THE UNIVERSITY OF CHICAGO

PHYSIOLOGICAL FUNCTION AND MECHANISM OF PRP22 IN REGULATING 3’ SPLICE SITE SELECTION

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

CODY ANTHONY HERNANDEZ

CHICAGO, ILLINOIS

MARCH 2022

Table of Contents

LIST OF FIGURES

LIST OF TABLES

This thesis is dedicated to all the students that deserved to be here but never got the opportunity

ACKNOWLEDGMENTS

Hernandez is the 15th most common last name in the United States and yet, I am the first in my family to be called Dr. Hernandez. For me, the Ph.D. has been difficult because of things that had little to nothing to do with science. I'd argue science was the easiest part. For the sake of science, I hope the future for marginalized groups will allow for more science and less survival in graduate school. Science needs it and we want it.

Thank you to my family and friends and specifically my Mom and Dad. Mom, you always told me to stand up for what I believe in and Dad, you taught me what hard work is. The sacrifices you've made for me are unmeasurable. I'm hoping I can pay that forward now. I love you both.

I'd like to thank my mentor, Dr. Jon Staley, for allowing me the freedom to pursue projects outside of the lab as well as within our lab. Jon, your emphasis on the presentation of science, written or an oral talk, is a tool I will carry with me for the rest of my life, and I appreciate the countless hours we spent honing that skill. Thank you to my committee for supporting me, especially at the end of my Ph.D. and during a pandemic. I'd also like to thank members of the Staley lab for their support. Thank you to the 8th floor for the memories, training, and countless hours of volleyball and soccer tennis. To my second lab, the Ruthen-lab, thank you for always being my biggest cheerleaders. Thank you!

Shoutout to the four most important people in my grad-school training, Dr.Yuen-Ling Chan, Dr. Nancy Schwartz, Dr. Mat Perez-Neut, and Amanda Keplinger. Ling, I'm forever indebted to you for the support, training, encouragement, and countless stories and techniques that you taught me. Nancy never has there been a person in this world that believes in me more than you. I felt that in every conversation we had, every presentation, and every conference we attended together. To that end, we (Dr. Perez, Dr. Schwartz, and Dr. Hernandez) had a hell of a

run, didn't we? Mat, I look back on the ideas and conversations we had and realize there is no other time in grad school have I felt I could be myself, scientifically and otherwise, and be appreciated for it. That's bois. To my partner, Amanda Keplinger, you put me on your back and carried me to the finish line of this dissertation. In addition to being an amazing partner, you are the reason I was able to finish. We did it!

Shoutout to my HHMI Gilliam Family, TAE, GRIT, SACNAS, and all my best friends (Dequantarius Speed, Eric McLean, Grace Schulz, Jimmy Elias, Josephine Tang, Kate Cavanaugh, Matthew McDonough) for always being there. Cheers to many more years with y'all!

There isn't enough space to appropriately thank all the people that have helped me to get here so I created a document with the pictures and memories thanking everyone.

https://drive.google.com/drive/folders/1oyVYFcC-CLcQr5fTMyHGjCMOzUVtRb4Q?usp=sharing

SUMMARY

In a key step in pre-mRNA splicing, the spliceosome recognizes the intronic 3' splice site (3'SS). Alternative usage of 3'SSs yields different mRNA isoforms that are essential for various cellular processes. How 3'SS selection is regulated to control gene expression is currently under investigation. In this study, I investigated the role of RNA secondary structure and the constitutive splicing factor Prp22, an RNA-dependent DExH ATPase, in 3' splice site choice. Previously, I provided evidence that Prp22 rejects suboptimal 3' splice sites by pulling on the splicing substrate. Here, I hypothesized that this pulling activity would also disrupt RNA secondary structure and thereby impact 3'SS selection, because RNA secondary structure can either repress a 3'SS by sequestering the site or activate a distal 3'SS by juxtaposing the 3'SS with the catalytic core of the spliceosome. To test our hypothesis in budding yeast, I assayed the impact of Prp22 on 3'SS selection in the context of RNA secondary structure in two native substrates. Indeed, Prp22p activated a 3' splice site within a stem loop and repressed a 3' splice site downstream of a stem loop. Mutational perturbation of the stems established that Prp22 destabilized the stems. Importantly, I observed that Prp22, through an impact on secondary structure, modulated alternative 3' splice site choice. Our findings indicate that wildtype Prp22 unwinds RNA secondary structure in the region of a 3' splice site and thereby activates a sequestered 3'SS and represses a distal 3'SS. This study identifies Prp22 as a potential target in the regulation of 3' splice site selection.

CHAPTER I

OVERVIEW OF PRE-mRNA SPLICING AND THE ROLES OF SPLICEOSOMAL DExD/H - BOX ATPASES DURING SPLICING

Brief History

Pre-mRNA splicing was first discovered in 1977 by the labs of Phil Sharp and Richard Roberts using biochemistry and electron microscopy[1,2].At the time, it was thought that one gene coded for a single protein, a dogma that would be shattered by the discoveries of Sharp and Roberts labs. Arguably one of the most important discoveries that seeded our understanding of splicing as I know it today came from a bold prediction by Joan Steitz. Two years after Sharp and Roberts discovery, Steitz and Michael Lerner discovered small ribonucleoprotein protein particles they termed snRNPs (pronounced "snurps") and suggested the particles were involved in the newly discovered phenomenon of splicing[3,4] . Shortly after, Steitz, Lerner, and other colleagues in the Steitz lab published an article titled "Are snRNPs involved in splicing?," which hypothesized the likely role of multiple snurps in splicing[5]. In 1981, Steitz lab would go on to show U1's role in binding the 5'SS using drosophila extract and several other biochemical experiments showed specific binding of the U1 snRNP to the 5'SS in other crude extracts[6]. A breakthrough in the field that allowed for functional characterization of the splicing process came when Hernandez and Keller published a protocol for studying splicing *in vitro*[7]. Indeed, in 1985 Brody and Abelson (yeast) and Grabowski and colleagues (human) used in vitro splicing assays and biochemical fractionation to show the 5 snRNPs work together to coordinate splicing[8,9].

Genetics and genetic screens in budding yeast are largely responsible for identification and characterization of various proteins and RNA elements involved in splicing. The groundwork for yeast genetics on splicing was laid by Leland Hartwell in 1967, before splicing was even discovered. Leland Hartwell identified 400 high temperature sensitive (*ts*) mutants, of which some he classified as being defective in producing mRNA[10]. Some of the more notable

mutants he identified would later be characterized by John Abelson's lab using newly developed *in vitro* splicing assays and come to be known as pre-mRNA processing (Prp) factors Prp2, Prp5, and Prp8[11,12]. Later work by Christine Guthrie's lab employed cold sensitive (*cs*) mutant screens as opposed to the previously employed *ts* screens. She identified 340 cold sensitive mutants that led to 12 genes being isolated. Of those 12 genes, four were helicases: Prp16, Prp22, Prp28, and Brr2. Together, the work of Guthrie and Abelson's labs provided a biochemical and genetic framework for studying splicing that would prove indispensable for the field of pre-mRNA splicing.

Chemistry of Splicing

In pre-mRNA splicing, two exons are ligated together concomitant with the removal of an intervening non-coding intron. This process is facilitated by a large ribonucleoprotein (RNP) complex, known as the spliceosome, which consists of 5 small nuclear RNAs (snRNAs) and almost 100 proteins[13]. During the first step of splicing, the 2' hydroxyl of the branch site adenosine attacks the phosphate of the 5' splice site (5'SS) to form a branched lariat intermediate (Figure 1.1)[14–18]. This process is the first chemical step of splicing and will be referred to herein as the "branching step" or the "first step." Following branching, the catalytic core is rearranged and the 3' hydroxyl of the 5' exon then attacks the phosphate of the 3' splice site (3'SS) to form messenger RNA (mRNA)[14,16–18]. This process is the second chemical step of splicing and will be referred to herein as "exon ligation" or the "second step." Sites of splicing chemistry are defined by short consensus sequences that are recognized by snRNA[16–22]. The mechanism for catalysis between these intronic elements were first proposed by Steitz and Steitz and referred to as the two-metal hypothesis[23]. The hypothesis posited that two-metal ions in the catalytic core of the spliceosome stabilize the transition states of the transesterification reaction[23]. This hypothesis

would later be supported by a series of biochemical analysis from the Staley and Piccirilli labs which showed both metals are coordinated by RNA to catalyze splicing between the intronic elements[15].

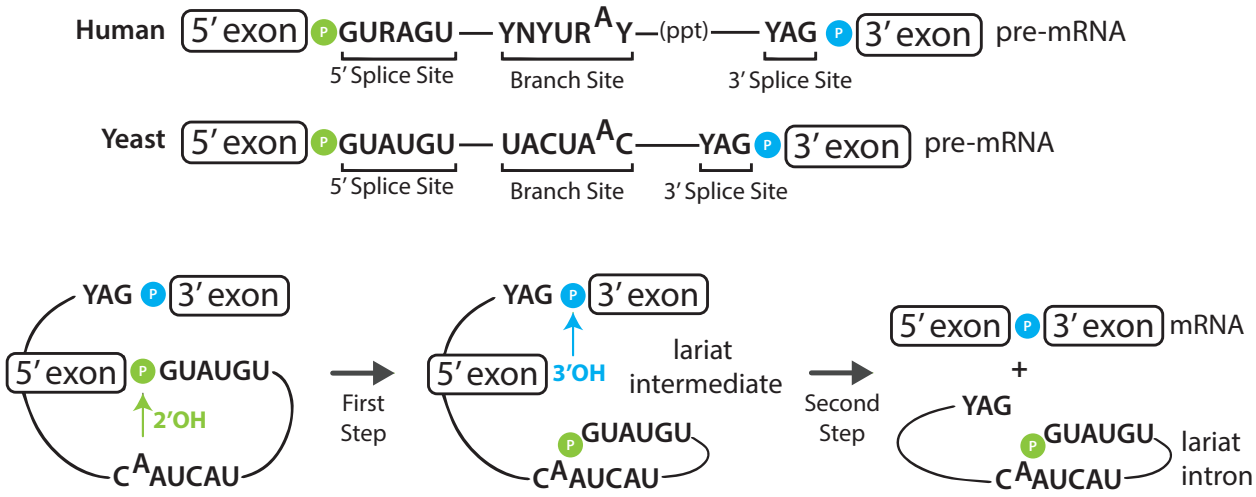


Figure 1. 1. Splicing occurs in two different transesterification steps between three intronic elements known as splice sites. Sites of splicing chemistry are defined by short consensus sequences[16–22]. The consensus sequences for yeast and human splice sites are shown. During the first step of splicing, the 2' hydroxyl of the branch site adenosine attacks the phosphate of the 5' splice site to form a branched lariat intermediate[14–18]. Following branching, the catalytic core is rearranged and the 3' hydroxyl of the 5' exon then attacks the phosphate of the 3' splice site to form messenger RNA (mRNA)[14,16–18]. Y is any pyrimidine, R is any purine, N is any nucleotide, ppt stands for poly pyrimidine track, the P in the blue and green circles stands for phosphate. This Figure was inspired by Staley and Guthrie 1998[24].

Features of the SF2 helicases

During splicing, the spliceosome undergoes several large scale remodeling events, rearrangements, and exchanges of protein and snRNA (small nuclear RNA) to convert a single pre-mRNA molecule to mRNA. Throughout the conversion of pre-mRNA to mRNA, the spliceosome has several discrete complexes that form. The formation of these splicing complexes can occur by the binding of a pre-formed snRNA-protein complex known as a snRNP (small nuclear ribonucleoproteins, Figure 1.2) or because of a remodeling event driven mainly by the SF2 helicase family. Generally, the role of SF2 helicases is to antagonize RNA-RNA or RNA-protein interactions by binding a target RNA substrate and displacing the protein or RNA bound to the target substrate via ATP-hydrolysis. In the context of the spliceosome, there are eight conserved SF2 Helicases between Saccharomyces cerevisiae (referred to herein as "budding yeast" or "yeast") and metazoans (Figure 1.2) and at least 5 additional helicases specific to metazoans[25–28].

The conserved features between DEAD-Box, DEAH-box, and Ski2-like helicases are two RecA domains of which, there are thirteen conserved motifs within the two domains. Motifs Ia, Ib, Ic, IV, IVa, V, and Vb bind the RNA, motifs Q, I, II, and VI bind the NTP and facilitate hydrolysis, and motif III and Va communicate the binding state to the NTP hydrolysis domains[29–31]. Differences between these helicases are often found in their variable N terminus which is believed to confer substrate specificity during splicing. Of the eight conserved SF2 helicases involved in splicing, three are DEAD-box helicases (Prp5, Sub2, and Prp28) involved in spliceosome assembly, four are DEAH-box helicases (Prp2, Prp16, Prp22, and Prp43) involved in activation of the spliceosome for catalysis and disassembly, and one Ski2-like helicases is omnipresent throughout the entirety of the splicing cycle with a major role early in splicing to

poise the spliceosome for catalysis, Brr2. Key differences between the DEAD-box family and the DEAH-box and Ski2-like family is the difference in processive unwinding and substrate binding requirements. For example, DEAH-Box helicases bind single stranded regions of RNA and translocate processively 3' to 5' along RNA whereas DEAD-box helicases can bind duplex or single stranded RNA regions and either promote dsRNA annealing (independent of ATP hydrolysis) or through one round of ATP hydrolysis, separates short duplexes.

A key feature in DEAH-box helicases that separates them from Ski2 like helicases is the presence of the OB-fold in the C-terminal domain of the helicase. For some DEAH-box proteins, it's thought that the OB-fold is the cite of stimulation by G-patch proteins. During activation of the spliceosome, specifically B_{act}, Prp2 interacts with the G-patch protein Spp2 through an interaction between Spp2 and Prp2's OB-fold and interactions with the RES complex[32,33]. G-patch proteins are known to activate the SF2 helicases and Spp2 has previously been shown to stimulate the ATPase activity of Prp2 in vitro[34,35] and in the case of activation of the spliceosome, Spp2 is required for the first step of splicing[36,37]. Cryo-EM structures have not confidently captured the interaction between Prp2 and Spp2 at high resolution but the compliment of modeling, biochemistry, and a crystal structure between Prp2 and Spp2 support the role of Spp2 in recruiting and activating Prp2 to the spliceosome prior to the first step of splicing[33,35,38,39]. A notable reason for the lack of density to assign Spp2 to its interacting partner Prp2 is the fact that Prp2 is located at the periphery of the spliceosome and density is often more diffused. Indeed, this is a common theme for helicases in the cryo-EM structures. A unifying theory for the diffuse signal of these helicases is the fact that they act from a distance to antagonize interactions within the catalytic core and therefore their density cannot always be easily assigned[40].

In summary, SF2 helicases are essential workhorses for the progression of the splicing cycle that exist in three different sub-families: DEAD-box, DEAH-box, and Ski2-like helicases. DEAD-box ATPases are not processive and function early in splicing whereas DEAH-box and Ski2-like helicases are processive and function at later stages in splicing. Additionally, DEAH-box helicases, can be recruited to the spliceosome and stimulated by interactions between the the OB-fold of the helicase and an interacting G-patch protein.

Spliceosome Assembly

The spliceosome begins assembling onto the pre-mRNA substrate through the interaction of U1 snRNA with the 5'SS followed by assembly of BBP-MUD2 (SF1 and U2AF in humans, respectively)[41,42] onto the branchpoint[43]. The interaction of these proteins and U1 with the substrate is known as the E-complex[44–46]. Through the action of Sub2 and Prp5 , both DEAD-box ATPases, the U2 snRNP binds the branch site via displacement of BP-MUD2 by Sub2[47–49] and confirmational rearrangements of U2 snRNA by Prp5[50–52]. Throughout the action of Sub2 and Prp5, U1 snRNP remains bound to the 5'SS. This complex is known as the A-complex[41,53,54]. Following A-complex formation is the joining of the pre-assembled U4/U6·U5 tri-snRNP to form the fully assembled pre-B complex[55,56]. In summary, two DEAD-box helicases, Prp5 and Sub2 are directly involved in spliceosome assembly.

Spliceosome Activation

To activate the assembled spliceosome, several rearrangements, namely the displacement of the U1 snRNP from the 5'ss by Prp28 (DEAD-box) to form the B-complex[57]. Prp28-dependent rearrangement leads to binding of the 5'SS to U6's ACAGAGA-box and the displacement of U4

from U6 by the Ski-2 like helicase, BRR2, which leads to to form the activated B complex (B_{act})[33,39,58,59]. In a final rearrangement before the branching reaction, Prp2(DEAH-box) displaces Hsh155 from the branch-site[33,39,60,61], base pairing between U6 and U2 snRNA juxtaposes the 5'SS and the branch site, and the catalytic core is formed. This complex is known as B*[36,55,62]. In summary, one DEAD-Box helicase (Prp28), one Ski2-like helicases (Brr2), and one DEAH-Box helicase (Prp2) is directly involved in spliceosome activation.

Spliceosome Catalysis and Disassembly

Following the liberation of the branch site adenosine by Prp2 is the first chemical reaction which results in C-complex which harbors the first-step products, a lariat intermediate and cleaved 5' exon[63,64]. C-complex is driven to C* (exon ligation confirmation) by Prp16 (DEAH-box) which promotes the following rearrangements within the catalytic core for docking of the 3'SS. Prp16 binds downstream of the branch site, translocates towards but not through its target[40], toggling of stem IIc confirmation of U2 snRNA to stem IIa[52,65–68], rotation of the branch helix by 75°[69–71] , and a rearrangement of U6's interaction with the 5'SS to form a non-canonical interaction between the U of the 5'SS (G**U**AUGU) and A51 of the ACAG**A**GA-box which creates a landing pad for G of the 3'SS YA**G** to π–π stack with A51 of the ACAG**A**GA-box[64,72–74] . Following the action of Prp16 is the joining of the second step factors Slu7 and Prp18 which subsequently recruit Prp22 (DEAH-box). Following association of the second step factors is the docking of the 3'SS. An optimal 3'ss docking event occurs through interactions between the 5'SS G+1 and G-1 of the 3'SS, the branch site adenosine and the A-2 of the 3'SS and a base stacking interaction between A51 and the 3'SS (Figure 1.3)[72–75]. The result of a successful docking event is catalysis which produces mRNA and excised lariat. This complex is known as

the P-complex[65]. The mRNA is then released through the action of Prp22 disrupting the interaction between U5 loop1 and the mRNA exons[76–79] . Following the action of Prp22, the residual complex harbors the snRNA and excised lariat. This complex is known as the intron lariat spliceosome (ILS) complex[80]. The ILS complex is disassembled after Ntr1 recruits Prp43 (DEAH-box) to act on the 3' end of U6[76,80–83].

In summary, Prp2 promotes the 1st step of catalysis followed by Prp16 evacuating the catalytic core to make room for the 3'SS. The action of Prp16 also creates a landing pad for the 3'SS to dock. In C*, the second step factors Slu7, Prp18, and Prp22 are recruited, and the resulting mRNA is released by Prp22. Lastly, Prp43 disassembles the spliceosome for snRNP recycling through recruitment and stimulation by Ntr1.

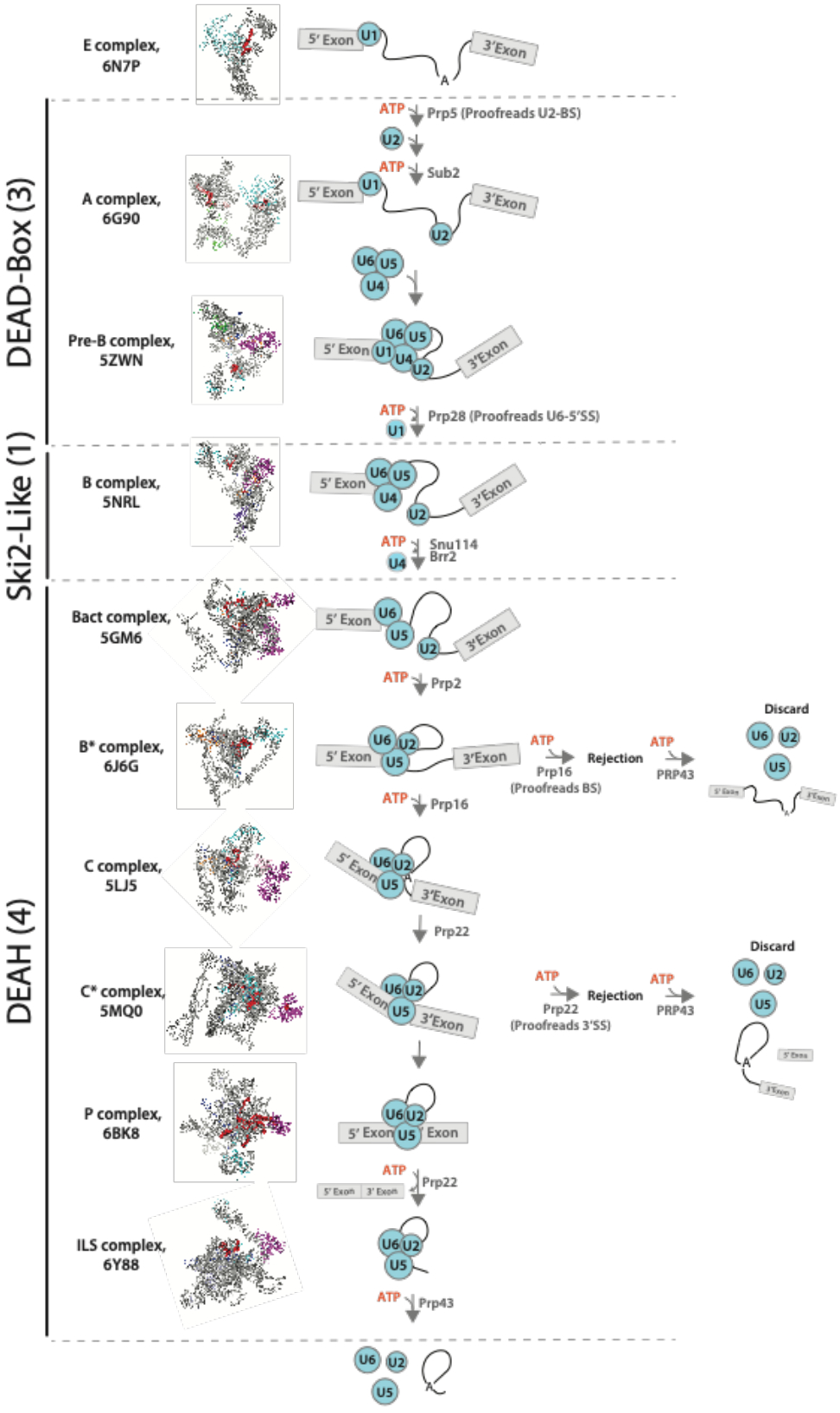


Figure 1. 2. The splicing pathway and the proofreaders. Helicases are bound at the periphery

Figure 1.2 (continued) of the spliceosome and function to promote the splicing pathway in addition to antagonizing on-pathway functions through proofreading. Cartoons depicting the snRNA's during splicing are shown to the right of the cryo-EM structures. Helicases are depicted on reaction arrows according to their action at that step. All helicases are shown in magenta and are oriented in a manner to highlight their position at the periphery of the spliceosome. The RNA substrates are shown as surface density colored red. The snRNA's are colored in each complex. Images were rendered using a paid subscription of pymol for each of the following complexes. E-complex (6N7P)[84], A-complex (6G90)[85], PreB-complex (5ZWN)[55], B-complex (5NRL)[86], Bact-complex (5GM6)[39], B*-complex (6J6G)[87], C-complex (5LJ5)[88], C*-complex (5MQ0)[69], P-complex (6BK8)[73] ,ILS-complex (5Y88)[76].

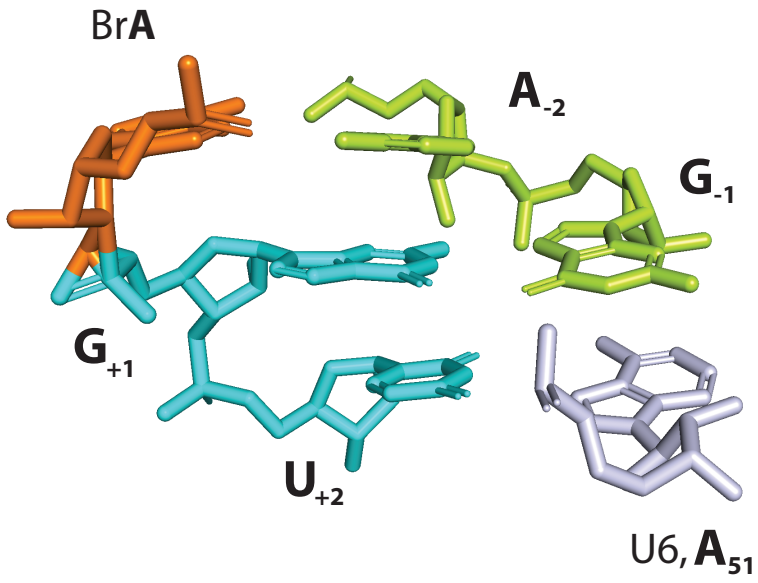


Figure 1. 3 The catalytic core of the second step of splicing. Shown is the catalytic core from the cryo-EM structure 6KB8[73]. All other proteins, snRNA and pre-mRNA substrate have been excluded for clarity.

Prp22 and 3'SS selection

Slu7 and Prp18 are required for recruitment of Prp22 to the spliceosome during the second step of splicing[89]. These proteins are required for Prp22 to be recruited to the spliceosome and have been shown to functionally interact with U5 snRNA[90,91]. Prp22 is required to promote exon ligation and in its absence, splicing can only occur at very short distances from the branch site (Figure 1.4)[78]. After exon ligation, Prp22 releases the mRNA from the spliceosome by translocating 3' to 5'[78,79]. This is the final step of splicing and renders the reaction irreversible.

Prp22 contains an essential N-terminal domain that is suggested to be required for interacting with the spliceosome as well as a C-terminal domain that is required for ATP hydrolysis and RNA binding[92]. Expressing either of these independent of the other is not sufficient to promote growth in budding yeast. However, expressing both fragments in trans rescues the otherwise lethal phenotype. Further, in vitro splicing analysis has shown that ATP hydrolysis mutants are unable to release mRNA but can promote exon ligation[92]. In contrast, mutants lacking parts of the N-terminal domain were limited in splicing but could release mRNA. This suggests Prp22 possesses a ATP-dependent and ATP-independent function[92]. Work from our lab characterizing the ATP-dependent role of Prp22 has shown that a sub-optimal 3'ss (UAc) substrate is rejected by Prp22 in a ATP-dependent manner[93]. More recently, our lab used *in vitro* splicing assays to reveal that Prp22 promotes alternative 3' splice site selection[40].

The importance of 3'SS selection is exemplified by the more than 2,000 humans introns that have a pair of 3'SS that are contiguous; these pairs are known as NAGNAGs (NAG corresponds to a single 3'SS; N correspond to any nucleotide)[94]. The differential use of these NAGNAGs (i.e., switching between the upstream and downstream NAGs) is regulated during

development and in a tissue-specific manner. Strong regulation occurs in 896 of these NAGNAGs, and alternative NAGNAG selection has been shown to alter the function of individual proteins. Further, 28% of SNPs located in NAGNAG sequences are found in disease-causing genes[95]. Of all alternative splicing events that maintain a reading frame, NAGNAG alternative splicing accounts for 20% of events, and alternative selection of 3'SSs separated by more than three nucleotides accounts for an additional 11%[94]. Yet, the mechanisms for regulating alternative 3'SS selection remain under investigation.

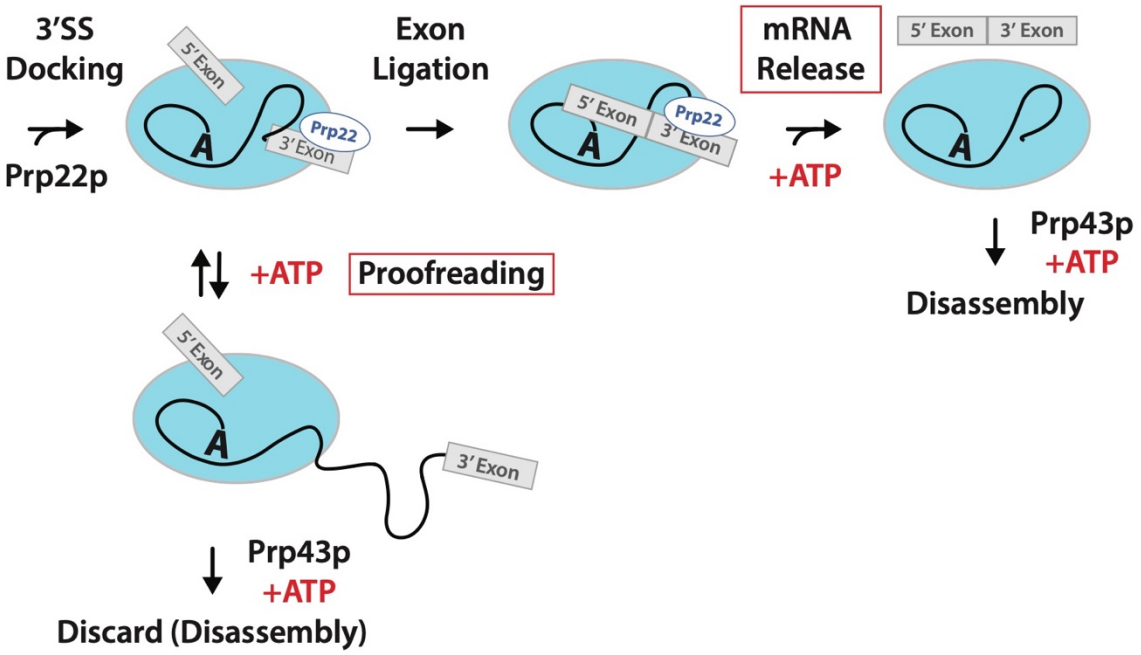


Figure 1.4. Known roles for Prp22 during pre-mRNA splicing. The large blue oval's represents the spliceosome and functions with red-box's around them are known Prp22p ATP-dependent remodeling steps. See text for more detail.

CHAPTER 2

THE DExH ATPASE PRP22 REGULATES THE ACCESIBILITY AND PROXIMITY OF 3' SPLICE SITES

Introduction

The 3'SS is defined by a dinucleotide AG and resides downstream of the branch point. An optimal 3'SS has a pyrimidine (denoted as Y) immediately upstream of the dinucleotide AG (**Y**AG), and in humans, the YAG is preceded by a polypyrimidine tract (PPT)[44,96]; together these elements recruit U2 auxiliary factor (U2AF), a heterodimer composed of U2AF2 and U2AF1[44,96,97]. U2AF1 binds the dinucleotide AG, and U2AF2 binds the PPT, helping to recruit SF1, which binds to the branch site[43,98–100]. SF1 and U2AF depart from the splicing substrate after the U2 snRNP is recruited to the branch point, so the 3'SS must be recognized again before exon ligation[41,43]. Recognition of the 3' splice site at this stage is dependent on the branching reaction. Branching links the conserved A of the branch site to the conserved GU of the 5' splice site, which together constitute the 3' splice site bindings site[72–75]. Because there is only a single catalytic core, after the branching reaction this linkage must be repositioned to allow for docking of the scissile 3'SS phosphate and the subsequent nucleophilic attack by the 5' exon[15].

The repositioning of the substrate within the catalytic core is carried out by the DEAH-box ATPase Prp16p, which pulls on the substrate to cause rearrangements within the catalytic core[40,69–71,101,102]. Subsequently, the binding of second step factors Slu7, Prp18, and Prp22 promote docking of the AG to the catalytic core[65,89,103,104]. The 3'SS docks with the branched linkage via two non-Watson-Crick base-pairing interactions. The 3'SS guanosine (YA**G**, G_{-1}) pairs with the 5'SS guanosine (G_{+1}), and the 3'SS adenosine (Y**A**G, A_{-2}) pairs with the branch point adenosine[72–75]. Additionally, U_{+2} of the 5'SS base pairs with a highly conserved adenosine in U6 snRNA leading to a base stacking interaction between U6 snRNA and G_{-1} of the 3'SS, an interaction that further stabilizes the interaction of the 3'SS with the catalytic core and positions G_{-1} for attack by the 5' exon[64,72–74]. The nucleotide that precedes the dinucleotide AG (**Y**$_{-3}$AG) of the 3'SS also appears to play a role in stabilizing a docked AG; the Y_{-3} position forms a non-

Watson-Crick interaction with a highly conserved glutamine in the alpha finger of Prp8[72]. Modeling substitutions of the pyrimidine for a purine suggest an adenosine would abolish the interaction with the alpha finger and a guanosine would sterically clash with the glutamine. Across the transcriptome of eukaryotes, YAG 3'SSs account for 95% of splicing events, while AAG 3' SSs account most of the remaining 5%[103].

In addition to the identity of the PPT and 3'SS, the proximity of an AG to the branch site also plays a critical role in determining 3'SS selection[105–107]. Previous studies in yeast have shown that 3'SS selection predominantly occurs within a window of 10 to 45 nts from the branshsite[103,108]. However, as is the case in humans, the spliceosome selects 3'SSs outside the 45 nt long window and, in some cases, selects distal AGs despite optimal AGs within the optimal 10-45 nt long window[109,110]. Selection of a distal 3'SS by the spliceosome is mediated by RNA secondary structure that brings the distal 3'SS into proximity with the branch site, and the same secondary structure can sequester and repress a more proximal 3'SS[108,111,112].

Notably, the role of RNA secondary structure in splicing extends beyond 3'SS selection. For example, in yeast, a secondary structure between the branshsite and the 5'SS has been shown to be essential for promoting juxtaposition of the branch site and 5'SS in vivo[113]. Conversely, RNA structure can occlude intronic elements, such as the 5'SS[114]. At the same time, RNA structure can be the target of regulatory factors such as the adenosine deaminases acting on RNA (ADAR) enzyme, which through RNA modification destabilizes RNA duplexes, in some cases unveiling functional splice sites[115]. Similarly, alternative splicing of *Tau*, a microtubule associated protein, is regulated by the destabilization of a stem loop by DDX5 to promote U1 binding[116]. Interestingly, the DEAD-box helicase DDX5 and its paralog DDX17 are thought to function as "master orchestrators" of splicing programs during cellular differentiation in humans

and have been proposed to influence splicing broadly by unwinding RNA secondary structure[117,118]. The discovery of DDX5-mediated unwinding of the RNA stem-loop of intron 10 in *Tau* was the first example of a helicase that regulates alternative splicing in humans.

Secondary structure in the pre-mRNA substrate and the spliceosome have been documented targets of DEAD and DExH-box helicases. In total, eight RNA helicases of the SF2 superfamily that are conserved from budding yeast to humans function during splicing, and six of the eight helicases promote the fidelity of splice site choice, in addition to driving the spliceosome forward along the pathway to mRNA forrmation[31,119,120]. One of the six helicases, Prp22p, promotes exon ligation in an ATP-independent manner[78] and in an ATP-dependent manner promotes 3'SS fidelity before exon ligation and mRNA release after exon ligation[40,70,72–74,79,93,121–124].

In promoting 3'SS fidelity, evidence implies that pulling by Prp22 on the splicing substrate can trigger rejection of a docked 3'SS in a manner that competes with exon ligation in a kinetic proofreading framework[93,119]. The mechanism by which Prp22 selectively rejects a suboptimal 3'SS remains under investigation, but two models have been proposed – the timer model and the sensor model[119,125,126]. In the timer model, Prp22 pulls on the substrate at a characteristic time; if mRNA has formed, then Prp22 releases the mRNA, but if mRNA has not yet formed, then Prp22 rejects the 3' splice site; in this model, suboptimal 3'SSs are distinguished as splicing slower. In the sensor model, Prp22 pulls on the substrate, before exon ligation, because docking of a suboptimal 3'SS activates Prp22 in some way, such as through conformational changes in the catalytic core or by virtue of less stable binding of the 3'SS to the catalytic core; in a similar manner, after exon ligation, the cleaved 3'SS would activate Prp22 for mRNA release. In either model, pulling before exon ligation can either lead to resampling of the

intron by the spliceosome for an optimal 3'SS or discard of the substrate by Prp43p in a ATP-dependent manner[127,128]. A similar kinetic proofreading framework has been invoked at other stages of splicing, as well as outside of splicing, such as during signal recognition by the signal recognition particle (SRP) during co-translational protein targeting[119,125].

In vitro splicing assays, crosslinking, and cryo-EM have all independently suggested that Prp22 binds at the periphery of the spliceosome and acts on its target from a distance, through an interaction with the 3' exon[40,72,77]. For example, our previous work established evidence that Prp22 binds to the 3' exon and translocates on the 3' exon toward but not through its targets implying Prp22p pulls the RNA substrate until tension builds and propagates through the substrate to either antagonize interactions between the 3'SS and the catalytic core or between the 5' exon of the mRNA and the spliceosome[40]. These two activities, pulling to proofread versus to release mRNA, differ only because the connectivity of the substrate has changes during exon ligation – before exon ligation, the downstream exon is linked to the 3' splice site but after exon ligation the downstream exon is linked to the 5' exon[93]. Herein I will refer to Prp22 as a pullase.

In addition to the role of Prp22p in fidelity and mRNA release, an additional function of Prp22p in promoting alternative 3'SS selection has been suggested[40]. Specifically, *in vitro* splicing assays of a chemically modified substrate with two alternative 3'SSs revealed that Prp22 can favor, in an ATP-dependent manner, selection of the downstream 3'SS, relative to the upstream 3'SS. While these results are intriguing, it has remained to be determined whether Prp22p impacts alternative splice site selection on physiological substrates *in vivo* and if so, then how.

In this study, I used budding yeast to investigate the role of Prp22 in alternative 3'SS selection in vivo. I first established that Prp22 impacts 3'SS selection of native substrates *in vivo*.

Then, I showed that Prp22 independently activates a 3'SS sequestered by a stem loop and represses a distal 3'SSs downstream of a stem loop. Finally, I found evidence that Prp22p achieves these functions by destabilizing RNA secondary structure. Together, the findings from this study and previous studies point toward a model for these activities in which Prp22 binds to a downstream exon, pulls the RNA substrate, and then disrupts RNA secondary structure, thereby impacting 3' splice site choice[40,69,72–74,93,121,123,129,130]. Thus, our results imply that Prp22 pulling can result not only in the release of mRNA and the rejection of a suboptimal 3'SS but also in the unwinding of RNA secondary structure during the search for a 3'SS. This new function of Prp22p has implications for the regulation of alternative 3'SS choice.

Results

Prp22 affects 3'SS selection in *APE2* and *DMC1 pre-mRNAs*

To test if Prp22p can affect alternative splicing of physiological substrates in vivo, I assayed splicing of substrates with multiple 3'SSs downstream of the branch point in the background of either wildtype *PRP22* or mutant *prp22* budding yeast cells. Specifically, I tested the splicing of *DMC1* and *APE2* transcripts[108]. *DMC1* has two 3'SSs: a proximal AAG (AG1) located 16 nts downstream of the branch point and a distal UAG (AG2) located 28 nts from the branch (Figure 2.1a). *APE2* has three 3'SSs: a proximal AAG (AG1), 30nt downstream of the branch; a medial AAG (AG2), 37 nts downstream of the branch; and a distal CAG (AG3), 55 nts from the branch point (Figure 2.1b)[108]. The endogenous *APE2* and *DMC1* substrates were assayed in the background of wildtype and *prp22-G810A* mutant yeast cells[93,131]. Although the unwinding, RNA binding, and NTPase activities of *prp22-G810A* (herein referred to as mutant or *prp22*) are not defined, this mutation i) confers a cold-sensitive defect characteristic of RNA unwinding mutants[131], ii) suppresses a suboptimal gAG 3'SS defect *in vivo*, indicating a loss in ATP-dependent fidelity[93], and iii) phenocopies the depletion of ATP in *in vitro* fidelity assays[93], altogether implying a defect in the ATP-dependent remodeling of the splicing substrate.

I found that Prp22p does affect 3'SS selection for both *DMC1* and *APE2* (Figure 2.1.). I isolated total RNA from either *PRP22* or mutant *prp22* yeast cells and used RT-PCR to assay the splicing status of *DMC1* and *APE2* transcripts. With *DMC1*, the spliceosome selected preferentially the proximal AG (AG1) in *PRP22* yeast cells and the distal AG (AG2) in mutant *prp22* yeast cells (Figure2.1a); specifically, the spliceosome selected AG1, relative to AG2, 5.4-fold more frequently in wildtype yeast cells than in mutant cells. With *APE2*, the spliceosome preferentially selected the proximal AG2 in wildtype yeast cells and the distal AG3 in mutant cells (Figure 2.1b); specifically, the spliceosome selected AG2, relative to AG3, 3-fold more

frequently in wildtype yeast cells than in mutant cells. In summary, with both *DMC1* and *APE2*, the spliceosome selected the 3'SS more proximal to the branchpoint in wildtype yeast cells and the distal 3'SS in *prp22* mutant yeast cells. I conclude that Prp22p influences alternative 3'SS selection *in vivo*.

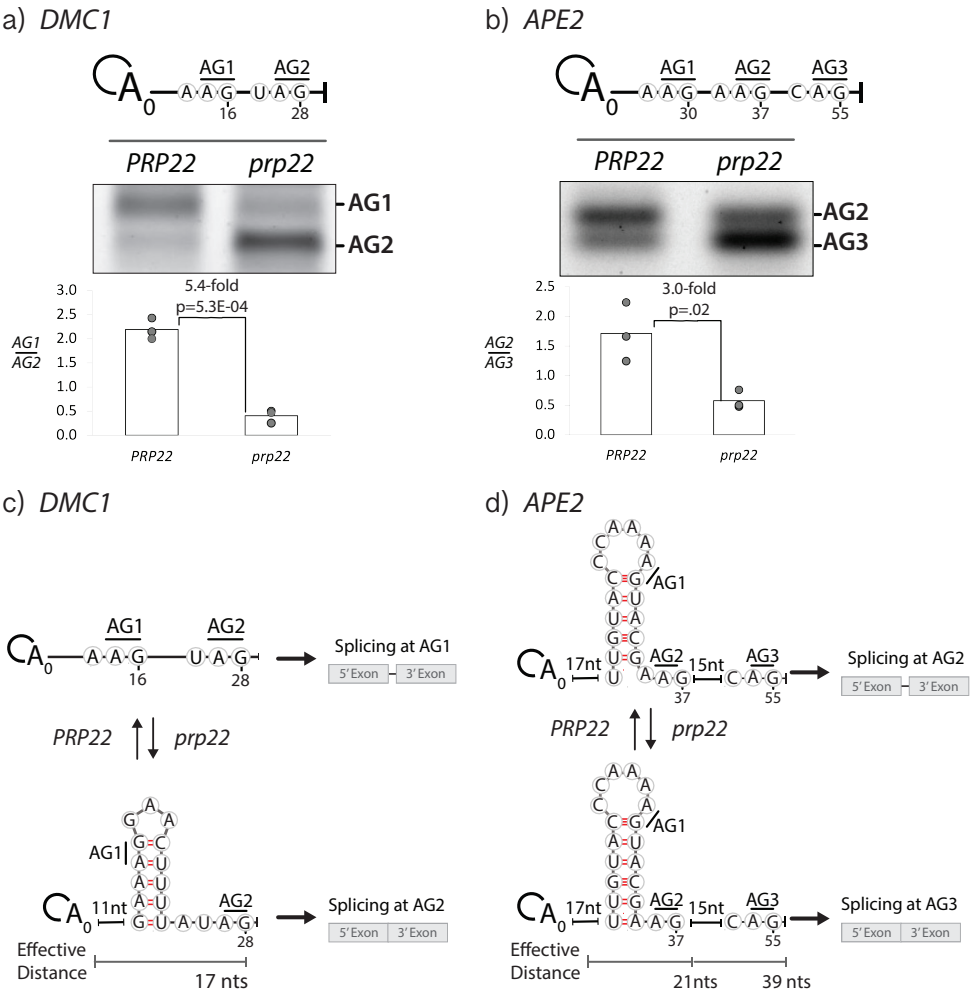


Figure 2.1. Prp22p promotes proximal 3'SS selection of the DMC1 and APE2 splicing substrates in vivo. (**a,b**) The positions of the 3'SSs in DMC1 **(a)** and APE2 **(b)** are shown in the context of the lariat intermediate; nucleotide numbers shown on the lariat represent the linear distance from the branch point to the indicated 3'SSs. Wildtype (PRP22) and mutant (prp22) strains were grown at 33°C, and total RNA was isolated for RT-PCR analysis of alternative mRNA isoforms resulting from alternative 3'SS choice in DMC1 **(a)** and APE2 **(b)**; mRNA products were separated on an agarose gel and visualized by ethidium bromide. The bar graphs show quantitation of 3'SS choice and illustrate data for three biological replicates. **(c,d)** The predicted secondary structure of DMC1 **(c)** and APE2 **(d)** according to RNAfold is shown[132,133], along with the interpreted splicing outcomes for the folded states as compared to the unfolded (DMC1) or less-folded states (APE2). The linear distance from the branch site to a 3' splice site is shown just below a 3' splice site, and the effective distance, after stem folding, is indicated at the bottom of the panel.

Although Prp22p proofreads suboptimal 3'SSs, and competing sites in *DMC1* and *APE2* include a suboptimal AAG, *PRP22* favors, rather than rejects, the sub-optimal splice site, arguing against a role for the proofreading function of *PRP22* in alternative 3' splice site choice[93]. While in the cases of *DMC1* and *APE2*, *PRP22* promoted a proximal 3'SS, our previous *in vitro* data indicated that Prp22p promoted a distal 3'SS in a *UBC4* substrate[40], ruling out a unidirectional mechanism for Prp22p in alternative 3'SS choice. Previous studies showed RNA secondary structure can modulate 3'SS selection by either sequestering and thereby repressing a 3'SS or promoting a non-sequestered, downstream 3'SS, by bringing the 3'SS in proximity to the branch point[108]. Therefore, as a helicase, Prp22p may impact 3'SS choice by destabilizing RNA secondary structure[78,122]. Indeed, RNA secondary structure predictions for our previous *UBC4* substrate, in which Prp22p favored a distal 3'SS, showed the distal 3'SS is sequestered within the RNA secondary structure of a stem loop and the upstream 3'SS is not, suggesting that Prp22p activated the distal 3' splice site by destabilizing the stem and thereby exposing the 3' splice site for docking and exon ligation (Figure 2.2).

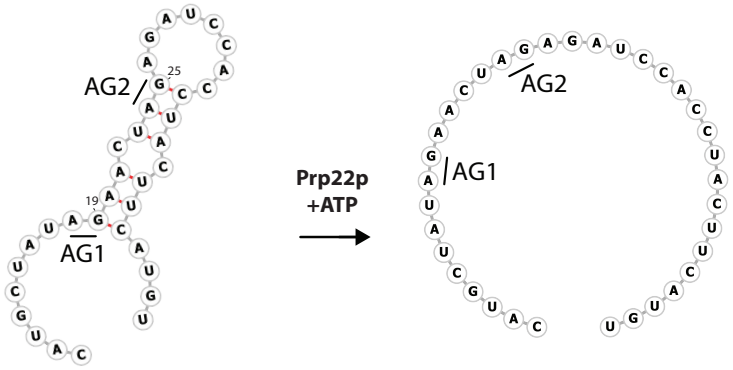


Figure 2.2. The UBC4 variant substrate previously characterized in Semlow et al.[40] can from a stem loop that preferentially sequesters the distal 3'SS. The proposed secondary structure in the intron between the branch site and 3'SS for UBC4 is illustrated on the left, according to predictions from RNAfold[132,133]. I hypothesize that ATP-hydrolysis by Prp22p leads to increased usage of the downstream UAG 3'SS (AG2), because Prp22p unwinds the stem. Nucleotide positions are numbered with respect to the branch site.

Consistent with a model in which Prp22p impacts 3'SS choice by destabilizing structure, *DMC1* and *APE2* can also form RNA secondary structure (stem loops) in the 3'SS region (Figure 2.1c-d)[108]. In the case of *DMC1*, the stem would sequester AG1 and bring AG2 in proximity to the branch point, shortening the effective distance from 28 to 17 nts. As I hypothesized for *UBC4* (Figure 2.2), I hypothesized that Prp22p unravels the stem in *DMC1,* exposing the sequestered AG1 and diminishing juxtaposition of AG2 with the branch point; consequently, the *prp22* mutant compromises unwinding, sequestering AG1 and bringing AG2 in proximity to the branchpoint, as observed (Figure 2.1c-d).

In the case of *APE2*, the stem would sequester AG1 by the loop (Figure 2.1d); modestly sequester AG2 at the base of the stem but also bring AG2 into proximity with the branch point, shortening the effective distance from 37 to 21 nts; and bring AG3 into proximity with the branch point, shortening the effective distance from 55 to 39 nts[108]. Indeed, disruption of the entire stem by mutation activates splicing at AG1, reduces splicing at AG2, and eliminates splicing at AG3 (Figure 2.3)[108]. As for *UBC4* and *DMC1*, for *APE2* I hypothesized simply that Prp22p destabilizes the stem, eliminating usage of AG3, reducing usage of AG2, and activating usage of AG1. Indeed, with the wild-type *APE2* intron, wild-type *PRP22* favors AG2 over AG3 (Figure2.1b and Figure2.1d), although *PRP22* does not favor AG1, likely because i) the stem is partially folded in wild-type *PRP22* cells (Figure 2.3a). Further aligning with the model, the *prp22* mutant favors AG3 over AG2 (Figure 2.1b), consistent with stabilization of the stem, sequestration of AG2, and shortening of the effective distance between AG3 and the branchpoint.

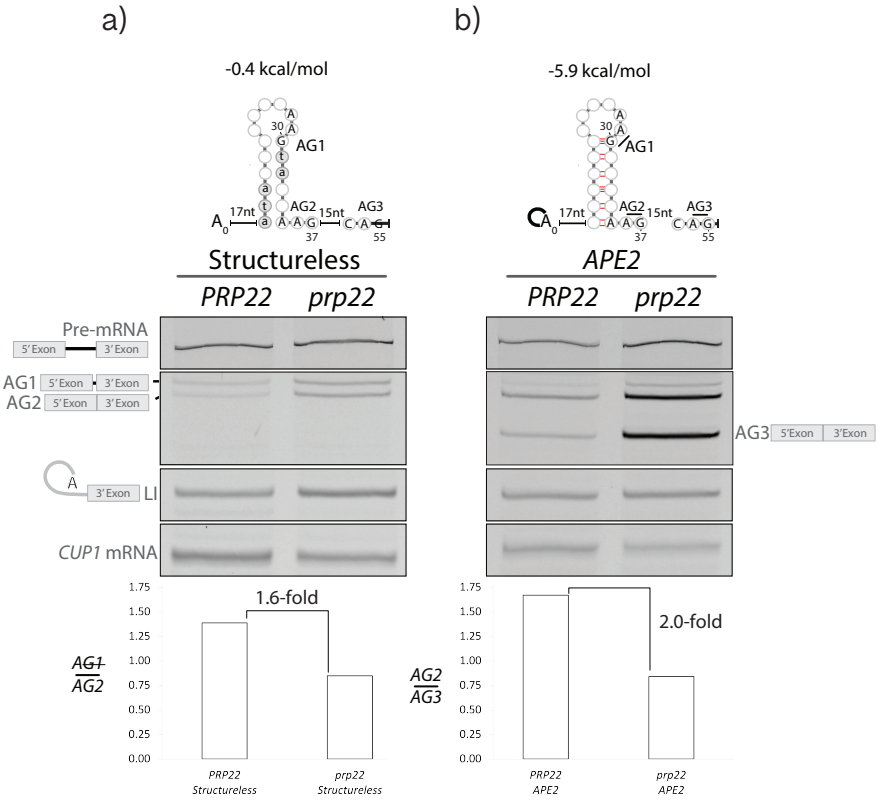


Figure 2.3 Usage of APE2's AG2 and AG3 is dependent on the formation of RNA secondary structure. (a, b) Structureless variant of APE2 were transformed into wildtype (PRP22) and mutant (prp22) strains and grown at 33°C, and then total RNA was isolated for primer extension analysis. The proposed secondary structures in the intron between the branch site and 3'SS are illustrated according to predictions from RNAfold[132,133]; predicted free energies of each secondary structure are indicated. Lowercase letters marked in dark grey indicate mutations designed to affect the stability of the stem. Splicing precursors, intermediates, and products of the reporters, separated on a denaturing gel, are visualized via a fluorescent Cy5 oligo used for primer extension. CUP1 mRNA is a loading control; LI stands for lariat intermediate. The ratio of AG1/LI is calculated as a reflection of the efficiency of exon ligation.

In summary, in both *DMC1* and *APE2,* splice site selection seems to be dependent on Prp22p-mediated destabilization of RNA secondary structure. Wildtype Prp22p appears to antagonize the stem-loops of *DMC1* and *APE2*, consequently leading to the availability of the otherwise sequestered 3'SSs. Conversely, in both *DMC1* and *APE2*, the Prp22p mutant appears to fail to destabilize the stem-loops, leading to splicing primarily at the 3'SSs downstream of the stem-loops. Thus, these data do support a model in which Prp22p disrupts secondary structure to impact 3'SS selection.

Prp22 activates a sequestered 3'SS

If Prp22p impacted 3' SS choice by destabilizing the stem loops in *DMC1* and *APE2*, then modulating the stability of the stems should similarly impact splice site choice and additionally modulate the impact of Prp22p on 3'SS choice. Decreasing the relative stability of the RNA secondary structure should shift splicing from the distal, downstream 3' splice site to the proximal, sequestered 3' splice site, given previous findings that splicing of distal, downstream 3'SSs is favored by folding of an upstream stem loop, whereas splicing of a 3'SSs sequestered in a stem loop is disfavored by folding of the stem loop[108]. Conversely, increasing the stability should increase usage of a downstream 3'SS and decrease usage of a sequestered 3'SS. To test these predictions, I re-engineered the established *ACT1-CUP1* splicing reporter system[134] by first swapping the *ACT1* intron out for the *DMC1* intron, and then I introduced mutations that either destabilized or stabilized the stem-loop (Figure 2.4). With the engineered reporter, wild-type *PRP22* promotes selection of AG1, and mutant *prp22* shifts splicing to AG2. Specifically, the *prp22* mutant decreased the usage of AG1 relative to AG2 by 2.1-fold (Figure 2.4a*)*. Thus, our reporter faithfully recapitulates the impact of *PRP22* on 3'SS selection in endogenous *DMC1*.

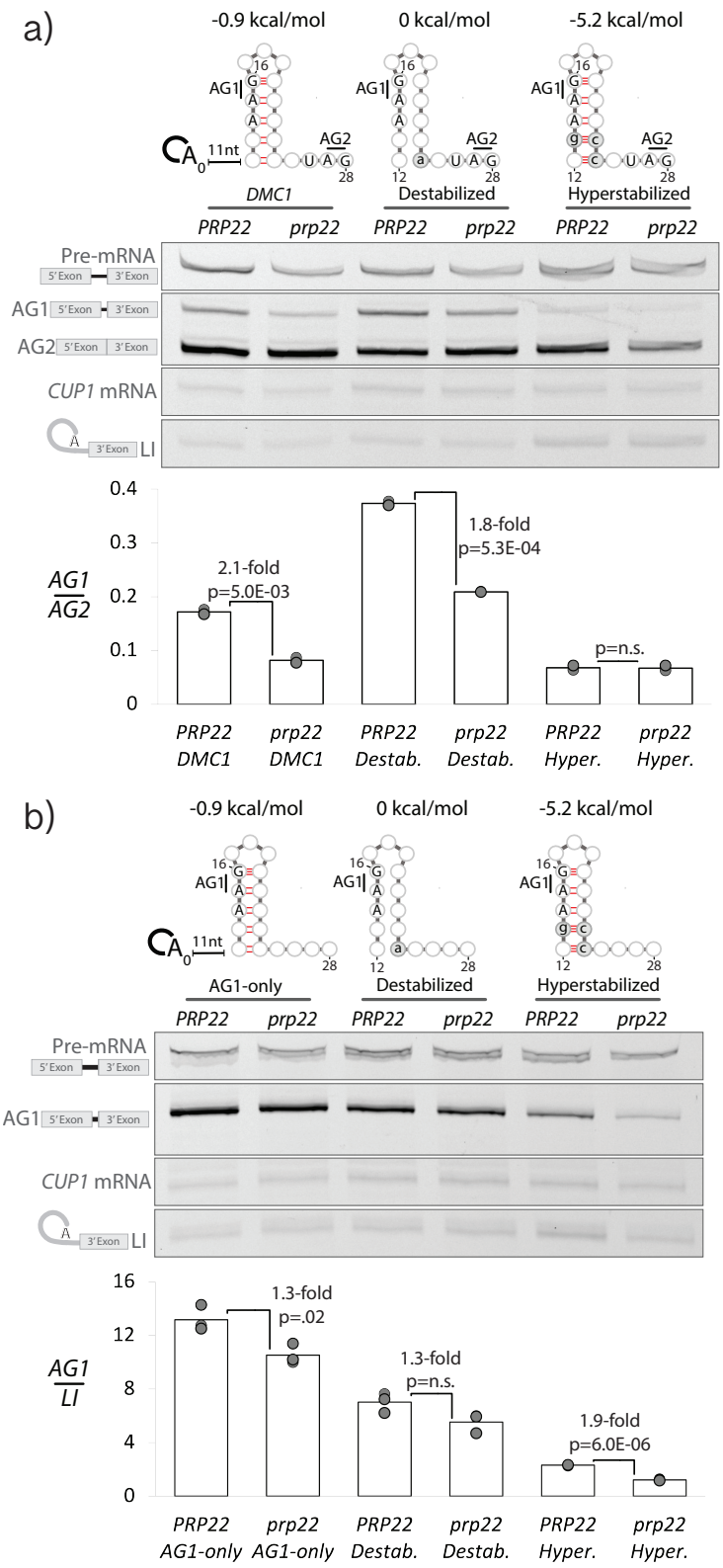


Figure 2.4. Prp22p activates a sequestered 3'SS. (a,b) Wildtype (*PRP22*) and mutant (*prp22*) strains were grown at 33°C, and total RNA was isolated for primer extension analysis.

Figure 2.4 (continued) The proposed secondary structures in the intron between the branch site and 3'SS for the *DMC1* **(a)** and AG1-only **(b)** series of reporters are illustrated according to predictions from RNAfold[132,133]; predicted free energies of each secondary structure are also indicated based on prediction from no structure is predicted for the destabilized reporters, although the potential to form four of the five bps in the stem remains. Lowercase letters marked in dark grey indicate mutations designed to affect the stability of the stem. In the AG1-only series **(b)**, AG2 is mutated from UAG to acG to prevent splicing at this site and competition with AG1. Splicing precursors, intermediates, and products, separated on a denaturing gel, were visualized via a fluorescent Cy5 oligo used for primer extension. *CUP1* mRNA is a loading control; LI stands for lariat intermediate. Each panel shows data from a single gel; irrelevant regions of the gel were omitted for clarity, as indicated by the gaps. The ratio of AG1/AG2 is calculated as a reflection of 3'SS splice site choice; the ratio of AG1/LI is calculated as a reflection of the efficiency of exon ligation. The data points reflect two **(a)** or three **(b)** biological replicates; the bars represent the average of the replicates; *p*-values were calculated using the student's t-test, two tailed distribution, two-sample equal variance (homoscedastic); "n.s." stands for "not significant".

To destabilize the stem-loop of *DMC1*, I mutated a nucleotide at the base of the stem-loop to decrease the propensity for the RNA substrate to fold (Figure 2.4a). The mutation changed the free energy of the structure from -0.9 kcal/mol in the *DMC1* reporter to 0 kcal/mol in the destabilized reporter, according to RNAfold[132,133]. To hyperstabilize the stem-loop, I enhanced base-pairing interactions at the base of the stem, which decreased the free energy to -5.2 kcal/mol (Figure 2.4a). I transformed these reporters into wildtype *PRP22* or mutant *prp22* yeast cells and then assayed for splicing via primer extension analysis.

As predicted, destabilization of the stem-loop increased splicing at AG1 relative to AG2 (AG1/AG2) for both wildtype *PRP22* and mutant *prp22* by 2.2-fold (p-value=8.5E-04) and 2.6-fold (p-value=1.5E-03), respectively (Figure 2.4a; cf. lanes 1 and 2 with 3 and 4, respectively*)*; nevertheless, in the destabilized reporter maximal splicing at AG1 relative to AG2 still requires *PRP22*, as the *prp22* mutant reduces AG1/AG2 by 1.8-fold, consistent with some degree of folding in the destabilized reporter. By contrast and also as predicted, hyperstabilization of the stem-loop in wildtype *PRP22* decreased splicing at AG1 relative to AG2 (AG1/AG2) by 2.5-fold (p-value=3.4E-03); curiously, in the *prp22* mutant hyperstabilization of the stem-loop did not significantly decrease splicing further at AG1 relative to AG2, perhaps because the stem loop in the unmutated *DMC1* substrate is already fully folded in the *prp22* mutant (Figure 2.4a; cf. lane 1 with 5 lane 2 with 6). Additionally, in the hyperstabilized reporter, the *prp22* mutant, relative to wildtype *PRP22*, reduced splicing at both AG1 and AG2, reducing total mRNA by 2-fold (p-value=.02), suggesting that hyperstabilization extends the stem, thereby occluding AG2, in addition to AG1 (Figure 2.4a; cf. lanes 5, 6; Figure 2.5d). Overall, these results provide evidence that a stem-loop does form in *DMC1* between the branch site and the 3'SS and that the stem-loop

sequesters AG1 and juxtaposes AG2 with the branch point, consistent with a role for *PRP22* in regulating 3'SS usage directly by unwinding the stem loop.

Because the 3'SSs in *DMC1* appear to compete (2.1a), Prp22p could formally affect alternative 3' splice site choice by either activating the sequestered 3' splice site, repressing the distal 3' splice site, or both. Thus, I tested whether Prp22p could, by unwinding RNA, activate the sequestered 3' splice site (AG1), independent of repressing the distal, downstream 3'SS (AG2). To isolate our analysis to the sequestered AG1, I mutated the downstream AG2, yielding the reporter "AG1-only" (Figure 2.4b). Then, I mutated AG1-only to destabilize or hyperstabilize the stem-loop, as above.

As expected, mutations to AG2 ablated splicing at that site (Figure 2.5a, 2.5b). A quantitative comparison of the AG1-only reporter with the unmutated *DMC1* reporter, showed that the overall levels of mRNA formation changed only modestly, implying the 3'SSs compete (1.4-fold p-value=1.6E-03; Figure 2.5a, 2.5b); indeed, splicing at AG1 increased in the AG1-only reporter, relative to the unmutated reporter, by 3.2-fold (p-value=8.4E-05; Figure 2.5a, 2.5b). These results verify that AG1 and AG2 compete with one another in *DMC1*.

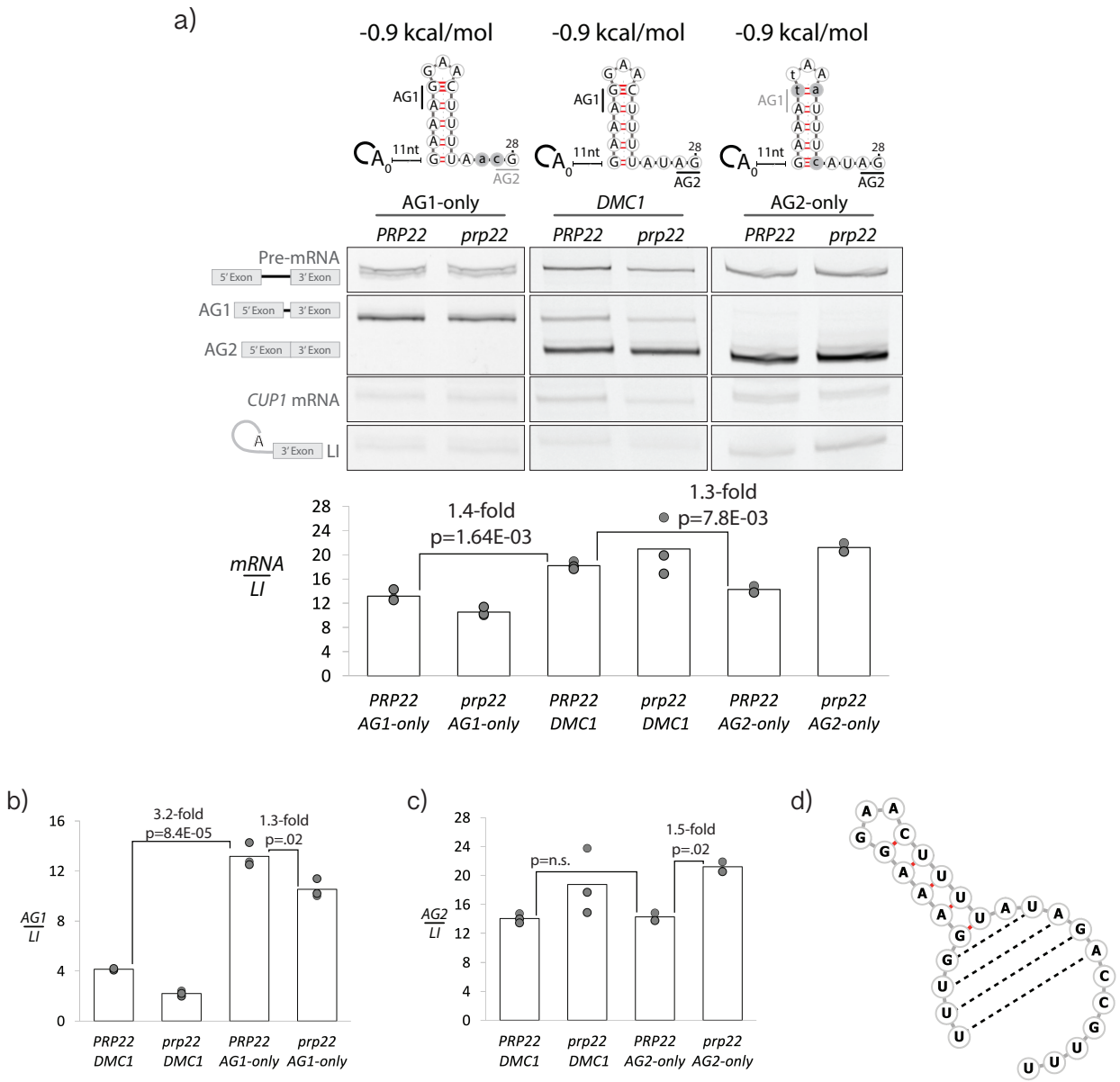


Figure 2.5. Evidence that the AG1-only substrate only splices at AG1 and that the 3'SSs in DMC1 compete. (a) Wildtype (PRP22) and mutant (prp22) strains, transformed with the indicated reporters, were grown at 33°C, and total RNA was isolated for primer extension analysis. The proposed secondary structures in the intron between the branch site and 3'SS for DMC1 is illustrated according to predictions from RNAfold[132,133]; no structure is predicted for the structureless reporters. Predicted free energies of each secondary structure are indicated. Lowercase letters marked in dark grey indicate mutations designed to eliminate splicing at AG1 or AG2, as well as to maintain the stability of the stem in AG2-only. A representative gel is shown (a) comparing the unmutated DMC1 reporter, having both AG1 and AG; the AG1-only reporter, and the AG2-only reporter. Splicing precursors, intermediates, and products of the

Figure 2.5 (continued) reporters, separated on a denaturing gel, are visualized via a fluorescent Cy5 oligo used for primer extension. CUP1 mRNA is a loading control; LI stands for lariat intermediate. The mRNA levels, relative to lariat intermediate levels (mRNA/LI), is plotted as a reflection of overall exon ligation efficiency (a), AG1/LI levels are plotted as a reflection of AG1 exon ligation efficiency (b), and AG2/LI levels are plotted as a reflection of AG2 exon ligation efficiency (c). The data points relflect three biological replicates (DMC1, AG1-only) or two biological replicates (AG2-only); the bars represent the average of the replicates; p-values were calculated using the student's t-test, two tailed distribution, two-sample equal variance (homoscedastic). (d) A potential extension of the stem-loop in DMC1 is illustrated. This extension would result in the sequestration of AG2, in addition to AG1, within the stem.

In the AG1-only reporter, the destabilizing mutations modestly decreased splicing at AG1 by 1.9-fold, for unclear reasons, but this decrease in splicing does not appear to reflect an unintended increase in structure (see below; Figure 2.4b). By contrast, the hyperstabilizing mutations decreased splicing at AG1, as expected, and substantially so, by 5.7-fold (p-value=4.0E-05; Figure 2.4b, cf. lanes 1 and 5). This result provides further evidence that AG1 is sequestered in the stem loop.

With the AG1-only reporter, mRNA formed at AG1 at similar levels in wild-type *PRP22* and mutant *prp22* yeast cells, differing by only 1.3-fold (Figure 2.4b). Because the *prp22* mutant increased AG1/AG2 by 2.1-fold in unmutated *DMC1* (Figure 2.4a), this result may seem surprising and suggest the possibility that Prp22p indirectly activates the sequestered stem by antagonizing the distal 3'SS, but our analysis of the hyperstabilized substrate below is inconsistent with such a mechanism. Instead, I infer that the sensitivity of the sequestered AG1 to Prp22p activity, in the *DMC1* reporter, in contrast to AG1-only, depends on competition between AG1 and AG2 (Figure 2.5; see Discussion).

With the destabilized AG1-only reporter, splicing at AG1 in mutant *prp22* versus wildtype *PRP22* showed no significant difference (Figure 2.4b). By contrast, with the hyperstabilized reporter, the *prp22* mutant decreased splicing by 1.9-fold (Figure 2.4b), indicating that maximal splicing at AG1 of the hyperstabilized, AG1-only reporter requires wildtype Prp22p function and that maximal splicing at AG1 in the *DMC1* reporter does not require repression of the distal, competing 3'SS. Importantly, the impact of mutant *prp22* on AG1 depends on the potential to form secondary structure, because with an unstructured, linearized variant of AG1-only, the *prp22* mutant by contrast has no significant effect (Figure

2.6). Altogether, these results provide evidence that the stem loop in *DMC1* can form and that Prp22p activates a sequestered 3'SS by destabilizing the stem.

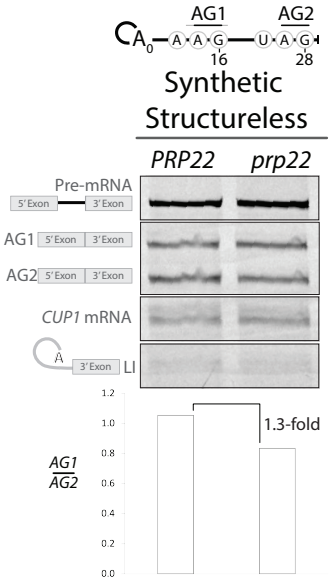


Figure 2.6. Prp22p does not impact 3'SS selection in a structureless variant of DMC1. The structureless variant was transformed into wildtype (PRP22) and mutant (prp22) strains and grown at 33°C, and then total RNA was isolated for primer extension analysis. The proposed secondary structures in the intron between the branch site and 3'SS are illustrated according to predictions from RNAfold[132,133]; predicted free energies of the secondary structure is zero; no structure is predicted for the destabilized reporters. Lowercase letters marked in dark grey indicate mutations designed to affect the stability of the stem. Splicing precursors, intermediates, and products of the reporters, separated on a denaturing gel, are visualized via a fluorescent Cy5 oligo used for primer extension. CUP1 mRNA is a loading control; LI stands for lariat intermediate.

While implicating a role for Prp22p in destabilizing the *DMC1* stem, our data also suggest a limit to this activity of Prp22p. Specifically, although with the hyperstabilized *DMC1* substrate, having both AG1 and AG2, *PRP22* was required for maximal splicing at AG1, as measured by AG1/LI, in the *PRP22* strain splicing at AG1 in the hyperstabilized substrate was lower than in the wild-type *DMC1* substrate (by 7.5-fold, p-value=8.7E-05; Figure 2.1a), implying a limit to the activity of Prp22p. Similarly, when the stem loop of the AG1-only *DMC1* substrate was hyperstabilized, wild-type *PRP22* was still required for maximal splicing at the sequestered AG1 (Figure 2.1b), but in the *PRP22* strain, splicing at AG1 in the hyperstabilized substrate, as noted above, was lower than in the wildtype AG1-only substrate (by 5.7-fold, p-value=4.0E-05; Figure 2.1b), again implying a limit to the activity of Prp22p. Such a limit for the destabilizing activity of Prp22p is consistent with the positive role secondary structure can play juxtaposing a 3'SS in proximity with a branch point. Altogether, our data indicate that splicing outcomes are not only impacted by RNA secondary structure but also the activity of Prp22p on such structure[108].

Prp22 antagonizes a distal 3'SS

Because splicing of both *APE2* and *DMC1* showed Prp22p promotes usage of a proximal 3'SS and repressed usage of a distal site (Figure 2.1), and further analysis of the sequestered 3'SS of *DMC1* showed Prp22p can independently activate the usage of a sequestered, proximal 3'SS (Figure 2.4), a question that emerged was: does Prp22p reduce splicing at a distal 3'SS site just because it activates a proximal 3'SS, or does Prp22p also play a role in antagonizing a distal 3'SS, independent of any impact of Prp22p on the upstream, proximal site. Given that the selection of a 3'SS that is downstream of the branch point by more than ~45nts is dependent on RNA secondary structure to juxtapose the 3'SS with the branchpoint[108,135], and given that Prp22p activates a 3'SS sequestered in a stem-loop of *DMC1* by destabilizing this structure, I predicted Prp22p would also independently antagonize a 3'SS downstream of a stem loop. Indeed, in a *DMC1* substrate that included only the downstream 3'SS AG2 (AG2-only), wildtype *PRP22* decreased splicing, relative to mutant *prp22*, by 1.5-fold (*p*-value=0.02; Figure 2.5a, 2.5c). Because the distal 3'SS in *APE2* is farther from the branchpoint (55 nts) than the distal 3' splice site (28 nts) in *DMC1*, such that the distal 3'SS in *APE2* should be more dependent on structure for juxtaposition with the branchpoint, I tested our predictions further in the context of *APE2*.

To investigate splicing of *APE2*, I again re-engineered the established *ACT1-CUP1* splicing reporter[134] by swapping the *ACT1* intron out for the *APE2* intron. To test whether the effect of promoting alternative splice site selection is due to its impact on the stem-loop of *APE2*, I introduced mutations that either destabilized, stabilized, or hyperstabilized the stem-loop (Figure 2.7a). To destabilize the stem-loop, I mutated a nucleotide in the middle of the stem-loop to decrease the propensity for the RNA substrate to fold. The mutation changed the free energy of

the structure from -5.9 kcal/mol in the wildtype *APE2* reporter to -4.5 kcal/mol in the destabilized reporter. To stabilize or hyperstabilize the stem-loop, I enhanced base-pairing interactions within the stem, which decreased the free energy to -8.1 kcal/mol and -10.3 kcal/mol, respectively. I transformed these reporters into wildtype *PRP22* or mutant *prp22* yeast cells and then assayed for splicing via primer extension analysis. With the engineered reporter, termed *APE2*, *PRP22* promoted selection of AG2, and the *prp22* mutant shifted splicing to AG3. Specifically, the *prp22* mutant increased usage of AG3 relative to AG2 by 2.0-fold (Figure 2.7a, 2.7b). Thus, our reporter faithfully recapitulates the impact of *PRP22* on 3' splice site selection in endogenous *APE2*.

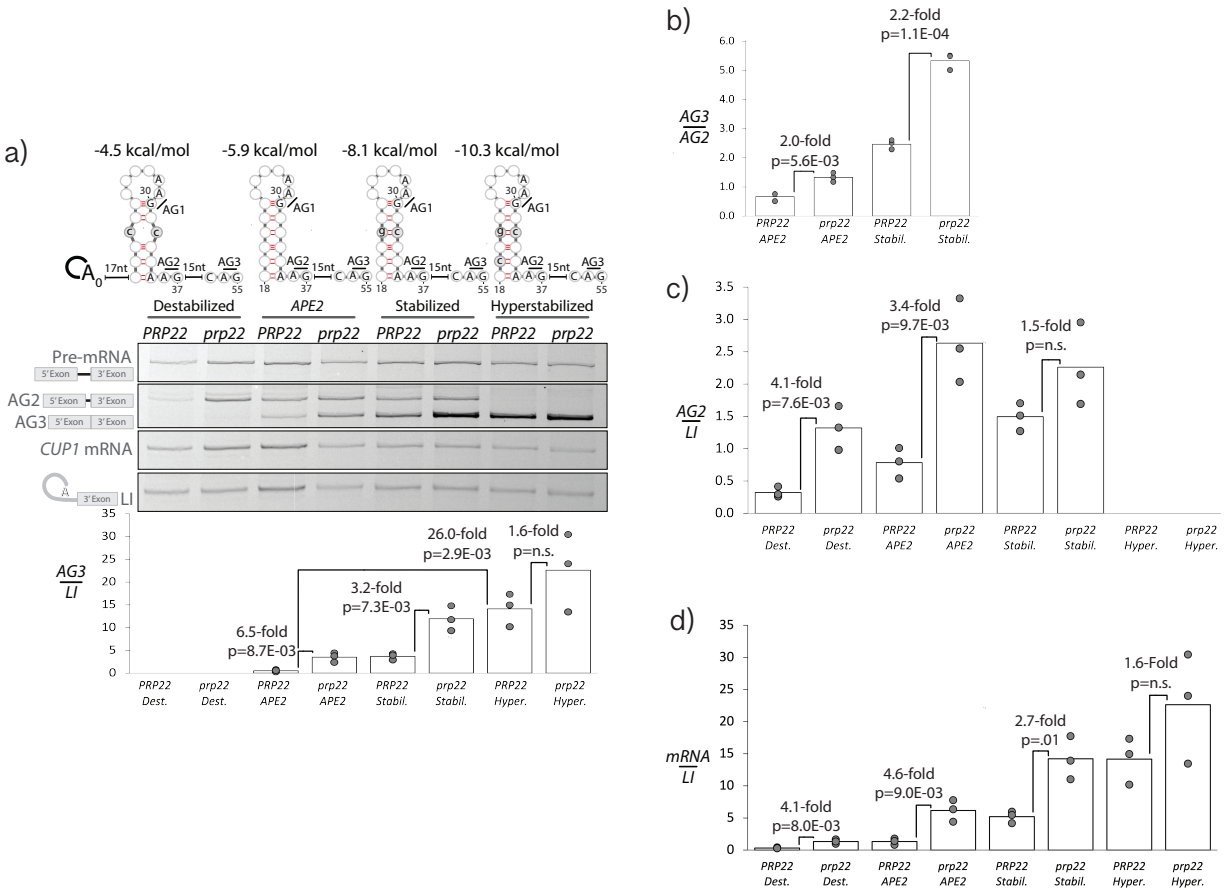


Figure 2.7. Prp22p antagonizes a distal 3'SS. (a,c) Wildtype (PRP22) and mutant (prp22) strains were grown at 33°C, and total RNA was isolated for primer extension analysis. The proposed secondary structures in the intron between the branch site and 3'SS for the APE2 **(a)** and AG3-only **(e)** series of reporters are illustrated according to predictions from RNAfold[132,133]; predicted free energies of each secondary structure are also indicated based on prediction from RNAfold[132,133]. Lowercase letters marked in dark grey indicate mutations designed to affect the stability of the stem. Splicing precursors, intermediates and products of the reporters, separated on a denaturing gel, are visualized via a fluorescent Cy5 oligo used for primer extension. CUP1 mRNA is a loading control; LI stands for lariat intermediate. Each panel shows data from a single gel; irrelevant regions of the gel were omitted for clarity, as indicated by the gaps. The ratio of AG3/LI is plotted as a reflection of the efficiency of the exon ligation

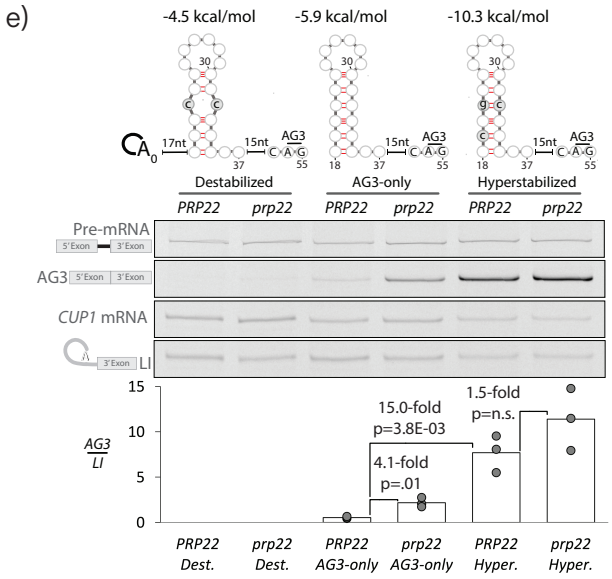


Figure 2.7 (continued) reaction. **(b)** The ratio of AG3/AG2, derived from **(a)**, is plotted as a reflection of 3'SS selection. The impact of PRP22 or stem loop destabilization/stabilization on splicing at AG2 or at all 3'SSs in APE2. (c,d) The data from **(a)** were replotted as AG2/LI (**c**) or mRNA/LI (**d**) as a reflection of the efficiency of splicing at AG2 or at AG1, AG2, and AG3 (**d**). In the AG3-only series **(e)**, AG1 and AG2 are mutated from UAG to acG to prevent splicing at this site and competition with AG3. The data points in all three panels reflect three biological replicates; the bars represent the average of the replicates; p-values were calculated using the student's t-test, two tailed distribution, two-sample equal variance (homoscedastic); "n.s." stands for "not significant".

As expected (c.f. Meyer *et al*. 2011[108]), destabilization of the *APE2* reporter inhibited splicing at AG3, relative to the lariat intermediate levels, to below background levels (Figure 2.7a)[108]. In contrast, stabilization of the reporter, relative to the *APE2* reporter, increased splicing at AG3, relative to lariat intermediate levels, by 6.8-fold (p-value=1.9-03, Figure 2.7a), and hyperstabilization of the reporter increased splicing at AG3, relative to *APE2*, even more, by a striking 26-fold (p-value=2.9E-03). In parallel, destabilization decreased splicing at AG2, relative to lariat intermediate levels, by 2.4-fold (p-value=3.0E-3), whereas stabilization increased splicing at AG2 by 3.8-fold (p-value=8.1E-03; Figure 2.7a, 2.7c), implying a role for the stem in also juxtaposing AG2 with the branch point, consistent with its position just downstream of the stem; hyperstabilization reduced splicing at AG2 to below background levels, suggesting that hyperstabilization of the stem sequestered AG2 (Figure 2.7a). As a consequence of these changes at both AG3 and AG2, destabilization decreased total mRNA formation, relative to the lariat intermediate levels, by 4.1-fold (p-value=.02), stabilization increased mRNA formation by 3.9-fold (p-value=3.0E-3), and hyperstabilization increased mRNA formation by 10.6-fold (p-value=3.8E-03), with the change in total mRNA from the destabilized reporter to the hyperstabilized reporter amounting to a striking 44-fold change (p-value=2.7E-03), implying that splicing at AG3 and AG2 compete with discard of the lariat intermediate (see Discussion). Note that a role for the stem in juxtaposing both AG2 and AG3 with the branch site is nevertheless consistent with a role for the stem in shifting splicing from AG2 to AG3, as in the *prp22* mutant, because folding of the stem loop will have a disproportionate effect on the more distal AG3; indeed stabilization of the stem increased splicing at AG3, relative to AG2, by 2.5-fold (p-value=6.7E-03 ; Figure 2.7a, 2.7b). Overall, these results provide compelling evidence that a

stem-loop does form in *APE2* and that the stem-loop increases the proximity of AG3, as well as AG2, to the branchpoint.

PRP22 not only modulated 3'SS usage and selection in the *APE2* reporter but also in the destabilized, and stabilized reporters. In the destabilized reporter, the *prp22* mutant increased splicing at AG2, relative to the lariat intermediate, by 6.5-fold (p-value=8.7E-03), implying that *PRP22* antagonizes the role of the stem in juxtaposing AG2 with the branch point (Figure 2.7a, 2.7c). In the stabilized reporter, as in the *APE2* substrate, the *prp22* mutant increased splicing at AG3, relative to AG2, by 2.2-fold (Figure 2.7a, 2.7c), providing further evidence that *PRP22* modulates 3'SS selection. The *prp22* mutant not only shifted splicing from AG2 to AG3 but also increased total mRNA, relative to lariat intermediate, in the *APE2*, destabilized, and stabilized reporters (Figure 2.7a, 2.7c), again implying that splicing at AG2 and AG3 competes with discard of the lariat intermediate. Taken together, these results suggest a general role for *PRP22* in destabilizing RNA secondary structure and thereby antagonizing distal 3'SS selection.

Because the 3'SSs in *APE2* appear to compete (Figure 2.1b), Prp22p could formally repress splicing at the downstream AG3 as an indirect consequence of activating the upstream AG2. Thus, to isolate our analysis to AG3, I mutated the upstream AG1 and AG2, yielding the reporter "AG3-only" (Figure 2.7e). Then, I mutated AG3-only to destabilize or hyperstabilize the stem-loop, as above. As expected, mutations to AG1 and AG2 precluded splicing at those sites and allowed us to focus our analysis on AG3 (Figure 2.8). A quantitative comparison of the AG3-only reporter with the unmutated, wild-type *APE2* reporter showed that the overall levels of mRNA formation were not maintained and instead decreased by 2.9-fold (Figure 2.8a, 2.8b), providing no evidence, surprisingly, that AG3 competes with the other 3'SSs; indeed, again comparing the AG3-only reporter with the wild-type reporter, the level of splicing at AG3,

relative to lariat intermediate levels, did not change significantly. Similarly, mutation of the AG3 3'SS in the wild-type reporter, yielding no-AG3, rather than maintaining mRNA levels, decreased overall mRNA levels modestly by 1.6-fold (Figure 2.8a, 2.8b); further, again comparing the no-AG3 reporter with the wild-type reporter, the level of splicing at AG1 and AG2, relative to lariat intermediate levels, did not change significantly. I infer, unexpectedly, that AG3 does not compete with the other 3'SSs in *APE2* (see Discussion).

In the AG3-only reporter, the destabilizing mutations decreased splicing at AG3 to below background levels, and the hyperstabilizing mutations increased splicing at AG3 by 15-fold (Figure 2.7e), establishing evidence that the stem loop juxtaposes AG3 with the branch point. With the AG3-only reporter, the *prp22* mutant, relative to wildtype *PRP22*, increased splicing at AG3 by 4.1-fold, establishing that Prp22p impacts AG3 usage, independent of an impact on other 3'SSs (Figure 2.7e). Altogether, these data support a role for Prp22p in regulating the juxtaposition of the distal 3'SS in *APE2* and doing so by unwinding secondary structure, consistent with the role of Prp22p in activating a sequestered 3'SS in *DMC1* by unwinding RNA.

At the same time, our data also suggest a limit to this destabilizing activity of Prp22p. When the stem loop of the wild-type *APE2* substrate was hyperstabilized (-10.3 kcal/mol), total splicing increased by 10.6-fold (p-value=3.7E-03), and when the stem loop of the AG3-only substrate was hyperstabilized, splicing at AG3 increased by 4.5-fold (p-value=3.6E-05), together implying a limit to the activity of Prp22p on the stem loop of *APE2*. Indeed, with both the hyperstabilized *APE2* substrate and the hyperstabilized AG3-only substrate, *PRP22* no longer significantly impacted AG3 usage (Figure 2.7a, 2.7e). Such a limit for the destabilizing activity of Prp22p with *APE2* parallels similar evidence for a limit for Prp22p with *DMC1*.

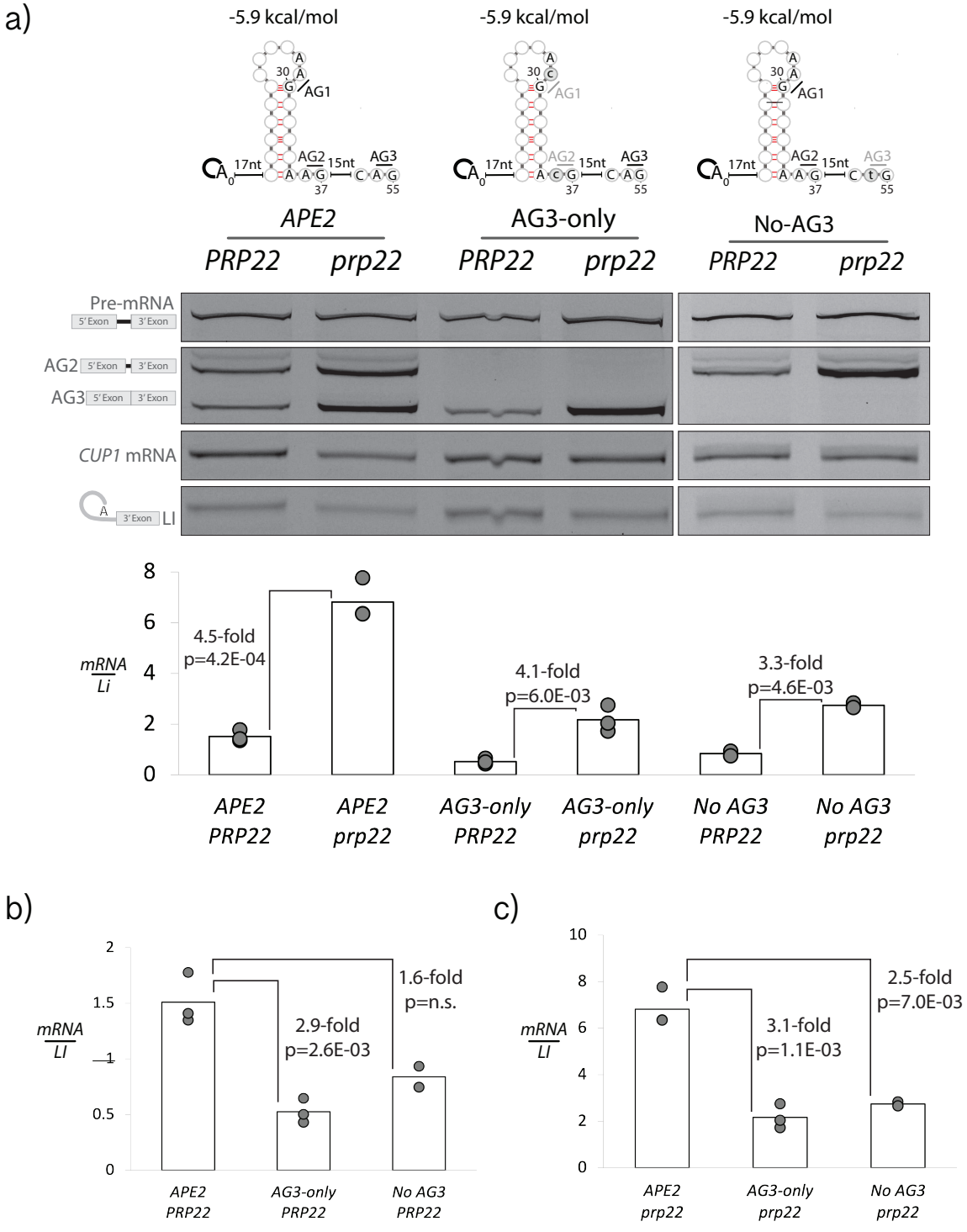


Figure 2.8. Evidence that the AG3-only substrate only splices at AG3 and that the 3'SSs in APE2 do not compete. (a-c) Wildtype (PRP22) and mutant (prp22) strains, transformed with the

Figure 2.8 (continued) indicated reporters, were grown at 33°C, and total RNA was isolated for primer extension analysis. The proposed secondary structures in the intron between the branch site and 3'SS for APE2, AG3-only, and no-AG3 are illustrated according to predictions from RNAfold[132,133]. Predicted free energies of each secondary structure are indicated. Lowercase letters marked in dark grey indicate mutations designed to eliminate splicing at AG1 and AG2 (AG3-only) or to eliminate splicing at AG3 (no-AG3). A representative gel is shown (a) comparing the splicing of the three indicate reporters. Splicing precursors, intermediates, and products of the reporters, separated on a denaturing gel, are visualized via a fluorescent Cy5 oligo used for primer extension. CUP1 mRNA is a loading control; LI stands for lariat intermediate. The mRNA levels, relative to lariat intermediate levels (mRNA/LI), are plotted as a reflection of overall exon ligation efficiency (a); these levels are replotted to facilitate a comparison between the reporters in either the wildtype PRP22 (b) or mutant prp22 (c) strains. The independent data points reflect three biological replicates (APE2, AG3-only) or two biological replicates (no-AG3). The bars represent the average of replications; pvalues were calculated using the student's t-test, two tailed distribution, two-sample equal variance (homoscedastic).

Material and Methods

Yeast strains and transformations. Plasmid transformations of competent yeast cells were executed as previously described[136]. Yeast strains used are described in ref.[93]. See Table 2.1 for specific strains and plasmids used for analysis. Parent strains were initially designed by Hiroshi Maita and referred to in Mayas et al. 2006. The parental genotype is shown in Table 2.1. Briefly, p358-Prp22 (bJPS1493) or p358-Prp22,g810a (bJPS1503) was independently shuffled into yeast strain yJPS1084, to create yJPS1085 or yJPS1088, respectively. These previously engineered strains were used for transforming reporter plasmids into.

Isolation of RNA. Yeast were grown in 2 mL single dropout -Leu media at 33 °C, rotating at 250 rpm, in Fisherbrand™ printed disposable culture tubes (14-956-9C) to an OD_{600} of 1.2-1.6 or in 48-well polypropylene 4.6 ml deep well plates (NEST Scientific, NST504002). When grown in plates, a small glass bead was included to promote aeration and to prevents clumping of yeast at the bottom of the plates. At harvest, 1.5 mL samples were spun down for 1 min at 13,000 rpm in uncapped tubes (Stellar scientific, T20-600), decanted, and immediately snap frozen using liquid nitrogen. Note: the type of tube matters, 2 mL flat bottom (as opposed to tapered) is required for sufficient RNA yield. Next, roughly 300 mL of 0.5 mm glass beads were added to the frozen pellets followed by 750 mL Trizol. Caps were then added, and samples were immediately placed at -80°C. To lyse the yeast cells, the samples were removed from the -80°C, immediately placed on a multi-tube adaptor (Scientific industries, SKU: 504-0234-00) and vortexed on high for about 10 mins. Samples were briefly placed in an ice water bath for 2-3 mins before 250 mL chloroform was added to the samples. Samples were then vortexed briefly and spun at room temperature on high speed for 3 mins. The aqueous (top layer) layer was then removed and added to a new tube containing at least 1 volume (~400 uL) chloroform. Samples were vortexed and spun at max speed for 3 mins. The aqueous (top layer) was then removed and

added to 1.2 volumes of isopropanol containing 1 μL 10 mg/mL glycogen. Samples were then spun at max speed for 30 mins in a refrigerated, 4 °C centrifuge. Supernatant was removed using a weak aspirator to ensure the pellet was not aspirated off with the liquid. Samples were quickly spun down again, and residual liquid was removed. Then, 1 mL 75% ethanol was added and then tubes were inverted and spun at max speed at room temperature for 5 mins. Samples were again aspirated and spun down and then residual liquid was removed, and pellets were air dried before adding 10 mL 50 °C doubly distilled H_20 (Quality Biologicals, 351-029).

Primer extension reaction. Reactions, totaling 20 μL, were assembled for primer extension. First, primers were annealed with RNA in a 10 μL annealing reaction (see below) at 65 °C before snap cooling in an ice water bath for at least 3 mins. Following the 3 min cold incubation, samples were equilibrated at room temperature for a minimum of 3 mins. During the equilibration, 10 μL RT mix (see below) was pre-heated at 47 °C for 3 mins before adding to the annealed reaction mix. Samples were then incubated at 42°C for 1 hr and then heat inactivated at 90°C for 5 min before 25 μL RNA loading buffer (see below) was added. Samples were then re-heated at 80 °C for 2 mins and then loaded and separated on a denaturing PAGE gel (dPAGE).

Annealing Reaction. The reaction included 40 mM Tris-HCl (pH=8.3), 48 mM NaCl, 8 mM DTT, 0.5 μM primer (JV121,Cy5-GGCACTCATGACCTTC-3' or pCody1, Cy5-AGCTACCACATTGGCAT-3'), 4 μg RNA, and water to bring the volume to 10 μL. Reactions were incubated at 65 °C in a thermal cycler for 3 mins before snap cooling in an ice water bath.

RT mix. The mix included 20 mM Tris (pH=8.3), 24 mM NaCl, 4 mM DTT, 12 mM MgOAc, 1.12 mM dNTP, 7 units AMV RT-HC (Promega, M9004), plus water up to a total volume of 10 μL. RT mix was then pre-heated at 47 °C for a minimum of 3 mins, maximum of 5 mins. RT mix

was then added to 10 mL annealing reaction, briefly vortexed, spun down lightly, and placed quickly at 42 °C.

RNA loading buffer. Loading buffer was comprised of 950 μL formamide, 25 mM EDTA, orange G (Sigma, O7252-25G)

dPAGE. Denaturing PAGE gels were cast with 7M urea, 8.5% polyacrylamide (19:1) gel, 1X TBE, pH=8.3 (89 mM Tris, 89 mM boric acid, 2 mM EDTA), 0.1% ammonium persulfate, 0.1% TEMED. The dPAGE gels were pre-run for 5 mins at 25 W. Following a 5 min pre-run, wells were cleared out and 15 μL of the total 45 mL were added to the gel and run for 1 hr at 25 W. Gels were imaged without disassembling the glass plates (Gel Company, GGV20L-5) on an Amersham Typhoon NIR Plus.

Secondary Structure Predictions. Predictions were executed as previously described (ref. [137]). Briefly, structural predictions and free energies were generated using RNAfold [132,133]. The sequences used to predict the structures were 7 nts downstream of the branchpoint and 7 nucleotides past the last 3'SS. The following is a brief overview of the conditions used for modeling the secondary structure and free energy:

Fold algorithms and basic options- minimum free energy (MFE) and partition function, avoid isolated base pairs

Advanced folding options- dangling energies on both sides of a helix in any case

RNA parameters- 2004, Turner model

Other Parameters- 33°C rescale energy parameters to given temperature (C)

Table 2.1 Strains and Oligos

Yeast Strains Used

Lab Strain	Genotype/Note
yJPS1084	MATa trp1-Δ1 ura3–52, IMT1 IMT2 imt3::TRP1 imt4::TRP1 leu2::hisG GAL+ prp22::KanMX4, covered by pRS316-PRP22
yJPS1085	yJPS1084 with p358-PRP22 shuffled in to replace pRS316-PRP22
yJPS1088	yJPS1084 with p358-prp22-G810A shuffled in to replace pRS316-PRP22

Primers used for RT-PCR

Target	Specific Step	Primer sequence
S288C_YER179W_DMC1	RT	TAGCGGGGATAAATCCCACC
S288C_YER179W_DMC1	PCR	ATGTCTGTTACAGGAACTGAGATCG
S288C_YER179W_DMC1	PCR	TTCCACTTTCACCTCACTTAACC
S288C_YKL157W_APE2	RT	TTAGCTCAATTTTGACTGAGCC
S288C_YKL157W_APE2	PCR	TCTGTTCACAGAAACTAGTCACG
S288C_YKL157W_APE2	PCR	AATCGTAGTGTAAAGGAACGAC

Primer Extension

Name	Sequence
JV121	Cy5-GGCACTCATGACCTTC-3’
pCody1	Cy5-AGCTACCACATTGGCAT-3’

Table 2.1 (continued)

Name of Reporters	Sequence
DMC1, Wt	GGTTCTTTCCCCCTTTCTTTTACTAACTAATAATTTGGAAAGGAACTTTTATAGACCGTTTTGTCAACAACAAGA
DMC1, Destabilized	GGTTCTTTCCCCCTTTCTTTTACTAACTAATAATTTGGAAAGGAACTTTaATAGACCGTTTTGTCAACAACAAGA
DMC1 Hyperstabilized	GGTTCTTTCCCCCTTTCTTTTACTAACTAATAATTTGGgAAGGAACTTccATAGACCGTTTTGTCAACAACAAGA
DMC1, AG1-only Destabilized	GGTTCTTTCCCCCTTTCTTTTACTAACTAATAATTTGGAAAGGAACTTTaAacGACCGTTTTGTCAACAACAAGA
DMC1, AG1-only	GGTTCTTTCCCCCTTTCTTTTACTAACTAATAATTTGGAAAGGAACTTTTAacGACCGTTTTGTCAACAACAAGA
DMC1, AG1-only Hyperstabilized	GGTTCTTTCCCCCTTTCTTTTACTAACTAATAATTTGGgAAGGAACTTccaacGACCGTTTTGTCAACAACAAGA
DMC1, AG2-only	GGTTCTTTCCCCCTTTCTTTTACTAACTAATAATTTGGAAttaAAaaTTcATAGACCGTTTTGTCAACAACAAGA
APE2, Wt	TCGTTACCGACCTTTGAGTTCTTACTAACTTACATTCTCTATTAATTGcACCCAAAAGTACGAAGGACAACTAAAATGACCAG TAAAACCCCAAATCGTGAAATTCTTC
APE2, Destabilized	TCGTTACCGACCTTTGAGTTCTTACTAACTTACATTCTCTATTAATTGcACCCAAAAGTcCGAAGGACAACTAAAATGACCAG TAAAACCCCAAATCGTGAAATTCTTC
APE2, Stabilized	TCGTTACCGACCTTTGAGTTCTTACTAACTTACATTCTCTATTAATTGgACCCAAAaGTcCGAAGGACAACTAAAATGACCAG TAAAACCCCAAATCGTGAAATTCTTC
APE2, Hyperstabilized	TCGTTACCGACCTTTGAGTTCTTACTAACTTACATTCTCTATTAATcGgACCCAAAAGTcCGAAGGACAACTAAAATGACCAG TAAAACCCCAAATCGTGAAATTCTTC
AG3-Only, Destabilized	TCGTTACCGACCTTTGAGTTCTTACTAACTTACATTCTCTATTAATTGCACCCAAACGTCCGACGGACAACTAAAATGACCAG TAAAACCCCAAATCGTGAAATTCTTC
AG3-Only	TCGTTACCGACCTTTGAGTTCTTACTAACTTACATTCTCTATTAATTGcACCCAAAcGTcCGAcGGACAACTAAAATGACCAGT AAAACCCCAAATCGTGAAATTCTTC
AG3-Only, Hyperstabilized	TCGTTACCGACCTTTGAGTTCTTACTAACTTACATTCTCTATTAATcGgACCCAAAcGTCCGAcGgaCAACTAAAATGACCAGT AAAACCCCAAATCGTGAAATTCTTC
APE2, No-AG3	TCGTTACCGACCTTTGAGTTCTTACTAACTTACATTCTCTATTAATTGTACCCAAAAGTACGAAGGACAACTAAAATGACCAt TAAAACCCCAAATCGTGAAATTCTTC
Structureless	GCTCGTTACCGACCTTTGAGTTCTTACTAACTTACATTCTCTATTAAATATACCCAAAAGATCGAAGGACAACTAAAATGAC CAGTAAAACCCCAAATCGTGAAATTCTTCCTGATAAT
Linear AG1	CCGACCTTTGAGTTCTTACTAACTTAAAATAAAAAATGAAAAtCTAAAAAAGAAAAATgAAAAAATGAATGAAAaaaAAATG AAAAAATGAAAAAAACCCCAAATCGTGAAATTCTTCC

CHAPTER 3

DISCUSSION AND SIGNIFICANCE

In the context of alternative splicing that maintains a reading frame, alternative 3'SS selection accounts for roughly 30% of all alternative splicing events[94]. Yet, in most cases, the mechanisms for regulating alternative 3'SSs remain undefined. In yeast, Prp22p has been shown to support 3'SS fidelity and, more recently, has been implicated in modulating selection of 3'SSs *in vitro*[40,93]. In the results presented in the previous chapter, I tested for a role for Prp22p in regulating 3'SS selection *in vivo*. I found that Prp22p does affect alternative 3'SS selection of endogenous substrates (Figure 2.1). Because the corresponding 3' splice site regions exhibited RNA secondary structure potential, I hypothesized that Prp22p regulates 3'SS selection by antagonizing RNA secondary structure, rather than promoting a leaky-scanning mechanism[40]. Indeed, I found evidence that Prp22p promotes 3'SS accessibility and docking to the catalytic core by unwinding a stem-loop (Figure 2.4). Further, I found that Prp22, again by unwinding a stem-loop, can inhibit juxtaposition of a downstream AG with the branchpoint, reducing splice site reactivity (Figure 2.7). Our data highlight a tug-of-war between pulling by Prp22p and folding of RNA secondary structure, with the outcome impacting alternative 3'SS selection.

These findings have important implications for the regulation of 3' splice site selection. In our analysis of splicing of *DMC1* transcribed from its endogenous locus, I were surprised to observe, in haploid mitotic cells, that splicing of AG1 relative to AG2 was 5.4-fold higher for *PRP22* relative to mutant *prp22* (Figure 2.1a), given that splicing at AG1 would not produce full-length, functional Dmc1p. The consequence of splicing at AG1 is the inclusion of an in frame UAG (AG2) stop codon which would truncate 48 amino acids from the C-terminus of Dmc1p. Notably, *DMC1* is a meiotic recombinase whose expression is strongly upregulated during meiosis[138]. I suggest that splicing at AG1 could serve as a safeguard for preventing expression of functional Dmc1p in haploid, non-meiotic cells. Importantly, I show that splicing

at AG1 at the endogenous locus as well as on a reporter is dependent on *PRP22* and that splicing at AG1 on a reporter, at least, is repressed by sequestration in a stem loop (Figures 2.1a, 2.1c).

Lastly, our analysis implies that meiosis requires a shift in splicing from AG1 to AG2 and that this shift could be accomplished by either repressing Prp22p or stabilizing the stem loop. Previous evidence suggests that the *APE2* intron includes an RNA thermosensor that regulates splicing by changes in temperature[137,139]. Specifically, at 23 °C, *APE2* can form stable RNA secondary structure that favors splicing at the distal AG3, whereas at 37 °C, the stem partly melts, consequently favoring splicing at the more proximal AG2[108]. Consistent with a role for RNA secondary structure in modulating the relative usage of AG2 and AG3, I found evidence that wildtype *PRP22* promotes splicing at AG2 and represses splicing at AG3 by antagonizing RNA secondary structure (Figures 2.1b, 2.1d). These observations establish that Prp22p, in addition to temperature, modulates the *APE2* RNA thermosensor.

Given that alternative 3' splice sites generally compete with one another (e.g., AG1 and AG2 in *DMC1*; Figure 2.5), I were surprised to find that the alternative 3' splice sites in *APE2* do not compete (Figure 2.8). This substrate is also unusual in that the majority, more than 90%, of the lariat intermediates appear to undergo discard and degradation, given that hyperstabilization of the stem loop increases mRNA levels by more than 10-fold, with no change in lariat intermediate levels (Figure 2.7). I propose that such a high probability of discard of the lariat intermediate precludes competition between the alternative 3' splice sites. Specifically, in the wild-type *APE2* substrate, the probability of a lariat intermediate splicing at AG2 would be 5% and at AG3 would be 5%, such that mutation of one site, such as AG2, would liberate 5% of the lariat intermediates, but these lariat intermediates would be >10-fold more likely to undergo discard and degradation than splicing at AG3, thereby precluding an increase in splicing at AG3

because of repressing splicing at AG2. More broadly, this model suggests a novel mechanism by which spliced isoform levels can be regulated independently of one another.

The most obvious experiment to do to test whether the rate of discard is competing with re-sampling prompted by Prp22 pulling is to deplete Prp43 via auxin degron to test, rapidly, whether the absence of Prp43 leads to changes in 3'SS selection. Depleting Prp43 rather than using an ATPase mutant since the mutant would still maintain the ability to bind the spliceosome and promote the discard confirmation would be essential for testing this hypothesis. Coupling this system with a bi-fluorescent splicing reporter such as the one used by Orengo and colleagues[140] and swapping in the intron of Ape2, for example, would be essential for looking at the ratios of sequestered to unsequestered 3'SSs. For clarity, FACS sorting should be used for quick optimization and real-time and reproducible analysis of the splicing events. Furthermore, it would be essential to titrate the auxin in the system to look at the effect of different amounts of Prp43. In addition, it would be worth considering using the destabilized version of GFP, d2GFP, whose half-life is dramatically shorter to ensure the signal is reasonably reporting on real time splicing. Likewise, a fluorescent protein with a similar maturation time and half-life should also be used to ensure ratios in fluorescence are reasonably proportional. Technical details on the fluorescent proteins can be found here https://www.fpbase.org/table/. Assuming the splicing of the sequestered 3'SS leads to the production of RFP and distal 3'SS usage leads to GFP, you'd expect depletion of Prp43 would lead to increased unwinding by Prp22 devoid of discard which would lead to usage of the sequestered 3'SS and increased RFP signal. If discard (the action of Prp43) is not limiting the usage of the sequestered 3'SS, you'd expect no increase in the RFP signal since splicing at the downstream site would be in frame with GFP. As a control, an auxin degron should be placed on Prp22 or Slu7 to set the baseline signal as a function of time since

the addition of auxin since splicing should not occur in the absence of Prp22. In optimizing the amount of time it takes to begin seeing a difference in signal as a function of second step splicing, these controls will be essential.

Although I have shown *in vivo* that Prp22p activates sequestered 3'SSs and antagonizes downstream 3'SSs by disrupting RNA secondary structure, our data also reveal a limit to this activity of Prp22p, consistent with the role for such secondary structure in juxtaposing a distal 3'SSs with the branch site[108,135]. Specifically, when the *DMC1* or AG1-only substrates or the *APE2* or AG3-only substrates were hyperstabilized, the ability of Prp22p to activate the sequestered 3'SSs was diminished (Figure 2.4, 2.7). At face value, these results are surprising, because previous *in vitro* helicase assays of Prp22p imply that Prp22p can separate duplexed RNA with stabilities greater than -60 kcal/mol[78,141], much more stable than the hyperstabilized stem loops of *DMC1* (-5.2 kcal/mol) and *APE2* (-10.3 kcal/mol). Note: the stability of the duplex's studied in Schwer 1998 and Tanaka & Schwer 2005 are calculated using the same approach as those outlined for calculating the stabilities of the stem-loops using in this study[78,141]. This difference could reflect differences in the assays; for example, in our splicing assays competing processes (i.e., lariat intermediate discard or exon ligation) limit the time for Prp22p-mediated unwinding, which precedence implies decreases in rate with increasing stem stability[141]. Indeed, with the hyperstabilized AG1-only *DMC1* substrate, relative to the wild-type AG1-only substrate, the decrease in splicing at AG1 was not accompanied by a concomitant increase in lariat intermediate, implying discard of the lariat intermediate and competition between Prp22p-mediated unwinding and Prp43p-mediated discard[127]. Further, with both the wild-type and AG3-only *APE2* substrates, Prp22p-mediated unwinding competes with splicing at AG3, and when the stem is hyperstabilized, Prp22p must fail to unwind the stem loop on the

timescale that AG3 is selected for splicing (Figure 2.7). In addition to differences in the assays, faster refolding of the hyperstabilized stem loops may mitigate the impact of Prp22p-mediated unwinding. Lastly, the structured substrates may be more stable than the secondary structure predictions suggest. Indeed, the cryo-EM structures of endogenous, post catalytic spliceosomes reveal that a stemlike structure between the branch point and 3'SS inserts into a tunnel that could function to stabilize RNA secondary structure during 3'SS selection[72,73].

Previously, I published evidence that Prp22p, as well as Prp16p, function as pullases to disrupt interactions between the spliceosome and the substrate from a distance, and cryo-EM structurers of the spliceosome have supported this model[142,143]. In the case of Prp22p, our evidence indicated that to disrupt the interaction between the spliceosome and the mRNA, Prp22p needed to translocate across the 3' exon but not across the 5' exon, the exon most stably bound to the spliceosome[40]. Importantly, these observations imply that Prp22p can translocate to the catalytic core, represented by the exon-exon junction. At the exon ligation stage, before a 3'SS is docked, the branch site of the lariat intermediate represents the catalytic core. Thus, our data suggest that in destabilizing the stem loops of *DMC1* and *APE2*, Prp22p can translocate through these structures, disrupting these stems by direct unwinding.

A role for a helicase in promoting alternative splicing through disruption of RNA secondary structure was first discovered through studies of the DEAD-box helicase DDX5 in humans[116]. DDX5 promotes inclusion of exon 10 in *Tau* by disrupting a stem-loop in the intron, a stem that would otherwise sequester the 5' splice site, to promote U1 binding to the 5'ss. Similar to the sequestered AG1 of *DMC1* (Figure 2.4), increasing the stability of the stem-loop decreased splicing of the sequestered 5'SS, and mutations that destabilized the stem-loop increased usage of the sequestered 5'SS. The stem-loop that occludes the 5'SS in *Tau* is remarkably conserved

between multiple species, and analysis of differences between species with single base mutations shows effects on 5'SS usage equivalent to that observed from mutational analysis of the stem-loop in humans[144]. Analysis of RNA structural stabilities around constitutive versus alternative 5'SSs and 3'SSs in humans, mice, and flies have shown alternative splice sites tend to fall within regions of more stable RNA secondary structures[145,146]. Further, in a fraction of these cases, the stable RNA structure around conserved alternative splice sites is itself conserved[146]. These observations suggest a general role for the regulation of RNA secondary structure in the regulation of alternative splicing. Whether Prp22p can antagonize RNA structures in humans, as it does in yeast, remains to be determined.

In yeast, the 3'SS is recognized during the exon ligation step of splicing[72]. In humans, after the 3'SS is first recognized during the earliest stages of spliceosome assembly, the 3'SS is recognized again during the second step of splicing by the same mechanism, so this newly discovered function of Prp22p in regulating 3'SS selection may be conserved between yeast and humans, especially when a pair of alternative 3'SSs are within 9 nucleotides of one another, pairs that are highly likely to share a branch point and undergo regulation at the exon ligation stage[94,109,147]. Indeed, even though U2AF1 can recognized a 3'SS early in spliceosome assembly, the spliceosome can splice at a different AG during exon ligation. For example, 3'SS selection in intron 2 of *Sex-lethal* in female fruit flies is regulated by SPF45 during the exon ligation step of splicing by directing the spliceosome to an AG that differs from the AG defined by U2AF1 during spliceosome assembly[148], demonstrating not only that the AG used for exon ligation can be different from that defined by U2AF1 but also that a trans-acting protein factor can influence AG selection during exon ligation. In this regard, it is notable that 20% of alternative splicing events that maintain the reading frame are regulated NAGNAGs, adjacent 3'SSs, which are

highly unlikely to be regulated by the action of U2AF during spliceosome assembly, given their close proximity[94,109,147] and commonly shared branch point. Indeed, analysis of NAG$(N)_x$NAGs, in *C. elegans*, where x=3,6, or 9, showed that knockdown of the U2AF1 ortholog did not lead to any changes in 3'SS selection, indicating a regulatory mechanism distinct from altered recruitment of U2AF1[147]. Further, the splicing of some human introns is independent of early recognition of the 3'SS AG; indeed, truncated lariat intermediates lacking sequences downstream of the branch site, including the 3'SS, stall after branching in human nuclear extracts and yet can still undergo exon ligation when a substrate with an AG is added in trans suggesting recognition of the 3'SS can occur independent of U2AF and during the second step of splicing[98,149–151]. Lastly, in humans, as in yeast, cryo-EM structures of the spliceosome have revealed Prp22p as the only ATPase bound to the splicing substrate downstream of the branch site and poised for pulling[69,152].

APPENDIX A

IN VITRO SPLICING ASSAYS OF STRUCTURED ACT1 SUBSTRATES REVEAL THE ATPASE ACTIVITY OF PRP22 IS NECESSARY FOR SPLICING OF A SEQUESTERED 3'SS

In attempting to test the role of Prp22 to regulate sub-optimal 3'ss usage, I inadvertently created substrates with the 3'ss sequestered in secondary structure, as predicted using RNAFOLD and mfold[132,133] and found an unexpected result. I found that 3' splice sites sequestered in stem loops can be activated solely by Prp22 but not by ATPase deficient Prp22. Interestingly, the function of Prp22 to activate 3' splice sites sequestered in stem loops had not previously been tested even though Prp22 is a helicase that was characterized as unwinding RNA. To more thoroughly test if this effect was due to Prp22, I made splicing extract devoid of endogenous Prp22 (Figure A.1a) and assayed for splicing of the structured substrate (Figure A.1b) in the presence of recombinant Prp22 or ATPase deficient Prp22. Note: To create splicing extract devoid of endogenous Prp22 I used the auxin degron system which is advantageous to commonly used immuno-depletion strategies because it i) does not require long incubations for depletion ii) is cost-effective and iii) allows for nearly 100% depletion of target protein in less than 45 minutes. The result of these assays, shown in Figure A.1c, showed that the depletion of Prp22p inhibited the second step of splicing only and not the first step (compare mRNA in lane 1, indicative of second step, and lariat intermediate, indicative of first step). Second, the mock depletion of yeast by adding auxin to yeast cells that had the auxin tag but lacked the protease for induced depletion were able to splice (compare mRNA in lane 2 to mRNA in lane1). These results indicate that I was able to deplete the essential second step splicing factor via the auxin degron system. Therefore, Prp22 is necessary for splicing of the structured substrate. Next, I added back recombinantly purified Prp22p to the depleted extract and assayed for splicing. For both the mock depleted extract and the depleted extract, splicing of the substrate occurred. These results suggest that the effect I are seeing because of Prp22 depletion is in fact Prp22-dependent. That is, Prp22 is necessary for splicing of a sequestered 3'SS. Next, since I now know that

Prp22p is required for splicing of the structured substrate in the depleted extract, I could now ask whether the ATPase function of Prp22 is required for splicing of a structured substrate as opposed to the mere presence of Prp22 (i.e. it's ATP-independent function). Indeed, addition of the dominant negative mutant of Prp22 which lacks ATPase activity was unable to splice the structured substrate. These findings, coupled with the findings in Chapter 2, suggest the ATP-dependent function of Prp22 is required for activating the sequestered 3' splice site in vitro, as it does in vivo. These findings and experimental approach set the stage for testing the pulling model in vitro.

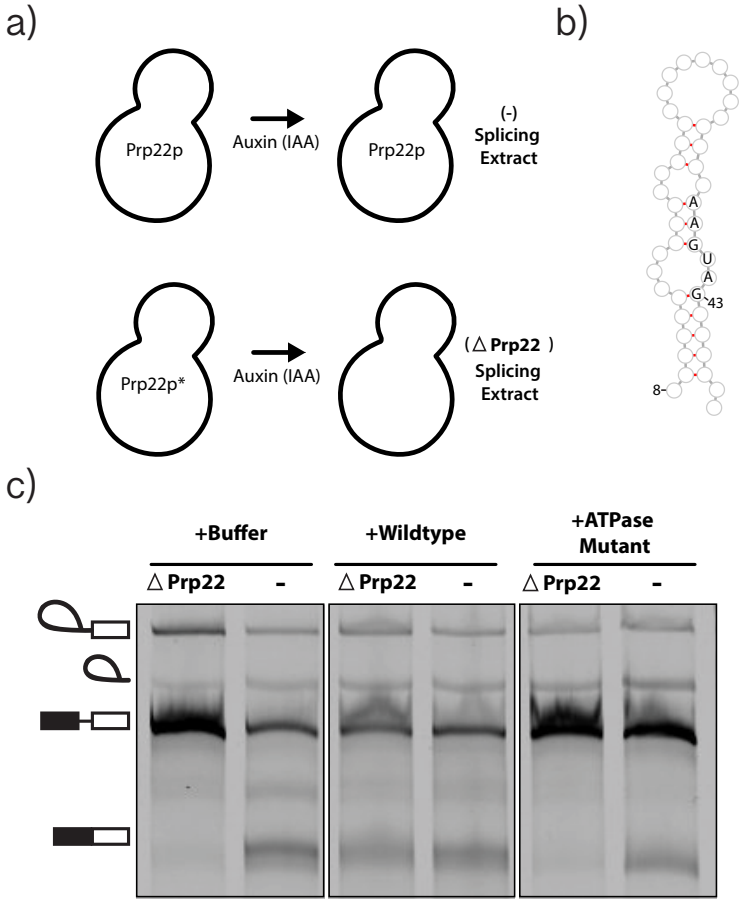


Figure A.1. Prp22p is necessary for splicing of a structured substrate. (a) Sole copies of Prp22 were tagged with the degron peptide sequence in cells that harbored the osTir1 protease (Prp22*) or in cells that did not (Prp22p). Cells were treated with auxin for 45' before being spun down and processed like normal splicing extract protocols. Splicing assays were carried out using the depleted or mock depleted extract on the structured substrate shown in (b). Products were visualized (c) after separation on a DPAGE gel via Cy5 signal on the pre-mRNA substrate.

APPENDIX B

PRELIMINARY ANALYSIS OF DHX8 FUNCTION IN MAMMALIAN CELLS

The importance of comprehending how the spliceosome searches for the 3'ss is directly relevant to understanding 3'ss usage, regulation, proteome plasticity, tissue specific function, and developmental regulation[94,95,147]. For example, in humans, there are over 2,000 genes that contain contiguous 3'ss sequences known as NAGNAG's. The use of the upstream or downstream NAG's is regulated during development and in a tissue-specific manner. Strong regulation (i.e., switching from upstream to downstream NAG) occurs in 1,460 of these genes and site selection has been shown to alter the function of individual proteins[94,95]. Further, 28% of SNPs located in NAGNAG sequences are linked to disease causing genes[95]. Currently, the mechanism for how these sites are regulated is unknown. However, given our findings that RNA structure can effect nearby 3'ss usage, I hypothesize that Prp22 is a target for regulating these sequences[40].

Studying the ATP-dependent function of Prp22 in vivo is complicated by its function as an essential splicing factor. Knocking out Prp22 in most eukaryotes is lethal. Historically, essential genes have been studied by identifying conditional mutants that act similar to wild-type at permissive temperatures but are mutant at lower temperatures. These mutants are typically identified by a screen or rational design. However, those tools are not available for Prp22 in human cells and so to circumvent this issue, I designed a FLP-FRT TREX cell line of presumed dominant negative mutations for Prp22 (Figure B.1). These mutations are presumed dominant negative because they are mutations to essential residues in conserved motifs such as the lysine in the conversed ATP-binding motif (G**K**T). For a more in depth review of the system, see[153].

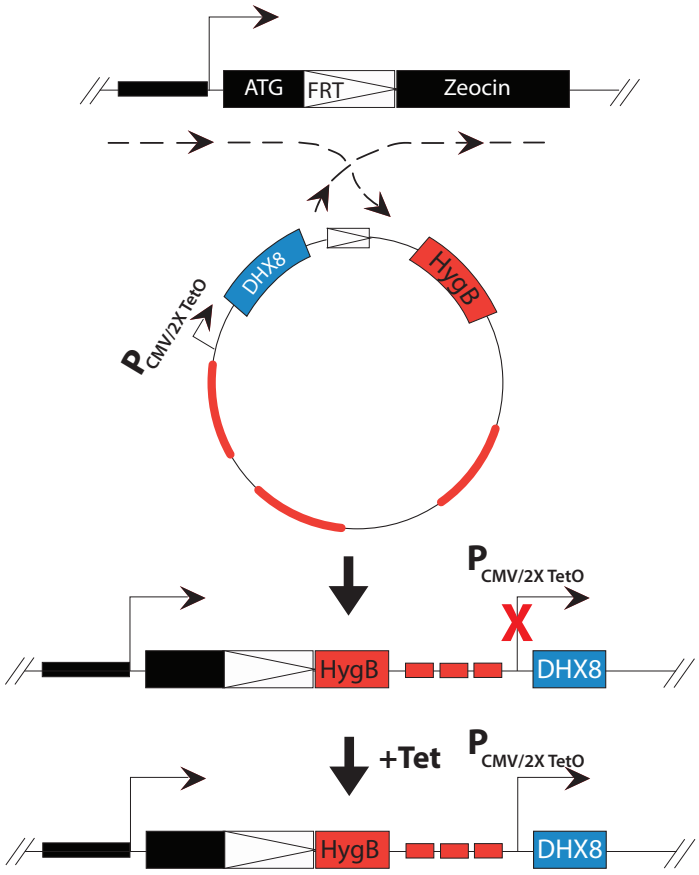


Figure B.1. The general scheme for integration of DHX8 into the genome using FLP recombinase. DHX8 is co-transfected into HEK293 cells with a recombinase that promotes integration into the FLP recombinase target (FRT) site. Positive clones are then selected for using Hygromycin B antibiotic. Not shown is the TetR protein which binds the TetO sites and represses expression of the downstream gene. Expression of said gene is induced by the addition of Tetracycline or the more potent Doxycycline which decreases the affinity of the TetR protein for the TetO sites which then allows for activity of the gene's expression.

Characterization of Dhx8 and Dhx8-K594A under the control of the Tet-on promoter

To gain a better understanding of the inducible system, and to characterize the effect of Dhx8-K594A, I carried out immunofluorescent staining (IF) and RNA-Fluorescent In-Situ Hybridization (RNA-Fish). In short, induction of mutant, but not wild type, led to a cellular and molecular phenotype suggesting mutant can outcompete endogenous Dhx8 *in vivo*. In the context of the cellular phenotype, induction of mutant at 24 hours led to an abnormal number of nuclei within a single cell and increased as a function of time. A molecular phenotype was seen in mutant cells upon analysis of nuclear speckles with probes against U2 snRNA provided by Dr. Jingyi Fei. Nuclear speckles, thought of as either storage or RNA processing centers, become diffuse in induced mutant cells but not in cells that are un-induced (Figure B.3B). While this finding could potentially be exciting, I hypothesize DHX8 is a fidelity factor and therefore the phenotype would need to be observed at an earlier time point to ensure the effect is due to the primary effect of Dhx8 on splicing rather than secondary effects of mis-spliced transcripts. Furthermore, it would be worth synchronizing the cells, releasing the block, and then visualizing the cells at different time points to ensure the analysis is robust and not a mixture of different cell states.

Profiling of the protein expression levels of induced protein relative to endogenous protein

To access the level of induced protein expression relative to endogenous Dhx8, I conducted a 24-hour time course and visualized the levels of induced Dhx8 protein. As seen in Figure B.2., the expression of induced protein (note: M2-Flag denotes induced DHX8 protein) in the nucleus, between 6 to 24 hours, increases linearly as a function of time. The mutation does not seem to alter translation. In the future, calibration of the minimal expression time relative to the onset of a mutant phenotype such as changes in splicing will be needed.

Changes in the epigenetic and transcriptional profile of cells as a function of DHX8's mutant phenotype

Given splicing occurs co-transcriptionally, I blotted for changes in post translational modifications known to control transcriptional dynamics; histone modification, H3K36me3, and phosphorylation of serine 5 on the C-terminal domain of RNA polymerase 2 (Pol2-Ser5). Both modifications are also thought to be enriched at the 3'ss. In addition, RNA-FISH was also carried out to look for increases in RNA in the nucleus (a known phenotype of Prp22 mutants) and changes in splicing speckles. RNA fish on DHX8-K594A mutant shows diffusion of U2snRNA granules in mutant induced cells indicative of changes in RNA processing.

While changes in the H3k36me3 and Pol2-Ser5 modifications have been shown to effect splicing by controlling the speed of Pol2, it is still unclear whether changes in splicing can affect the regulation of these sites. Excitingly, my preliminary data suggests there is a change between mutant and wild type (Figure B.3). The data suggests these modifications are regulated by Dhx8's ATPase function, either directly or indirectly. Notably, enzymatic digest of chromatin to di-nucleosomes followed by fractionation of chromatin has shown that Pol2-Ser5 and H3k36me3 appear enriched in the same fraction after separation via glycerol gradient suggesting the two modifications are, at face value, concurrent[159]. Furthermore, mass-spec analysis of IP'd Pol2-Ser5 revealed a highly enriched Dhx8p. Interestingly, in a conversation with Dr. Jean Beggs about my results, she noted that inducing expression of a lethal dominant negative overactive ATPase mutant of Prp22, I764A, leads to H3k4me3 levels like wildtype but that induced expression of a dominant negative mutant of Prp22p with decreased ATPase activity leads to significant drops in H3K4me3 near promoters of genes with introns. IP of Prp22 precipitates with SET1, the writer of H3k4me3 and treatment of the IP with RNase does not affect the

pulldown. Taken together, these results suggest strong preliminary data that DHX8 may be signaling an active spliceosome. Note: Both mutants tested by Dr. Beggs were independently designed and validated (see Table B.1.)

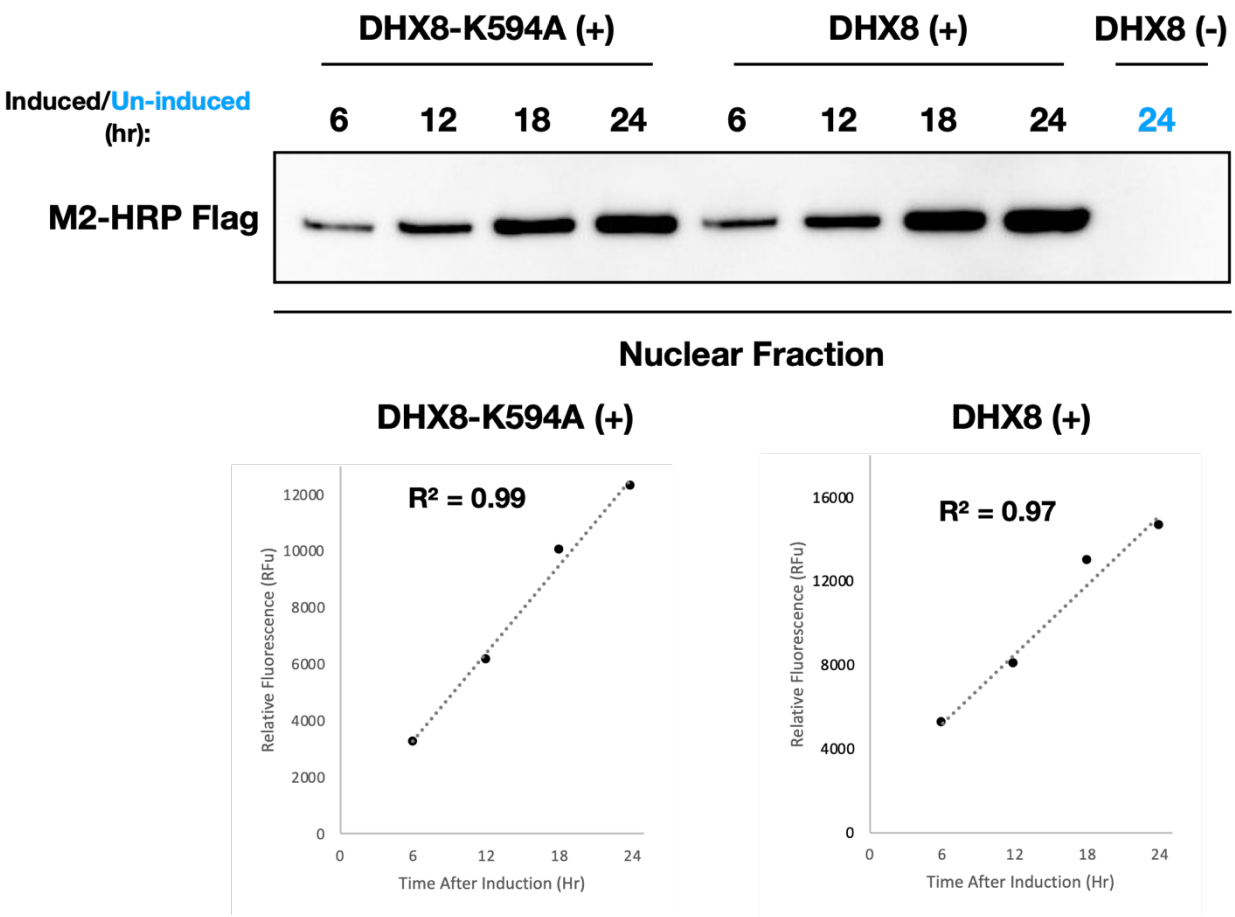


Figure B.2. Linear expression of induced DHX8 HEK293 engineered cells with dox. HEK293 FLP-FRT Trex cells were co-transfected with either pcDNA5-DHX8 and pOG44 or pcDNA5-DHX8(K594A) and pOG44 and single clonal populations were selected after 20-30 days of hygromycin selection. Monoclonal populations were independently grown to about 50% confluency and induced at the same time with freshly prepared and filter sterilized Doxycline. The final concentration of dox was 5ng/uL. Individual time points were taken at the annotated time.

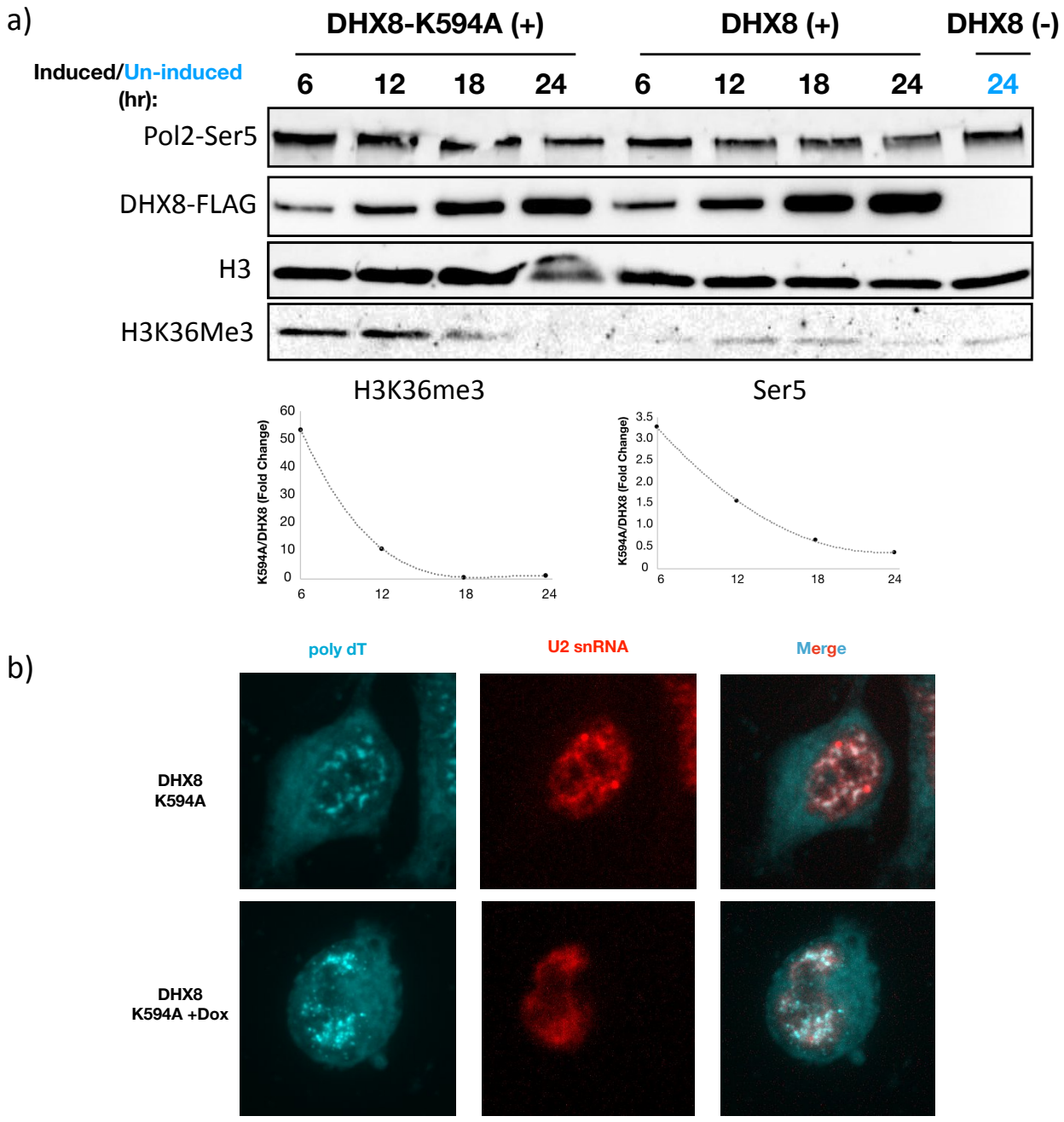


Figure B.3. Induction of mutant K594A causes changes in the epigenetic profile of the cell in addition to altering the dynamics of splicing speckles. (a) See Figure B.2. for induction conditions and explanation of experimental conditions. Mouse monoclonal anti-Pol II CTD, phospho Ser5, MBL international, Cat# MABI0603; Mouse monoclonal anti-H3, MBL international, Cat# MABI0301; Rabbit monoclonal Tri-Methyl-Histone H3 (Lys36), (D5A7) XP® Rabbit mAb, Cell Signaling, #4909 Cat#4909S; Monoclonal ANTI-FLAG® M2-

Figure B.3 (continued) Peroxidase (HRP) antibody produced in mouse, SIGMA, Cat#A8592. (b) Poly dT is labeling total RNA, U2snRNA is labeling U2 snRNA. Probes provided by Dr. Jingyi Fei[160].

Recent meta-analysis from encode confirms my initial hypothesis that DHX8 is in fact a bonafide proofreading factor in humans, mice, and flies. In fact, cryptic splicing of the 3'SS is most specific for the knockdown of DHX8[161]. That is, knockdown of other factors leads to a higher number of mis-splicing events but only the knockdown of DHX8 is as specific to 3'SS events. Whether this is due to DHX8's effect on RNA structure between the branch and 3'SS or just proofreading sub-optimal sites remains to be determined. To more specifically interrogate DHX8, I propose using a library of rationally designed sequence variants between the branch site and 3'SS[162] in the background of either wildtype Tet-inducible DHX8, DHX8(mutant of choice), or an empty vector (for a direct comparison to no changes in DHX8 expression but still in the presence of doxycycline). The use of this system, which also has barcodes on the 5'exon would allow extremely deep sequencing of the changes in splicing without requiring exhaustive library preparations. The use of only one genomic region would also be highly advantageous and minimize secondary effects from chromatin states (euchromatin vs. heterochromatin for example). This would all also be able to be carried out with DHX15 (Prp43), DHX38 (Prp16), BRR2 (snRNP200), and Aquarius (mammalian specific helicase). The cell lines that were initially made but lost due to an electrical issue with the tissue culture freezer, are listed in Table B.1. Transfection and re-generating these cells lines should take roughly 2 months. These plasmids will need to be re-introduced into HEK293 FLP-FRT TREX cells or used for transient expression in cell lines lacking the Tet repressor. The initial library took one month to design, clone, and sequence validate, two months to generate positive cell lines, and one month of data analysis (data shown in these appendix). Id also recommend a quicker solution being the use of mRNA. I designed these constructs with this in mind. The constructs contain a T7 promoter. Id recommend using alpha-globin as the 5'UTR or generating a linear 5'UTR that is minimally 40

nt long. A T2A site downstream of, for example, DHX8, and followed by a fluorescent protein would also be essential for visualizing the transfection and rapidly optimizing transfection conditions.

Importantly, in correspondence with Dr. Eran Segal, he agreed to send us the cell lines or collaborate to carry out this project. Dr. Segal has also constructed libraries for screening 5'SS and sequences in the exon. Each of these libraries would be of high interest for screening the helicases of the spliceosome. Again, this approach is superior (cost-wise and data-wise) to any other approach I currently have for analysis of helicases on fidelity.

As a second aim, I propose the following. Inducing expression (CRISPR-ON) or decreasing expression (CRISPR-OFF)[163] of DHX8 to test whether the potential changes at NAGNAG's are due to changes in expression of DHX8 in different cell types. For clarity, inducing expression in cells with low expression of DHX8 and decreasing expression of cells with High DHX8. Note that the GTEx portal (https://gtexportal.org/home/) provides tissue expression analysis of genes on a graphically rich interface that does not require coding. Preliminarily, I'd recommend referencing the NAGNAG data from Bradley et al.[94] from different tissue types and use that as a guide to design NAGNAG primers. I would then select high and low expression lines for analysis that are commercially available. For example, DHX8 expression is slated from highest to lowest expression as: Thyroid> Breast> Adrenal> Muscle> Kidney> Liver. I'd order these cell lines from ATCC and conduct the experiments I highlighted above to look for changes in NAGNAG. Note: the specifics of the guide RNA for ON and OFF switches are provided in the CRISPR ON-OFF paper and has already been optimized for most protein coding genes in the genome[163]. Furthermore, I have received the recently developed plasmids and protocols from Dr. James Nuñez and have received formal permission for our lab

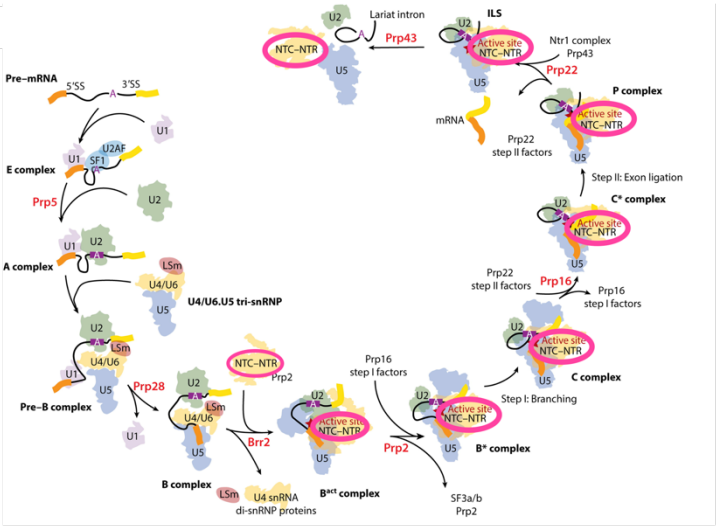
to use them. These experiments, though secondary to the initial set of experiments described, could be done in parallel.

Rapid IP of Prp19 for sequencing splicing intermediates, and mRNA

To profile spliceosomes, I developed a rapid method for IP of spliceosomes, which captures splicing intermediates and mRNA. To do this, I tagged Prp19, a tetramer that joins the spliceosome before the first step of splicing and is present through exon ligation, using a newly developed tag, CL7, which binds its target antigen with a $K_d < 10^{-14}$ M (Figure B.4)[164]. The choice of tagging Prp19 is two-fold, it's present for all catalytic steps in splicing and the fact that it's a tetramer increases the possibility of a successfully IP. Note, the tag was engineered to be exposed and is not anticipated to interfere with binding to the spliceosome. Without optimization, I was able to harvest induced cells, pulldown in-tact spliceosome by IP of CL7-Prp19 (Note: Prp8 is a core splicing component and is indicative of an assembled spliceosome) and elute spliceosomes via protease cleavage (3 min) in less than two hours. With optimization, or by using a liquid handler, this approach would be quicker. The workflow is outlined in Figure B.5. I designed, cloned, and developed the idea. I created the cell lines with Matthew McDonough. This approach could be used for general spliceosome profiling or could be coupled to analysis of a helicase of interest via a bi-directional inducible promoter. Cells assayed under specific conditions could be assayed rapidly and capture all targets independent of non-sense mediated decay. Furthermore, Cl7-Prp19 mRNA could be used to transfected into patients' samples, or other cell types for the purpose of profiling spliceosome bound species independent of ribosomal RNA or other RNA that often plague chromatin enriched RNA preps, for example. I'd recommend altering the 5'UTR and 3'UTR of the construct to Beta-globin(5'UTR) and tandem repeats of alpha globin (3'UTR)[165]. An interesting application for this development

would be as a diagnostic for non-sense mediated decay. Nonsense mutations currently account for "10%–15% of all genetic lesions leading to disease, causing nearly 1000 serious genetic disorders[166,167]." Determining transcripts that could potentially be NMD targets is not trivial given the low abundance species are often degraded. Given the breakthrough development of suppressor tRNA's, the profiling of NMD targets in research and development, or as a diagnostic, have become of the upmost importance. This approach provides the solution.

a)



b)

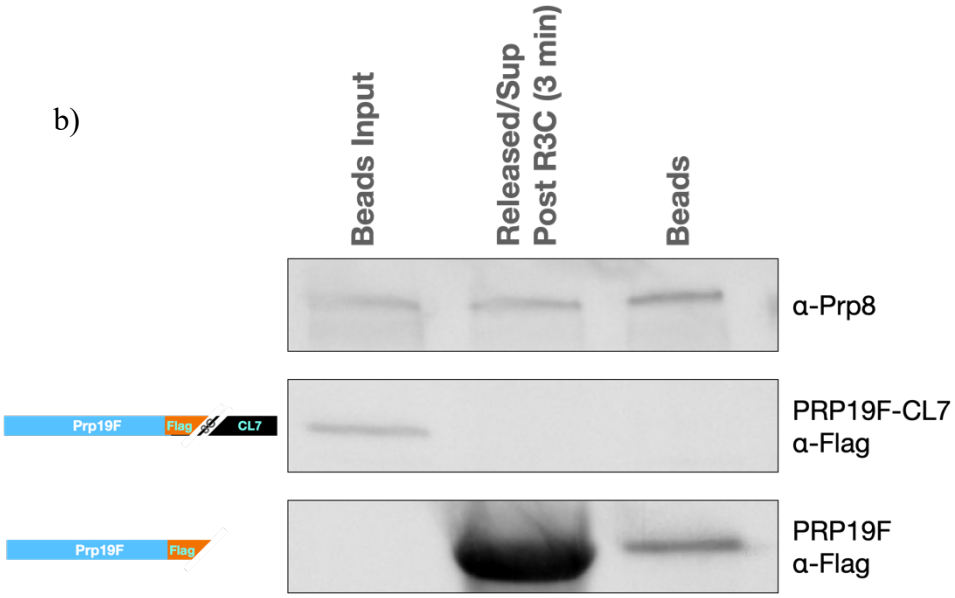


Figure B.4. Prp19 is present throughout catalysis and serves as a handle for profiling spliceosomes. (a) Prp19 joins the splicing cycle prior to the first step of splicing and is present throughout exon ligation. The image in **(a)** is pulled from[142]. **(b)** Following lysis "input," Prp19 lysate was incubated on beads for one hour before being washed with a modified version of RIPA buffer and eluted via R3C protease. The sample was spun down and the supernatant was collected "Release/Sup Post R3C). Pelleted "beads" were assessed for unreleased complex. PRP8 Recombinant Rabbit Monoclonal Antibody (JG57-38), ThermoFisher, Catalog # MA5-34736, α-Flag, Cell Signaling Catalog #14793, Clone D6W5B

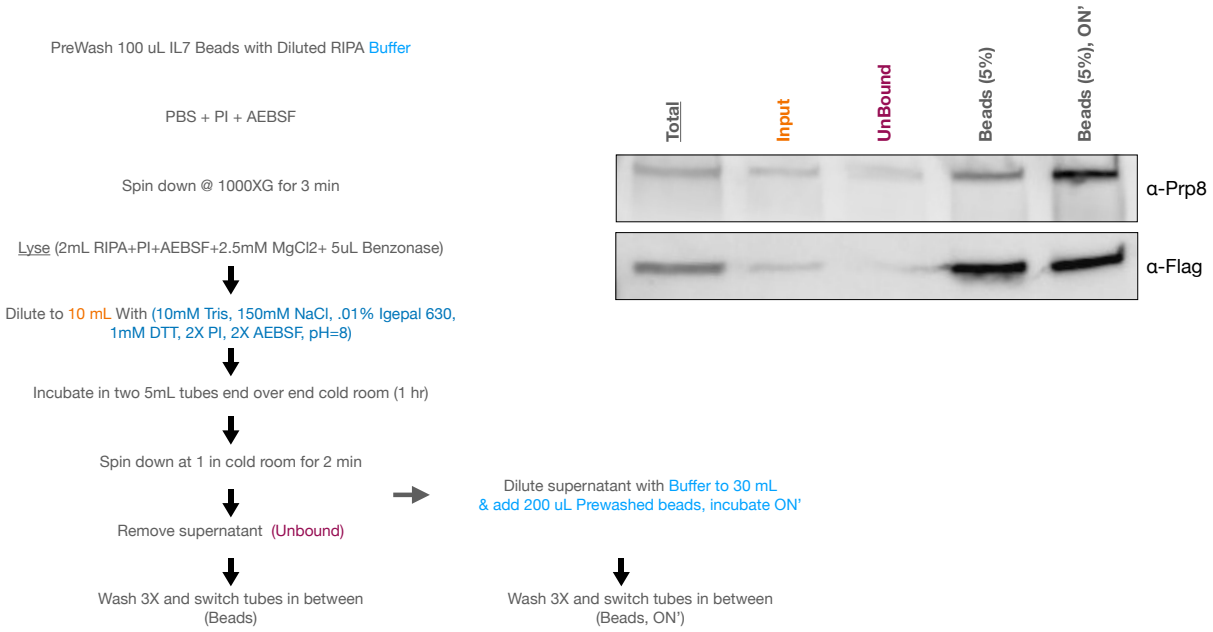


Figure B.5. Workflow/Protocol for isolation of Prp19 from HEK293 cells tagged with CL7-Prp19 under the control of a Tet-On promoter. Induction at 1ug/mL final concentration (fresh Doxycycline hyclate made up in water) added dropwise to 70% confluent Hek293 Trex cells (Thermo #R71007) harboring inducible pcDNA5-Prp19F-Flag-R3C-CL7 gene. About 20 million cells were pelletd and resuspend in 2mL RIPA*(150mM NaCl, 1% Igepal-630 (NP-40), 0.5% Sodium deoxycholate, 0.1% SDS, 50 mM Tris, 1mM EDTA, 2X protease inhibitor, 2mM DTT, 2X AEBSF 2.5 mM MgCl2, pH=7.5 @ 4C +10µL Benzonase (E1014-25KU, Sigma)). 100 µL out of 2 mL (5%) saved for western, labeled "**Total**". 8 mL of buffer* (10mM Tris pH=8, 150mM NaCl, .01% Igepal 630, 2mM DTT, 2X AEBSF, 2.5 mM MgCl2 added to lysate. 100 µL (1%) out of 10mL saved for western "**Input**." After 1hr incubation on 100µL IL7 beads, tubes spun down and 100µL supernatant was removed and saved for western "**Unbound**." The remainder of the supernatant was collected and further diluted to 30 mL in buffer lacking MgCl2* (10mM Tris pH=8, 150mM NaCl, .01% Igepal 630, 2mM DTT, 2X AEBSF) and incubated overnight with 100 µL of IL7 beads **"Beads, ON'**. After 3 washes with Buffer (10mM Tris pH=8, 150mM NaCl, .01% Igepal 630, 2mM DTT, 2X AEBSF), the original beads were saved for downstream analysis "**Beads**." PRP8 Recombinant Rabbit Monoclonal Antibody (JG57-38), ThermoFisher, Catalog # MA5-34736, α-Flag, Cell Signaling Catalog #14793, Clone D6W5B

Table B.1. Plasmids Generated

Gene	Modification	N-Tag	C-Tag
DHX8	K594A		Flag
DHX8	T839A		Flag
DHX8	H688A		Flag
DHX8	T847V		Flag
DHX8	I846A		Flag
DHX8	T847A		Flag
DHX8	WT		Flag
DHX8	WT-CL7 Tag	CL7-R3C	Flag
DHX8	WT-3XHA	3XHA	Flag
DHX8	K594A-CL7 Tag	CL7-R3C	Flag
DHX8	K594A-3XHA	3XHA	Flag
DHX8-K594A	T847V		Flag
DHX8-K594A	T847A		Flag
DHX8-K594A	D685A		Flag
DHX8-K594A	H688A		Flag
DHX8-K594A	I846A		Flag
DHX8-K594A	T839A		Flag
DHX8-K594A	K594A		Flag
DHX15	Q846A		Flag
DHX15	T421V		Flag
DHX15	T167A		Flag
DHX15	T429V		Flag
DHX15	WT		Flag
DHX38	WT		Flag
DHX38	T805V		Flag
DHX38	T813V		Flag
DHX38	G560A		Flag
DHX38	T562A		Flag
DHX38	K561A		Flag

Table B.1 (continued)

Gene	Modification	N-Tag	C-Tag
DBR1	Wt		Flag
DBR1	N84A		Flag
DBR1	H85S		Flag
DBR1	H85A		Flag
DBR1	H226A		Flag
DBR1	WT	CL7-R3C	Flag
DBR1	N84A	CL7-R3C	Flag
DBR1	H85S	CL7-R3C	Flag
DBR1	H85S	CL7-R3C	Flag
DBR1	H226A	CL7-R3C	Flag
DBR1	WT	3XHA-TEV	Flag
PRP19	WT	3XHA-TEV	Flag
PRP19	WT	CL7	Flag
GFP	AQR	GFP	

Table B.2. Oligos used to make mutants

Name	Sequence
CMV -450	TTTGGCAGTACATCAATGGGCG
DHX15-900	ATGAAGCCTGTAAGAGAATAAAGCG
DHX15-1575	ATTTTATGGATCCACCAGCTCC
DHX15 -2100	ATGGACAGATTTAATTTGCCTCG
DHX15-Q468A ADB F	GGTGACAGCTATTAGTAAAGCTTCAGCTCAGGCCAGGGCTGGTCGAGCT GGACGTACCAGACC
DHX15-Q468A ADB R	GGTCTGGTACGTCCAGCTCGACCAGCCCTGGCCTGAGCTGAAGCTTTAC TAATAGCTGTCACC
DHX15-Q468A ADL F	TCGAGCTGGACGTACCAGACC
DHX15-Q468A ADL R	TGAAGCTTTACTAATAGCTGTCACC
DHX15-T167A AD F	TGGTTGGTGAGACTGGGTCTGGTAAAGCCACACAGATTCCACAGTGGTG TGTGGAG
DHX15-T167A AD R	CTCCACACACCACTGTGGAATCTGTGTGGCTTTACCAGACCCAGTCTCAC CAACCA
DHX15-T167A ADLin F	ATTCCACAGTGGTGTGTGGAG
DHX15-T167A ADLin R	TTTACCAGACCCAGTCTCACC
DHX15-T421V Uncoup F	GGAGCAATTGGAAGAAAGGTAGTTGTGTCAGTGAACATAGCAGAGACG TCTTTGACAATAGATGG
DHX15-T421V Uncoup R	CCATCTATTGTCAAAGACGTCTCTGCTATGTTCACTGACACAACTACCTT TCTTCCAATTGCTCC
DHX15-T421V UncoupLin F	AACATAGCAGAGACGTCTTTGAC
DHX15-T421V UncoupLin R	AACTACCTTTCTTCCAATTGCTCC
DHX15-T429V UncoupRB F	GGTAGTTGTGTCAACTAACATAGCAGAGACGTCTTTGGTGATAGATGGT GTGGTGTTTGTGATTGATCC
DHX15-T429V UncoupRB R	GGATCAATCACAAACACCACACCATCTATCACCAAAGACGTCTCTGCTA TGTTAGTTGACACAACTACC
DHX15-T429V UncoupRBLin F	ATGGTGTGGTGTTTGTGATTGATCC
DHX15-T429V UncoupRBLin R	AAGACGTCTCTGCTATGTTAGTTGAC

Table B.2 (continued)

Name	Sequence
DHX38-600	ATGGAGGGTCAGAGCGTAGC
DHX38-1200	ATCAATGAGGATAACGAGCG
DHX38-1550	ATGAAGAGAAAGAGCGAAGC
DHX38-2190	AAGCAGTCCTTGCAGGTGC
DHX38-2850	AACATGCTCAACTCTATGTATCAG
DHX38-3360	ATTACATAGTGTATCACGAGTTGG
DHX38-560/562 ADLin F	TACCTGCATGAAGATGGTTACACG
DHX38-560/562 ADLin R	AACCACGATCACGATGCTG
DHX38-G560A AD F	CAGCATCGTGATCGTGGTTGGGGAGACGGGGAGTGCCAAGACCACTCA GCTGACGCAGTACCTGCATGAAGATGGTTACACG
DHX38-G560A AD R	CGTGTAACCATCTTCATGCAGGTACTGCGTCAGCTGAGTGGTCTTGGCAC TCCCCGTCTCCCCAACCACGATCACGATGCTG
DHX38-K561A AD F	CAGCATCGTGATCGTGGTTGGGGAGACGGGGAGTGGTGCCACCACTCAG CTGACGCAGTACCTGCATGAAGATGGTTACACG
DHX38-K561A AD R	CGTGTAACCATCTTCATGCAGGTACTGCGTCAGCTGAGTGGTGGCACCA CTCCCCGTCTCCCCAACCACGATCACGATGCTG
DHX38-T562A AD F	CAGCATCGTGATCGTGGTTGGGGAGACGGGGAGTGGTAAGGCCACTCA GCTGACGCAGTACCTGCATGAAGATGGTTACACG
DHX38-T562A AD R	CGTGTAACCATCTTCATGCAGGTACTGCGTCAGCTGAGTGGCCTTACCAC TCCCCGTCTCCCCAACCACGATCACGATGCTG
DHX38-T805V Uncoup F	TGGCGTTCGGAAGTGCATCGTTGCCGTGAATATTGCCGAGACGTCTCTC ACTGTTGAC
DHX38-T805V Uncoup R	GTCAACAGTGAGAGACGTCTCGGCAATATTCACGGCAACGATGCACTTC CGAACGCCA
DHX38-T805V UncoupLin F	AATATTGCCGAGACGTCTCTCAC
DHX38-T805V UncoupLin R	AACGATGCACTTCCGAACG
DHX38-T813V UncoupRB F	CCACCAATATTGCCGAGACGTCTCTCGTGGTTGACGGCATCATGTTTGTT ATC

Table B.2 (continued)

Name	Sequence
DHX38-T813V UncoupRB Lin F	TTGACGGCATCATGTTTGTTATCG
DHX38-T813V UncoupRB Lin R	AGAGACGTCTCGGCAATATTGG
DHX8-500	TTCTGTTCGGACCATGTTGG
DHX8-1100	AAGATCTAAACCCAAATAGACGGC
DHX8-1500	AACAAACACTGGGTTGACCC
DHX8-2000	ATCATCATGTTGGACGAGGC
DHX8-2500	TTGTGATTGCCACCAATATCG
DHX8-3000	
K594A Fwd	CCTGATTGTCATTGGTGAGACAGGATCTGGAGCCACAACACAGATCACC CAGTACCTGG
K594A Rev	CCAGGTACTGGGTGATCTGTGTTGTGGCTCCAGATCCTGTCTCACCAATG ACAATCAGG
K594A Lin Fwd	AACACAGATCACCCAGTACCTGG
K594A Lin Rev	TGTCTCACCAATGACAATCAGG
K594A Lin Rev2	AGATCCTGTCTCACCAATGAC
DHX8-685 Lin F	AATTCACACTGATGTGCTCTTTGG
DHX8-685_Lin R	ATGATCGCGTACTGAGTGAGG
DHX8-D685A ADF	CCTCACTCAGTACGCGATCATCATGTTGGCCGAGGCACATGAGAGGACA ATTCACACTGATGTGCTCTTTGG
DHX8-D685A ADR	CCAAAGAGCACATCAGTGTGAATTGTCCTCTCATGTGCCTCGGCCAACA TGATGATCGCGTACTGAGTGAGG
DHX8-H688A ADF	CCTCACTCAGTACGCGATCATCATGTTGGACGAGGCAGCCGAGAGGACA ATTCACACTGATGTGCTCTTTGG
DHX8-H688A ADR	CCAAAGAGCACATCAGTGTGAATTGTCCTCTCGGCTGCCTCGTCCAACA TGATGATCGCGTACTGAGTGAGG
DHX8-I846A UncoupHypF	CCACCAATATCGCAGAGACATCGGCCACTATTGATGGTATCTACTATGT GGTGG
DHX8-I846A UncoupHypR	CCACCACATAGTAGATACCATCAATAGTGGCCGATGTCTCTGCGATATT GGTGG

Table B.2 (continued)

Name	Sequence
DHX8-T839A Uncoup F	AGGCAGCAGAAAGGTTGTGATTGCCGCCAATATCGCAGAGACATCG
DHX8-T839A Uncoup R	CGATGTCTCTGCGATATTGGCGGCAATCACAACCTTTCTGCTGCCT
DHX8-T839A UncoupLin F	ATCGCAGAGACATCGCTGAC
DHX8-T839A UncoupLin R	AATCACAACCTTTCTGCTGCC
DHX8-T847 Uncoup-Hyper, Lin	ATTGATGGTATCTACTATGTGGTGG
DHX8-T847 Uncoup-Hyper, Lin	ATGTCTCTGCGATATTGGTGG
DHX8-T847A Uncou-HypBri F	CCACCAATATCGCAGAGACATCGCTGGCCATTGATGGTATCTACTATGT GGTGG
DHX8-T847A Uncoup-HypR	CCACCACATAGTAGATACCATCAATGGCCAGCGATGTCTCTGCGATATT GGTGG

Table B.2 (continued)

Name	Sequence
DHX8-T847V Uncoup-HypF	CCACCAATATCGCAGAGACATCGCTGGTGATTGATGGTATCTACTATGT GGTGG
DHX8-T847V Uncoup-HypR	CCACCACATAGTAGATACCATCAATCACCAGCGATGTCTCTGCGATATT GGTGG
pcDNA-DHX# Lin R1	
pcDNA5-DHX# Lin F1	TACGCGGCCGCTCGAGGATTATAAGG
pcDNA5-DHX# Lin F2	TCGAGGATTATAAGGATGACGACG
pcDNA5-DHX# Lin R2	ATCGCGGCGGCAGATCTCC
pENTR-DHX# Lin R1	TATCGTCGTCATCCTTATAATCCTCG
pENTR-DHX# Lin F1	TACGACTCACTATAGGGAGAGGATCCGG
Pentr-DHX# Lin F2	ATCCGGTACCGAGGAGATCTGC
pENTR-DHX# Lin R2	AATTTATCGTCGTCATCCTTATAATCC
Rev, C term seq pcDNA#	AAAGGACAGTGGGAGTGGC

APPENDIX C

DELINEATING THE ROLE OF THE N AND C-TERMINAL OF PRP22

Previous genetic, biochemical, and structural analysis have been conflicted as to the role of Prp22's N-terminal and C-terminal domain. Analysis of Prp22's functional domains, namely the S1 RNA binding domain in the N-terminus, revealed that the first 261 amino acids of Prp22 (region containing S1) is dispensable for growth, has no effect on it's ATPase activity, and no effect on its unwinding of duplex RNA. Further analysis revealed that removing 420 amino acids from the N-terminus of Prp22 increased total unwinding of duplex RNA by Prp22 compared to wildtype. However, *in vitro* splicing assays in which Prp22 was depleted and truncated rPrp22 was added back to extract revealed that the truncations of the first 260 amino acids were indistinguishable from wildtype but the addition of recombinant protein truncated beyond 260 amino acids were unable to promote splicing to wildtype levels. Note: each protein tested in the depleted extract for rescue of the exon ligation defect were all able to unwind duplex RNA at similar or higher amounts relative to wildtype, suggesting the enzyme has maintained function. To determine whether the effect of the truncation was due to the inability for Prp22 to bind the spliceosome, the authors stalled spliceosomes after exon ligation and added back the truncated protein to determine if the truncated regions were unable to release the mRNA and allow for excised lariat release. Beyond the 260 amino acid truncations, the truncated proteins were unable to rescue the phenotype and release the mRNA. These results suggest that the N-terminus targets binding of Prp22 to the spliceosome. To further test the idea that the N-Terminus of Prp22 is specific for recruitment of Prp22 to the spliceosome, and to test whether the N-Terminus alone was sufficient to promote exon ligation (ATP-independent), a visiting summer student (Jessica Priest) and I purified the N-terminus (1-480) of Prp22 using Ni-NTA and added the fragment to splicing extract and assayed for changes in splicing. I expected that if the N-terminus alone could promote exon ligation that I would see no change in mRNA but a dramatic increase in excised

lariat since the absence of Prp22's ATPase activity would prevent release of mRNA which would presumably block the release of the excised lariat. However, if the N-terminus was specific to only recruitment of Prp22, I expected addition of the N-terminus would block splicing since the full length Prp22 would not be able to engage with the spliceosome. I found that the addition of the N-terminus blocked exon ligation in a concentration dependent manner (Figure C.1). These results are supported by the observation that the expression of the N-terminus (1-480) is unable to support growth but, when expressed in trans with the C-terminus (466-1145) is able to support growth in a Prp22 deletion background[78]. Taken together, these findings provided compelling biochemical data to support the role for Prp22's N-terminus in promoting binding to the spliceosome.

Recent structural analysis of the P-complex revealed, in two of the three structures solved in budding yeast, that the N-terminus of Prp22 extends into the core of the spliceosome. However, a different structure doesn't observe the NTD's extension into the spliceosome and instead shows the CTD extending into the catalytic core. Notable differences in the isolation of the structures for Cryo-EM are as follows. Both structures that observe the N-terminus extending into the core were assembled on a model substrate in vitro and the structure that observed the CTD, was purified from a mixture of splicing complexes from whole cell extract. To discern whether the CTD of Prp22 is required for splicing, I created plasmids harboring Prp22 truncated by Δ4, Δ15, Δ32, Δ64, and Δ95 amino acids (Figure C.2.). Of the five truncations, only two were viable after selection, Δ4 and Δ15. The Δ4 and Δ15 truncations had cold sensitive growth defects, characteristic of a helicase deficiency, with the most pronounced sensitivity being the Δ15 truncation. Functional analysis of the truncations revealed that, like the presumed ATPase mutant of Prp22 (Figure 2.2.), the truncated cs mutants promoted downstream 3'SS selection.

These results suggest the C-terminal domain plays a role in its ATPase activity which would, for the first time, shed light on the mechanism of Prp22's activation.

The structure shown in Figure (C.3) shows the CTD of Prp22 wedged between the intronic elements, Prp8, the mRNA and what was later identified by Dr. Max Wilkinson as FYV1. Note well: This information has not been published but I was given permission to include for the purpose of explaining this data.

A trivial explanation for the phenotype would be that the truncation of the CTD has removed the nuclear localization signal. Computational predictions of Prp22's NLS provides two "confident" NLS predictions[168–170]. The NLS closest to the CTD truncations would be slightly affected by only the Δ32 truncation as 2/9 amino acids would have been truncated. (QSRKKAKIII, "I" indicates the truncated region). It's therefore likely that the lethal phenotype observed in the Δ32, Δ64, and Δ95 are due to removal of the NLS. In the future, this result will need to be further validated by first checking localization of the mutant's using immunofluorescence or fusing GFP to the N-terminus of Prp22 and visualizing using microscopy. Or, just fusing a known yeast NLS to the N-terminus of Prp22 and verifying the phenotype. If these findings are shown to be specific to Prp22's second step function, it has implications for understanding the fidelity mechanism of DExH-box helicases given the longstanding question of the role of the signature N and C terminal domains of these proteins. Interestingly, splicing variants in eukaryotes lacking these domains do exist and it'd be exciting to test whether this an evolved mechanism for splice site regulation in different tissue types.

Splicing Extract Protocol

2.5X media recipe:
Prepare YPDA in 6L Flasks
312.5 g of YPDA in 2400 mL DI water
Autoclave for 15' to avoid carmelizing the glucose
2.5 X Splicing Extract Prep
1. Prepare 4x AGK buffer, chill to 4° C.
***Add DTT Just prior to using
Cool the 1 liter bottles for harvesting
2. Harvest cells at OD600= 4.0-4.4*. Place into 1 L Bottles. Spin 5000 rpm for 14 min.
*OD600 @ 5.8 has worked but i dont recomend it.
Example: Harvested
1510-YSSC1-Prp19HA. 4.32
3. For 2.5 Liters of the 2.5X media I **resuspended in 500 mL AGK Buffer (total) and collected into one bottle.
Spin 5000 rpm for 14 min
4. **Resuspend in 500 mL AGK Buffer (total)
Spin 5000 rpm for 14 min
5. ***Resuspend in AGK + Protease inhibiot tablet (1 tablet in 50 mL = 1X)
Equation for total vol to resuspend in: 10.15 * (OD Harvested/4.2)
**Note: I resuspend on rotator. This takes approximately 3' if you use 100 mls of whatever liquid you are resuspending with. More than 100 mL's of liquid takes longer for pellet to reuspend. Doing it this way prevents bubbles and makes resuspending 6X6L Bottles MUCH EASIER/QUICKER/REPRODUCIBLE.
***Fore the last resuspension with small volume i use serelogical to move the pellet around for about 30" and put it at 4 on shaker to resuspend for 5-10'.
There should be no chunks when everything is fully resuspended.
Example: Final Resuspension
1510-YSSC1-Prp19HA. 4.32 -> 10.44 mL
6. Drip suspension into liq N2 using a serological. Store the tube at –80 ° C until ready for next step (or continue on if one wants).
7. Prepare to grind the cells in Liq N2. Ball Mill @ 10 hz, 3' (5 times) or Freezer mill using the same settings.
8. Use a pre-chilled spatula to scoop the powder into a pre-chilled oakridge tube (for SS34 rotor).
9. Thaw the powder rapidly in 16° C incubator at 100 rpm (Lay the tube flat on shaker) and then place immediately on ice.
10. Spin at 17K rpm for 30" in a SS34 rotor, 4° C. Pre-chill ultracentrifuge and rotor for step 14.
11. Remove supernatant and place into prechilled oakridge tube and spin again for 30'.
12. Remove supernatant and place in a 70 Ti tube.
13. Spin at 37K rpm (~150,000g) for 1 hour and 20 min in 70 Ti rotor, 4° C . Be sure to prechill rotor and ultracentrifuge for about 30 minutes prior to spinning.
14. Remove as much of the middle phase as possible (top fuzzy layer is lipid/protein and bottom layer is insoluble junk, such as membranes etc and contains RNases, be careful!).

15. Dialyze against 3 x 2L of buffer D*, 30',30',45'
*Always Add DTT Just before you use the buffer
16. Aliquot (100 uL) and freeze in liq N2.

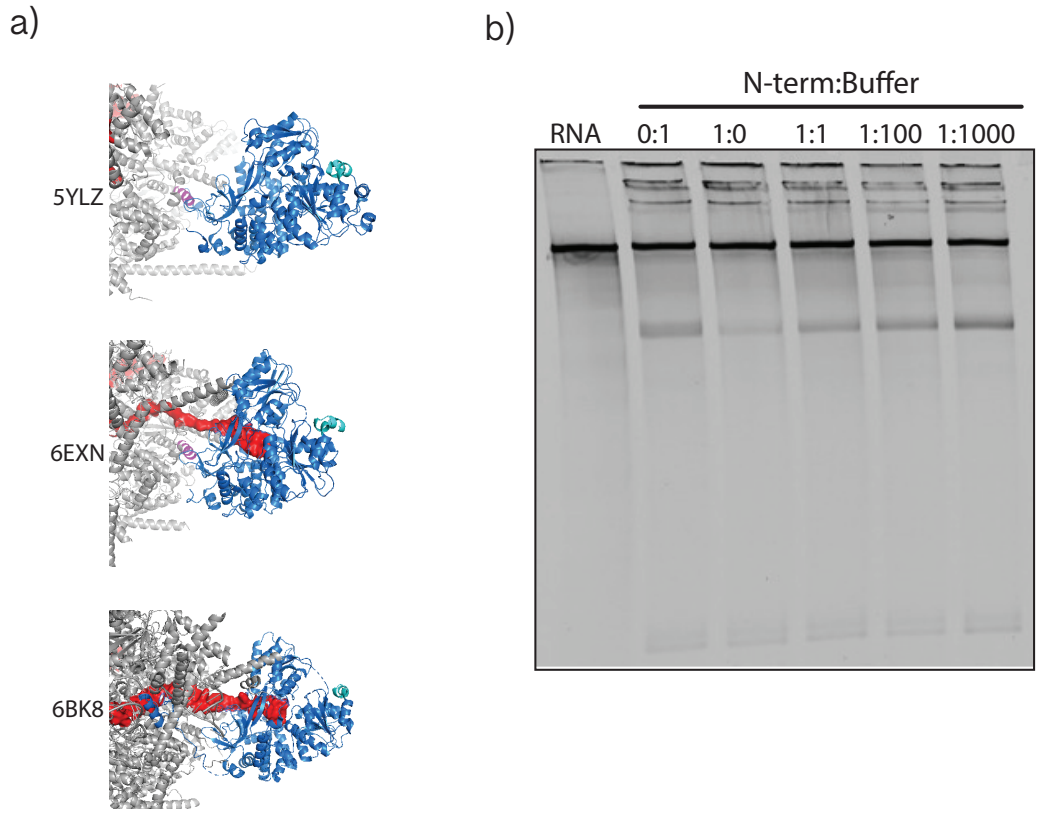
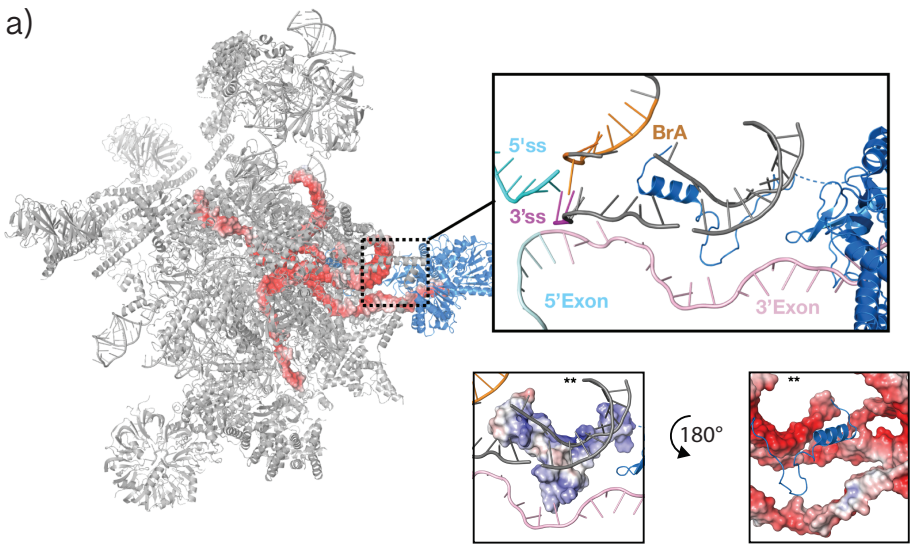


Figure C.1. P-complex structures differ in the assignment of the NTD and CTD of Prp22. **(a)** The three yeast P-complexes identified by their PDB code on the left differ by placement of the NTD in the 5YLZ and 6EXN and the CTD in 6BK8 within the core of the post catalytic complex. Dark blue is Prp22, light blue serves as a fiduciary mark for the region around amino acid 480 in Prp22, Magenta is the NTD and Blue is the CTD. mRNA substrate is represented as surface density in red. **(b)** Using in vitro splicing assays, I titrated the N-terminus of Prp22 (1-480) from 5ng/uL (50 mM Tris–HCl pH 8.0, 250 mM NaCl, 2 mM DTT, 1 mM EDTA, 0.1% Triton X-100) and assayed for splicing using a fluorescent Cy5 Actin pre-mRNA substrate.



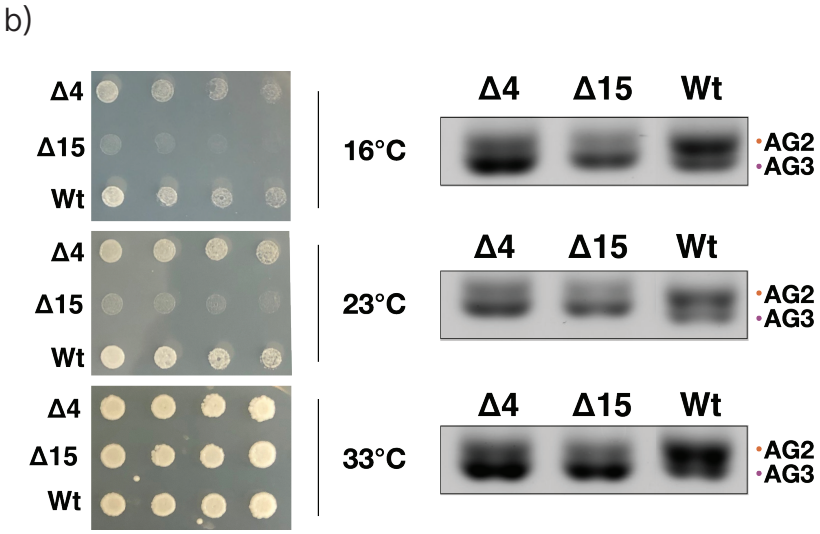


Figure C.2. The CTD of Prp22 may be essential for regulating it's ATPase function.
(a) The yeast P-complexes (6BK8)[73] purified from endogenous particles reveals a tunnel for the region between the branch and the 3'SS to reside and also shows the CTD of Prp22p buried near the catalytic core. I assayed for the functionality of this domain by designing various truncations and swapping them into yeast **(b).** Note: The yeast in **(b)** are overgrown. Temperature sensitive mutants were then assayed for differences in Ape2 splicing (see Figure 2.1 for Ape2 RNA structure) via RT-PCR at different growth conditions.

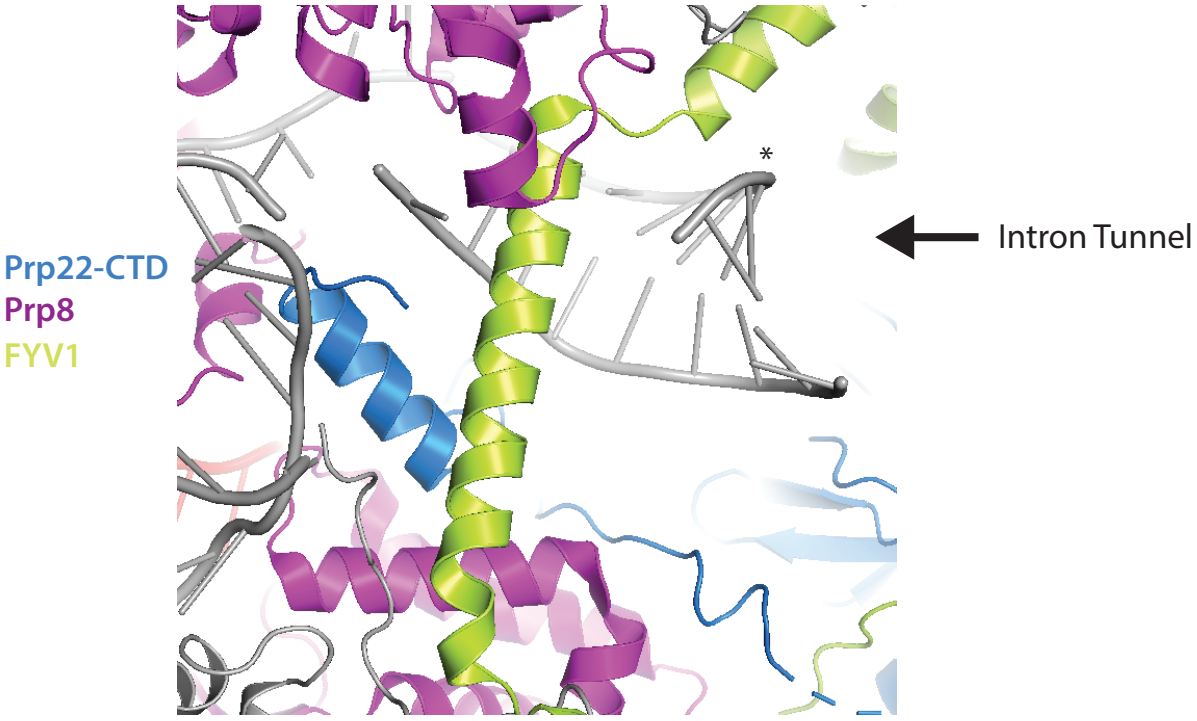


Figure C.3. The latch for securing Prp22 may be the protein FYV1. The yeast P-complex (6BK8)[73]. The intron tunnel is seen wedged between FYV1 (green) and Prp22 (blue). Both Prp22 and FYV1 and holstered by Prp8.

References

1. Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* **12**, 1–8 (1977).
2. Berget, S. M., Moore, C. & Sharp, P. A. Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci.* **74**, 3171–3175 (1977).
3. Lerner, M. R. & Steitz, J. A. Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *Proc. Natl. Acad. Sci.* **76**, 5495–5499 (1979).
4. Bringmann, P. & Lührmann, R. Purification of the individual snRNPs U1, U2, U5 and U4/U6 from HeLa cells and characterization of their protein constituents. *EMBO J.* **5**, 3509–3516 (1986).
5. Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. & Steitz, J. A. Are snRNPs involved in splicing? *Nature* **283**, 220–224 (1980).
6. Mount, S. M. & Steitz, J. A. Sequence of U1 RNA from Drosophila melanogaster: implications for U1 secondary structure and possible involvement in splicing. *Nucleic Acids Res.* **9**, 6351–6368 (1981).
7. Hernandez, N. & Keller, W. Splicing of in vitro synthesized messenger RNA precursors in HeLa cell extracts. *Cell* **35**, 89–99 (1983).
8. Brody, E. & Abelson, J. The ‘Spliceosome’: Yeast Pre-Messenger RNA Associates with a 40S Complex in a Splicing-Dependent Reaction. *Science* **228**, 963–967 (1985).
9. Grabowski, P. J., Seiler, S. R. & Sharp, P. A. A multicomponent complex is involved in the splicing of messenger RNA precursors. *Cell* **42**, 345–353 (1985).
10. Hartwell, L. H. Macromolecule Synthesis in Temperature-sensitive Mutants of Yeast. *J. Bacteriol.* **93**, 1662–1670 (1967).

11. Lustig, A. J., Lin, R. J. & Abelson, J. The yeast RNA gene products are essential for mRNA splicing in vitro. *Cell* **47**, 953–963 (1986).
12. Vijayraghavan, U., Company, M. & Abelson, J. Isolation and characterization of pre-mRNA splicing mutants of Saccharomyces cerevisiae. *Genes Dev.* **3**, 1206–1216 (1989).
13. Kastner, B., Will, C., Stark, H. & Lührmann, R. Structural Insights into Nuclear pre-mRNA Splicing in Higher Eukaryotes. *Cold Spring Harb. Perspect. Biol.* **11**, a032417 (2019).
14. B Ruskin, M. G., AR Krainer, T. Maniatis. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. *Cell* **38**, 317–331 (1984).
15. Fica, S. M. *et al.* RNA catalyses nuclear pre-mRNA splicing. *Nature* **503**, 229–234 (2013).
16. JR Rodriguez, M. R., CW Pikielny. In vivo characterization of yeast mRNA processing intermediates. *Cell* **39**, 603–610 (1984).
17. RA Padgett, P. S., MM Konarska, PJ Grabowski, SF Hardy. Lariat RNAs as intermediates and products in the splicing of messenger RNA precursors. *Science* **225**, 898–903 (1984).
18. Domdey, H. *et al.* Lariat structures are in vivo intermediates in yeast pre-mRNA splicing. *Cell* **39**, 611–621 (1984).
19. Konarska, M. M., Grabowski, P. J., Padgett, R. A. & Sharp, P. A. Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors. *Nature* **313**, 552–557 (1985).
20. Aebi, M., Hornig, H., Padgett, R. A., Reiser, J. & Weissmann, C. Sequence requirements for splicing of higher eukaryotic nuclear pre-mRNA. *Cell* **47**, 555–565 (1986).
21. Zeitlin, S. & Efstratiadis, A. In vivo splicing products of the rabbit beta-globin pre-mRNA. *Cell* **39**, 589–602 (1984).

22. Ruskin, B. & Green, M. R. Role of the 3' splice site consensus sequence in mammalian pre-mRNA splicing. *Nature* **317**, 732–734 (1985).

23. TA Steitz, J. S. A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci USA* **90**, 6498–6502 (1993).

24. JP Staley, C. G. Mechanical devices of the spliceosome: motors, clocks, springs and things. *Cell* **92**, 315–326 (1998).

25. Irion, U. & Leptin, M. Developmental and cell biological functions of the Drosophila DEAD-box protein Abstrakt. *Curr. Biol.* **9**, 1373–1381 (1999).

26. CHAN, C. C. *et al.* eIF4A3 is a novel component of the exon junction complex. *RNA* **10**, 200–209 (2004).

27. CL Will, R. L., H. Urlaub, T. Achsel, M. Gentzel, M. Wilm. Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. *EMBO J* **21**, 4978–88 (2002).

28. Sam, M. *et al.* Aquarius, a novel gene isolated by gene trapping with an RNA-dependent RNA polymerase motif. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* **212**, 304–317 (1998).

29. De Bortoli, F., Espinosa, S. & Zhao, R. DEAH-Box RNA Helicases in Pre-mRNA Splicing. *Trends Biochem. Sci.* **46**, 225–238 (2021).

30. O Cordin, J. B. RNA helicases in splicing. *RNA Biol* **10**, 83–95 (2012).

31. Jarmoskaite, I. & Russell, R. RNA helicase proteins as chaperones and remodelers. *Annu. Rev. Biochem.* **83**, 697–725 (2014).

32. Hamann, F. *et al.* Structural analysis of the intrinsically disordered splicing factor Spp2 and its binding to the DEAH-box ATPase Prp2. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 2948–2956 (2020).

33. Rauhut, R. Molecular architecture of the Saccharomyces cerevisiae activated spliceosome. *Science* **353**, 1399–1405 (2016).

34. Bohnsack, K. E., Ficner, R., Bohnsack, M. T. & Jonas, S. Regulation of DEAH-box RNA helicases by G-patch proteins. *Biol. Chem.* **402**, 561–579 (2021).

35. Warkocki, Z. *et al.* The G-patch protein Spp2 couples the spliceosome-stimulated ATPase activity of the DEAH-box protein Prp2 to catalytic activation of the spliceosome. *Genes Dev* **29**, 94–107 (2015).

36. J Roy, J. W., K. Kim, JR Maddock, JG Anthony. The final stages of spliceosome maturation require Spp2p that can interact with the DEAH box protein Prp2p and promote step 1 of splicing. *RNA* **1**, 375–90 (1995).

37. Bao, P., Höbartner, C., Hartmuth, K. & Lührmann, R. Yeast Prp2 liberates the 5' splice site and the branch site adenosine for catalysis of pre-mRNA splicing. *RNA N. Y. N* **23**, 1770–1779 (2017).

38. Hamann, F. *et al.* Structural analysis of the intrinsically disordered splicing factor Spp2 and its binding to the DEAH-box ATPase Prp2. *Proc. Natl. Acad. Sci.* **117**, 2948–2956 (2020).

39. C Yan, Y. S., R. Wan, R. Bai, G. Huang. Structure of a yeast-activated spliceosome at 3.5-Å resolution. *Science* **353**, 904–911 (2016).

40. Semlow, D. R., Blanco, M. R., Walter, N. G. & Staley, J. P. Spliceosomal DEAH-Box ATPases Remodel Pre-mRNA to Activate Alternative Splice Sites. *Cell* **164**, 985–998 (2016).

41. Kistler, A. L. & Guthrie, C. Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for sub2, an essential spliceosomal ATPase. *Genes Dev* **15**, 42–9 (2001).

42. Q Wang, B. R., L. Zhang, B. Lynn. A BBP-Mud2p heterodimer mediates branchpoint recognition and influences splicing substrate abundance in budding yeast. *Nucleic Acids Res* **36**, 2787–98 (2008).

43. Berglund, J. A., Abovich, N. & Rosbash, M. A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev.* **12**, 858–867 (1998).

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

45. Y Zhuang, A. W. A compensatory base change in U1 snRNA suppresses a 5′ splice site mutation. *Cell* **46**, 827–835 (1986).

46. Wong, M. S., Kinney, J. B. & Krainer, A. R. Quantitative Activity Profile and Context Dependence of All Human 5' Splice Sites. *Mol. Cell* **71**, 1012-1026.e3 (2018).

47. Ruby, S. W., Chang, T. H. & Abelson, J. Four yeast spliceosomal proteins (PRP5, PRP9, PRP11, and PRP21) interact to promote U2 snRNP binding to pre-mRNA. *Genes Dev* **7**, 1909–1925 (1993).

48. Kistler, A. L. & Guthrie, C. Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for Sub2, an essential spliceosomal ATPase. *Genes Dev.* **15**, 42–49 (2001).

49. H Shen, M. G., X. Zheng, J. Shen, L. Zhang, R. Zhao. Distinct activities of the DExD/H-box splicing factor hUAP56 facilitate stepwise assembly of the spliceosome. *Genes Dev* **22**, 1796–803 (2008).

50. CL O'Day, J. A., G. Dalbadie-McFarland. The Saccharomyces cerevisiae Prp5 protein has RNA-dependent ATPase activity with specificity for U2 small nuclear RNA. *J Biol Chem* **271**, 33261–7 (1996).

51. R Perriman, M. A., I. Barta, GK Voeltz, J. Abelson. ATP requirement for Prp5p function is determined by Cus2p and the structure of U2 small nuclear RNA. *Proc Natl Acad Sci U A* **100**, 13857–62 (2003).

52. Perriman, R. J. & Ares, M. Rearrangement of competing U2 RNA helices within the spliceosome promotes multiple steps in splicing. *Genes Dev* **21**, 811–820 (2007).

53. Liang, W.-W. & Cheng, S.-C. A novel mechanism for Prp5 function in prespliceosome formation and proofreading the branch site sequence. *Genes Dev.* **29**, 81–93 (2015).

54. Perriman, R. & Ares, M. Invariant U2 snRNA Nucleotides Form a Stem Loop to Recognize the Intron Early in Splicing. *Mol. Cell* **38**, 416–427 (2010).

55. Bai, R., Wan, R., Yan, C., Lei, J. & Shi, Y. Structures of the fully assembled Saccharomyces cerevisiae spliceosome before activation. *Science* **360**, 1423–1429 (2018).

56. Charenton, C., Wilkinson, M. E. & Nagai, K. Mechanism of 5' splice site transfer for human spliceosome activation. *Science* **364**, 362–367 (2019).

57. JP Staley, C. G. An RNA switch at the 5′ splice site requires ATP and the DEAD-box protein Prp28p. *Mol Cell* **3**, 55–64 (1999).

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

59. PL Raghunathan, C. G. RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Curr Biol* **8**, 847–855 (1998).

60. EJ Silverman, R. L., A. Maeda, J. Wei, P. Smith, JD Beggs. Interaction between a G-patch protein and a spliceosomal DEXD/H-box ATPase that is critical for splicing. *Mol Cell Biol* **24**, 10101–10 (2004).

61. SH Kim, R. L. Spliceosome activation by PRP2 ATPase prior to the first transesterification reaction of pre-mRNA splicing. *Mol Cell Biol* **16**, 6810–9 (1996).

62. Warkocki, Z. Reconstitution of both steps of Saccharomyces cerevisiae splicing with purified spliceosomal components. *Nat Struct Mol Biol* **16**, 1237–1243 (2009).

63. AJ Newman, C. N. U5 snRNA interacts with exon sequences at 5′ and 3′ splice sites. *Cell* **68**, 743–754 (1992).

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

65. Ohrt, T. *et al.* Molecular dissection of step 2 catalysis of yeast pre-mRNA splicing investigated in a purified system. *RNA* **19**, 902–915 (2013).

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

67. Hilliker, A. K., Mefford, M. A. & Staley, J. P. U2 toggles iteratively between the stem IIa and stem IIc conformations to promote pre-mRNA splicing. *Genes Dev* **21**, 821–834 (2007).

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

69. Fica, S. M. *et al.* Structure of a spliceosome remodelled for exon ligation. *Nature* **542**, 377–380 (2017).

70. Yan, C., Wan, R., Bai, R., Huang, G. & Shi, Y. Structure of a yeast step II catalytically activated spliceosome. *Science* **355**, 149–155 (2017).

71. Bertram, K. *et al.* Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature* **542**, 318–323 (2017).

72. Wilkinson, M. E. *et al.* Postcatalytic spliceosome structure reveals mechanism of 3'-splice site selection. *Science* **358**, 1283–1288 (2017).

73. Liu, S. *et al.* Structure of the yeast spliceosomal postcatalytic P complex. *Science* **358**, 1278–1283 (2017).

74. Bai, R., Yan, C., Wan, R., Lei, J. & Shi, Y. Structure of the Post-catalytic Spliceosome from Saccharomyces cerevisiae. *Cell* **171**, 1589-1598.e8 (2017).

75. Parker, R. & Siliciano, P. G. Evidence for an essential non-Watson–Crick interaction between the first and last nucleotides of a nuclear pre-mRNA intron. *Nature* **361**, 660–662 (1993).

76. Wan, R., Yan, C., Bai, R., Lei, J. & Shi, Y. Structure of an Intron Lariat Spliceosome from Saccharomyces cerevisiae. *Cell* **171**, 120-132.e12 (2017).

77. Zhang, X. *et al.* Structures of the human spliceosomes before and after release of the ligated exon. *Cell Res.* **29**, 274–285 (2019).

78. Schwer, B. & Gross, C. H. Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. *EMBO J.* **17**, 2086–2094 (1998).

79. Wagner, J. D., Jankowsky, E., Company, M., Pyle, A. M. & Abelson, J. N. The DEAH-box protein PRP22 is an ATPase that mediates ATP-dependent mRNA release from the spliceosome and unwinds RNA duplexes. *EMBO J.* **17**, 2926–2937 (1998).

80. Tsai, R.-T. *et al.* Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2. *Genes Dev.* **19**, 2991–3003 (2005).

81. Tanaka, N., Aronova, A. & Schwer, B. Ntr1 activates the Prp43 helicase to trigger release of lariat-intron from the spliceosome. *Genes Dev.* **21**, 2312–2325 (2007).

82. JE Arenas, J. A. Prp43: An RNA helicase-like factor involved in spliceosome disassembly. *Proc Natl Acad Sci U A* **94**, 11798–802 (1997).

83. Toroney, R., Nielsen, K. H. & Staley, J. P. Termination of pre-mRNA splicing requires that the ATPase and RNA unwindase Prp43p acts on the catalytic snRNA U6. *Genes Dev.* **33**, 1555–1574 (2019).

84. Li, X. *et al.* A unified mechanism for intron and exon definition and back-splicing. *Nature* **573**, 375–380 (2019).

85. Plaschka, C., Lin, P.-C., Charenton, C. & Nagai, K. Prespliceosome structure provides insights into spliceosome assembly and regulation. *Nature* **559**, 419–422 (2018).

86. C Plaschka, K. N., P. C. Lin. Structure of a pre-catalytic spliceosome. *Nature* **546**, 617–621 (2017).

87. Wan, R., Bai, R., Yan, C., Lei, J. & Shi, Y. Structures of the Catalytically Activated Yeast Spliceosome Reveal the Mechanism of Branching. *Cell* **177**, 339-351.e13 (2019).

88. Galej, W. Cryo-EM structure of the spliceosome immediately after branching. *Nature* **537**, 197–201 (2016).

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

90. Crotti, L. B. & Horowitz, D. S. Exon sequences at the splice junctions affect splicing fidelity and alternative splicing. *Proc Natl Acad Sci USA* **106**, 18954–18959 (2009).

91. Crotti, L. B., Bacikova, D. & Horowitz, D. S. The Prp18 protein stabilizes the interaction of both exons with the U5 snRNA during the second step of pre-mRNA splicing. *Genes Dev* **21**, 1204–1216 (2007).

92. Schneider, S. & Schwer, B. Functional Domains of the Yeast Splicing Factor Prp22p *. *J. Biol. Chem.* **276**, 21184–21191 (2001).

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

94. Bradley, R. K., Merkin, J., Lambert, N. J. & Burge, C. B. Alternative Splicing of RNA Triplets Is Often Regulated and Accelerates Proteome Evolution. *PLOS Biol.* **10**, e1001229 (2012).

95. Hiller, M. *et al.* Single-nucleotide polymorphisms in NAGNAG acceptors are highly predictive for variations of alternative splicing. *Am. J. Hum. Genet.* **78**, 291–302 (2006).

96. Sheth, N. *et al.* Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Res.* **34**, 3955–3967 (2006).

97. Zamore, P. D., Patton, J. G. & Green, M. R. Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* **355**, 609–614 (1992).

98. Wu, S., Romfo, C. M., Nilsen, T. W. & Green, M. R. Functional recognition of the 3′ splice site AG by the splicing factor U2AF35. *Nature* **402**, 832–835 (1999).

99. Zorio, D. A. R. & Blumenthal, T. Both subunits of U2AF recognize the 3′ splice site in Caenorhabditis elegans. *Nature* **402**, 835–838 (1999).

100. Merendino, L., Guth, S., Bilbao, D., Martínez, C. & Valcárcel, J. Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3’ splice site AG. *Nature* **402**, 838–841 (1999).

101. Schwer, B. & Guthrie, C. PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. *Nature* **349**, 494–499 (1991).

102. B Schwer, C. G. A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. *EMBO J* **11**, 5033–5039 (1992).

103. Horowitz, D. S. The mechanism of the second step of pre-mRNA splicing. *Wiley Interdiscip. Rev. RNA* **3**, 331–350 (2012).

104. Smith, D. J., Query, C. C. & Konarska, M. M. "Nought may endure but mutability": spliceosome dynamics and the regulation of splicing. *Mol Cell* **30**, 657–666 (2008).

105. Akerman, M. & Mandel-Gutfreund, Y. Does distance matter? Variations in alternative 3′ splicing regulation. *Nucleic Acids Res.* **35**, 5487–5498 (2007).

106. Cellini, A., Felder, E. & Rossi, J. J. Yeast pre-messenger RNA splicing efficiency depends on critical spacing requirements between the branch point and 3′ splice site. *EMBO J* **5**, 1023–1030 (1986).

107. Luukkonen, B. G. & Séraphin, B. The role of branchpoint-3′ splice site spacing and interaction between intron terminal nucleotides in 3′ splice site selection in Saccharomyces cerevisiae. *EMBO J* **16**, 779–792 (1997).

108. Meyer, M., Plass, M., Pérez-Valle, J., Eyras, E. & Vilardell, J. Deciphering 3′ss Selection in the Yeast Genome Reveals an RNA Thermosensor that Mediates Alternative Splicing. *Mol. Cell* **43**, 1033–1039 (2011).

109. Zeng, Y. *et al. Profiling of Nascent Lariat Intermediates Reveals Key Genetic Determinants of the Timing of Human Co-transcriptional Splicing*. 2021.10.18.464728 https://www.biorxiv.org/content/10.1101/2021.10.18.464728v1 (2021) doi:10.1101/2021.10.18.464728.

110. Pineda, J. M. B. & Bradley, R. K. Most human introns are recognized via multiple and tissue-specific branchpoints. *Genes Dev.* **32**, 577–591 (2018).

111. Deshler, J. O. & Rossi, J. J. Unexpected point mutations activate cryptic 3′ splice sites by perturbing a natural secondary structure within a yeast intron. *Genes Dev* **5**, 1252–1263 (1991).

112. Goguel, V., Wang, Y. & Rosbash, M. Short artificial hairpins sequester splicing signals and inhibit yeast pre-mRNA splicing. *Mol. Cell. Biol.* **13**, 6841–6848 (1993).

113. Charpentier, B. & Rosbash, M. Intramolecular structure in yeast introns aids the early steps of in vitro spliceosome assembly. *RNA* **2**, 509–522 (1996).

114. Xu, B., Meng, Y. & Jin, Y. RNA structures in alternative splicing and back-splicing. *WIREs RNA* **12**, e1626 (2021).

115. Lehmann, K. A. & Bass, B. L. The importance of internal loops within RNA substrates of ADAR1. *J. Mol. Biol.* **291**, 1–13 (1999).

116. Kar, A. *et al.* RNA Helicase p68 (DDX5) Regulates tau Exon 10 Splicing by Modulating a Stem-Loop Structure at the 5′ Splice Site. *Mol. Cell. Biol.* **31**, 1812–1821 (2011).

117. Dardenne, E. *et al.* RNA Helicases DDX5 and DDX17 Dynamically Orchestrate Transcription, miRNA, and Splicing Programs in Cell Differentiation. *Cell Rep.* **7**, 1900–1913 (2014).

118. Lin, C., Yang, L., Yang, J. J., Huang, Y. & Liu, Z.-R. ATPase/helicase activities of p68 RNA helicase are required for pre-mRNA splicing but not for assembly of the spliceosome. *Mol. Cell. Biol.* **25**, 7484–7493 (2005).

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

120. Cordin, O., Hahn, D. & Beggs, J. D. Structure, function and regulation of spliceosomal RNA helicases. *Curr. Opin. Cell Biol.* **24**, 431–438 (2012).

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

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

123. McPheeters, D. S. & Muhlenkamp, P. Spatial organization of protein-RNA interactions in the branch site-3′ splice site region during pre-mRNA splicing in yeast. *Mol Cell Biol* **23**, 4174–4186 (2003).

124. Fica, S. M., Oubridge, C., Wilkinson, M. E., Newman, A. J. & Nagai, K. A human postcatalytic spliceosome structure reveals essential roles of metazoan factors for exon ligation. *Science* **363**, 710–714 (2019).

125. Burgess, S. M. & Guthrie, C. Beat the clock: paradigms for NTPases in the maintenance of biological fidelity. *Trends Biochem. Sci.* **18**, 381–384 (1993).

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

127. Mayas, R. M., Maita, H., Semlow, D. R. & Staley, J. P. Spliceosome discards intermediates via the DEAH box ATPase Prp43p. *Proc Natl Acad Sci USA* **107**, 10020–10025 (2010).

128. Pandit, S., Lynn, B. & Rymond, B. C. Inhibition of a spliceosome turnover pathway suppresses splicing defects. *Proc Natl Acad Sci USA* **103**, 13700–13705 (2006).

129. Strittmatter, L. M. *et al.* psiCLIP reveals dynamic RNA binding by DEAH-box helicases before and after exon ligation. *Nat. Commun.* **12**, 1488 (2021).

130. McPheeters, D. S., Schwer, B. & Muhlenkamp, P. Interaction of the yeast DExH-box RNA helicase prp22p with the 3' splice site during the second step of nuclear pre-mRNA splicing. *Nucleic Acids Res.* **28**, 1313–1321 (2000).

131. Schwer, B. & Meszaros, T. RNA helicase dynamics in pre-mRNA splicing. *EMBO J* **19**, 6582–6591 (2000).

132. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol. Biol. AMB* **6**, 26 (2011).

133. Hofacker, I. L. RNA secondary structure analysis using the Vienna RNA package. *Curr Protoc Bioinforma.* (2009).

134. Lesser, C. F. & Guthrie, C. Mutational analysis of pre-mRNA splicing in Saccharomyces cerevisiae using a sensitive new reporter gene, CUP1. *Genetics* **133**, 851–863 (1993).

135. Plass, M., Codony-Servat, C., Ferreira, P. G., Vilardell, J. & Eyras, E. RNA secondary structure mediates alternative 3′ss selection in Saccharomyces cerevisiae. *RNA* **18**, 1103–1115 (2012).

136. Gietz, R. D. & Schiestl, R. H. Frozen competent yeast cells that can be transformed with high efficiency using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2**, 1–4 (2007).

137. M Meyer, J. V., M. Plass, J. Pérez-Valle, E. Eyras. Deciphering 3′ss selection in the yeast genome reveals an RNA thermosensor that mediates alternative splicing. *Mol Cell* **43**, 1033–1039 (2011).

138. Bishop, D. K., Park, D., Xu, L. & Kleckner, N. DMC1: A meiosis-specific yeast homolog of E. coli recA required for recombination, synaptonemal complex formation, and cell cycle progression. *Cell* **69**, 439–456 (1992).

139. Yassour, M. *et al.* Ab initio construction of a eukaryotic transcriptome by massively parallel mRNA sequencing. *Proc Natl Acad Sci USA* **106**, 3264–3269 (2009).

140. Orengo, J. P., Bundman, D. & Cooper, T. A. A bichromatic fluorescent reporter for cell-based screens of alternative splicing. *Nucleic Acids Res.* **34**, e148 (2006).

141. N Tanaka & B Schwer. Characterization of the NTPase, RNA-binding, and RNA helicase activities of the DEAH-Box splicing factor Prp22. *Biochemistry* **44**, 9795–9803 (2005).

142. Wilkinson, M. E., Charenton, C. & Nagai, K. RNA Splicing by the Spliceosome. *Annu. Rev. Biochem.* **89**, 359–388 (2020).

143. Semlow, D. R., Blanco, M. R., Walter, N. G. & Staley, J. P. Spliceosomal DEAH-Box ATPases Remodel Pre-mRNA to Activate Alternative Splice Sites. *Cell* **164**, 985–998 (2016).

144. Grover, A. *et al.* 5′ Splice Site Mutations in tau Associated with the Inherited Dementia FTDP-17 Affect a Stem-Loop Structure That Regulates Alternative Splicing of Exon 10. *J. Biol. Chem.* **274**, 15134–15143 (1999).

145. Zhang, J., Kuo, C. J. & Chen, L. GC content around splice sites affects splicing through pre-mRNA secondary structures. *BMC Genomics* **12**, 90 (2011).

146. Shepard, P. J. & Hertel, K. J. Conserved RNA secondary structures promote alternative splicing. *RNA* **14**, 1463–1469 (2008).

147. Ragle, J. M., Katzman, S., Akers, T. F., Barberan-Soler, S. & Zahler, A. M. Coordinated tissue-specific regulation of adjacent alternative 3′ splice sites in C. elegans. *Genome Res.* **25**, 982–994 (2015).

148. Lallena, M. J., Chalmers, K. J., Llamazares, S., Lamond, A. I. & Valcárcel, J. Splicing Regulation at the Second Catalytic Step by Sex-lethal Involves 3′ Splice Site Recognition by SPF45. *Cell* **109**, 285–296 (2002).

149. Anderson, K. & Moore, M. Bimolecular exon ligation by the human spliceosome. *Science* **276**, 1712–1716 (1997).

150. Chen, S., Anderson, K. & Moore, M. J. Evidence for a linear search in bimolecular 3′ splice site AG selection. *Proc Natl Acad Sci USA* **97**, 593–598 (2000).

151. Guth, S., Martínez, C., Gaur, R. K. & Valcárcel, J. Evidence for Substrate-Specific Requirement of the Splicing Factor U2AF35 and for Its Function after Polypyrimidine Tract Recognition by U2AF65. *Mol. Cell. Biol.* **19**, 8263–8271 (1999).

152. Zhang, X. *et al.* An Atomic Structure of the Human Spliceosome. *Cell* **169**, 918-929.e14 (2017).

153. Szczesny, R. J. *et al.* Versatile approach for functional analysis of human proteins and efficient stable cell line generation using FLP-mediated recombination system. *PLOS ONE* **13**, e0194887 (2018).

154. Kim, S., Kim, H., Fong, N., Erickson, B. & Bentley, D. L. Pre-mRNA splicing is a determinant of histone H3K36 methylation. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 13564–13569 (2011).

155. Kraushaar, D. C. *et al.* Genome-wide incorporation dynamics reveal distinct categories of turnover for the histone variant H3.3. *Genome Biol.* **14**, R121 (2013).

156. Venkatesh, S. *et al.* Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. *Nature* **489**, 452–455 (2012).

157. Nojima, T. *et al.* Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* **161**, 526–540 (2015).

158. Chathoth, K. T., Barrass, J. D., Webb, S. & Beggs, J. D. A splicing-dependent transcriptional checkpoint associated with prespliceosome formation. *Mol. Cell* **53**, 779–790 (2014).

159. Nojima, T. *et al.* RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing. *Mol. Cell* **72**, 369-379.e4 (2018).

160. Fei, J. *et al.* Quantitative analysis of multilayer organization of proteins and RNA in nuclear speckles at super resolution. *J. Cell Sci.* **130**, 4180–4192 (2017).

161. Li, M. *AQR Plays a Key Role in Splicing Fidelity in Higher Organisms*. https://www.researchsquare.com/article/rs-20141/v1 (2020) doi:10.21203/rs.3.rs-20141/v1.

162. Mikl, M., Hamburg, A., Pilpel, Y. & Segal, E. Dissecting splicing decisions and cell-to-cell variability with designed sequence libraries. *Nat. Commun.* **10**, 4572 (2019).

163. Nuñez, J. K. *et al.* Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. *Cell* **184**, 2503-2519.e17 (2021).

164. Vassylyeva, M. N. *et al.* Efficient, ultra-high-affinity chromatography in a one-step purification of complex proteins. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E5138–E5147 (2017).

165. Asrani, K. H. *et al.* Optimization of mRNA untranslated regions for improved expression of therapeutic mRNA. *RNA Biol.* **15**, 756–762 (2018).

166. Mort, M., Ivanov, D., Cooper, D. N. & Chuzhanova, N. A. A meta-analysis of nonsense mutations causing human genetic disease. *Hum. Mutat.* **29**, 1037–1047 (2008).

167. Porter, J. J., Heil, C. S. & Lueck, J. D. Therapeutic promise of engineered nonsense suppressor tRNAs. *WIREs RNA* **12**, e1641 (2021).

168. Kosugi, S., Hasebe, M., Tomita, M. & Yanagawa, H. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 10171–10176 (2009).

169. Kosugi, S. *et al.* Six classes of nuclear localization signals specific to different binding grooves of importin alpha. *J. Biol. Chem.* **284**, 478–485 (2009).

170. Kosugi, S. *et al.* Design of peptide inhibitors for the importin alpha/beta nuclear import pathway by activity-based profiling. *Chem. Biol.* **15**, 940–949 (2008).