutations that destabilize or hyperstabilize U2 stem IIc, respectively (Hilliker, Mefford, and Staley 2007; Perriman and Ares 2007); and gAG, which has a growth defect at 30 °C that is exacerbated by mutations that destabilize U2 stem IIc and suppressed by mutations that stabilize U2 stem IIc relative U2 stem IIa (Hilliker, Mefford, and Staley 2007). To test the effect of modulating the Waston-Crick base-pairing at U2-G100:U2-C59 on these exon ligation deficient alleles, we again used primer extension and determined differences in exon ligation by taking the ratio of the mRNA exon ligation product over the LI exon ligation reactant.

Mutations that destabilize the U2-G100:U2-C59 base pair had minimal effect on exon ligation of a wildtype *ACT1-CUP1* substrate (**Fig. 3.4A**). A C59g mutation had an insignificant decrease in mRNA-to-LI ratio, whereas a G100c mutation had a modest but

significant decrease in mRNA-to-LI ratio. These data likely reflect the effect these mutations have on the U2 stem IIc to U2 stem IIa transitions, as C59g would destabilize both U2 stem IIc and U2 stem IIa, whereas G100c would destabilize U2 stem IIc relative to U2 stem IIa to potentially disrupt structural toggling between these two intramolecular structures along the splicing pathway. Restoring Watson-Crick base pairing by combining U2-C59g with U2-G100c restored exon ligation to a level seen with wildtype U2 snRNA. These data suggest these mutations in U2 snRNA impart defects in exon ligation. The modest effects seen with wildtype *ACT1-CUP1* suggest exon ligation of this transcript is not limited by this structural toggling between U2 stem IIc and U2 stem IIa.

Curiously, modulation of Watson-Crick base pairing at this position had an even subtler effect on exon ligation of a wildtype *ACT1-CUP1* reporter combined with a *prp16-302* mutant. Although exon ligation of *ACT1-CUP1* decreased by 12-fold in *prp16-302* strains shifted to 16 °C for 2 hours, neither U2-C59g nor U2-G100c suppressed this exon ligation defect (**Fig. 3.4A**). We initially expected mutations that destabilize U2 stem IIc would suppress the exon ligation defect of *prp16-302*, as similar mutations suppress the cold sensitive growth defect of *prp16-302* strains (Hilliker, Mefford, and Staley 2007; Perriman and Ares 2007). In the context of our results, *prp16-302* likely confers a growth defect by antagonizing splicing of an endogenous transcript that is sensitive to Prp16p driving the spliceosome out of the U2 stem IIc state. Conversely, our data suggest wildtype *ACT1-CUP1* is not significantly limited by U2 snRNA toggling. In comparison, *ACT1-CUP1* reporters containing gAG exhibited much stronger exon ligation defects at 30 °C with wildtype U2 snRNA, and these defects were modestly but significantly worsened in the context of U2-C59g or U2-G100c (**Fig. 3.4A**). These data are consistent with previous data showing disruption of U2 stem IIc exacerbates copper growth defects of strains with

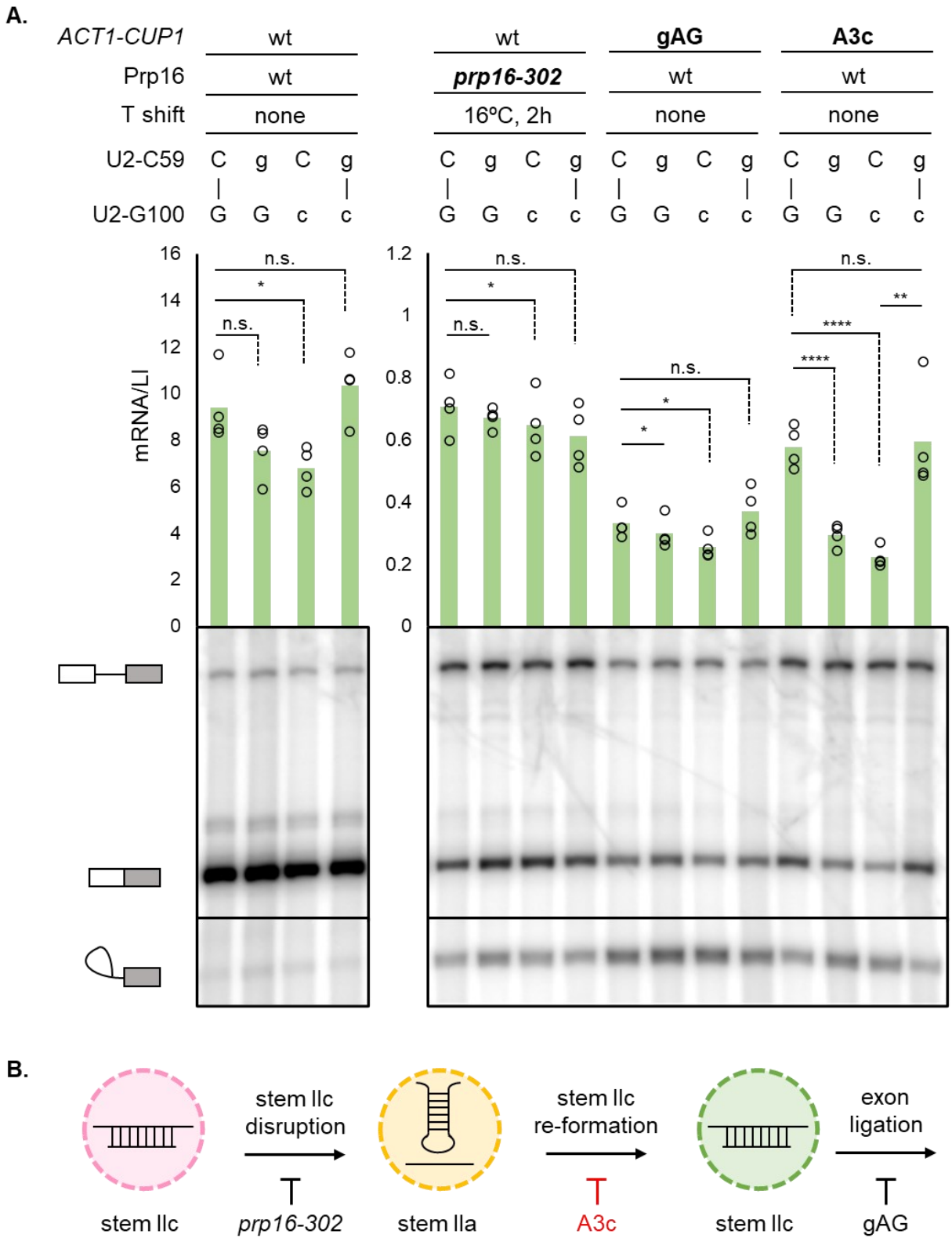

Figure 3.4: A3c antagonizes the step of U2 stem IIc re-formation upon entry into the exon ligation conformation.

Figure 3.4: A3c antagonizes the step of U2 stem IIc re-formation upon entry into the exon ligation conformation (continued)

A. Primer extension analysis of *in vivo* splicing in the context of mutations that modulate stability of U2 stem IIc. Destabilizing U2 stem IIc through introduction of either C59g or G100c leads to a reduction of exon ligation efficiency of wildtype, gAG, and A3c substrates in the context of wildtype *PRP16*, but does not affect splicing of a wildtype substrate in a *prp16-302* mutant. Exacerbation is greater for A3c substrates. Restoring Watson-Crick base pairing by combining C59g and G100c restores exon ligation efficiency. Significance was calculated by two-tailed student's t-test. * $P < 0.05$, ** $P < 0.01$, **** $P < 0.0001$, *n.s.* $P > 0.05$.

B. Overview of U2 snRNA rearrangement between the branching and exon ligation conformation during the Prp16p-dependent and Prp16p-independent rearrangement.

gAG *ACT1-CUP1* reporters (Hilliker, Mefford, and Staley 2007). Combining U2-G100c and U2-C59g restored exon ligation to levels seen with wildtype U2 snRNA, consistent with our observations with wildtype *ACT1-CUP1*. Taken together, these data indicate U2 stem IIc re-formation is limiting for gAG transcripts.

In the context of wildtype U2 snRNA, exon ligation of *ACT1-CUP1* transcripts containing A3c at 30 °C exhibited a decrease similar to *prp16-302* after a 16 °C shift (**Fig. 3.4A**). Strikingly, this exon ligation defect was exacerbated to a great extent with either U2-C59g or U2-G100c, with a two-fold decrease in mRNA-to-LI ratio when compared to wildtype U2 snRNA. Similar to both wildtype and gAG *ACT1-CUP1* transcripts, this decrease was moderately worse with the U2-G100c mutation when compared to U2-C59g, and exon ligation levels were restored to levels seen with wildtype U2 snRNA when both point mutations were combined. In comparison to *prp16-302,* the exacerbation of the exon ligation defect of A3c by U2-C59g or U2-G100c point mutations were significant, suggesting that A3c does not affect the Prp16p-dependent rearrangement. Instead, the A3c parallels the phenotypes of gAG, suggesting U2 stem IIc re-formation is also limiting for A3c. However, the magnitude of exacerbation by U2-C59g or U2-G100c also suggests

U2 stem IIc destabilization has a synergistic effect on A3c, indicating U2 stem IIc reformation and A3c may affect the same step of the splicing pathway (**Fig. 3.4B**).

Extended pairing between the 5' splice site and U6 snRNA would reposition key residues in the catalytic core to antagonize 3'SS binding

Previous characterization shows A3c introduces an additional base pair to the 5'SS:U6 helix (Konarska *et al.* 2006), which led to a model where A3c antagonizes exon ligation by preventing transition out of the branching conformation. However, recent cryo-EM structures of the spliceosome immediately after exon ligation (P complex) have illuminated the importance of residues of the 5'SS and U6 snRNA for stable 3'SS binding and proper positioning of the 3'SS reactant (Fig 1C) (Bai *et al.* 2017; Wilkinson *et al.* 2017; Liu *et al.* 2017)**.** Thus, introducing a Watson-Crick base pair between position +3 of the 5'SS and U6-G50 may instead antagonize exon ligation by impairing proper 3'SS binding.

Modelling of the consequences of the A3c mutation support this view of compromised 3' splice site binding. In the P complex, the base of G-1 of the 3'SS stacks onto residue U6-A51, which in turn is stabilized by hydrogen bonding to U+2 of the 5'SS (**Fig. 3.5A**). G-1 of the 3'SS is further stabilized through hydrogen bonding of its Hoogsteen face to the Watson-Crick face of G+1 of the 5'SS. Cryo-EM structures suggest these interactions are facilitated by the breaking of a non-Watson-Crick base pair between A+3 and U6-G50 which is present in the branching conformation, but absent from exon ligation conformation (Fica et al. 2017). Given these observations, we reasoned that the introduction of a Watson-Crick base-pair between 5'SS-A3c and U6-G50 may interfere with proper coordination of these key 3' splice site-orienting residues, antagonizing stable 3'SS binding. To examine this possibility, we extended the 5'SS:U6 duplex by an

additional Watson-Crick base pair between U6-G50 and a C residue at position +3 of the 5'SS *in silico* in a model RNA-RNA helix. We then overlaid this helix onto the 5'SS:U6 interaction in a cryo-EM structure of the P complex (Liu *et al.* 2017) (**Fig. 3.5A**) to examine how the backbones of the 5'SS and U6 snRNA deviated and what impact such deviation may have on the residues that interact with the 3'SS.

Extension of the 5'SS:U6 duplex perturbs the backbone of U6 modestly, although the base of U6-G50 is rotated significantly to allow base pairing at its Watson-Crick face (**Fig. 3.5A**). This movement is accompanied by a shift in the backbone of the 5'SS at position +3 by 5.2 Å to align the Watson-Crick faces of A3c and U6-G50, which would reposition U+2 and, consequently, G+1. As the backbone of U6 appears to remain in place, we don't expect as drastic of a movement of U6-A51 or G-1 of the 3'SS. Instead, 3'SS stabilization would be compromised by shifting G+1 of the 5'SS out of position so it cannot stably hydrogen bond to G-1 of the 3'SS, and U+2 of the 5'SS may also be shifted out of position to prevent proper orientation of U6-A51. The shift of the backbone of 5'SS could further extend the 5'SS:U6 duplex, promoting formation of a Watson-Crick base pair between U+2 of the 5'SS and U6-A51, which would further shift the nitrogenous base of U6-A51 by 10 Å to disfavor stacking onto G-1 altogether (**Fig. 3.5B**). If increased extension of the 5'SS:U6 helix due to introduction of A3c does move these residues out of position to inhibit or prohibit stable binding of the 3'SS, then A3c would confer an exon ligation defect by reducing the population of spliceosomes that can catalyze exon ligation.

The A3c 5'SS mutation is sensitive to the distance separating the 3'SS from the BS

Binding of the 3'SS to the catalytic core requires the region of the RNA substrate downstream of the BS to loop out of and back into the spliceosome, allowing the 3'SS to

bind to the 2'-5' phosphodiester bond linkage between 5'SS-G+1 and brA (**Fig. 3.1D**) (Meyer et al. 2011; Pérez-Valle and Vilardell 2012; S. Liu et al. 2017b; Wilkinson et al. 2017; Bai et al. 2017). This looping requirement suggests the distance between the BS and 3'SS impacts the time it takes for the spliceosome to initially bind the 3'SS, given the mean first passage time (MFLT) for the 3' end of an RNA polymer to loop back onto its 5' end is proportional to its persistence length (Toan et al. 2006). Notably, the number of single-stranded nucleotides between the BS and 3'SS across all introns in budding yeast is clustered around a relatively short effective length of 25 nucleotides (Meyer et al. 2011), implying that pre-mRNA splicing favors relatively short BS to 3'SS distances. This relatively short BS to 3'SS distance is also seen in human introns (Zeng et al. 2021), and a long BS to 3'SS distance also contributes to antagonizing exon ligation of *TER1* pre-mRNA in the fission yeast *S. pombe* (Kannan *et al.* 2013). Conservation of short BS to 3'SS distances imply a long BS to 3'SS is detrimental to pre-mRNA splicing and, together with structural and genetic data, imply this detriment is due to delaying 3'SS docking.

Our structural analysis above also suggests A3c may also delay 3'SS docking by altering the 3'SS binding site, which can have several non-mutually exclusive effects: First, the spliceosome may sample a conformation that allows for 3'SS binding at a lower frequency than spliceosomes with a wildtype 5'SS. Second, in spliceosomes that can bind the 3'SS, initial binding of the 3'SS may be compromised, leading to a slower on-rate of docking. Third, after docking, the 3'SS may be more likely to dissociate, leading to a faster off-rate. Additionally, although the spliceosome can re-sample the 3'SS after it is undocked (Semlow et al. 2016), destabilizing 3'SS binding could also impair the efficiency of this re-sampling. Lastly, after binding, the 3'SS may not be properly orientated to allow for efficient catalysis (see below) (Wilkinson et al. 2017). Any of these effects can

antagonize exon ligation by limiting the length of time the 3'SS is bound in the proper orientation to below that necessary for successful exon ligation catalysis. Since long BS to 3'SS distances appear to delay 3'SS docking, we hypothesized shortening the BS to 3'SS distance would reduce the time it takes for initial binding of, or re-sampling after dissociation of, the 3'SS which would, in turn, alleviate the exon ligation defect of A3c.

As a first test of this hypothesis, we deleted several nucleotides of the intron just downstream of the BS in our *ACT1-CUP1* construct (**Fig. 3.5C**). These deletions disrupt the intramolecular RNA-RNA structure between the BS and 3'SS within *ACT1* intron in our *ACT1-CUP1* construct (Meyer et al. 2011). Deletion of the first 7 nucleotides after the BS (Δ7), predicted to completely disrupt intramolecular structure and increase the effective BS to 3'SS distance from 26 to 36 nucleotides, exacerbated the exon ligation defect of A3c, reducing the mRNA-to-LI ratio by 70%. Conversely, deleting the first 28 nucleotides past the BS (Δ28), shortens the effective BS to 3'SS distance from 26 to 16 nucleotides, led to a slight alleviation of the exon ligation defect of A3c, increasing the mRNA-to-LI ratio by 25% (**Fig. 3.5C**). Interestingly, a similar effect is seen with substrate bearing a wildtype 5'SS, where exon ligation efficiency worsened with the Δ7 substrate but improved to near-wildtype levels with the Δ28 substrate. These data suggest intramolecular structure is necessary to shorten the effective BS to 3'SS distance to allow for efficient exon ligation, as implicated in many yeast introns (Meyer et al. 2011; Pérez-Valle and Vilardell 2012; S. Liu et al. 2017a). The observation that exon ligation is improved even further in A3c by shortening the effective BS to 3'SS from 26 to 16 nucleotides supports our hypotheses that A3c confers a defect in 3'SS binding, and that shortening the BS to 3'SS distance alleviates this defect.

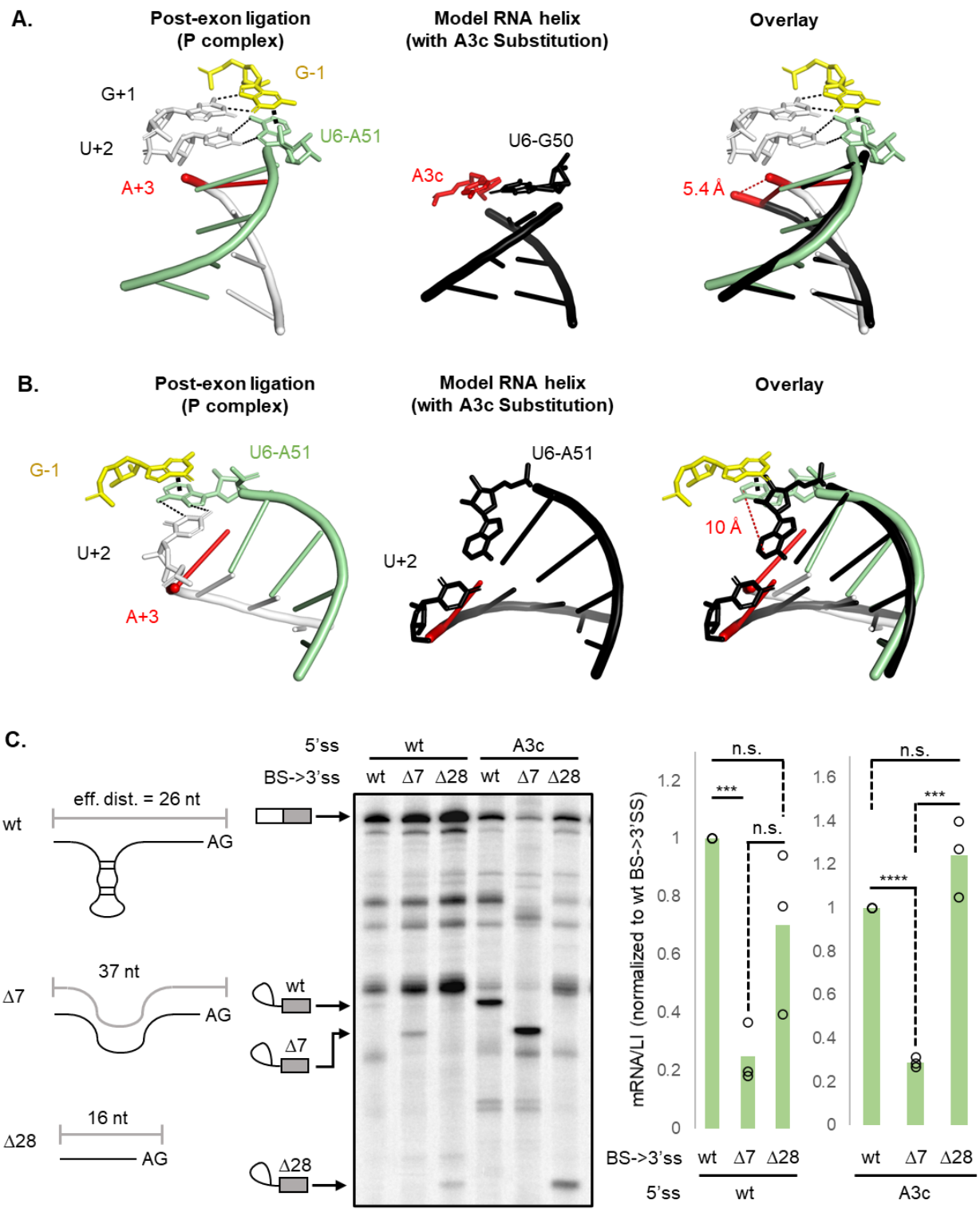

Figure 3.5: Introduction of a Watson-Crick base pair between A3c and U6-G50 distorts the 3'SS binding pocket.

Figure 3.5: Introduction of a Watson-Crick base pair between A3c and U6-G50 distorts the 3'SS binding pocket (continued)

A. The backbone of the 5'SS would need to shift 5.2 Å to accommodate hydrogen bonding between the Watson-Crick faces of A3c and U6-G50, which would move 5'SS-G1 and 5'SS-U2 out of position to stabilize 3'SS-G3 through hydrogen bonding with 5'SS-G1 and base stacking onto U6-A51.

B. A3c could promote base pairing between 5'SS-U2 and U6-A51, which would move the base of U6-A51 by 10 Å to disfavor stacking onto 3'SS-G3.

C. Removing seven nucleotides between the BS and 3'SS disrupts intramolecular structure within the intron, increasing the effective BS to 3'SS distance and antagonizing exon ligation. Deleting 28 nucleotides shortens the effective distance and restores exon ligation efficiency.

In a more systematic test of whether the exon ligation defect of A3c can be alleviated by shortening the BS to 3'SS distance, we replaced the intronic sequence between the BS and 3'SS with stretches of adenine residues, varying from 6 to 30 adenosine nucleotides long in 3 nucleotide (nt) increments, to create BS to 3'SS distances from 9 to 33 nt (**Fig. 3.6**). Exon ligation of both wildtype and A3c 5'SS sequences were modulated with changing BS to 3'SS distances. Qualitatively, both wildtype and A3c exhibited optimal exon ligation around 15 to 18 nt, as evidenced by strong mRNA accumulation accompanied by weak LI bands. Exon ligation decreased with increasing BS to 3'SS distance, with decreases in mRNA accumulation and concomitant increases in LI band intensity, consistent with previous studies showing exon ligation is decreased with distal 3'SS (Beate Schwer and Gross 1998; Meyer et al. 2011; R. Kannan et al. 2013; Ohrt et al. 2013). Unexpectedly, exon ligation also decreased with BS to 3'SS distances shorter than 15 nt, implying 3'SS sampling also becomes increasingly difficult below a certain threshold *in vivo*. A plot of exon ligation for A3c substrates relative to BS to 3'SS

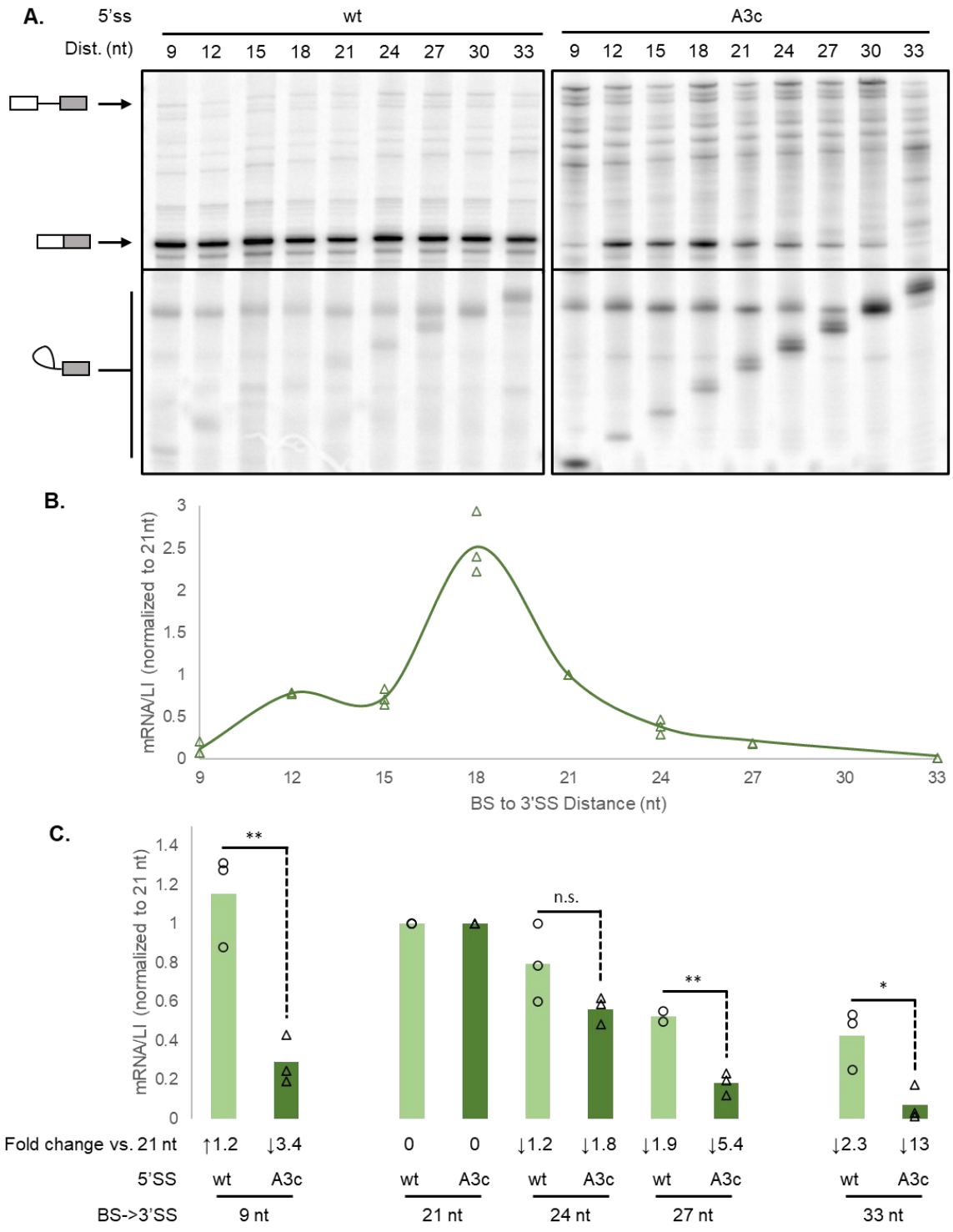

Figure 3.6: BS to 3'SS distance modulates exon ligation efficiency

A. Exon ligation efficiency is antagonized at both very short and long distances for substrates containing a wildtype or A3c 5'SS.

Figure 3.6: BS to 3'SS distance modulates exon ligation efficiency (continued)

B. A3c is highly sensitized to the BS to 3'SS distance, with a narrow window of high exon ligation efficiency at 18 nt.

C. Comparison of exon ligation efficiencies between substrates with wildtype 5'SS or A3c. Exon ligation efficiencies were normalized to the exon ligation efficiency at 21 nt for each series of substrates. A3c substrates exhibit a markedly sharper decline in exon ligation efficiency at 9 nt and over distances longer than 21 nt when compared to substrates bearing a wildtype 5'SS.

distance revealed a steep drop-off with BS to 3'SS distances shorter or longer than 18 nt, indicating deviations in BS to 3'SS distance modulate exon ligation efficiency.

Due to greater exon ligation efficiency, we could not reliably quantitate all substrates containing a wildtype 5'SS because LI levels either fell below a quantifiable level or because the LI band overlapped with another band corresponding to a premature stop associated with mRNA levels. Consequently, to ensure reliable quantitation, we limited our analysis of wildtype 5'SS substrates to BS to 3'SS distances of 9, 21, 24, 27, and 33 nt and compared exon ligation levels to A3c substrates at these same distances. While qualitatively it appears exon ligation efficiency decreased at 9 nt and 33 nt for both wildtype and A3c substrates (**Fig. 3.6A**), quantitation of exon ligation for A3c relative to 21 nt showed a markedly stronger decrease at 9 nt and 33 nt, (10-fold and 50-fold decrease, respectively), as compared to wildtype (1.3-fold and 3-fold decrease, respectively, Fig. 4B). Additionally, while exon ligation efficiency of substrates with a wildtype 5'SS decreased with increasing distance from 21 to 33 nt, the exon ligation efficiency of A3c substrates showed a markedly steeper fall-off across these same distances. These data indicate that A3c narrows the window for the

The exon ligation defect of the A3c 5' splice site mutation is alleviated by ATPase mutations in PRP22 and *PRP43*

Prp22p proofreads aberrant 3'SS sequences in an ATP-dependent manner. Prp22 acts as a "pullase," binding to the periphery of the spliceosome and applying force from a distance to move the 3'SS out of the catalytic core (Semlow et al. 2016). After Prp22p-mediated proofreading, Prp43p acts to disassemble the spliceosome to allow for discard of the and 5' exon and LI (Mayas *et al.* 2010), which is debranched by Dbr1 and degraded (Liu *et al.* 2007). Becuase A3c appears to destabilize 3'SS docking, Prp22p acts at the stage of exon ligation by pulling on the substrate (Ohrt *et al.* 2013; Schwer and Gross 1998; Frank and Guthrie 1992), and Prp43p acts to disassemble spliceosomes after Prp22p proofreading activity, we hypothesized the exon ligation defect could be suppressed by ATPase mutants of Prp22p and Prp43p.

To test this hypothesis, we transformed *ACT1-CUP1* reporters bearing A3c or, as a positive control for proofreading, the gAG 3'SS mutation into yeast containing wildtype *PRP22* or ATPase mutant *prp22-R805A* and separately wild-type *PRP43* or ATPase mutant *prp43-Q423N* and assayed for exon ligation efficiency using primer extension. As shown previously, *prp22-R805A* strongly alleviated the exon ligation defect of gAG (**Fig. 3.7**) (Rabiah M Mayas, Maita, and Staley 2006). The *prp22-R805A* mutant also alleviated the exon ligation defect of A3c, although to a lesser extent. In both instances, the increase in calculated exon ligation efficiency is due to an increase in mRNA band intensity with a concomitant decrease in LI band intensity. The exon ligation defects of both gAG and A3c were alleviated by *prp43-Q423N*, though unlike with *prp22-R805A* the magnitude of alleviation was similar between the two substrates. In contrast to *prp22-R805A*, this increase in calculated exon ligation efficiency is driven mainly by a decrease in LI band intensity, suggesting decreased discard of LI occurs in the context of a *prp43-Q423N* mutation. These data, along with our structural analyses above showing deformation of

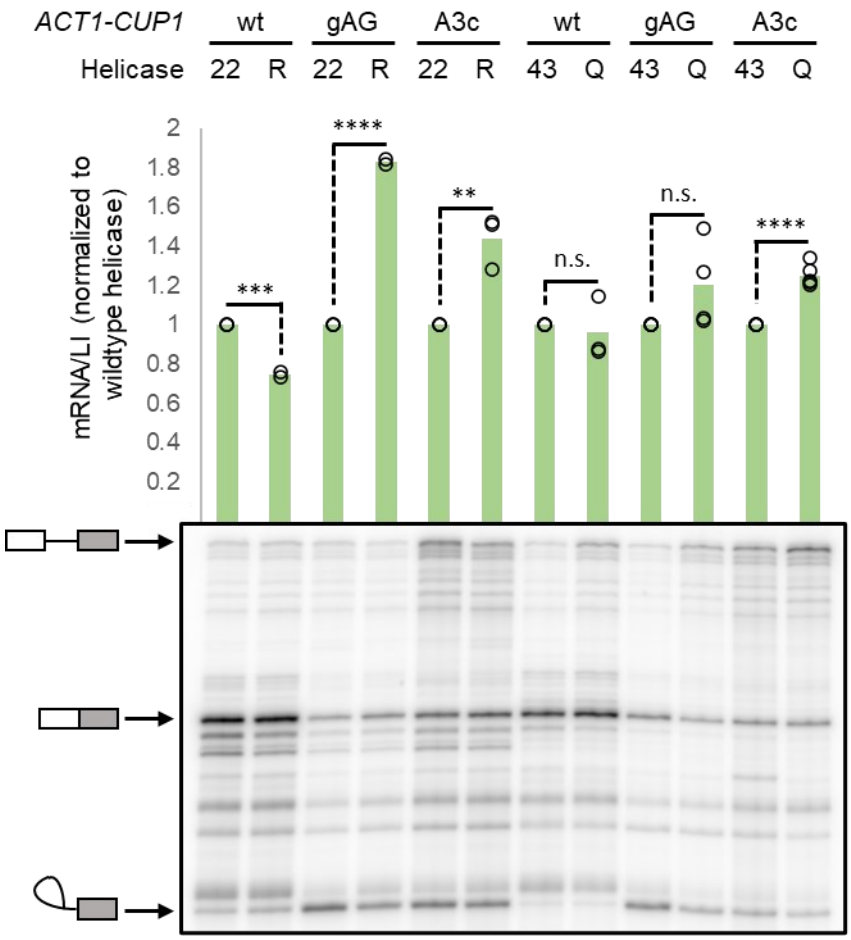

Figure 3.7: ATPase mutants of Prp22p and Prp43p alleviate the exon ligation defect of A3c.

Exon ligation efficiency of mutant helicases were normalized to the exon ligation efficiency of the wildtype helicase for each *ACT1-CUP1* substrate. Prp22p and Prp43p ATPase mutants alleviate the exon ligation defect of both gAG and A3c. Prp22p alleviates the exon ligation defect of A3c to a lesser extent than gAG, and a Prp43p mutant alleviates the exon ligation defects of both A3c and gAG to a similar extent.

the 3'SS binding site (Figs. 3A, B), suggest Prp22p and Prp43p cooperate to reject and discard A3c substrates that reach the exon ligation state.

However, our structural analyses also suggest A3c alters the positioning of multiple residues necessary for stabilizing 3'SS binding in the catalytic core (**Fig. 3.5A**) and could nucleate an interaction involving U+2 that would further disallow productive 3'SS binding (Fig 3B). The interaction between A3c and U6-G50 would occur at the end of the

5'SS:U6 helix, making this base pair susceptible to breathing and allowing sampling of a conformation competent for 3'SS binding. This is consistent with A3c having a temperature dependence effect on exon ligation efficiency, with lower temperatures that would stabilize hydrogen bonding correlating with worse exon ligation efficiency (Konarska, Vilardell, and Query 2006). These observations, as well as our data showing A3c affects the U2 stem IIc re-formation step upstream of 3'SS binding (**Fig. 3.4**) and narrows the allowable BS to 3SS distance for exon ligation (**Fig. 3.6B, C**), implies a subpopulation of A3c substrates do not proceed to the 3'SS binding step and, therefore, is not susceptible to Prp22p-mediated proofreading. Since A3c triggers discard of intermediates prior to exon ligation in some fission yeasts (Ram Kannan et al. 2015), we reasoned that Prp43p may directly proofread the subset of A3c substrates that fail to proceed to the 3'SS docking step.

Mathematical modeling suggests a discard step between branching and exon ligation imposes an upper limit on BS to 3'SS distance

To test whether a splicing intermediate discard step just upstream of 3'SS binding would disfavor mRNA formation and decrease exon ligation efficiency, we created a mathematical model of the catalytic stages of the splicing pathway (**Fig. 3.8A**) With this model we focused on indicators of exon ligation efficiency, total LI and total mRNA levels at steady state, as well as on steps along the splicing pathway that would modulate these levels. The model includes the rate of branching catalysis (k_{br}), which would generate LI from pre-mRNA, the rate of branching catalysis (k_{br}), which would generate LI from pre-mRNA, as well as the rate of Prp16p-dependent remodeling (k_{16}), which would initiate rearrangement out of the branching conformation and undock the LI from the catalytic

core. As the binding of the 3'SS is coincident with binding of protein factors required for exon ligation, these factors bind after Prp16p ATPase activity (Ohrt et al. 2013), and our data above suggests 3'SS docking is coincident with the Prp16p-independent step of stem IIc re-formation, we collapsed this value into a single rate (k_{dock}). We also modeled in the rate of exon ligation (k_{el}), which generates mRNA, and collapsed the rate of Prp22p-dependent mRNA release and mRNA decay into one rate (k_{decay}). We then placed a LI discard step ($k_{discard}$), which that acts on the population of undocked LI, in between the Prp16p-dependent remodeling step and the 3'SS docking step. We then adjusted the discard rate to determine its effect on total mRNA and LI levels.

Assuming all steps are at equilibrium, we derived equations for the concentrations of LI bound to the spliceosome in the branching (LI_b), undocked (LI_u), and docked states (LI_d), to allow for calculation of total LI concentration at steady state:

$$[LI_b] = \frac{k_{br}\,[pre]}{k_{16}}$$

$$[LI_u] = \frac{k_{16}[LI_b]}{k_{dock} + k_{discard}}$$

$$[LI_d] = \frac{k_{dock}\,[LI_u]}{k_{el}}$$

$$[LI_{total}] = k_{br}[pre]\left(\frac{1}{k_{16}} + \frac{1}{k_{dock} + k_{discard}}\left(1 + \frac{k_{dock}}{k_{el}}\right)\right)$$

We also determined the concentration of mRNA and took the ratio of mRNA-to-total LI to determine levels of exon ligation. Interestingly, pre-mRNA concentration and branching rate cancel out in this scheme, making the level of exon ligation dependent only on the rates of Prp16p activity, discard, docking, and exon ligation:

$$[mRNA] = k_{el}[LI_d] = \frac{k_{br}[pre]}{k_{decay}\left(1 + \frac{k_{discard}}{k_{dock}}\right)}$$

$$\frac{[mRNA]}{[LI_{total}]} = \frac{1}{k_{decay}\left(\frac{1}{k_{el}} + \frac{1}{k_{dock}}\left(1 + \frac{k_{dock} + k_{discard}}{k_{16}}\right)\right)}$$

Since most of these rates have not been directly measured in the spliceosome, we assigned the rates in these equations to values that would best correspond to actual rates. For time to branching, we used the median number of nt past the 3'SS in which co-transcriptional lariat intermediates to arise in introns of median length in HEK293 cells (107 nt) divided by the average rate of RNA polymerase II (1.7 kb/min) (Zeng et al. 2021; Fong et al. 2014). For time to exon ligation, we took the median number of nt past the 3'SS where exon-exon junctions appear in HEK293 cells (154 nt) (Reimer et al. 2020) divided by 1.7 kb/min and subtracted the time to branching calculated above. We converted both time to branching and time to exon ligation to k_{br} and k_{el} by converting to milliseconds and taking the inverse of each value. We also assumed the rates of Prp16p and Prp22p ATPase activity are slower than of k_{br} and k_{el}, as these helicases antagonize splicing of substrates with chemical modifications that are understood to slow the rates of branching and exon ligation, respectively(Koodathingal et al. 2010; Semlow et al. 2016). Specifically, we assigned the rates of k_{16} and k_{22} to 10% slower than the rates of k_{br} and k_{el}. We then assigned values of k_{dock} equal to the inverse of the mean first looping time (MFLT) of a single-stranded RNA with a distance of n nucleotides(Toan et al. 2006), assuming the persistence length is equal to the length of a single ribonucleotide:

$$k_{dock} = n^{-2.2}$$

We also assigned the rate of discard to be equal to a rate where k_{dock} at a given nucleotide position n would yield 90% total mRNA at steady state:

$$[mRNA]_{max} = \frac{k_{br}[pre]}{k_{decay}}$$

$$0.9\left(\frac{k_{br}[pre]}{k_{decay}}\right) = \frac{k_{br}[pre]}{k_{decay}\left(1 + \frac{k_{discard}}{k_{dock}\ @\ n\ nt}\right)}$$

$$0.9 = \frac{1}{\left(1 + \frac{k_{discard}}{k_{dock}\ @\ n\ nt}\right)}$$

$$k_{discard} = \frac{k_{dock}\ @\ n\ nt}{9}$$

When plotting [total mRNA] vs. number of nucleotides, we find that [total mRNA] is held at a constant value across the range of nucleotide distances we tested (**Fig. 3.8B**). This result is unsurprising, given mRNA yield relies on the amount of LI that progress through exon ligation. Lack of a LI discard step would ensure all LI generated by branching would eventually be converted to mRNA. Consistent with this logic, when setting $k_{discard}$ to a non-zero value, we saw a negative correlation between [total mRNA] and increasing BS to 3'SS distances. We first set $k_{discard}$ to coincide with 90% [total RNA] at 25 nt, which is where the effective BS to 3'SS distances are clustered in budding yeast and one nucleotide downstream of where we begin to see a divergence between relative exon ligation of substrates with wildtype or A3c 5'SS (**Fig. 3.6C**). At a $k_{discard}$ of 25 nt, we see a decrease in [total mRNA] with increasing BS to 3'SS distances (**Fig. 3.8B**). This result is consistent with our data above, showing a decrease in exon ligation efficiency for both wildtype and A3c 5'SS substrates, and may explain why longer effective BS to 3'SS distances are underrepresented across the intronome in several yeast species (Meyer et

al. 2011; Pérez-Valle and Vilardell 2012). When we slow the rate of discard to allow for half-maximal mRNA formation at 45 nt, we see a broadening of the distribution of [total mRNA] relative to longer nucleotide distances when compared to a $k_{discard}$ of 25 nt.

When we calculate exon ligation by taking the mRNA-to-LI ratio, we see a drop-off in exon ligation with increasing distance with all discard rates (**Fig. 3.8C**). However, we see a slight decrease in the mRNA-to-LI ratio with a fast discard rate relative to no discard. When the discard rate is slowed to allow efficient exon ligation at 45 nt, exon ligation increases at increased distances. These changes are largely driven by accumulation of LI at longer distances (**Fig 3.8D, E**). With all three discard parameters, we see a constant increase in LI levels with increasing nucleotide distance. However, in the absence of discard, LI accumulates exponentially, whereas LI accumulation levels begin to level off at 250 nt and 300 nt for fast and slow discard rates, respectively. When the discard rate is slowed to allow for optimal mRNA formation at 45 nt, LI accumulated to levels three-fold higher than when the discard rate is set to allow optimal formation of mRNA at 25 nt. This variability in LI accumulation was driven entirely by undocked LI, as levels of docked LI were the same for all discard parameters tested (**Fig. 3.8F**). These data suggest a LI discard step is necessary to prevent LI accumulation, and that preventing LI accumulation has a trade-off of forming lower levels of mRNA at increasing BS to 3'SS distances.

Prp43p antagonizes splicing of substrates with longer BS to 3'SS distances

Given the possibility that Prp43p could play a role in proofreading substrates that poorly engage the 3'SS, such as A3c, we decided to test this model directly. We transformed *ACT1-CUP1* reporters containing BS to 3'SS distances of 21 and 27 nt into

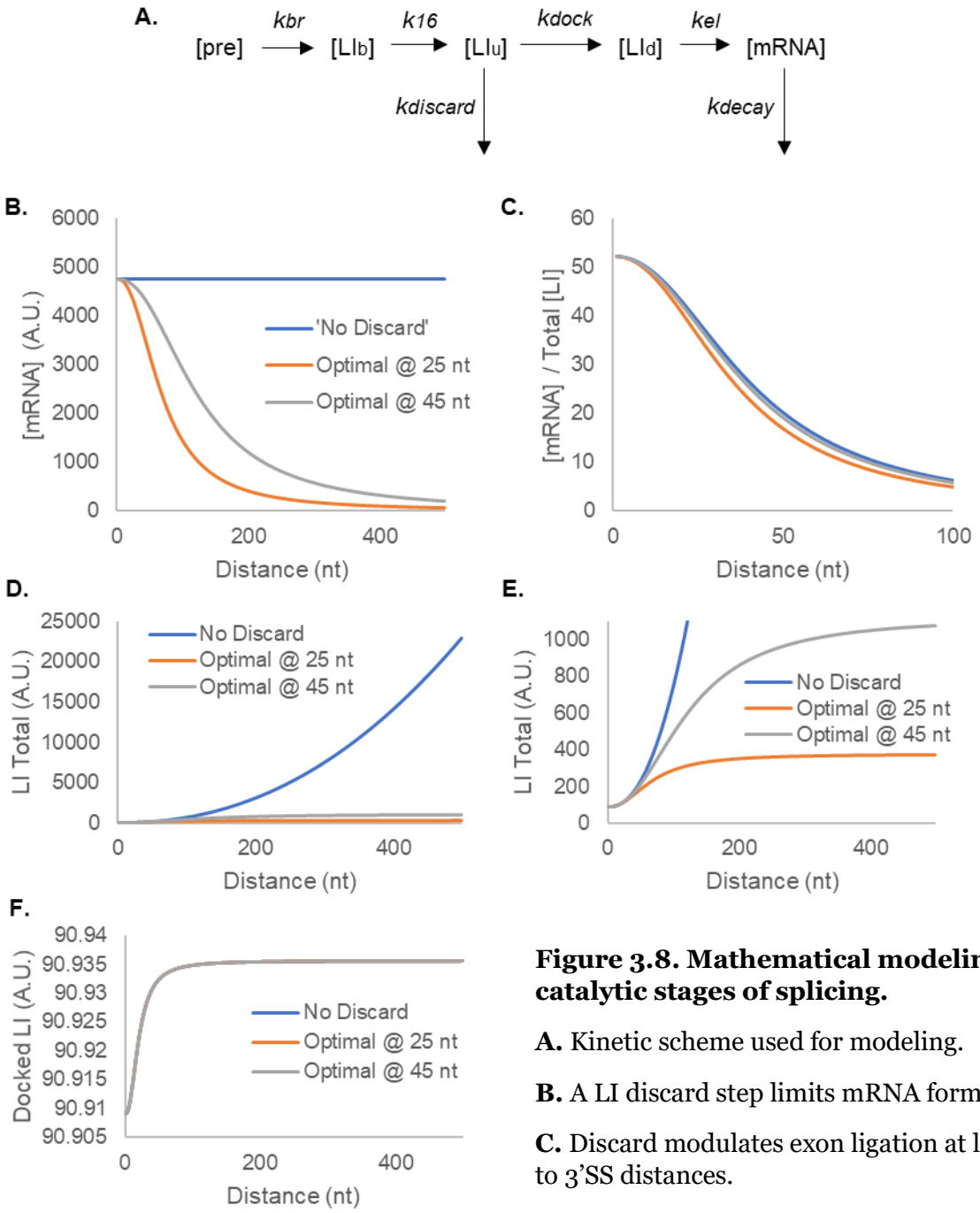

Figure 3.8. Mathematical modeling of the catalytic stages of splicing.

A. Kinetic scheme used for modeling.

B. A LI discard step limits mRNA formation.

C. Discard modulates exon ligation at longer BS to 3'SS distances.

D, E. Accumulation of total LI is higher with slowed or no discard.

F. Docked LI levels are not determined by discard rate.

strains containing either wildtype *PRP43* or *prp43-Q423N* and compared exon ligation efficiency using primer extension. The *PRP43* strain, as expected, exhibits a drop-off in exon ligation efficiency when the BS to 3'SS distance is extended from 21 to 27 nt (**Fig. 3.9A**). In the *prp43-Q423N* strain, exon ligation efficiency of both 21 and 27 nt are increased, relative to same distances in wt strain. Specifically, exon ligation increases 1.8-fold ±0.1 and 2.8-fold ± 0.7 with *prp43-Q423N* at BS to 3'SS distances of 21 and 27 nt, respectively. Both the increase in overall exon ligation efficiency, as well as the increase in fold exon ligation at extended distances, implies Prp43p ATPase activity increasingly antagonizes exon ligation of LIs as BS to 3'SS distance increases, as suggested by our mathematical modeling. These data are consistent with a role of Prp43p in disassembling spliceosomes stalled at the step of 3'SS docking.

Discussion

In this study, we sought to determine how an A3c mutation in the 5'SS antagonizes the exon ligation step of pre-mRNA splicing. In doing so, we uncovered a novel role for the DEAH-box ATPase Prp43p in disassembling spliceosomes that are deficient in 3'SS binding. We found that mutations that destabilize the intramolecular U2 stem IIc structure exacerbate the exon ligation defect of A3c (**Figs. 3.3, 3.4**), suggesting A3c impacts the Prp16-independent step of spliceosome rearrangement. Through *in silico* modeling of A3c into the 5'SS:U6 snRNA duplex (**Fig. 3.5**), we hypothesized A3c would impair 3'SS binding to the spliceosome. Consistent with this hypothesis, we found A3c narrowed the allowable BS to 3'SS distance (**Figs. 3.5, 3.6**), suggesting A3c does indeed negatively impact 3'SS binding. We also found that ATPase mutations of Prp22p and Prp43p can suppress the exon ligation defect of A3c (**Fig. 3.7**). Using mathematical

modeling of the catalytic stages of splicing, we hypothesized that a discard step between the Prp16p-dependent remodeling step and 3'SS binding could limit the allowable BS to 3'SS distance within an intron (**Fig 3.8**). Finally, we show that a Prp43p ATPase mutant also increases the allowable BS to 3'SS distance for exon ligation *in vivo* (**Fig. 3.9**). These data suggest that Prp43p is proofreading factor that discriminates against substrates with mutations that impart inefficiencies in 3'SS docking.

The 5'SS:U6 interaction has implications for 3'SS binding and usage

Our *in silico* structural analyses in combination with yeast genetics suggest that the interaction between the 5'SS and U6 snRNA has implications for 3'SS binding. Introduction of a Watson-Crick base pair between position +3 of the 5'SS residue G50 of U6 snRNA likely causes several shifts in residues responsible for 3'SS stabilization (**Fig. 3.5A, B**). This suggests A3c either destabilizes 3'SS binding (**Fig. 3.5A, B**), rendering the 3'SS susceptible to increased undocking by Prp22p (**Fig. 3.7**), and/or prevents binding of the 3'SS altogether and renders the spliceosome susceptible to Prp43p mediated discard (**Fig. 3.9**). Notably, exon ligation efficiency decreases at lower temperatures with A3c substrates (Konarska, Vilardell, and Query 2006). If A3c only conferred weakened 3'SS binding, lower temperatures should stabilize 3'SS binding and increase exon ligation efficiency. However, the converse is seen, suggesting lower temperatures may instead increase the stability of the A3c base-pairing interaction with U6 snRNA which would, in turn, inhibit binding of the 3'SS altogether.

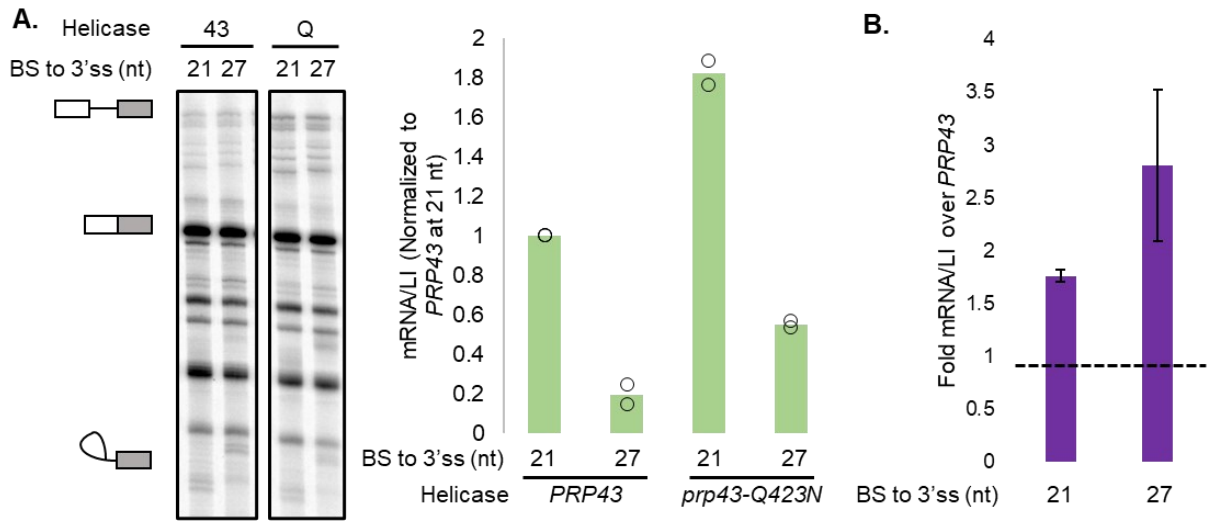

Figure 3.9: Prp43p antagonizes exon ligation of distal BS.

A. Primer extension of substrates bearing a wildtype 5'SS and BS to 3'SS distances of 21 and 27 nt. Exon ligation efficiencies were normalized to the exon ligation efficiency at 21 nt with wildtype Prp43p. A Prp43p ATPase mutant increases exon ligation efficiency at both 21 and 27 nt when compared to wildtype Prp43p.

B. Fold increase of exon ligation efficiency with a Prp43p ATPase mutant over wildtype Prp43p increases with increasing BS to 3'SS distance.

A3c introduces a Watson-Crick base pair at the end of the 5'SS:U6 helix. Importantly, base pairs at the end of RNA-RNA helices are susceptible to breathing. A3c substrates may therefore still be able to sample conformations that allow for 3'SS binding. In the budding yeast gene *RPL30,* which also contains an A3c substitution, a second substitution at position +4 (U4a) disrupts a canonical base-pairing interaction between the 5'SS and U6 snRNA (Eng and Warner 1991). This secondary substitution alleviates the exon ligation defect of A3c (Konarska, Vilardell, and Query 2006). Introduction of U4a likely prevents the nucleation and/or stability of Watson-Crick base pairing between 5'SS-A3c and U6-G50, thereby allowing the 5'SS and U6 snRNA to adopt conformations that more readily dock the 3'SS. Interestingly, co-variation between A3c and substitutions at position +4 that would disrupt 5'SS:U6 base-pairing is also seen across several 5'SS sequences in higher eukaryotes (Roca et al. 2008).

In a recent analysis of donor (5'SS) and acceptor (3'SS) splice site sequences, deviations at the 5'SS alter th e probability of selecting either a proximal or distal 3'SS in "NAGNAG" introns that contain tandem 3'SS sequences (Hujová et al. 2021). Deviations at position +3 that introduced a C or a U (A3c or A3u), which both have the potential to base pair with U6-G50 (Konarska, Vilardell, and Query 2006), antagonized the selection of an upstream suboptimal AAG 3'SS sequence in favor of a downstream optimal CAG 3'SS splice site (Hujová et al. 2021). The reduced usage of suboptimal 3'SS in the context of A3c and A3u likely reflects a deficiency in 3'SS binding, which would impose a greater requirement for selecting an optimal 3'SS sequence (e.g., CAG over AAG). Taken together, these analyses suggest substitutions in the 5'SS that alter its interaction with U6 snRNA antagonize exon ligation by disrupting the ability for the 3'SS to stably bind in the catalytic core.

Altering the identity of the 5'SS or the BS antagonizes exon ligation of distal 3'SSs

Our data show A3c narrows the usable window of BS to 3'SS distances (**Fig. 3.6**). Narrowing of this window likely reflects a 3'SS binding defect, in which the spliceosome spends decreased time in a conformation that allows stable 3'SS docking. We propose above that this is due to A3c driving the spliceosome into a conformation where key residues responsible for stable 3'SS binding are shifted out of position. These data rationalize why an A3c mutation in the *ACT1-CUP1* reporter and A3c substitutions in the *TER1* introns of *S. cryophilus* and *S. octosporus* block exon ligation (Ram Kannan et al. 2015). However, in *S. pombe*, the *TER1* intron is also suboptimal for exon ligation but does not include the A3c substitution at the 5'SS (R. Kannan et al. 2013). Instead, features such as an uncharacteristically strong BS sequence and a moderately long BS to 3'SS

distance confer a block to exon ligation during *TER1* splicing in *S. pombe*. Additionally, as seen with the A3c *ACT1-CUP1* substrate in our study, *TER1* in *S. pombe* also had a particularly narrow window of optimal BS to 3'SS distances for efficient exon ligation, suggesting this substrate may also have an underlying 3'SS binding defect.

The BS of *TER1* may impart this underlying 3'SS binding defect. The BS of *TER1* is unusual in that it carries a "strong" UACUAAC sequence compared to the more degenerate BS sequences in the *S. pombe* intronome. This sequence is more complementary to the GUAGUA box in U2 snRNA where the BS binds before branching. This increased complementarity was initially thought to impede the Prp16p-dependent transition out of the branching conformation, since Prp16p can separate the U2:BS duplex *in vitro* to allow for re-binding of the U2 snRNA to an alternative BS prior to branching catalysis (Semlow et al. 2016). However, cryo-EM structures of the spliceosome captured after Prp16p activity (C. Yan et al. 2017; Fica et al. 2017) and immediately after exon ligation (S. Liu et al. 2017a; Wilkinson et al. 2017) suggest the U2:BS duplex persists throughout the splicing cycle. Instead, Prp16p may catalyze a more subtle remodeling of the BS to allow for stable 3'SS binding. The third residue of the BS consensus sequence, which is a C in *S. cerevisiae*, forms a triplet interaction with residue U+2 of the 5'SS in the branching conformation (Galej et al. 2016). This triplet interaction is broken to allow U+2 to form an interaction with the Hoogsteen face of U6-A51 in the post-Prp16p complex (Fica et al. 2017), which appears to be important for proper positioning of U6-A51 (S. Liu et al. 2017a). Biochemical data show mutations at U6-A51 confer a strong block to exon ligation *in vitro* (Fabrizio and Abelson 1990), and structural data show U6-A51 stacks onto G-1 of the 3'SS to properly position the phosphate reactant for efficient exon ligation (Wilkinson et al. 2017). Thus, stabilizing the triplet interaction between U+2 and the

upstream C of the BS would antagonize a rearrangement necessary for proper 3'SS positioning within the catalytic core. Consistent with this, introduction of a U-to-C mutation at the third position of the BS of *FCP1*, which has a "weak" sequence of AAUUAAC that is more characteristic of *S. pombe* introns, confers a strong exon ligation defect similar to that seen with the UACUAAC BS of *TER1* (Kannan *et al.* 2013).

We propose above that the U+2:U6-A51 interaction would also be disrupted through the introduction of A3c, which would move the backbone of the 5'SS by 5.4 Å and move residue U+2 out of position to hydrogen bond to U6-A51 (**Fig. 3.5A**) and could nucleate a Watson-Crick base pair between U+2 and U6-A51 to further shift the base of U6-A51 by 10 Å (**Fig. 3.5B**). In this sense A3c in *ACT1-CUP1* in *S. cerevisiae* and the U-to-C substitution in the BS of *TER1* in *S. pombe* would both antagonize proper positioning of U6-A51, which would disfavor a conformation that allows for proper 3'SS binding and positioning. Thus, our data showing A3c narrows the allowable BS to 3'SS distance for optimal exon ligation mirrors the data from *TER1* in *S. pombe* because both substrates contain substitutions that antagonize 3'SS binding. Importantly, an ATPase mutation in *S. pombe* Prp43p also allows increased exon ligation of *TER1* (R. Kannan et al. 2013), indicating a general role for Prp43p ATPase activity in proofreading 3' splice site docking.

Prp22p binding commits the spliceosome to exon ligation by antagonizing recruitment of Prp43p

Prp22p has been shown to proofread both the 3'SS as well as mutations at G+1 and brA that alter residues required for 3'SS docking (Rabiah M Mayas, Maita, and Staley 2006). Mutations of these residues that reduced exon ligation can be suppressed by mutations in Prp22p that abrogate ATPase or helicase activity. We show the ATPase

mutant *prp22-R805A* also suppresses A3c, suggesting that increase exon ligation of A3c substrates may be due to disabling Prp22p proofreading activity. However, our discovery that Prp43p also proofreads A3c, as well as antagonizes distal 3'SS, suggests an orthogonal mechanism for *prp22-R805A* in alleviating the exon ligation defect of A3c.

Increased residence time of Prp22p on the spliceosome may also alleviate the exon ligation defect of A3c by inhibiting the binding of Prp43p. Recruitment of Prp43p requires binding of the dimeric Ntr1p-Ntr2p complex to the spliceosome (R.-T. Tsai et al. 2007). Cryo-EM structures of the post-exon ligation spliceosome indicate Ntr2p binding is sterically Slu7 (S. Liu et al. 2017a; Xiaofeng Zhang et al. 2019), an exon ligation factor that recruits Prp22p to the spliceosome (S.-A. James, Turner, and Schwer 2002). These data are consistent with previous biochemical studies showing that the presence of Slu7p antagonizes Ntr2p association with the spliceosome (Chen et al. 2013). Importantly, Slu7p remains bound to the spliceosome after exon ligation in the absence of Prp22p ATPase activity, and a dominant-negative *prp22-Q804A* mutation also antagonizes association of Prp43p (S.-A. James, Turner, and Schwer 2002). Release of excised lariat introns, a process dependent on Prp43p-mediated spliceosome disassembly, is also blocked by an ATPase mutant of Prp22p (Toroney *et al.* 2019), further suggesting that increased residence time of Prp22p blocks Prp43p activity.

Prp22p has also been shown to have an ATP-independent role in promoting exon ligation, with Prp22p being necessary for exon ligation of substrates that have long BS to 3'SS distances *in vitro* (Beate Schwer and Gross 1998). This ATP-independent role of Prp22p is likely due to stabilizing binding of Prp18p and Slu7p to the spliceosome, both of which also promote exon ligation of substrates with distal 3'SS (Frank and Guthrie 1992; Ohrt et al. 2013) that may be susceptible to Prp43p-mediated disassembly (Fig.

3.8). These observations support a model in which Prp22p binding commits the spliceosome to exon ligation by stabilizing Slu7p and Prp18p association to the spliceosome, preventing Prp43p-mediated disassembly by antagonizing Ntr2p binding. In the case of Prp22p-mediated proofreading, the ATPase activity of Prp22p would displace both itself (Beate Schwer and Gross 1998) and Slu7p (S.-A. James, Turner, and Schwer 2002), allowing for Ntr1p-Ntr2p binding for Prp43p recruitment to disassemble spliceosomes bound to a suboptimal substrate.

Prp43p as a timer for spliceosomal rearrangements

Our data showing *prp43-Q423N* allows higher exon ligation efficiencies at longer BS to 3'SS distances (**Fig. 3.9**) suggests Prp43p competes with the rate of 3'SS docking. Notably, 3'SS docking can only occur after branching is catalyzed, which creates a 2'-5' phosphodiester bond between the G+1 and brA that creates a platform to stabilize 3'SS binding (S. Liu et al. 2017a; Wilkinson et al. 2017). Prp16p ATPase activity is also necessary to reposition the G+1 and brA to allow for the juxtaposition of the 3'SS reactant with the catalytic core (**Fig. 3.5D**) (B. Schwer and Guthrie 1992; Semlow et al. 2016). Like Prp22p, Prp16p binds to the spliceosome prior to branching and can play an ATPase-independent role in promoting branching catalysis at suboptimal BS sequences (Tseng, Liu, and Cheng 2011). Prp16p residence on the spliceosome also inhibits binding of Ntr2 to prevent Prp43p recruitment (Chen et al. 2013), and Prp16p ATPase activity leads to Prp16p dissociation from the spliceosome (Chung et al. 2019). Thus, much like Prp22p ATPase activity during 3'SS proofreading, Prp16p ATPase activity after branching initiates rearrangements necessary for 3'SS binding while at the same time rendering the spliceosome susceptible to Ntr2 binding and Prp43p-mediated disassembly.

These observations suggest a model in which Prp43p generally acts on spliceosomes that are deficient in adopting a catalytic conformation after a helicase-mediated rearrangement. This model is consistent with a recent study showing that human Prp43p (DHX15) acts during spliceosome assembly to disassemble spliceosomes that are deficient in BS binding by U2 snRNA (Maul-Newby et al. 2021). In this case, DHX15 acts after an unspecified DEAH-box ATPase which removes SF3B to allow for BS sampling. This activity is similar to the way Prp16p and Prp22p ATPase activity allow re-sampling of the intron for BS and 3'SS, respectively (Semlow et al. 2016). Notably, re-sampling is limited to a short distance from the splice sites chosen before Prp16p or Prp22p. As Prp43p can act after Prp16p and Prp22p ATPase activity during proofreading (R. M. Mayas et al. 2010; Koodathingal et al. 2010; Toroney, Nielsen, and Staley 2019), Prp43p may impose a limit on the amount of time spent re-sampling the substrate. While re-sampling the intron can allow for productive splicing in instances where the spliceosome errantly binds a site near the correct splice site, increased sampling of the intron could lead to greater errors in splicing. Here, Prp43p would limit this re-sampling time to prevent erroneous splicing.

In the case of 3'SS binding, our data show Prp43p also limits the time available for 3'SS sampling after Prp16p activity (**Fig. 3.9**), much like DHX15 limits the time available for BS sampling during spliceosome assembly (Maul-Newby et al. 2021). In both instances, the spliceosome is placed in a state that is devoid of factors that would prevent Ntr1p-Ntr2p binding and Prp43p-mediated disassembly. Increased residence time in these Prp43p-sensitive states would increase the chance that the spliceosome would be disassembled prior to binding of the 3'SS or BS. In this sense, proofreading of the spliceosome may be akin to the initial kinetic proofreading scheme proposed by J.J.

Hopfield for translation fidelity (Hopfield 1974), where DEAH-box ATPases drive the spliceosome into a state where incorporation of a given splice site into the catalytic core is in direct competition with timing of spliceosome disassembly by Prp43p.

Prp43p may have exerted evolutionary pressure on intron structure and spliceosome composition

In our study, we find Prp43p imposes a limit on BS to 3'SS distances in the budding yeast *S. cerevisiae*. While BS to 3'SS distances in *S. cerevisiae* can be as long as 165 nt, the majority of introns form intramolecular structure in this region to shorten the "effective distance" between the BS and 3'SS to 25 nt (Meyer et al. 2011). Interestingly, our assays show an optimal distance of 18 nt when replacing the BS to 3'SS region with poly-A stretches (**Fig. 3.6**). However, given stretches of poly-A residues may form intramolecular base-base stacking interactions that would alter the persistence length of this ssRNA region (Seol et al. 2007), clustering of BS to 3'SS distances in the budding yeast genome around 25 nt may reflect the actual optimal distance for exon ligation. As detailed above, Prp43p imposes an upper limit on BS to 3'SS, suggesting budding yeast have evolved intramolecular RNA-RNA structure in introns with long BS to 3'SS distances to ensure efficient splicing of these substrates. Notably, intramolecular structure in the BS to 3'SS region of introns is also evolutionarily conserved across several yeast species (Meyer et al. 2011).

While yeast introns are relatively short, introns in higher eukayotes such as humans span several kilobases long, creating an even greater pressure to evolve a mechanism to ensure BS to 3'SS distances are short enough to allow for exon ligation. Although human introns are orders of magnitude longer than yeast introns, their BS to

3'SS distances also cluster around a 25 nt (Zeng et al. 2021). In humans, 3'SS recognition is concurrent with initial BS recognition during spliceosome assembly through the factor U2AF1 (Wu et al. 1999; Wahl, Will, and Lührmann 2009). Interestingly, U2AF1 lacks an ortholog in yeast, suggesting that humans and yeasts independently evolved mechanisms to ensure a short BS to 3'SS distance is present to ensure efficient 3'SS binding prior to Prp43p activity.

Prp43p disassembly of spliceosomes stalled in 3'SS binding would ensure an active pool of snRNPs for splicing of optimal substrates

Prp43p was first identified as a disassembly factor for the spliceosome that acts after exon ligation to promote additional rounds of pre-mRNA splicing (Arenas and Abelson 1997). Prp43p ATPase mutations not only prevent degradation of the excised lariat intermediate (Martin, Schneider, and Schwer 2002), but also lead to accumulation of pre-mRNA through preventing recycling of snRNP components (Arenas and Abelson 1997). Thus, Prp43p activity is necessary to allow for efficient splicing across the intronome. In the case of spliceosomes stalled along the splicing pathway, such as those stalled in 3'SS binding, Prp43p would act to disassemble unproductive spliceosomes to allow for reassembly of the snRNP components onto newly-transcribed introns that can be efficiently spliced (**Fig. 3.9**). In this sense, Prp43p acts as a surveillance factor,

ensuring snRNP components aren't sequestered on aberrant introns but instead are readily available to ensure quick and accurate genome-wide pre-mRNA splicing.

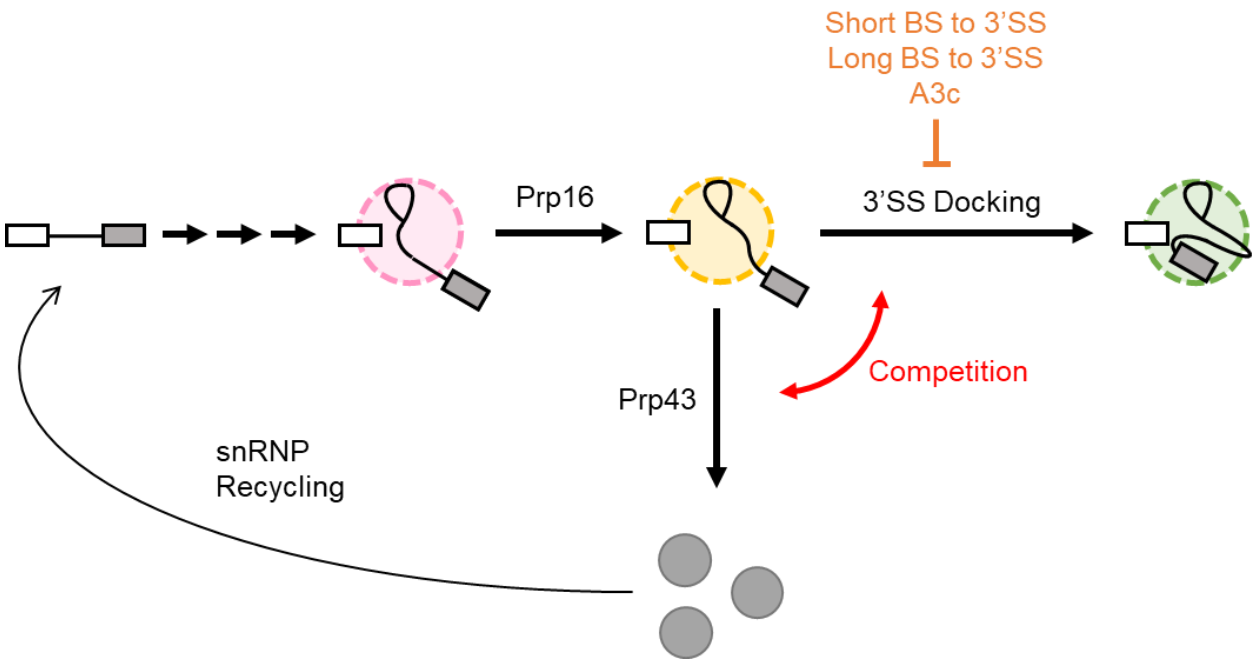

Figure 3.10: A model for Prp43p-mediated disassembly of spliceosomes stalled at the 3'SS docking step.

Kinetic competition between Prp43p ATPase activity and the rate of stable 3'SS docking antagonizes spliceosomes deficient in 3'SS docking. Disassembly of stalled spliceosomes allows recycling of snRNPs, allowing for reassembly onto newly transcribed introns to maintain efficient genome-wide splicing.

CHAPTER 4

CONCLUSIONS AND BROAD PERSPECTIVES

In this work, we examined the mechanisms by which DEAH-box ATPases ensure splicing fidelity and promote progression along the pre-mRNA splicing pathway. In characterizing disease-correlated mutations in the DEAH-box ATPase Prp16, we uncovered a conserved N-terminal motif that mediates an autoinhibitory interaction that regulates the spliceosomal activity of Prp16. Mutations in the N-terminal motif that interfere with this interaction lead to increased discard of BS mutations prior to the first catalytic step of splicing. In characterizing a mutation in the 5'SS of the intron, we uncovered a novel role for the spliceosomal DEAH-box ATPase Prp43 in discarding splicing intermediates that are deficient in 3'SS binding. These findings uncover novel mechanisms of regulating pre-mRNA splicing and increase our knowledge of how the activity of DEAH-box ATPases are leveraged to maintain faithful gene expression.

Rethinking the mechanisms of DEAH-box ATPase-mediated splicing fidelity

The DEAH-box ATPases Prp16 and Prp22 have long been known to play dual roles in splicing progression and fidelity (Semlow and Staley 2012). Genetic and biochemical data suggest Prp16 and Prp22 proofread through a kinetic proofreading model to discard suboptimal substrates prior to catalysis of branching or exon ligation, respectively (S. M. Burgess and Guthrie 1993; Rabiah M Mayas, Maita, and Staley 2006; Koodathingal et al. 2010). The suboptimal substrates that are discarded contain mutations in splice site sequences that would destabilize RNA-RNA interactions between the intron and the snRNA components they bind to, particularly with regards to branch site binding to the U2 snRNA component of the spliceosome (S. M. Burgess and Guthrie 1993). This observation, in addition the observation that Prp16 has the capacity to disengage the

branch site from U2 snRNA to activate alternative branch site usage (Semlow et al. 2016), led to a model in which Prp16 and Prp22 proofread by disengaging the intron from their snRNA binding sites. Given this model, mutations that weaken interactions between the intron and snRNA would be more likely to be disengaged by Prp16 or Prp22.

However, cryo-EM structures have illuminated functions for these branch site and 3'SS residues in coordinating reactants during both the branching and exon ligation catalytic steps. Importantly, the branch site residues found to be most sensitive to Prp16-mediated proofreading are found to participate in tertiary interactions important for stabilizing the branch point reactant in the catalytic core (S. M. Burgess and Guthrie 1993; Galej et al. 2016). Additionally, Prp22 not only discards 3'SS with suboptimal sequences, but also discards mutations in the 5'SS and the BS that play important roles in 3'SS binding (Rabiah M Mayas, Maita, and Staley 2006; Wilkinson et al. 2017; S. Liu et al. 2017a). Additionally, while the residues most sensitive to Prp16- and Prp22- dependent proofreading are more conserved between budding yeast and the degenerate splice site sequences found in metazoans. These observations lead to a subtle distinction in the kinetic proofreading model, where Prp16 and Prp22 do not proofread weakened binding of the intron to the spliceosome but instead proofread defects in the catalytic core.

Importantly, these DEAH-box ATPases bind to the periphery of the spliceosome and initiate rearrangements though pulling on the 3' end of the substrate (Semlow et al. 2016; Galej et al. 2016; Zhan et al. 2018; S. Liu et al. 2017a; Wilkinson et al. 2017). Prp16 and Prp22 also have a general role in removing reactants from the catalytic core both during proofreading and when promoting splicing progression (Semlow et al. 2016). These data suggest Prp16 and Prp22 may act as timers, with set rates of activity that

disengage reactants from the catalytic core (Semlow and Staley 2012). Mutations that slow or block transesterification chemistry are therefore more susceptible to discard, since chemistry cannot proceed prior to Prp16 or Prp22 activity (Koodathingal and Staley 2013). However, a separate but not mutually exclusive model for proofreading would be the "sensor" model, where a suboptimal substrate promotes a faster rate of Prp16 or Prp22 activity. How Prp16 or Prp22 could "sense" defects in formation of the catalytic core remains an open question. While these helicases bind to the periphery of the spliceosome, far away from the catalytic core of the spliceosome, their recruitment relies on recruitment factors such as Cwc25 and Yju2 for Prp16 or Slu7 and Prp18 for Prp22. Notably, these recruitment factors have domains that reach deep into the catalytic core (Galej et al. 2016; Fica et al. 2017; C. Yan et al. 2017; S. Liu et al. 2017a; Wilkinson et al. 2017; Zhan et al. 2018), which may sense changes or malformations within the catalytic core and, in turn, promote Prp16 or Prp22 ATPase activity.

Prp16 and Prp22 cooperate with the disassembly factor Prp43 to fully discard rejected substrates through spliceosome disassembly (R. M. Mayas et al. 2010; Koodathingal et al. 2010; Toroney, Nielsen, and Staley 2019). Although Prp43 is canonically a disassembly factor (Arenas and Abelson 1997), a previous study found that grossly mutated 3'SS sequences may be rejected by Prp43 in a Prp22-independent manner (R. M. Mayas et al. 2010). Here, we show that Prp43 also possesses a proofreading function, antagonizing exon ligation of intermediates that are deficient in binding of the 3'SS (Chapter 3). Our data parallels the observations seen previously, as grossly mutated 3'SS are likely deficient in initial 3'SS binding (R. M. Mayas et al. 2010). Notably, the 3'SS binding site is generated by the branching reaction (Fica et al. 2017),

which properly positions the G+1 and brA residues to interact with the G-1 and A-2 residues of the 3'SS (S. Liu et al. 2017a; Wilkinson et al. 2017). In this sense, pre-mRNA substrates that are rejected by Prp16-mediated proofreading are also deficient in 3'SS docking and are thus rendered susceptible to Prp43-mediated proofreading.

Prp43 has also been found to proofread U2 snRNP addition during spliceosome assembly (Maul-Newby et al. 2021), likely acting after an upstream DEAH-box ATPase-mediated rearrangement but before the ATPase DDX46. Similarly, the Staley Lab has preliminary evidence that Prp43 also acts to discard spliceosomes rejected by Prp2 (Jennifer Moore and Rebecca Toroney, unpublished data). Here, I find Prp43 acts between the ATPase activities of Prp16 and Prp22 to antagonize spliceosomes delayed in entry to the exon ligation conformation. In this sense, Prp43 may act as a timer of spliceosomal rearrangements, acting on spliceosomes that are stalled in forward progression. This mechanism of proofreading is akin to the kinetic proofreading scheme proposed by JJ Hopfield in translation fidelity. In the spliceosome, DEAH-box ATPases such as Prp2 and Prp16 initiate a spliceosomal rearrangement, displacing factors that prevent Ntr2 binding and placing the spliceosome in a state that is susceptible to Prp43-mediated discard (see Chapter 3). Progression into the next step of splicing recruits factors that disallow Prp43 recruitment, such as Prp16 or Slu7 (Chen et al. 2013). In this sense, irreversible discard of the spliceosome by Prp43 is in kinetic competition with the spliceosome adopting a catalytic conformation, much like tRNA disassociation is in competition with tRNA incorporation in the ribosome (Hopfield 1974).

Implications of helicase dysregulation on splice site choice

In addition to proofreading and initiating rearrangements in the spliceosome, Prp16 and Prp22 have ATP-independent roles in promoting splicing catalysis. An ATPase-dead mutant of Prp16 has been shown to promote branching of a suboptimal brC mutation (Tseng, Liu, and Cheng 2011), and Prp22 has been shown to promote splicing of substrates with distal 3'SS (Beate Schwer and Gross 1998). While these ATP-independent roles have been attributed to stabilizing the catalytic core, Prp16 and Prp22 residence time may instead prevent Prp43-dependent spliceosome disassembly (see above). These data suggest recruitment of DEAH-box ATPases is important for mediating splice site usage.

Prp16 and Prp22 have also been shown to mediate splice site choice. Prp16 activates alternative splice site choice *in vitro* (Semlow et al. 2016), and Prp22 allows for usage of downstream 3'SS both *in vitro* (Semlow et al. 2016) and *in vivo* (Ragle et al. 2015; Cody Hernandez, in prep). As alternative 3'SS choice is prevalent in AS events, efficient recruitment of Prp22 can alter AS through altering 3'SS usage. Alternative BS usage can also alter AS, and in diseases such as myelodysplastic disorders alternative 3'SS usage is activated through alternative BS selection (Alsafadi et al. 2016; Tang et al. 2016; Carrocci et al. 2017). In these diseases, alternative BS selection occurs as a result of altered recruitment of the helicase Prp5, which proofreads the BS early in the splicing cycle.

Prp16 is implicated in proofreading of the BS at the catalytic stage of splicing in an ATPase-dependent manner. Dysregulation of Prp16 ATPase activity has been shown to relax BS fidelity in budding yeast, which may also alter BS usage patterns in higher eukaryotes with more degenerate BS sequences. As mentioned above, I have found

evidence for an autoinhibitory interaction in Prp16, and disruption of this interaction leads to a "hyperfidelity" phenotype similar to that seen with some SF3b1 mutations that are orthologous to those alter BS usage in humans (Carrocci et al. 2017).

Importantly, BS selection can impact 3'SS usage, as I demonstrate in Chapter 3 of this thesis. Increased distance between the BS and 3'SS leads to decreased efficiency of exon ligation (see Chapter 3) because introns are delayed in 3'SS binding. Recently, sequencing of co-transcriptional LI in the Staley Lab have revealed 3'SS selection is coupled to BS selection in HEK293 cells (Zeng et al. 2021). While this coupling likely occurs upstream of branching by a protein complex consisting of the branch point binding factor SF3b1 and the 3'SS binding factor U2AF1 (also known as U2AF35) (Wahl, Will, and Lührmann 2009), co-transcriptional LI sequencing also revealed a subset of introns that do not require U2AF1 for branching (Zeng et al. 2021). This would uncouple 3'SS selection from BS selection, rendering these introns more susceptible to aberrant 3'SS usage as a consequence of altered BS selection. While this hypothesis is compelling, more examination is required to determine whether dysregulation of Prp16 autoinhibition does indeed lead to altered 3'SS usage through aberrant BS selection in humans.

Consequences of LI discard prior to exon ligation

In this work, I show that Prp43 disassembles spliceosomes that are bound to introns that are deficient in 3'SS binding. A notable example of this is the *TER1* intron in fission yeasts (R. Kannan et al. 2013; Ram Kannan et al. 2015), which leverages this premature spliceosome disassembly to generate telomerase RNA. More endogenous introns in different yeast species were found to be discarded after branching catalysis

through the use of spliceosome footprinting (Burke et al. 2018). One of these "incompletely spliced" transcripts is the *BDF2* transcript in *S. cerevisiae*. Characterization of *BDF2* splicing shows that lariat intermediates become extended upon deletion of the debranchase Dbr1 and the 5' to 3' exonuclease Rat1. Dbr1 is necessary for debranching of excised lariat introns (Chapman and Boeke 1991), which cleaves the 2'-5' phosphodiester bond between brA and the 5' splice site. Notably, Rat1 is also involved in transcription termination by acting on 5' ends of nascent RNA generated by poly-A site cleavage (Kim et al. 2004). These data suggest that LI discarded through Prp43-mediated spliceosome assembly are degraded co-transcriptionally, and this degradation leads to transcription termination of incompletely spliced transcripts. Indeed, spliceosome profiling also shows increased 3' tail lengths of RNAs when ATPase activity of Prp43 is disabled (Burke et al. 2018). I also have preliminary evidence that Dbr1-mediated co-transcriptional degradation is necessary to allow for expression of the gene downstream of *TER1* in the fission yeast *S. pombe* (see Appendix A). In this sense, pre-mRNA splicing feeds back onto RNA transcription, prompting termination and degradation of transcripts that have failed to undergo exon ligation to prevent aberrant gene expression.

APPENDIX A

EARLY EVIDENCE FOR A SPLICEOSOME-DEPENDENT

PATHWAY FOR TRANSCRIPTION TERMINATION

Introduction

Discard of splicing intermediates prior to exon ligation has been noted for certain substrates in yeast species. Notably, the *TER1* pre-mRNA in fission yeast species require 3' end processing by the spliceosome to generate telomerase RNA (R. Kannan et al. 2013; Ram Kannan et al. 2015). Intronic features within ensure early termination of pre-mRNA splicing prior to exon ligation, triggering discard of mature telomerase RNA encoded by the free 5' exon. Both the free 5' exon product, as well as a branched LI species, are products of the branching catalytic step of splicing. As both branching and exon ligation occur co-transcriptionally (Reimer et al. 2020; Zeng et al. 2021), and exon ligation is required to fully excise the lariat intron from nascent RNA, disassembly of the spliceosome prior to exon ligation would have the consequence of LI species remaining tethered to RNA polymerase II. Consistent with this, blocking co-transcriptional spliceosome disassembly using a catalytically dead mutant of Prp43p has been shown to increase the tail length of RNA species (Burke et al. 2018).

Co-transcriptional disassembly of the spliceosome would render LI susceptible to Dbr1p, an enzyme that "debranches" lariat introns by cleaving the 2'-5' linkage between the 5' splice site and brA (Chapman and Boeke 1991). This activity is necessary for turnover of lariat intermediates by exposing a free 5' end that allows for exonuclease degradation. Dbr1p has also been shown to stabilize LI in budding yeast *in vivo* (Query and Konarska 2004). The observations that Prp43p protects excised lariat introns from degradation (Martin, Schneider, and Schwer 2002; Toroney, Nielsen, and Staley 2019), that Dbr1p is also necessary for LI degradation, and that blocking Prp43p-dependent co-transcriptional spliceosome disassembly leads to elongated nascent RNA associated with

the spliceosome (Burke et al. 2018) all suggest that co-transcriptional spliceosome disassembly renders nascent LI renders susceptible to Dbr1p-mediated degradation.

Notably, Dbr1p would generate a free 5' RNA end that can be acted on by 5'-to-3' exonucleases. The 5'-to-3' exonuclease Rat1p has been shown to act co-transcriptionally on nascent RNA after cleavage at the poly-A site. Co-transcriptional degradation by Rat1p leads to disruption of the interaction between DNA and RNA polymerase II, leading to transcription termination in a mechanism termed "torpedo transcription termination." (Kim et al. 2004) . A similar mechanism of transcription termination has been observed after co-transcriptional decapping of nascent RNA, which exposes the 5' end of nascent RNA to the 5'-to-3' exonuclease Xrn2. Like Rat1p in yeast, Xrn2 activity on nascent RNA leads to early transcription termination in HeLa and HEK293. Since co-transcriptional Dbr1p activity would also expose a free 5' end, 5'-to-3' degradation of nascent RNA can also lead to transcription termination of LI discarded by Prp43p.

To determine what impact co-transcriptional debranching of LI would have on transcription termination, we looked to the fission yeast *S. pombe*. The *TER1* transcript in *S. pombe* undergoes branching but not exon ligation (R. Kannan et al. 2013). Additionally, NET-seq shows very low RNA Pol II occupancy past of the branch site of *TER1* (**Fig A.1**, *SPNCRNA.214*) (Wery et al. 2018), suggesting that RNA pol II stops transcription just after the spliceosome catalyzes branching. Just downstream of *TER1* is *SPAC16A10.03c*, which is a paralog of the vacuole targeting factor Vps11p (Sajiki et al. 2009). Given the close proximity of these genes, we hypothesized that blocking Dbr1p activity would lead to readthrough of *TER1* transcription, which would dysregulate *SPAC16A10.03c* expression.

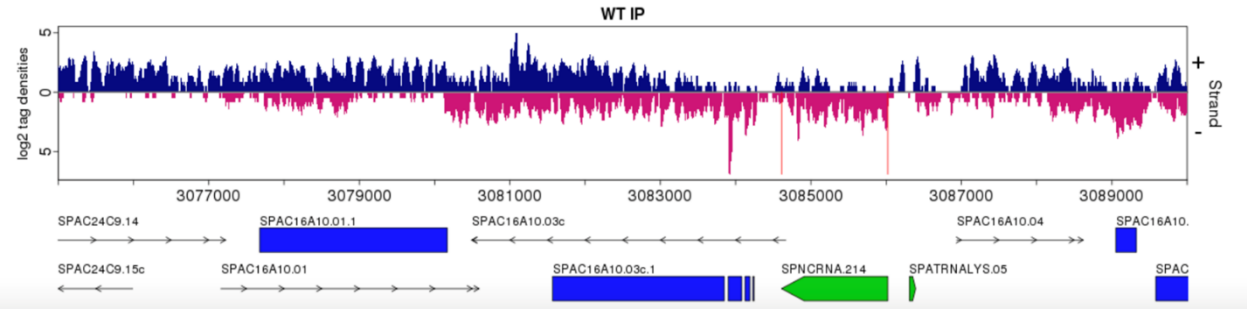

Figure A.1. Native elongating transcript sequencing (NET-seq) data show low occupancy on *TER1* after its branch site consensus sequence.

TER1 is denoted as SPNCRNA.214. RNA polymerase II occupancy is inferred by the number of reads on the minus strand (pink).

Materials and Methods

Vacuole targeting assay: Wildtype, *dbr1Δ*, and *spac16a10.03cΔ* were grown in 6 mL YE5S media to log phase. Six µL of FM4-64FX dye was added, and cultures were incubated for 30 minutes at 30 °C. Cells were spun down, washed with YE5S, and incubated for another 30 minutes at 30 °C prior to imaging.

Results

In wildtype cells, normal vacuole targeting of the FM4-64FX was seen, as indicated by the presence of a number small, circular structures (**Fig. A.2**). Wildtype cells also exhibited normal morphology. In contrast, *dbr1Δ* cells had a notable vacuole targeting defect, as evidenced by high levels of FM4-64FX within the cytoplasm. *dbr1Δ* cells also had morphological defects such as elongated cell types and the presence of multiple nuclei (**Fig. A.2**), indicative of a cell cycle defect. Of these phenotypes, the vacuole targeting defect seen in *dbr1Δ* is also present in *spac16a10.03cΔ*, supporting our hypothesis that disabling co-transcriptional Dbr1p activity may downregulate proper expression of *SPAC16A10.03C*. While these data are preliminary, they suggest Dbr1p activity on *TER1*

LI discarded by Prp43p may be necessary for proper transcription termination of the *TER1* gene to allow for proper expression of *SPAC16A10.03C*.

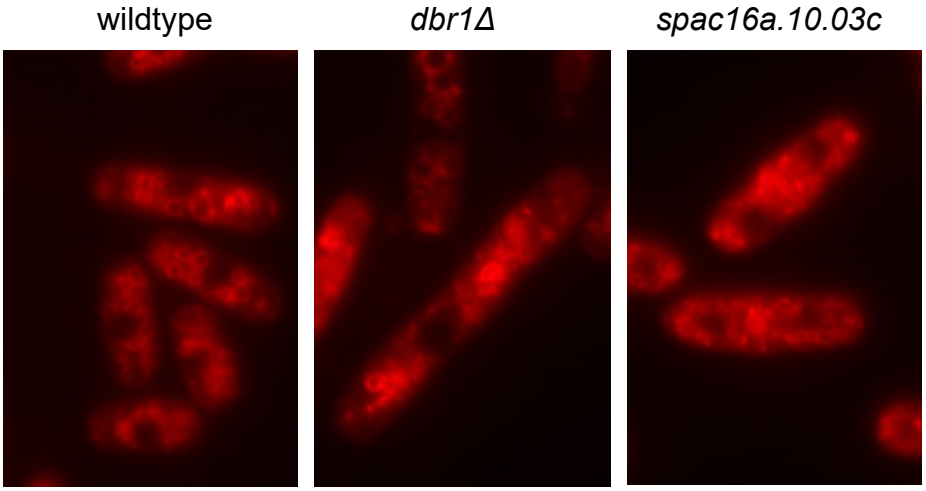

Figure A.2. Discard of *TER1* pre-mRNA in *S. pombe* after branching but before exon ligation may be important for faithful expression of *SPAC16A.10.3C*.

FM4-64FX staining of *S. pombe* strains suggest a slight vacuole targeting defect in *dbr1*Δ that is phenocopied by *SPAC16A10.03c.1* deletion.

REFERENCES

Absmeier, Eva, Karine F. Santos, and Markus C. Wahl. 2019. "Molecular Mechanism Underlying Inhibition of Intrinsic ATPase Activity in a Ski2-like RNA Helicase." *Structure*, December. https://doi.org/10.1016/j.str.2019.11.014.

Absmeier, Eva, Jan Wollenhaupt, Sina Mozaffari-Jovin, Christian Becke, Chung-Tien Lee, Marco Preussner, Florian Heyd, et al. 2015. "The Large N-Terminal Region of the Brr2 RNA Helicase Guides Productive Spliceosome Activation." *Genes & Development*, December, genesdev;gad.271528.115v1. https://doi.org/10.1101/gad.271528.115.

Ajmal, Muhammad, Muhammad Imran Khan, Kornelia Neveling, Yar Muhammad Khan, Maleeha Azam, Nadia Khalida Waheed, Christian P Hamel, et al. 2014. "A Missense Mutation in the Splicing Factor Gene *DHX38* Is Associated with Early-Onset Retinitis Pigmentosa with Macular Coloboma." *Journal of Medical Genetics* 51 (7): 444–48. https://doi.org/10.1136/jmedgenet-2014-102316.

Alsafadi, Samar, Alexandre Houy, Aude Battistella, Tatiana Popova, Michel Wassef, Emilie Henry, Franck Tirode, et al. 2016. "Cancer-Associated SF3B1 Mutations Affect Alternative Splicing by Promoting Alternative Branchpoint Usage." *Nature Communications* 7 (1): 10615. https://doi.org/10.1038/ncomms10615.

Arenas, J. E., and J. N. Abelson. 1997. "Prp43: An RNA Helicase-like Factor Involved in Spliceosome Disassembly." *Proceedings of the National Academy of Sciences* 94 (22): 11798–802. https://doi.org/10.1073/pnas.94.22.11798.

Bai, Rui, Ruixue Wan, Chuangye Yan, Qi Jia, Jianlin Lei, and Yigong Shi. 2021. "Mechanism of Spliceosome Remodeling by the ATPase/Helicase Prp2 and Its Coactivator Spp2." *Science* 371 (6525). https://doi.org/10.1126/science.abe8863.

Bai, Rui, Chuangye Yan, Ruixue Wan, Jianlin Lei, and Yigong Shi. 2017. "Structure of the Post-Catalytic Spliceosome from Saccharomyces Cerevisiae." *Cell* 171 (7): 1589-1598.e8. https://doi.org/10.1016/j.cell.2017.10.038.

Bailey, Timothy L., James Johnson, Charles E. Grant, and William S. Noble. 2015. "The MEME Suite." *Nucleic Acids Research* 43 (W1): W39–49. https://doi.org/10.1093/nar/gkv416.

Blencowe, Benjamin J. 2006. "Alternative Splicing: New Insights from Global Analyses." *Cell* 126 (1): 37–47. https://doi.org/10.1016/j.cell.2006.06.023.

Burgess, S. M., and C. Guthrie. 1993. "A Mechanism to Enhance MRNA Splicing Fidelity: The RNA-Dependent ATPase Prp16 Governs Usage of a Discard Pathway for Aberrant Lariat Intermediates." *Cell* 73 (7): 1377–91. https://doi.org/10.1016/0092-8674(93)90363-u.

Burgess, Sean, Joseph R. Couto, and Christine Guthrie. 1990. "A Putative ATP Binding Protein Influences the Fidelity of Branchpoint Recognition in Yeast Splicing." *Cell* 60 (5): 705–17. https://doi.org/10.1016/0092-8674(90)90086-T.

Burke, Jordan E., Adam D. Longhurst, Daria Merkurjev, Jade Sales-Lee, Beiduo Rao, James J. Moresco, John R. Yates, Jingyi Jessica Li, and Hiten D. Madhani. 2018. "Spliceosome Profiling Visualizes Operations of a Dynamic RNP at Nucleotide Resolution." *Cell* 173 (4): 1014-1030.e17. https://doi.org/10.1016/j.cell.2018.03.020.

Buskin, Adriana, Lili Zhu, Valeria Chichagova, Basudha Basu, Sina Mozaffari-Jovin, David Dolan, Alastair Droop, et al. 2018. “Disrupted Alternative Splicing for Genes Implicated in Splicing and Ciliogenesis Causes PRPF31 Retinitis Pigmentosa.” *Nature Communications* 9 (1): 4234. https://doi.org/10.1038/s41467-018-06448-y.

Carmel, Liran, and Michal Chorev. 2012. “The Function of Introns.” *Frontiers in Genetics* 3: 55. https://doi.org/10.3389/fgene.2012.00055.

Carrocci, Tucker J., Douglas M. Zoerner, Joshua C. Paulson, and Aaron A. Hoskins. 2017. “SF3b1 Mutations Associated with Myelodysplastic Syndromes Alter the Fidelity of Branchsite Selection in Yeast.” *Nucleic Acids Research* 45 (8): 4837–52. https://doi.org/10.1093/nar/gkw1349.

Chapman, K. B., and J. D. Boeke. 1991. “Isolation and Characterization of the Gene Encoding Yeast Debranching Enzyme.” *Cell* 65 (3): 483–92. https://doi.org/10.1016/0092-8674(91)90466-c.

Chen, Hsin-Chou, Chi-Kang Tseng, Rong-Tzong Tsai, Che-Sheng Chung, and Soo-Chen Cheng. 2013. “Link of NTR-Mediated Spliceosome Disassembly with DEAH-Box ATPases Prp2, Prp16, and Prp22.” *Molecular and Cellular Biology* 33 (3): 514–25. https://doi.org/10.1128/MCB.01093-12.

Chung, Che-Sheng, Chi-Kang Tseng, Yung-Hua Lai, Hui-Fang Wang, Andrew J Newman, and Soo-Chen Cheng. 2019. “Dynamic Protein–RNA Interactions in Mediating Splicing Catalysis.” *Nucleic Acids Research* 47 (2): 899–910. https://doi.org/10.1093/nar/gky1089.

Company, Mahshid, Jaime Arenas, and John Abelson. 1991. “Requirement of the RNA Helicase-like Protein PRP22 for Release of Messenger RNA from Spliceosomes.” *Nature* 349 (6309): 487–93. https://doi.org/10.1038/349487a0.

Cordin, Olivier, Daniela Hahn, Ross Alexander, Amit Gautam, Cosmin Saveanu, J. David Barrass, and Jean D. Beggs. 2014. “Brr2p Carboxy-Terminal Sec63 Domain Modulates Prp16 Splicing RNA Helicase.” *Nucleic Acids Research* 42 (22): 13897–910. https://doi.org/10.1093/nar/gku1238.

Couto, J. R., J. Tamm, R. Parker, and C. Guthrie. 1987. “A Trans-Acting Suppressor Restores Splicing of a Yeast Intron with a Branch Point Mutation.” *Genes & Development* 1 (5): 445–55. https://doi.org/10.1101/gad.1.5.445.

Crick, Francis. 1970. “Central Dogma of Molecular Biology.” *Nature* 227 (5258): 561–63. https://doi.org/10.1038/227561a0.

Darman, Rachel B., Michael Seiler, Anant A. Agrawal, Kian H. Lim, Shouyong Peng, Daniel Aird, Suzanna L. Bailey, et al. 2015. “Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3′ Splice Site Selection through Use of a Different Branch Point.” *Cell Reports* 13 (5): 1033–45. https://doi.org/10.1016/j.celrep.2015.09.053.

Dillman, Allissa A., David N. Hauser, J. Raphael Gibbs, Michael A. Nalls, Melissa K. McCoy, Iakov N. Rudenko, Dagmar Galter, and Mark R. Cookson. 2013. “MRNA Expression, Splicing and Editing in the Embryonic and Adult Mouse Cerebral Cortex.” *Nature Neuroscience* 16 (4): 499–506. https://doi.org/10.1038/nn.3332.

Eng, Francis J., and Jonathan R. Warner. 1991. “Structural Basis for the Regulation of Splicing of a Yeast Messenger RNA.” *Cell* 65 (5): 797–804. https://doi.org/10.1016/0092-8674(91)90387-E.

Fabrizio, P., and J. Abelson. 1990. “Two Domains of Yeast U6 Small Nuclear RNA Required for Both Steps of Nuclear Precursor Messenger RNA Splicing.” *Science* 250 (4979): 404–9. https://doi.org/10.1126/science.2145630.

Fairman-Williams, Margaret E, Ulf-Peter Guenther, and Eckhard Jankowsky. 2010. “SF1 and SF2 Helicases: Family Matters.” *Current Opinion in Structural Biology* 20 (3): 313–24. https://doi.org/10.1016/j.sbi.2010.03.011.

Ferrari, Luca, Riccardo Stucchi, Katerina Konstantoulea, Gerarda van de Kamp, Renate Kos, Willie J. C. Geerts, Laura S. van Bezouwen, et al. 2020. “Arginine π-Stacking Drives Binding to Fibrils of the Alzheimer Protein Tau.” *Nature Communications* 11 (1): 571. https://doi.org/10.1038/s41467-019-13745-7.

Fica, Sebastian M., Chris Oubridge, Wojciech P. Galej, Max E. Wilkinson, Xiao-Chen Bai, Andrew J. Newman, and Kiyoshi Nagai. 2017. “Structure of a Spliceosome Remodelled for Exon Ligation.” *Nature* 542 (7641): 377–80. https://doi.org/10.1038/nature21078.

Fica, Sebastian M., Nicole Tuttle, Thaddeus Novak, Nan-Sheng Li, Jun Lu, Prakash Koodathingal, Qing Dai, Jonathan P. Staley, and Joseph A. Piccirilli. 2013. “RNA Catalyses Nuclear Pre-MRNA Splicing.” *Nature* 503 (7475): 229–34. https://doi.org/10.1038/nature12734.

Fong, Nova, Hyunmin Kim, Yu Zhou, Xiong Ji, Jinsong Qiu, Tassa Saldi, Katrina Diener, Ken Jones, Xiang-Dong Fu, and David L. Bentley. 2014. “Pre-MRNA

Splicing Is Facilitated by an Optimal RNA Polymerase II Elongation Rate." *Genes & Development* 28 (23): 2663–76. https://doi.org/10.1101/gad.252106.114.

Fourmann, Jean-Baptiste, Olexandr Dybkov, Dmitry E Agafonov, Marcel J Tauchert, Henning Urlaub, Ralf Ficner, Patrizia Fabrizio, and Reinhard Lührmann. 2016. "The Target of the DEAH-Box NTP Triphosphatase Prp43 in Saccharomyces Cerevisiae Spliceosomes Is the U2 SnRNP-Intron Interaction." *ELife* 5 (April): e15564. https://doi.org/10.7554/eLife.15564.

Frank, D, and C Guthrie. 1992. "An Essential Splicing Factor, SLU7, Mediates 3' Splice Site Choice in Yeast." *Genes & Development* 6 (11): 2112–24. https://doi.org/10.1101/gad.6.11.2112.

Fromont-Racine, Micheline, Jean-Christophe Rain, and Pierre Legrain. 1997. "Toward a Functional Analysis of the Yeast Genome through Exhaustive Two-Hybrid Screens." *Nature Genetics* 16 (3): 277–82. https://doi.org/10.1038/ng0797-277.

Galej, Wojciech P., Max E. Wilkinson, Sebastian M. Fica, Chris Oubridge, Andrew J. Newman, and Kiyoshi Nagai. 2016. "Cryo-EM Structure of the Spliceosome Immediately after Branching." *Nature* 537 (7619): 197–201. https://doi.org/10.1038/nature19316.

Gietz, R., and Robin Woods. 2002. "Transformation of Yeast by Lithium Acetate/Single-Stranded Carrier DNA/Polyethylene Glycol Method." *Methods in Enzymology* 350 (February): 87–96. https://doi.org/10.1016/S0076-6879(02)50957-5.

Haselbach, David, Ilya Komarov, Dmitry E. Agafonov, Klaus Hartmuth, Benjamin Graf, Olexandr Dybkov, Henning Urlaub, Berthold Kastner, Reinhard Lührmann, and Holger Stark. 2018. "Structure and Conformational Dynamics of the Human

Spliceosomal Bact Complex." *Cell* 172 (3): 454-464.e11. https://doi.org/10.1016/j.cell.2018.01.010.

He, Yangzi, Gregers R Andersen, and Klaus H Nielsen. 2010. "Structural Basis for the Function of DEAH Helicases." *EMBO Reports* 11 (3): 180–86. https://doi.org/10.1038/embor.2010.11.

He, Yangzi, Jonathan P. Staley, Gregers Rom Andersen, and Klaus H. Nielsen. 2017. "Structure of the DEAH/RHA ATPase Prp43p Bound to RNA Implicates a Pair of Hairpins and Motif Va in Translocation along RNA." *RNA* 23 (7): 1110–24. https://doi.org/10.1261/rna.060954.117.

Hilliker, A. K., M. A. Mefford, and J. P. Staley. 2007. "U2 Toggles Iteratively between the Stem IIa and Stem IIc Conformations to Promote Pre-MRNA Splicing." *Genes & Development* 21 (7): 821–34. https://doi.org/10.1101/gad.1536107.

Hong, Xin, Douglas G. Scofield, and Michael Lynch. 2006. "Intron Size, Abundance, and Distribution within Untranslated Regions of Genes." *Molecular Biology and Evolution* 23 (12): 2392–2404. https://doi.org/10.1093/molbev/msl111.

Hopfield, J. J. 1974. "Kinetic Proofreading: A New Mechanism for Reducing Errors in Biosynthetic Processes Requiring High Specificity." *Proceedings of the National Academy of Sciences* 71 (10): 4135–39. https://doi.org/10.1073/pnas.71.10.4135.

Hoskins, Aaron A., and Melissa J. Moore. 2012. "The Spliceosome: A Flexible, Reversible Macromolecular Machine." *Trends in Biochemical Sciences* 37 (5): 179–88. https://doi.org/10.1016/j.tibs.2012.02.009.

Hujová, Pavla, Přemysl Souček, Lenka Radová, Michal Kramárek, Tatiana Kováčová, and Tomáš Freiberger. 2021. "Nucleotides in Both Donor and Acceptor Splice Sites Are Responsible for Choice in NAGNAG Tandem Splice Sites." *Cellular and*

Molecular Life Sciences 78 (21): 6979–93. https://doi.org/10.1007/s00018-021-03943-2.

Irimia, Manuel, Robert J. Weatheritt, Jonathan D. Ellis, Neelroop N. Parikshak, Thomas Gonatopoulos-Pournatzis, Mariana Babor, Mathieu Quesnel-Vallières, et al. 2014. "A Highly Conserved Program of Neuronal Microexons Is Misregulated in Autistic Brains." *Cell* 159 (7): 1511–23. https://doi.org/10.1016/j.cell.2014.11.035.

Jackson, Richard J., Christopher U. T. Hellen, and Tatyana V. Pestova. 2010. "The Mechanism of Eukaryotic Translation Initiation and Principles of Its Regulation." *Nature Reviews Molecular Cell Biology* 11 (2): 113–27. https://doi.org/10.1038/nrm2838.

James, Philip, John Halladay, and Elizabeth A. Craig. 1996. "Genomic Libraries and a Host Strain Designed for Highly Efficient Two-Hybrid Selection in Yeast." *Genetics* 144 (4): 1425–36.

James, Shelly-Ann, William Turner, and Beate Schwer. 2002. "How Slu7 and Prp18 Cooperate in the Second Step of Yeast Pre-MRNA Splicing." *RNA* 8 (8): 1068–77. https://doi.org/10.1017/S1355838202022033.

Jankowsky, Eckhard. 2011. "RNA Helicases at Work: Binding and Rearranging." *Trends in Biochemical Sciences* 36 (1): 19–29. https://doi.org/10.1016/j.tibs.2010.07.008.

Jumper, John, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, et al. 2021. "Highly Accurate Protein Structure Prediction with AlphaFold." *Nature* 596 (7873): 583–89. https://doi.org/10.1038/s41586-021-03819-2.

Kang, Hyun-Seo, Carolina Sánchez-Rico, Stefanie Ebersberger, F. X. Reymond Sutandy, Anke Busch, Thomas Welte, Ralf Stehle, et al. 2020. “An Autoinhibitory Intramolecular Interaction Proof-Reads RNA Recognition by the Essential Splicing Factor U2AF2.” *Proceedings of the National Academy of Sciences* 117 (13): 7140–49. https://doi.org/10.1073/pnas.1913483117.

Kannan, R., S. Hartnett, R. B. Voelker, J. A. Berglund, J. P. Staley, and P. Baumann. 2013. “Intronic Sequence Elements Impede Exon Ligation and Trigger a Discard Pathway That Yields Functional Telomerase RNA in Fission Yeast.” *Genes & Development* 27 (6): 627–38. https://doi.org/10.1101/gad.212738.112.

Kannan, Ram, Rachel M. Helston, Richard O. Dannebaum, and Peter Baumann. 2015. “Diverse Mechanisms for Spliceosome-Mediated 3′ End Processing of Telomerase RNA.” *Nature Communications* 6 (1): 6104. https://doi.org/10.1038/ncomms7104.

Kim, Minkyu, Nevan J. Krogan, Lidia Vasiljeva, Oliver J. Rando, Eduard Nedea, Jack F. Greenblatt, and Stephen Buratowski. 2004. “The Yeast Rat1 Exonuclease Promotes Transcription Termination by RNA Polymerase II.” *Nature* 432 (7016): 517–22. https://doi.org/10.1038/nature03041.

Konarska, Maria M., Josep Vilardell, and Charles C. Query. 2006. “Repositioning of the Reaction Intermediate within the Catalytic Center of the Spliceosome.” *Molecular Cell* 21 (4): 543–53. https://doi.org/10.1016/j.molcel.2006.01.017.

Koodathingal, Prakash, Thaddeus Novak, Joseph A. Piccirilli, and Jonathan P. Staley. 2010. “The DEAH Box ATPases Prp16 and Prp43 Cooperate to Proofread 5′ Splice Site Cleavage during Pre-MRNA Splicing.” *Molecular Cell* 39 (3): 385–95. https://doi.org/10.1016/j.molcel.2010.07.014.

Koodathingal, Prakash, and Jonathan P. Staley. 2013. “Splicing Fidelity: DEAD/H-Box ATPases as Molecular Clocks.” *RNA Biology* 10 (7): 1073–79. https://doi.org/10.4161/rna.25245.

Latif, Zahid, Imen Chakchouk, Isabelle Schrauwen, Kwanghyuk Lee, Regie Lyn P. Santos-Cortez, Izoduwa Abbe, Anushree Acharya, et al. 2018. “Confirmation of the Role of DHX38 in the Etiology of Early-Onset Retinitis Pigmentosa.” *Investigative Ophthalmology & Visual Science* 59 (11): 4552–57. https://doi.org/10.1167/iovs.18-23849.

Ledoux, Sarah, and Christine Guthrie. 2016. “Retinitis Pigmentosa Mutations in Bad Response to Refrigeration 2 (Brr2) Impair ATPase and Helicase Activity.” *Journal of Biological Chemistry* 291 (23): 11954–65. https://doi.org/10.1074/jbc.M115.710848.

Leeds, N. B., E. C. Small, S. L. Hiley, T. R. Hughes, and J. P. Staley. 2006. “The Splicing Factor Prp43p, a DEAH Box ATPase, Functions in Ribosome Biogenesis.” *Molecular and Cellular Biology* 26 (2): 513–22. https://doi.org/10.1128/MCB.26.2.513-522.2006.

Lesser, C. F., and C. Guthrie. 1993. “Mutational Analysis of Pre-MRNA Splicing in Saccharomyces Cerevisiae Using a Sensitive New Reporter Gene, CUP1.” *Genetics* 133 (4): 851–63.

Li, Heng, Avril Coghlan, Jue Ruan, Lachlan James Coin, Jean-Karim Hériché, Lara Osmotherly, Ruiqiang Li, et al. 2006. “TreeFam: A Curated Database of Phylogenetic Trees of Animal Gene Families.” *Nucleic Acids Research* 34 (suppl_1): D572–80. https://doi.org/10.1093/nar/gkj118.

Liu, Li, Charles C Query, and Maria M Konarska. 2007. “Opposing Classes of Prp8 Alleles Modulate the Transition between the Catalytic Steps of Pre-MRNA Splicing.” *Nature Structural & Molecular Biology* 14 (6): 519–26. https://doi.org/10.1038/nsmb1240.

Liu, Shiheng, Xueni Li, Lingdi Zhang, Jiansen Jiang, Ryan C. Hill, Yanxiang Cui, Kirk C. Hansen, Z. Hong Zhou, and Rui Zhao. 2017a. “Structure of the Yeast Spliceosomal Postcatalytic P Complex.” *Science* 358 (6368): 1278–83. https://doi.org/10.1126/science.aar3462.

———. 2017b. “Structure of the Yeast Spliceosomal Postcatalytic P Complex.” *Science* 358 (6368): 1278–83. https://doi.org/10.1126/science.aar3462.

Lorenz, Ronny, Stephan H. Bernhart, Christian Höner zu Siederdissen, Hakim Tafer, Christoph Flamm, Peter F. Stadler, and Ivo L. Hofacker. 2011. “ViennaRNA Package 2.0.” *Algorithms for Molecular Biology* 6 (1): 26. https://doi.org/10.1186/1748-7188-6-26.

Madhani, H. D., and C. Guthrie. 1994a. “Randomization-Selection Analysis of SnRNAs in Vivo: Evidence for a Tertiary Interaction in the Spliceosome.” *Genes & Development* 8 (9): 1071–86. https://doi.org/10.1101/gad.8.9.1071.

———. 1994b. “Genetic Interactions between the Yeast RNA Helicase Homolog Prp16 and Spliceosomal SnRNAs Identify Candidate Ligands for the Prp16 RNA-Dependent ATPase.” *Genetics* 137 (3): 677–87.

Madhani, Hiten D., and Christine Guthrie. 1992. “A Novel Base-Pairing Interaction between U2 and U6 SnRNAs Suggests a Mechanism for the Catalytic Activation of the Spliceosome.” *Cell* 71 (5): 803–17. https://doi.org/10.1016/0092-8674(92)90556-R.

Maeder, Corina, Alan K Kutach, and Christine Guthrie. 2009. “ATP-Dependent Unwinding of U4/U6 SnRNAs by the Brr2 Helicase Requires the C Terminus of Prp8.” *Nature Structural & Molecular Biology* 16 (1): 42–48. https://doi.org/10.1038/nsmb.1535.

Manning, Kassie S., and Thomas A. Cooper. 2017. “The Roles of RNA Processing in Translating Genotype to Phenotype.” *Nature Reviews Molecular Cell Biology* 18 (2): 102–14. https://doi.org/10.1038/nrm.2016.139.

Martin, Arnold, Susanne Schneider, and Beate Schwer. 2002. “Prp43 Is an Essential RNA-Dependent ATPase Required for Release of Lariat-Intron from the Spliceosome.” *Journal of Biological Chemistry* 277 (20): 17743–50. https://doi.org/10.1074/jbc.M200762200.

Maul-Newby, Hannah M., Angela N. Amorello, Turvi Sharma, John H. Kim, Matthew S. Modena, Beth Prichard, and Melissa S. Jurica. 2021. “A Model for DHX15 Mediated Disassembly of A-Complex Spliceosomes.” https://doi.org/10.1101/2021.09.10.459862.

Mayas, R. M., H. Maita, D. R. Semlow, and J. P. Staley. 2010. “Spliceosome Discards Intermediates via the DEAH Box ATPase Prp43p.” *Proceedings of the National Academy of Sciences* 107 (22): 10020–25. https://doi.org/10.1073/pnas.0906022107.

Mayas, Rabiah M, Hiroshi Maita, and Jonathan P Staley. 2006. “Exon Ligation Is Proofread by the DExD/H-Box ATPase Prp22p.” *Nature Structural & Molecular Biology* 13 (6): 482–90. https://doi.org/10.1038/nsmb1093.

Mayerle, Megan, and Christine Guthrie. 2018. “Brr2 Is a Splicing Fidelity Factor.” Preprint. Molecular Biology. https://doi.org/10.1101/354514.

Mazin, Pavel, Jieyi Xiong, Xiling Liu, Zheng Yan, Xiaoyu Zhang, Mingshuang Li, Liu He, et al. 2013. “Widespread Splicing Changes in Human Brain Development and Aging.” *Molecular Systems Biology* 9: 633. https://doi.org/10.1038/msb.2012.67.

Mercer, Tim R., Michael B. Clark, Stacey B. Andersen, Marion E. Brunck, Wilfried Haerty, Joanna Crawford, Ryan J. Taft, Lars K. Nielsen, Marcel E. Dinger, and John S. Mattick. 2015. “Genome-Wide Discovery of Human Splicing Branchpoints.” *Genome Research* 25 (2): 290–303. https://doi.org/10.1101/gr.182899.114.

Merrick, William C., and Graham D. Pavitt. 2018. “Protein Synthesis Initiation in Eukaryotic Cells.” *Cold Spring Harbor Perspectives in Biology* 10 (12): a033092. https://doi.org/10.1101/cshperspect.a033092.

Meyer, Markus, Mireya Plass, Jorge Pérez-Valle, Eduardo Eyras, and Josep Vilardell. 2011. “Deciphering 3′ss Selection in the Yeast Genome Reveals an RNA Thermosensor That Mediates Alternative Splicing.” *Molecular Cell* 43 (6): 1033–39. https://doi.org/10.1016/j.molcel.2011.07.030.

Moore, Melissa J., and Nick J. Proudfoot. 2009. “Pre-MRNA Processing Reaches Back to Transcription and Ahead to Translation.” *Cell* 136 (4): 688–700. https://doi.org/10.1016/j.cell.2009.02.001.

Mozaffari-Jovin, Sina, Traudy Wandersleben, Karine F. Santos, Cindy L. Will, Reinhard Lührmann, and Markus C. Wahl. 2013. “Inhibition of RNA Helicase Brr2 by the C-Terminal Tail of the Spliceosomal Protein Prp8.” *Science* 341 (6141): 80–84. https://doi.org/10.1126/science.1237515.

Nguyen, Thi Hoang Duong, Wojciech P. Galej, Xiao-chen Bai, Chris Oubridge, Andrew J. Newman, Sjors H. W. Scheres, and Kiyoshi Nagai. 2016. “Cryo-EM Structure of the Yeast U4/U6.U5 Tri-SnRNP at 3.7 Å Resolution.” *Nature* 530 (7590): 298–302. https://doi.org/10.1038/nature16940.

Niño, C. A., L. Hérissant, A. Babour, and C. Dargemont. 2013. “MRNA Nuclear Export in Yeast.” *Chemical Reviews* 113 (11): 8523–45. https://doi.org/10.1021/cr400002g.

Nues, Rob W van, and Jean D Beggs. 2001. “Functional Contacts With a Range of Splicing Proteins Suggest a Central Role for Brr2p in the Dynamic Control of the Order of Events in Spliceosomes of Saccharomyces Cerevisiae,” 17.

Ohrt, T., P. Odenwalder, J. Dannenberg, M. Prior, Z. Warkocki, J. Schmitzova, R. Karaduman, et al. 2013. “Molecular Dissection of Step 2 Catalysis of Yeast Pre-MRNA Splicing Investigated in a Purified System.” *RNA* 19 (7): 902–15. https://doi.org/10.1261/rna.039024.113.

Pan, Qun, Ofer Shai, Leo J. Lee, Brendan J. Frey, and Benjamin J. Blencowe. 2008. “Deep Surveying of Alternative Splicing Complexity in the Human Transcriptome by High-Throughput Sequencing.” *Nature Genetics* 40 (12): 1413–15. https://doi.org/10.1038/ng.259.

Pérez-Valle, Jorge, and Josep Vilardell. 2012. “Intronic Features That Determine the Selection of the 3′ Splice Site: Intronic Features That Determine the Selection of the 3′ Splice Site.” *Wiley Interdisciplinary Reviews: RNA* 3 (5): 707–17. https://doi.org/10.1002/wrna.1131.

Perriman, R. J., and M. Ares. 2007. “Rearrangement of Competing U2 RNA Helices within the Spliceosome Promotes Multiple Steps in Splicing.” *Genes & Development* 21 (7): 811–20. https://doi.org/10.1101/gad.1524307.

Query, Charles C, and Maria M Konarska. 2004. “Suppression of Multiple Substrate Mutations by Spliceosomal Prp8 Alleles Suggests Functional Correlations with Ribosomal Ambiguity Mutants.” *Molecular Cell* 14 (3): 343–54. https://doi.org/10.1016/S1097-2765(04)00217-5.

Raghunathan, Pratima L., and Christine Guthrie. 1998. “RNA Unwinding in U4/U6 SnRNPs Requires ATP Hydrolysis and the DEIH-Box Splicing Factor Brr2.” *Current Biology* 8 (15): 847–55. https://doi.org/10.1016/S0960-9822(07)00345-4.

Ragle, James Matthew, Sol Katzman, Taylor F. Akers, Sergio Barberan-Soler, and Alan M. Zahler. 2015. “Coordinated Tissue-Specific Regulation of Adjacent Alternative 3′ Splice Sites in *C. Elegans*.” *Genome Research* 25 (7): 982–94. https://doi.org/10.1101/gr.186783.114.

Ramakrishnan, V. 2002. “Ribosome Structure and the Mechanism of Translation.” *Cell* 108 (4): 557–72. https://doi.org/10.1016/S0092-8674(02)00619-0.

Reimer, Kirsten A., Claudia Mimoso, Karen Adelman, and Karla M. Neugebauer. 2020. “Rapid and Efficient Co-Transcriptional Splicing Enhances Mammalian Gene Expression.” Preprint. Molecular Biology. https://doi.org/10.1101/2020.02.11.944595.

Reitz, Diedre, Jennifer Grubb, and Douglas K. Bishop. 2019. “A Mutant Form of Dmc1 That Bypasses the Requirement for Accessory Protein Mei5-Sae3 Reveals

Independent Activities of Mei5-Sae3 and Rad51 in Dmc1 Filament Stability." *PLOS Genetics* 15 (12): e1008217. https://doi.org/10.1371/journal.pgen.1008217.

Roca, Xavier, Andrew J. Olson, Atmakuri R. Rao, Espen Enerly, Vessela N. Kristensen, Anne-Lise Børresen-Dale, Brage S. Andresen, Adrian R. Krainer, and Ravi Sachidanandam. 2008. "Features of 5'-Splice-Site Efficiency Derived from Disease-Causing Mutations and Comparative Genomics." *Genome Research* 18 (1): 77–87. https://doi.org/10.1101/gr.6859308.

Rose, Alan B. 2019. "Introns as Gene Regulators: A Brick on the Accelerator." *Frontiers in Genetics* 9: 672. https://doi.org/10.3389/fgene.2018.00672.

Ruby, S. W., T. H. Chang, and J. Abelson. 1993. "Four Yeast Spliceosomal Proteins (PRP5, PRP9, PRP11, and PRP21) Interact to Promote U2 SnRNP Binding to Pre-MRNA." *Genes & Development* 7 (10): 1909–25. https://doi.org/10.1101/gad.7.10.1909.

Sajiki, Kenichi, Mitsuko Hatanaka, Takahiro Nakamura, Kojiro Takeda, Mizuki Shimanuki, Tomoko Yoshida, Yuichiro Hanyu, Takeshi Hayashi, Yukinobu Nakaseko, and Mitsuhiro Yanagida. 2009. "Genetic Control of Cellular Quiescence in S. Pombe." *Journal of Cell Science* 122 (9): 1418–29. https://doi.org/10.1242/jcs.046466.

Schmeing, T. Martin, and V. Ramakrishnan. 2009. "What Recent Ribosome Structures Have Revealed about the Mechanism of Translation." *Nature* 461 (7268): 1234–42. https://doi.org/10.1038/nature08403.

Schneider, Susanne, Hans-Rudolf Hotz, and Beate Schwer. 2002. "Characterization of Dominant-Negative Mutants of the DEAH-Box Splicing Factors Prp22 and Prp16

*.” *Journal of Biological Chemistry* 277 (18): 15452–58. https://doi.org/10.1074/jbc.M112473200.

Schwer, B., and C. Guthrie. 1992. “A Conformational Rearrangement in the Spliceosome Is Dependent on PRP16 and ATP Hydrolysis.” *The EMBO Journal* 11 (13): 5033–39. https://doi.org/10.1002/j.1460-2075.1992.tb05610.x.

Schwer, Beate, and Christian H. Gross. 1998. “Prp22, a DExH-Box RNA Helicase, Plays Two Distinct Roles in Yeast Pre-MRNA Splicing.” *The EMBO Journal* 17 (7): 2086–94. https://doi.org/10.1093/emboj/17.7.2086.

Schwer, Beate, and Christine Guthrie. 1991. “PRP16 Is an RNA-Dependent ATPase That Interacts Transiently with the Spliceosome.” *Nature* 349 (6309): 494–99. https://doi.org/10.1038/349494a0.

Semlow, Daniel R., and Jonathan P. Staley. 2012. “Staying on Message: Ensuring Fidelity in Pre-MRNA Splicing.” *Trends in Biochemical Sciences* 37 (7): 263–73. https://doi.org/10.1016/j.tibs.2012.04.001.

Semlow, Daniel R., Mario R. Blanco, Nils G. Walter, and Jonathan P. Staley. 2016. “Spliceosomal DEAH-Box ATPases Remodel Pre-MRNA to Activate Alternative Splice Sites.” *Cell* 164 (5): 985–98. https://doi.org/10.1016/j.cell.2016.01.025.

Seol, Yeonee, Gary M. Skinner, Koen Visscher, Arnaud Buhot, and Avraham Halperin. 2007. “Stretching of Homopolymeric RNA Reveals Single-Stranded Helices and Base-Stacking.” *Physical Review Letters* 98 (15): 158103. https://doi.org/10.1103/PhysRevLett.98.158103.

Singh, Sameer, Arnaud Vanden Broeck, Linamarie Miller, Malik Chaker-Margot, and Sebastian Klinge. 2021. “Nucleolar Maturation of the Human Small Subunit

Processome." *Science* 373 (6560): eabj5338. https://doi.org/10.1126/science.abj5338.

Spingola, Marc, Leslie Grate, David Haussler, and Manuel Ares. 1999. "Genome-Wide Bioinformatic and Molecular Analysis of Introns in Saccharomyces Cerevisiae." *RNA* 5 (2): 221–34. https://doi.org/10.1017/S1355838299981682.

Staley, Jonathan P, and Christine Guthrie. 1998. "Mechanical Devices of the Spliceosome: Motors, Clocks, Springs, and Things." *Cell* 92 (3): 315–26. https://doi.org/10.1016/S0092-8674(00)80925-3.

———. 1999. "An RNA Switch at the 5′ Splice Site Requires ATP and the DEAD Box Protein Prp28p." *Molecular Cell* 3 (1): 55–64. https://doi.org/10.1016/S1097-2765(00)80174-4.

Tanaka, Naoko, and Beate Schwer. 2005. "Characterization of the NTPase, RNA-Binding, and RNA Helicase Activities of the DEAH-Box Splicing Factor Prp22." *Biochemistry* 44 (28): 9795–9803. https://doi.org/10.1021/bi050407m.

Tang, Qing, Susana Rodriguez-Santiago, Jing Wang, Jia Pu, Andrea Yuste, Varun Gupta, Alberto Moldón, Yong-Zhen Xu, and Charles C. Query. 2016. "SF3B1/Hsh155 HEAT Motif Mutations Affect Interaction with the Spliceosomal ATPase Prp5, Resulting in Altered Branch Site Selectivity in Pre-MRNA Splicing." *Genes & Development* 30 (24): 2710–23. https://doi.org/10.1101/gad.291872.116.

Toan, N. M., D. Marenduzzo, P. R. Cook, and C. Micheletti. 2006. "Depletion Effects and Loop Formation in Self-Avoiding Polymers." *Physical Review Letters* 97 (17): 178302. https://doi.org/10.1103/PhysRevLett.97.178302.

Toroney, Rebecca, Klaus H. Nielsen, and Jonathan P. Staley. 2019. "Termination of Pre-MRNA Splicing Requires That the ATPase and RNA Unwindase Prp43p Acts on

the Catalytic SnRNA U6." *Genes & Development* 33 (21–22): 1555–74. https://doi.org/10.1101/gad.328294.119.

Tsai, Rong-Tzong, Ru-Huei Fu, Fu-Lung Yeh, Chi-Kang Tseng, Yu-Chieh Lin, Yu-hsin Huang, and Soo-Chen Cheng. 2005. "Spliceosome Disassembly Catalyzed by Prp43 and Its Associated Components Ntr1 and Ntr2." *Genes & Development* 19 (24): 2991–3003. https://doi.org/10.1101/gad.1377405.

Tsai, R.-T., C.-K. Tseng, P.-J. Lee, H.-C. Chen, R.-H. Fu, K.-j. Chang, F.-L. Yeh, and S.-C. Cheng. 2007. "Dynamic Interactions of Ntr1-Ntr2 with Prp43 and with U5 Govern the Recruitment of Prp43 To Mediate Spliceosome Disassembly." *Molecular and Cellular Biology* 27 (23): 8027–37. https://doi.org/10.1128/MCB.01213-07.

Tseng, C.-K., H.-L. Liu, and S.-C. Cheng. 2011. "DEAH-Box ATPase Prp16 Has Dual Roles in Remodeling of the Spliceosome in Catalytic Steps." *RNA* 17 (1): 145–54. https://doi.org/10.1261/rna.2459611.

Vernon, Robert McCoy, Paul Andrew Chong, Brian Tsang, Tae Hun Kim, Alaji Bah, Patrick Farber, Hong Lin, and Julie Deborah Forman-Kay. 2018. "Pi-Pi Contacts Are an Overlooked Protein Feature Relevant to Phase Separation." Edited by Yibing Shan. *ELife* 7 (February): e31486. https://doi.org/10.7554/eLife.31486.

Vuong, Celine K., Douglas L. Black, and Sika Zheng. 2016. "The Neurogenetics of Alternative Splicing." *Nature Reviews Neuroscience* 17 (5): 265–81. https://doi.org/10.1038/nrn.2016.27.

Wahl, Markus C., Cindy L. Will, and Reinhard Lührmann. 2009. "The Spliceosome: Design Principles of a Dynamic RNP Machine." *Cell* 136 (4): 701–18. https://doi.org/10.1016/j.cell.2009.02.009.

Wan, Ruixue, Rui Bai, and Yigong Shi. 2019. “Molecular Choreography of Pre-MRNA Splicing by the Spliceosome.” *Current Opinion in Structural Biology*, Catalysis and Regulation ● Protein Nucleic Interactions, 59 (December): 124–33. https://doi.org/10.1016/j.sbi.2019.07.010.

Wang, Eric T., Rickard Sandberg, Shujun Luo, Irina Khrebtukova, Lu Zhang, Christine Mayr, Stephen F. Kingsmore, Gary P. Schroth, and Christopher B. Burge. 2008. “Alternative Isoform Regulation in Human Tissue Transcriptomes.” *Nature* 456 (7221): 470–76. https://doi.org/10.1038/nature07509.

Wang, Guey-Shin, and Thomas A. Cooper. 2007. “Splicing in Disease: Disruption of the Splicing Code and the Decoding Machinery.” *Nature Reviews Genetics* 8 (10): 749–61. https://doi.org/10.1038/nrg2164.

Wang, Wenyan, and Bruce A. Malcolm. 1999. “Two-Stage PCR Protocol Allowing Introduction of Multiple Mutations, Deletions and Insertions Using QuikChangeTM Site-Directed Mutagenesis.” *BioTechniques* 26 (4): 680–82. https://doi.org/10.2144/99264st03.

Wang, Yan, and Christine Guthrie. 1998. “PRP16, a DEAH-Box RNA Helicase, Is Recruited to the Spliceosome Primarily via Its Nonconserved N-Terminal Domain.” *RNA* 4 (10): 1216–29. https://doi.org/10.1017/S1355838298980992.

Wang, Yan, John D.O. Wagner, and Christine Guthrie. 1998. “The DEAH-Box Splicing Factor Prp16 Unwinds RNA Duplexes in Vitro.” *Current Biology* 8 (8): 441–51. https://doi.org/10.1016/S0960-9822(98)70178-2.

Werner, Michael S., Matthew A. Sullivan, Rohan N. Shah, Rangarajan D. Nadadur, Adrian T. Grzybowski, Vasiliy Galat, Ivan P. Moskowitz, and Alexander J. Ruthenburg. 2017. “Chromatin-Enriched LncRNAs Can Act as Cell-Type Specific

Activators of Proximal Gene Transcription." *Nature Structural & Molecular Biology* 24 (7): 596–603. https://doi.org/10.1038/nsmb.3424.

Wery, Maxime, Camille Gautier, Marc Descrimes, Mayuko Yoda, Hervé Vennin-Rendos, Valérie Migeot, Daniel Gautheret, Damien Hermand, and Antonin Morillon. 2018. "Native Elongating Transcript Sequencing Reveals Global Anti-Correlation between Sense and Antisense Nascent Transcription in Fission Yeast." *RNA* 24 (2): 196–208. https://doi.org/10.1261/rna.063446.117.

Wilkinson, Max E., Clément Charenton, and Kiyoshi Nagai. 2020. "RNA Splicing by the Spliceosome." *Annual Review of Biochemistry* 89 (1): 359–88. https://doi.org/10.1146/annurev-biochem-091719-064225.

Wilkinson, Max E., Sebastian M. Fica, Wojciech P. Galej, Christine M. Norman, Andrew J. Newman, and Kiyoshi Nagai. 2017. "Postcatalytic Spliceosome Structure Reveals Mechanism of 3′–Splice Site Selection." *Science* 358 (6368): 1283–88. https://doi.org/10.1126/science.aar3729.

Wlodaver, A. M., and J. P. Staley. 2014. "The DExD/H-Box ATPase Prp2p Destabilizes and Proofreads the Catalytic RNA Core of the Spliceosome." *RNA* 20 (3): 282–94. https://doi.org/10.1261/rna.042598.113.

Wu, Shaoping, Charles M. Romfo, Timothy W. Nilsen, and Michael R. Green. 1999. "Functional Recognition of the 3′ Splice Site AG by the Splicing Factor U2AF35." *Nature* 402 (6763): 832–35. https://doi.org/10.1038/45590.

Xu, Yong-Zhen, Catherine M. Newnhem, Sei Kameoka, Tao Huang, Maria M. Konarska, and Charles C. Query. 2004. "Prp5 Bridges U1 and U2 SnRNPs and Enables Stable U2 SnRNP Association with Intron RNA." *The EMBO Journal* 23 (2): 376–85. https://doi.org/10.1038/sj.emboj.7600050.

Xu, Yong-Zhen, and Charles C. Query. 2007. “Competition between the ATPase Prp5 and Branch Region-U2 SnRNA Pairing Modulates the Fidelity of Spliceosome Assembly.” *Molecular Cell* 28 (5): 838–49. https://doi.org/10.1016/j.molcel.2007.09.022.

Yan, Chuangye, Ruixue Wan, Rui Bai, Gaoxingyu Huang, and Yigong Shi. 2017. “Structure of a Yeast Step II Catalytically Activated Spliceosome.” *Science* 355 (6321): 149–55. https://doi.org/10.1126/science.aak9979.

Yan, D, and M Ares. 1996. “Invariant U2 RNA Sequences Bordering the Branchpoint Recognition Region Are Essential for Interaction with Yeast SF3a and SF3b Subunits.” *Molecular and Cellular Biology* 16 (3): 818–28. https://doi.org/10.1128/MCB.16.3.818.

Yang, Fei, Xiu-Ye Wang, Zhi-Min Zhang, Jia Pu, Yu-Jie Fan, Jiahai Zhou, Charles C. Query, and Yong-Zhen Xu. 2013. “Splicing Proofreading at 5′ Splice Sites by ATPase Prp28p.” *Nucleic Acids Research* 41 (8): 4660–70. https://doi.org/10.1093/nar/gkt149.

Zeng, Yi, Huilin Zeng, Benjamin J. Fair, Aiswarya Krishnamohan, Yichen Hou, Johnathon M. Hall, Alexander J. Ruthenburg, Yang I. Li, and Jonathan P. Staley. 2021. “Profiling of Nascent Lariat Intermediates Reveals Key Genetic Determinants of the Timing of Human Co-Transcriptional Splicing.” https://doi.org/10.1101/2021.10.18.464728.

Zhan, Xiechao, Chuangye Yan, Xiaofeng Zhang, Jianlin Lei, and Yigong Shi. 2018. “Structure of a Human Catalytic Step I Spliceosome.” *Science* 359 (6375): 537–45. https://doi.org/10.1126/science.aar6401.

Zhang, Xiaochang, Ming Hui Chen, Xuebing Wu, Andrew Kodani, Jean Fan, Ryan Doan, Manabu Ozawa, et al. 2016. “Cell-Type-Specific Alternative Splicing Governs Cell Fate in the Developing Cerebral Cortex.” *Cell* 166 (5): 1147-1162.e15. https://doi.org/10.1016/j.cell.2016.07.025.

Zhang, Xiaofeng, Xiechao Zhan, Chuangye Yan, Wenyu Zhang, Dongliang Liu, Jianlin Lei, and Yigong Shi. 2019. “Structures of the Human Spliceosomes before and after Release of the Ligated Exon.” *Cell Research* 29 (4): 274–85. https://doi.org/10.1038/s41422-019-0143-x.

Zhao, Chen, Deepti L. Bellur, Shasha Lu, Feng Zhao, Michael A. Grassi, Sara J. Bowne, Lori S. Sullivan, et al. 2009. “Autosomal-Dominant Retinitis Pigmentosa Caused by a Mutation in SNRNP200, a Gene Required for Unwinding of U4/U6 SnRNAs.” *The American Journal of Human Genetics* 85 (5): 617–27. https://doi.org/10.1016/j.ajhg.2009.09.020.

Zhou, Zhaolan, and Robin Reed. 1998. “Human Homologs of Yeast Prp16 and Prp17 Reveal Conservation of the Mechanism for Catalytic Step II of Pre-MRNA Splicing.” *The EMBO Journal* 17 (7): 2095–2106. https://doi.org/10.1093/emboj/17.7.2095.