Uncovering Unknown Pathways and Roles of Splicing Factors in Spliceosome Activation

By

Xingyang Fu

A dissertation submitted in partial fulfillment of

the requirement for the degree of

Doctor of Philosophy

(Chemistry)

at the

UNIVERSITY OF WISCONSIN-MADISON

2022

Date of final oral examination: 05/05/2022

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

Aaron A. Hoskins, Associate Professor, Biochemistry and Chemistry

Randall Goldsmith, Associate Professor, Chemistry

Ci Ji Lim, Assistant Professor, Biochemistry

Thomas Record, Professor, Biochemistry and Chemistry

Abstract

Pre-mRNA splicing is one of the most important RNA processing steps in eukaryotic gene expression. A megaDalton enzyme, the spliceosome, is responsible for catalyzing the splicing process, during which non-coding introns on pre-mRNA are removed and coding exons are ligated. A significant number of disease-causing mutations impact splicing. Thus, understanding splicing mechanism is important for understanding how mis-splicing can occur in diseases and devising therapeutic approaches to correct mis-splicing.

Decades of efforts in understanding splicing mechanism by genetic, biochemical, structural and single molecule approaches have yielded a relatively complete model of how spliceosome-mediated splicing occurs. However, many molecular details remain elusive. For example, one transition within the spliceosome activation process, during which the catalytic center is created, involves the exchange of approximately 50 factors. Here, I present my efforts to understand the complicated spliceosome activation process.

To study the pathway of activation, I employed Colocalization Single Molecule Spectroscopy (CoSMoS) to monitor the dynamics of different factors (Lsm2-8, B complex proteins, and Snu66; Chapters 2 and 3) involved in the activation process in real-time during splicing reactions. Their recruitment and release patterns were analyzed relative to activation landmark events (U4 snRNP release and NTC recruitment) and combined to form a typical pathway for activation. Kinetic analysis reveals the lifetimes of spliceosome intermediates with characteristic compositions along the activation pathway. Putting together the patterns of dynamic events and kinetic analysis, I construct a kinetic map for the *in vitro* spliceosome activation process. Novel pathways and intermediates can be identified from the map, potentially illuminating future research directions.

To better understand the role of different factors involved in activation, I carried out genetic characterization of Spp381, one of the B complex proteins (Chapter 4). Preliminary mutational analysis on Spp381 allowed me to identify helical regions and residues that are essential in a context-dependent manner. Future efforts are needed to better understand whether these *spp381* mutants affect cell growth via disrupting spliceosome activation specifically. Finally, I contributed to studies of the splicing factor Ecm2 (Chapter 5). Ecm2 is recruited to the spliceosome during activation and functions in later catalysis steps. A genetic approach was taken to assay the synthetic effects of combining *ecm2* mutants with a variety of previously well-characterized catalysis-related mutants in splicing factors. Our results support a model that Ecm2 facilitates the formation and stabilization of the catalytic site in the 1st step as well as promotes the catalysis in the 2nd step.

Acknowledgements

I am extremely grateful for all of the support that I received over the past six years. To begin with, I would like to thank my advisor, Aaron Hoskins, for welcoming me into his lab and leading me through an invaluable research experience. Aaron's advice has long been instrumental to me for continuing to pursue rigorous and solid arguments and motivating me to become an enthusiastic scientist. I would also like to thank my committee members, Dr. Thomas Record and Dr. Randall Goldsmith for their wonderful classes and academic support in committee meetings. I also want to thank Dr. Ci Ji Lim for constructive feedback on my manuscripts.

I am also very grateful to be a member of tri-lab group. Exposing me to different areas of research broadened my horizon and provided me with sources of inspiration. I would especially like to thank Dr. David Brow and Dr. Samuel Butcher for their provoking questions and great suggestions on my work.

Many thanks to the past and present members of the Hoskins Lab. I have picked up many skills and received tremendous feedback from all of you. Finishing my dissertation could not be possible without your help. Thanks to Ian, Josh, Tucker and Sarah for your help when I initially joined the lab. Thanks to Josh Paulson for preparing many reagents used for my experiments. Thanks to Clarisse and Harpreet for your help from almost every aspect over many years in the lab. Thanks to Karli, Kathy, Sierra, David, Junqiao and Ye for many helpful discussions on my work. Special thanks to Junqiao and Ye for your company outside the lab. I also would like to thank my undergraduate student, Lukas Voigts who has contributed directly to my work.

Finally, I want to thank my friends and family for their unconditional mental support. My parents have supported me always no matter what my choice is. My grandfather planted a seed of science in me when I was a child. I always feel greatly indebted to him. I also want to thank my friends I met since my freshman year in college and friends I met here in Madison. We went on

trips together across the United States and skied down the mountains together. When I am depressed, they would always be on my side and relieve my nerves. I lastly would like to thank some of the greatest philosophers and writers on my mind. Only in their words, I was able to find keys to puzzle of the life and gain strength again.

Table of Contents

List of Tables and FIgures

Chapter 1: Introduction

Chapter 2: Identification of Transient Intermediates During Spliceosome Activation by Single Molecule Fluorescence Microscopy

Chapter 5: Saccharomyces cerevisiae Ecm2 Modulates the Catalytic Steps of pre-mRNA Splicing

Chapter One

Introduction

Pre-mRNA Splicing is a Spliceosome-mediated, Nucleotide-level Precision Process.

In the typical expression of a gene, the genetic information is transcribed from the genomic DNA to the intermediate mRNA and finally is translated into protein (Orphanides and Reinberg, 2002). Eukaryotic precursor messenger RNA (pre-mRNA) splicing is a mRNA processing step that occurs either coupled to transcription or after the generation of nascent mRNA molecules (Proudfoot et al., 2002). It involves the removal of non-coding regions (introns) and the ligation of coding regions (exons) (**Fig. 1.1A**). A megaDalton ribonucleoprotein (RNP) enzyme, the spliceosome, is required to carry out this process with nucleotide-level precision (Wahl et al., 2009).

Cis-acting splicing signals are found in the introns: the 5' splice site (5'SS), branch point site (BPS), and 3' splice site (3'SS) (**Fig. 1.1A**). The sequences from these sites are relatively conserved and are repeatedly recognized during splicing. The 5'SS and 3'SS define the 5' and 3' intron-exon boundaries, respectively. Most studied organisms showed conserved “GU” and “AG” dinucleotides at the 5'SS and 3'SS, respectively (Roy and Irimia, 2014). The evolutionarily conserved adenosine nucleotide at the BPS (Branch Point Adenosine, BPA) localizes closer to the 3'SS. It is left unpaired prior to catalysis and the oxygen atom of its 2' hydroxyl is used for the nucleophilic attack on the phosphodiester bond at the 5' exon/intron boundary via a S_N2 mechanism (**Fig. 1.1C**) (Moore and Sharp, 1993). The first transesterification reaction is known as the branching reaction due to the formation of a lariat-shaped intron-3' exon RNA. In the same manner, the oxygen of the 3' hydroxyl of the last nucleotide of the 5' exon attacks the phosphodiester bond at the 3' exon/intron boundary to produce ligated exons and an excised lariat intron.

The chemical mechanism of splicing by the spliceosome is shared by self-splicing Group II introns. The Group II introns have been proposed to be ancestral to spliceosomal introns (Lambowitz and Zimmerly, 2011). The catalytic centers of the Group II intron and spliceosome

are highly similar, consisting of two Mg^{2+} ions coordinated by the negatively charged phosphate backbone of RNA catalytic triplex (**Fig. 1.1B**). The Mg^{2+} ions help catalyze the two transesterification reactions by promoting the formation of the transition states in the S_N2-type nucleophilic substitution reaction (**Fig. 1.1C**). For the spliceosome, the catalytic triplex is composed of two small nuclear RNAs, the U2 and U6 snRNAs (**Fig. 1.1B**). For the splicing of the Group II intron, the triplex is formed by organizing tertiary RNA structures among domains within the intron itself (Lambowitz and Zimmerly, 2011). The process of forming the spliceosome active site and how its assembly is guided by the spliceosomal RNP complex are of great research interest.

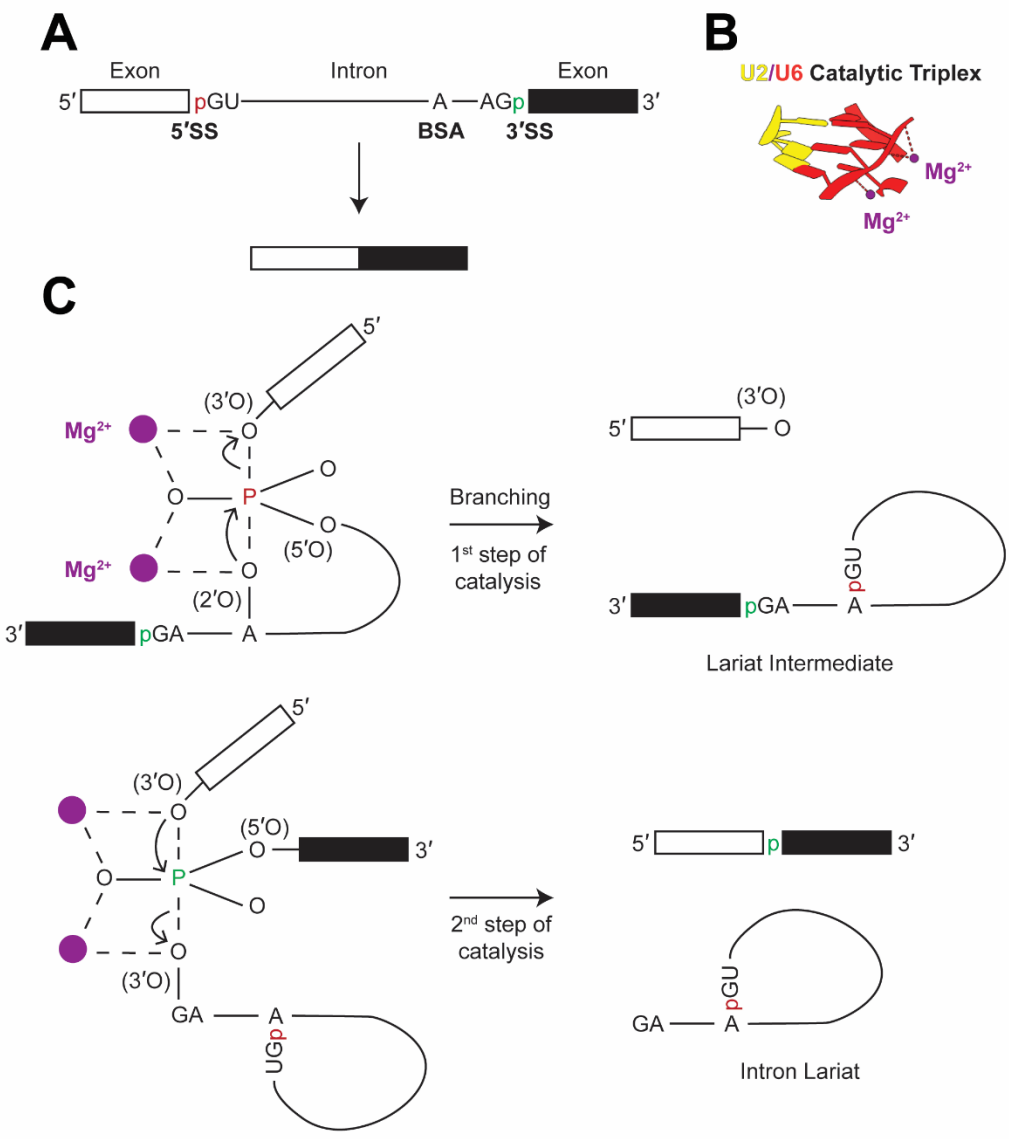


Figure 1.1 Chemistry of Spliceosome-mediated Splicing.

(**A**) A simple model for the pre-mRNA splicing process. Nucleotide sequences within the intron, including the 5'SS (GU), 3'SS (AG), and BSA (Branch Site Adenosine), are recognized during the splicing process. Introns are removed at the 5'SS and 3'SS boundaries with nucleotide-level precision concomitant with exon ligation. Scissile phosphates at the 5' and 3' intron-exon boundaries are indicated as red and green, respectively. (**B**) The catalytic center for spliceosome-mediated splicing involves a U2/U6 snRNA catalytic triplex and two coordinated magnesium ions.

(PDB: 5GM6) (**C**) The two transesterification steps of splicing are mediated by two magnesium ions at the catalytic center. (Top) In the 1st step of catalysis (known as branching), the 2' oxygen of the BPA attacks the scissile phosphate at the 5' intron-exon boundary, resulting in cleaved 5' exon and lariat-3' exon intermediate. (Bottom) In the 2nd step of catalysis, 3' oxygen at the end of 5' exon attacks the scissile phosphate at the 3' intron-exon boundary, resulting in ligated exons and intron lariat. Figure is modeled after (Fica et al., 2014).

Disrupting Splicing Leads to a Myriad of Genetic Diseases.

Mis-splicing often occurs in cells and can result in a myriad of consequences. In some cases, mis-splicing will produce premature termination codons in the mRNA transcripts. These transcripts can be detected during translation via the surveillance pathway made of UPF proteins, resulting in the non-sense mediated decay (NMD) of the mis-spliced transcripts (Baker and Parker, 2004). In some cases, gene products of the mis-spliced transcripts can lead to disease states of cells. It is estimated that 15% to 60% disease-causing mutations impact splicing (Kwan et al., 2008; López-Bigas et al., 2005).

Splicing-related mutations can cause changes in splicing via affecting *cis*-acting or *trans*-acting elements (Singh and Cooper, 2012). *Cis*-acting elements include the splicing signals within the intron and the splicing regulatory sequences from both introns and exons (e.g., exonic splicing enhancers or silencers (ESEs or ESSs) or intronic splicing enhancers or silencers (ISEs or ISSs). The regulatory sequences often function in aiding the recruitment of auxiliary factors. *Trans*-acting elements include the spliceosomal proteins and auxiliary factors (splicing regulatory (SR) and heterogeneous nuclear RNP (hnRNP) proteins) that can recognize the enhancer or silencer elements.

Understanding the mechanism of splicing carried out by the spliceosome is significant for understanding how mis-splicing can occur in diseases and devising therapeutic approaches to correct mis-splicing.

Saccharomyces cerevisiae Represents a Simple Model Organism for Studying Pre-mRNA Splicing.

Introns exists in all eukaryotic organisms. While all introns provide recognition sites for the splicing process, introns from different organisms also exhibit various architectures. To illustrate this, introns from *Homo sapiens* (hs, human) and *Saccharomyces cerevisiae* (sc, yeast) will be compared.

The density of introns in protein-coding genes is vastly different between yeast and humans. Only 283 out of the ~6,000 protein coding genes in yeast contain introns with a density of ~1.0 intron/gene (7 yeast genes contain two introns) (Wongpalee and Sharma, 2014). In contrast, less than 700 out of 20,000-25,000 human protein coding genes are intronless (Louhichi et al., 2011). Of the >97% of human genes containing introns, each gene has a much higher intron density of 7.8 intron/gene than in yeast (Sakharkar et al., 2004). Thus, while the human splicing machinery is required to splice almost every human gene, on which each contains an average of eight introns, the yeast splicing machinery is needed for the splicing of merely 5% yeast genes, with one intron per gene (**Fig. 1.2A**).

The size ratio of an average intron to exon adds another layer of difference to the splicing mechanism. Splicing generally starts with the recognition of core splicing signals by spliceosomal components. Then splicing signals are brought together for later stages of splicing. Due to the nature of long introns (typical ~3,000 nt) and short exons (~80% are shorter than 200 nt) in human multi-exon pre-mRNAs, it is hypothesized that human splicing is initiated by formation of exon definition complex that associates the BPS and 5'SS from two adjacent introns across the short exon (Berget, 1995) (**Fig. 1.2B**). Because the exon definition complex cannot carry out splicing, a switch to intron definition complex needs to occur in later splicing steps (Schneider et al., 2010). In contrast, a typical yeast pre-mRNA with a single intron makes more frequent use of an intron

definition complex by bringing together splicing signals within the same intron. Thus, the relative size of intron to exon can play a role in determining the splicing mechanism.

One last important difference in intron architecture is the conservation of sequences at the recognition sites. Because the sequences at the BPS and 5'SS in human introns are more degenerate than in yeast introns (**Fig. 1.2C**) (Coelho and Smith, 2014), the process of selecting splice sites in human pre-mRNAs is much more complicated than in yeast. About 95% of human pre-mRNAs can undergo alternative splicing (Pan et al., 2008; Wang et al., 2008), which is a process of making multiple gene products from the same gene via use of different splice sites. Due to the presence of highly conserved splicing signals within yeast introns, alternative splicing rarely occurs in yeast pre-mRNAs. In a sum, splicing in yeast represents a simplified model for studying the fundamental mechanism of pre-mRNA splicing.

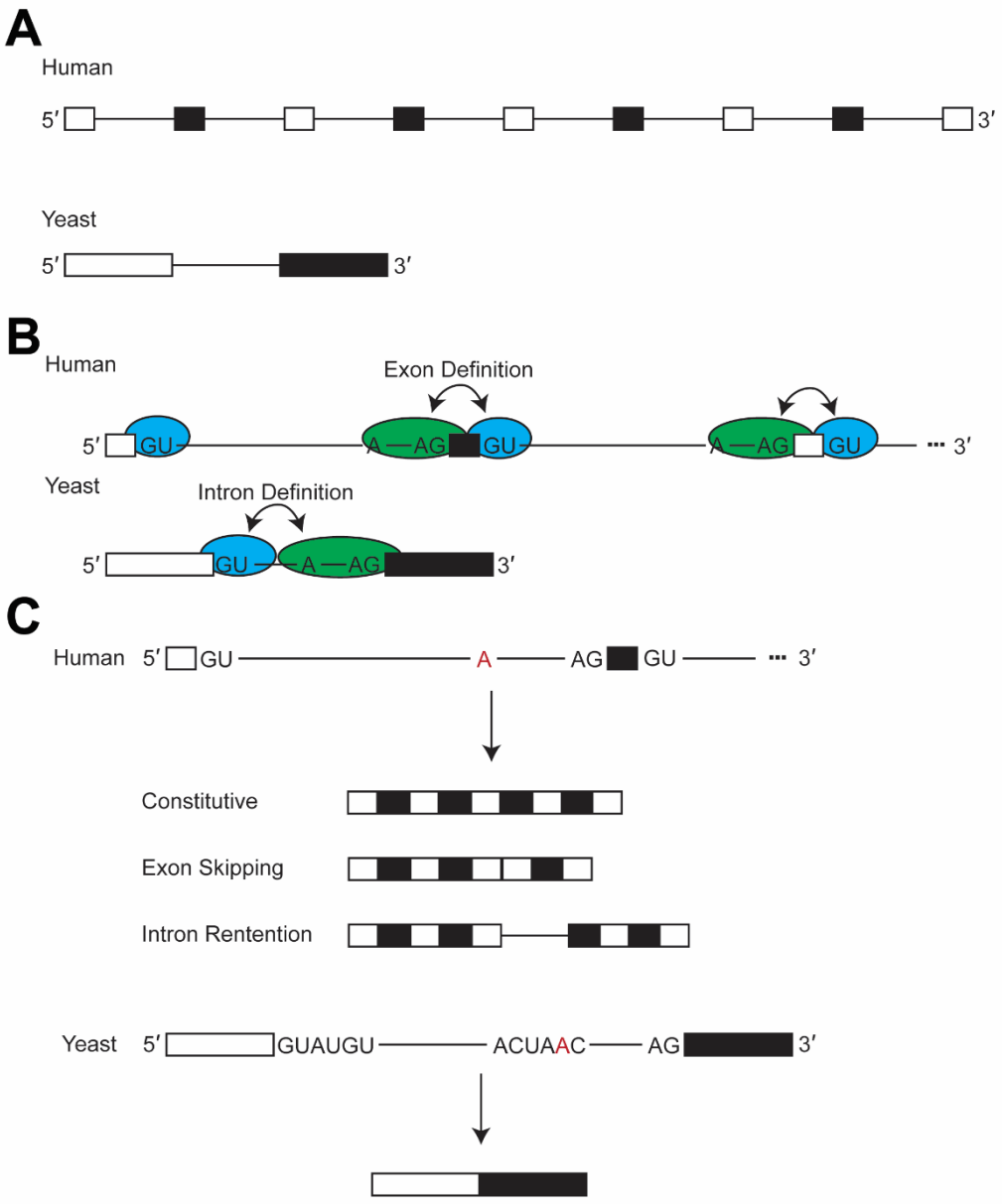


Figure 1.2 Comparison of Intron Architectures in Yeast and Humans.

(**A**) An average human pre-mRNA has eight introns and nine exons. An average yeast pre-mRNA has one intron and two exons. (**B**) In humans, the greater intron-to-exon size ratio has been proposed to lead to exon definition. Intron definition complexes tend to form for smaller introns as in yeast. (**C**) Low sequence conservation at splice sites for human pre-mRNAs correlates with alternative splice site usage, such as exon skipping and intron retention. Highly conserved sequences in yeast pre-mRNAs correlate with robust splicing site usage.

The Spliceosome and the Current Model for Its Splicing Mechanism

Spliceosomal Components

Decades of study of the megaDalton-sized spliceosome have revealed a relatively complete picture of the spliceosome-catalyzed splicing on simple model pre-mRNAs (Wahl ***et al.***, 2009). The building blocks of spliceosomes are five small nuclear ribonucleoproteins (snRNPs) and a number of individual proteins and protein complexes (such as NineTeen Complex, NTC). Each small nuclear ribonucleoprotein consists of one small nuclear RNA (snRNA) and several protein factors. While U1 snRNP and U2 snRNP join the spliceosome as individual subcomplexes, the U4/U6.U5 tri-snRNP particle joins the spliceosome as a single unit, which is pre-assembled from individual snRNPs.

Both the yeast and human splicing machineries have been studied extensively. Comparing their components reveals that a larger number of proteins exist in human splicing machinery (~170) than in yeast splicing machinery (~90) (Will and Lührmann, 2011). Almost every yeast spliceosomal protein has an evolutionary conserved orthologue in human, so they are expected to share many similar core functions in splicing. Even so, many differences exist between the two. The increased number of proteins found in the human spliceosome is proposed to afford extra flexibility when encountering the degenerate splicing signals in human transcripts.

Splicing Cycle Model

A splicing cycle model has been constructed to describe the stepwise assembly, activation, catalysis, disassembly and recycling of yeast and human spliceosomes Human splicing model is mostly conserved with yeast splicing model, with slight component differences. For the sake of this thesis, a simplified yeast splicing model is presented (**Fig. 1.3**). At the beginning of spliceosome assembly, the U1 snRNP and Mud2/Msl5 protein heterodimer recognize the 5'SS and BPS, respectively, leading to the formation of a commitment complex (named E complex).

Then, U2 snRNP displaces Mud2/Msl5 at the BPS to form a pre-spliceosome complex (the A complex). Subsequently, the U4/U6.U5 tri-snRNP joins the spliceosome to form a fully assembled pre-B complex. After all the snRNPs have assembled on the intron (pre-B complex), the activation stage begins. The 5'SS is switched from base pairing with the U1 snRNA to base pairing with the ACAGA sequence (upstream of ACAGA sequence in yeast) of the U6 snRNA, resulting in the release of U1 snRNP and the formation of pre-catalytic B spliceosome (named B complex). The subsequent transition from B to B^{ACT} involves the exchange of ~50 splicing factors, including U4 snRNP release and recruitment of a large protein-only assembly called the NTC. This requires disruption of U4/U6 base pairing, which in turn allows the formation of the U2/U6 catalytic triplex in B^{ACT}. In the last step of activation, the BPS is juxtaposed with the 5'SS in the B^{*} spliceosome. The spliceosome is then ready to carry out the first transesterification reaction at the catalytic center. In these catalytic stages, the spliceosome carries out the first step of splicing (branching, from B^{*} to C), then is further remodeled for the second step (from C to C^{*}), and lastly carries out the second step of splicing to form a post-splicing, product complex spliceosome (P complex). In the final disassembly and recycling stage, P complex, which contains both splicing products, will first release the mRNA (ligated exons) and form the intron-lariat spliceosome (ILS complex). Lastly, the ILS complex is disassembled, and its components are recycled for further rounds of splicing.

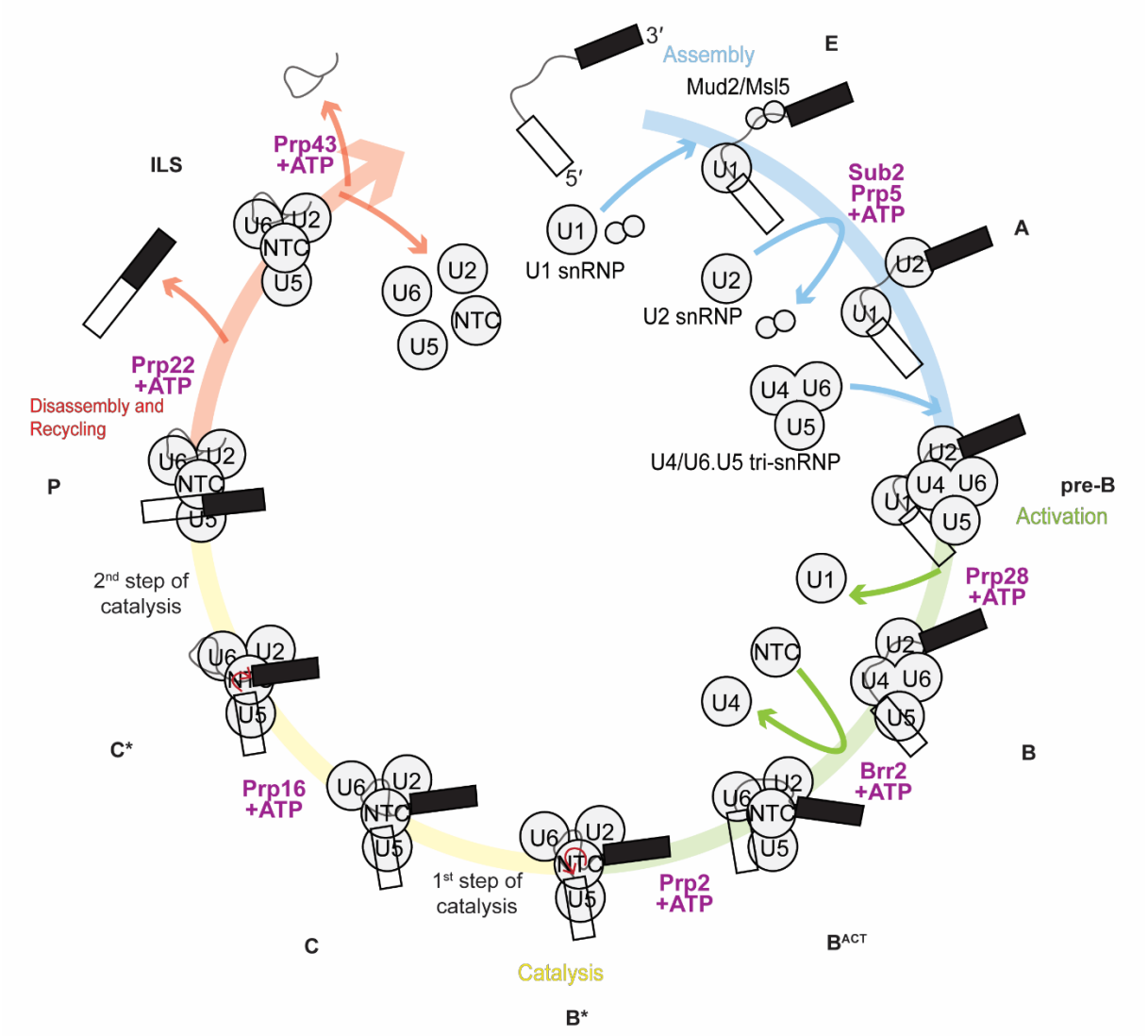


Figure 1.3 A Simplified Yeast Splicing Cycle.

The spliceosome needs to undergo assembly, activation, catalysis and disassembly and recycling to carry out splicing on a model pre-mRNA with a single intron. Briefly, the U1 snRNP, U2 snRNP and U4/U6.U5 tri-snRNP will sequentially assemble on a pre-mRNA undergoing the transition from commitment complex (E) to pre-spliceosome complex (A) to a fully assembled complex (pre-B). Transition from the E to A complex is driven by the ATPases Sub2 and Prp5 via replacing the Mud2/Msl5 heterodimer with the U2 snRNP. In the subsequent activation steps, the Prp28 ATPase facilitates the transfer of the 5'SS from the U1 snRNP to the U6 snRNA, resulting in the release of U1 and formation of the B complex. The transition to B^{ACT} is driven by Brr2 helicase

activity. The U4 snRNP is released, and the NTC is recruited to stabilize the nascent catalytic center. Prp2 activity triggers the remodeling of the spliceosome to bring branch point adenosine and 5'SS closer to the catalytic center, resulting in the formation of B* complex. Catalysis stage starts with the attack of BPA on the scissile phosphate at the 5' intron-exon boundary (red arrow; 1st step of catalysis) and forms C complex. Then C complex is remodeled by Prp16 activity to bring the end of 5' exon closer to the scissile phosphate at the 3' intron-exon boundary. In the transition from C* to P complex (red arrow; 2nd step of catalysis), mRNA and intron lariat are produced and retained in the product complex. With Prp22 and Prp43 activity, the products are released sequentially, and the ILS spliceosome is disassembled. The factors are recycled for other rounds of splicing. Figure is modeled after (Plaschka et al., 2019).

Principles of The Splicing Process

Compiling the details of the above steps of the splicing cycle leads to the discovery of working principles fundamental to spliceosomes (Wahl *et al.*, 2009). Firstly, spliceosome-catalyzed splicing is a highly compositionally and conformationally dynamic process. Subcomplexes and splicing factors join to and release from the spliceosome stepwise to ensure the accurate selection of splice sites, the proper formation of catalytic center, and the remodeling around active site. For instance, the step for forming the active sites (from B to B^{ACT} transition) involves the exchange of ~50 factors (**Fig. 1.3, 1.5**). It is of great interest in understanding how the dynamic and ordered recruitment and release of splicing factors are coordinated during the activation and other splicing stages. Secondly, the spliceosome uses an extremely intricate network of protein-protein, RNA-RNA, and protein-RNA interactions to ensure the progression of splicing. Thus, understanding the evolution of these interactions in different transition steps is also critical for understanding the mechanism of splicing (as discussed further in Section 1.6). Thirdly, a major driving force in the splicing cycle arises from the action of ATPases or helicases in the Ski2-like (Brr2), DEAD-box (Sub2, Prp5, Prp28), and DEAH-box (Prp2, Prp16, Prp22, Prp43) families (**Fig. 1.3**). These proteins can either unwind RNA/RNA duplex or translocate on single-stranded RNA to initiate structural rearrangements. Understanding how their actions are regulated or coupled to other events in splicing is also of great interest.

Complicated Transitions in Dynamic Spliceosome Activation

Unlike other ribonucleoprotein complexes such as the ribosome, the spliceosome does not have a pre-formed catalytic center (Wahl *et al.*, 2009). The catalytic triplex in the spliceosome is formed by the interaction between a short three-nucleotide long helix ($U2^{21-23}$/$U6^{59-61}$ helix Ib) and another three nucleotides, two of which are from the last $G^{52}A^{53}$ in the $ACGAG^{52}A^{53}$ region of the U6 snRNA and the third bulged U^{80} nucleotide from the internal stem-loop region (ISL, 62-85nt) of U6 (**Fig. 1.4B**).

In the spliceosome assembly stage, the residues for constituting the catalytic triplex remain separate (**Fig. 1.4C**). The corresponding $U2^{21-23}$ nucleotides cannot be identified in the cryo-EM structure of fully assembled pre-B spliceosome intermediate. However, its flanking sequences appear to be stabilized. The downstream branch site helix (BSH, $U2^{34-39}$/BPS) is held by Hsh155 (a U2 snRNP protein) and the upstream $U2^{3-13}$/$U6^{92-102}$ helix II may be stabilized by the Lsm2-8 ring (a U6 snRNP protein complex). Elsewhere in the structure, $U6^{56-80}$ remains in an inactive conformation by base pairing with the U4 snRNA to form U4/U6 stem I and stem II (**Fig. 1.4A, C**). During activation, the U4 snRNA is released by the action of the ATP-dependent helicase Brr2 to allow the formation of the $U6^{62-85}$ ISL, $U2^{26-30}$/$U6^{54-58}$ helix Ia, and $U2^{21-23}$/$U6^{59-61}$ helix Ib (**Fig. 1.4B, D**). The catalytic triplex is finally scaffolded by the largest spliceosome protein, Prp8 (279kDa, a U5 snRNP protein). The whole activation process can be broken into three transition steps, catalyzed in sequence by the ATPases Prp28, Brr2 and Prp2. In each transition step, dynamic compositional exchanges and structural rearrangements occur.

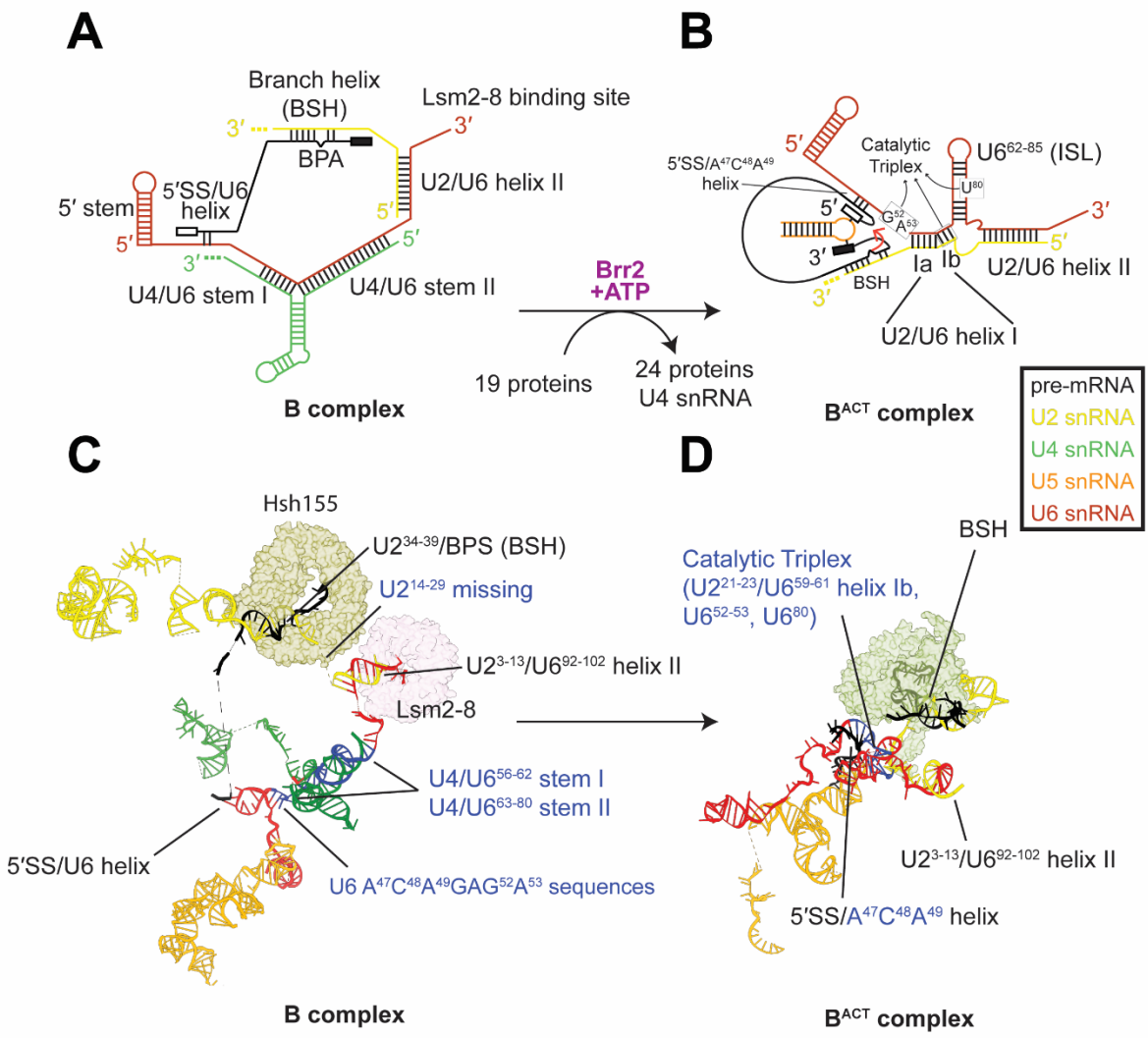


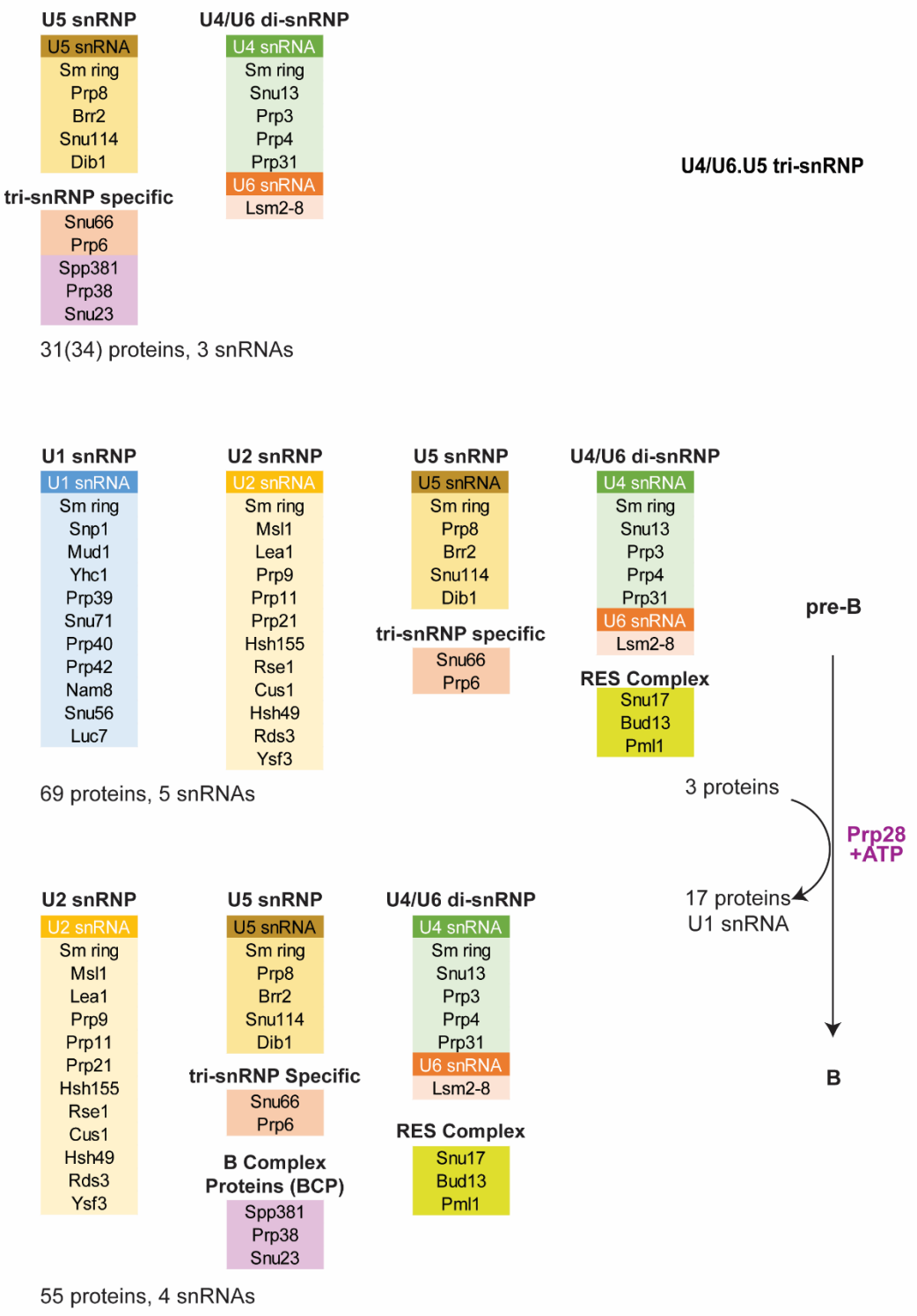
Figure 1.4 Spliceosome RNA Network Evolution from the Pre-catalytic to Catalytic State.

(A-B) Secondary structure representation of RNA networks from the pre-catalytic (B complex) to catalytic state (B^{ACT} complex). Components of U2/U6 catalytic triplex in B^{ACT} remain inactive in B complex. U6 nucleotides base pair with U4 in the pre-catalytic complex. After U4 snRNA release, the U6 ISL and U2/U6 helix Ia and Ib can form, which is required for forming catalytic triplex. **(C-D)** RNA tertiary structure representation of RNA networks from the pre-catalytic (B complex; PDB: 5NRL) to catalytic state (B^{ACT} complex; PDB: 5GM6). In B complex, the branch site helix (BSH) is held by Hsh155 (a U2 snRNP protein) and U2/U6 helix II is held by Lsm2-8. U2 nucleotides between BSH and helix II are missing from the structure in B complex (**C**). After U4 snRNA release,

the U2 snRNA is relocated proximal to the U6 snRNA for forming the U2/U6 catalytic triplex (**D**). In B^{ACT}, the 5'SS/$A^{47}C^{48}A^{49}$ duplex is positioned next to the catalytic triplex and the BSH remains held by Hsh155 away from the catalytic center. Labels for regions related to the catalytic triplex nucleotides are in blue.

Classes of Proteins Involved in Activation

Components of the spliceosome can be classified by either their presence in an isolatable subcomplex or by the step in which they are recruited and/or function during splicing (**Fig. 1.5**). The major subcomplexes are the snRNPs (U1, U2, U4, U5, and U6) and the NTC. The yeast tri-snRNP is assembled from U4/U6 di-snRNP and U5 snRNP, along with tri-snRNP specific proteins (e.g. Snu66 and Prp6 in yeast). Proteins from these major subcomplexes are recruited to the spliceosome together as part of the snRNP. On the other hand, the recruitment of step-specific factors is less defined. For instance, the nineteen related factors (NTR) are recruited at the same transition as the NTC (**Fig. 1.5,** B to B^{ACT} step). However, they are immunoprecipitated with NTC proteins at different sub-stoichiometric amounts, indicating that their recruitment may be different from the NTC core proteins (Hogg et al., 2010; Ohi et al., 2002). In another case, B-specific factors (yeast Prp38, Snu23, Spp381) are thought to be recruited together with the tri-snRNP in yeast but not in humans where they join at a later step (Agafonov et al., 2016; Bertram et al., 2017; Nguyen et al., 2016; Ulrich and Wahl, 2017). The dynamics and roles of these step-specific factors are less understood than the major subcomplexes. Understanding this class of proteins can provide further mechanistic insights into the spliceosome activation.


U5 snRNP
U5 snRNA
Sm ring
Prp8
Brr2
Snu114
Dib1
tri-snRNP specific
Snu66
Prp6
Spp381
Prp38
Snu23
31(34) proteins, 3 snRNAs
U4/U6 di-snRNP
U4 snRNA
Sm ring
Snu13
Prp3
Prp4
Prp31
U6 snRNA
Lsm2-8
U4/U6.U5 tri-snRNP
U1 snRNP
U1 snRNA
Sm ring
Snp1
Mud1
Yhc1
Prp39
Snu71
Prp40
Prp42
Nam8
Snu56
Luc7
U2 snRNP
U2 snRNA
Sm ring
Msl1
Lea1
Prp9
Prp11
Prp21
Hsh155
Rse1
Cus1
Hsh49
Rds3
Ysf3
69 proteins, 5 snRNAs
U5 snRNP
U5 snRNA
Sm ring
Prp8
Brr2
Snu114
Dib1
tri-snRNP specific
Snu66
Prp6
U4/U6 di-snRNP
U4 snRNA
Sm ring
Snu13
Prp3
Prp4
Prp31
U6 snRNA
Lsm2-8
RES Complex
Snu17
Bud13
Pml1
pre-B
3 proteins
Prp28
+ATP
17 proteins
U1 snRNA
B
U2 snRNP
U2 snRNA
Sm ring
Msl1
Lea1
Prp9
Prp11
Prp21
Hsh155
Rse1
Cus1
Hsh49
Rds3
Ysf3
U5 snRNP
U5 snRNA
Sm ring
Prp8
Brr2
Snu114
Dib1
tri-snRNP Specific
Snu66
Prp6
B Complex
Proteins (BCP)
Spp381
Prp38
Snu23
U4/U6 di-snRNP
U4 snRNA
Sm ring
Snu13
Prp3
Prp4
Prp31
U6 snRNA
Lsm2-8
RES Complex
Snu17
Bud13
Pml1
55 proteins, 4 snRNAs

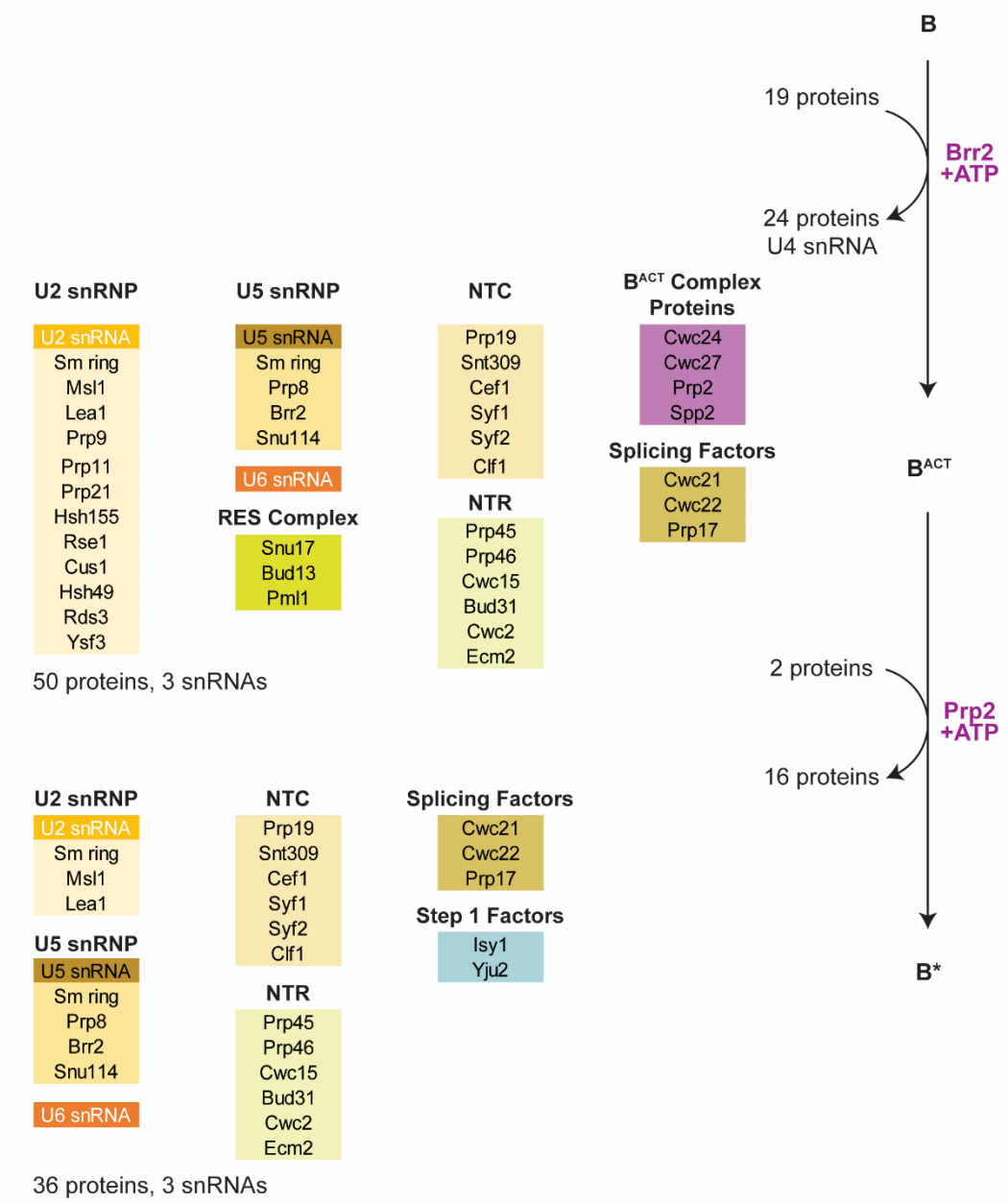


Figure 1.5 Overall Compositional Changes during Spliceosome Activation from the pre-B to B^{ACT} Complex.

Compositions of each activation intermediate are listed based on classes. Number of factors in each intermediate are indicated below the list. Transitions between intermediates and associated exchanges during the transition are shown on the right of the list. Tri-snRNP compositions are also included for comparison. Briefly, from pre-B to B complex, the U1 snRNP is released and B

complex proteins (BCP) are recruited. Note: BCP are identified in tri-snRNP, but not in pre-B complex. The transition from B to B^{ACT} represents a more complicated step in splicing, involving the exchange of about 50 factors. From B^{ACT} to B^{*}, a subset of U2 snRNP proteins, the RES complex, and B^{ACT} complex proteins are released while step 1 factors are recruited.

The Early Activation Stage – Prp28-mediated Transition from pre-B to B Complex

After the recruitment of the tri-snRNP to the pre-spliceosome, the pre-B complex undergoes a transition to the pre-catalytic B complex, mediated by the ATPase Prp28 (**Fig. 1.6A-B**). Prp28 destabilizes the U1/5'SS duplex to facilitate 5'SS transfer to the U6 snRNA (near the $A^{47}C^{48}A^{49}G^{50}A^{51}G^{52}A^{53}$ sequence, yeast numbering) and U1 snRNP release. In later steps, a base pairing interaction between the 5'SS and $A^{47}C^{48}A^{49}$ will be established and remain throughout the activation process. This interaction will then be brought to the active site when the nearby U6 nucleotides $G^{52}A^{53}$ become part of the catalytic triplex.

The Prp28 binding site is subsequently occupied by the B complex proteins (BCP; Prp38, Snu23, scSpp381/hsMFAP1). The site is structurally close to the 5'SS/U6 duplex. Significant differences arise when comparing further details of the human and yeast cryo-EM structures. Some yeast cryo-EM structures of the pre-B complex, particularly those obtained at low concentrations of ATP, lack the evolutionarily conserved and essential Prp28 and Sad1 proteins, which can be found in the human pre-B intermediate (Bai et al., 2018; Charenton et al., 2019). These same compositional differences are also found when comparing the yeast tri-snRNP to the human tri-snRNP (Agafonov *et al.*, 2016; Nguyen *et al.*, 2016).

Comparing the locations of other essential factors in the yeast and human spliceosome reveals several additional differences. In human cryo-EM structures, Brr2 locates away from its substrate in the pre-B intermediate (and tri-snRNP). In the transition to the B intermediate, Brr2 relocates and is loaded onto the U4 snRNA binding site (**Fig. 1.6B**). The initial location of Brr2 seems to be stabilized by the presence of human Prp28 and Sad1 (**Fig. 1.6B, D**). In contrast, Brr2 has already loaded onto the U4 substrate in the yeast tri-snRNP and remains unchanged in pre-B and B intermediate structures (**Fig. 1.6A**). Another disparity is the content of the exon channel in different complex structures. The description of a protein channel for the exon was initially proposed based on the observation of 5' exon extending away from the active site and

being accommodated by a tunnel created between the $Prp8^{NTD}$ and $Prp8^{Large}$ [including Reverse Transcriptase (RT), Thumb/X, Linker, and Endonuclease (En)] domains in structures of catalytic spliceosomes (Galej et al., 2016). In structures of human spliceosomes during the early stages of activation, the exon channel is first occupied by a protein strand of $hPrp28^{259-295}$ (acting as an exon RNA mimic) to the 5' exon (**Fig. 1.6B, D**). Unusually, the exon channel observed in cryo-EM structures of yeast spliceosomes prior to activation and the tri-snRNP is occupied by the 5' stem of the U6 snRNA (**Fig. 1.6A, C**). This indicates that there potentially needs to be a switch of the U6 5' stem for the 5' exon inside the tunnel during the stage of yeast activation. However, no evidence can support a molecular mechanism for this process up to now. Alternatively, it is possible that these structures from yeast do not represent catalytically competent intermediates.

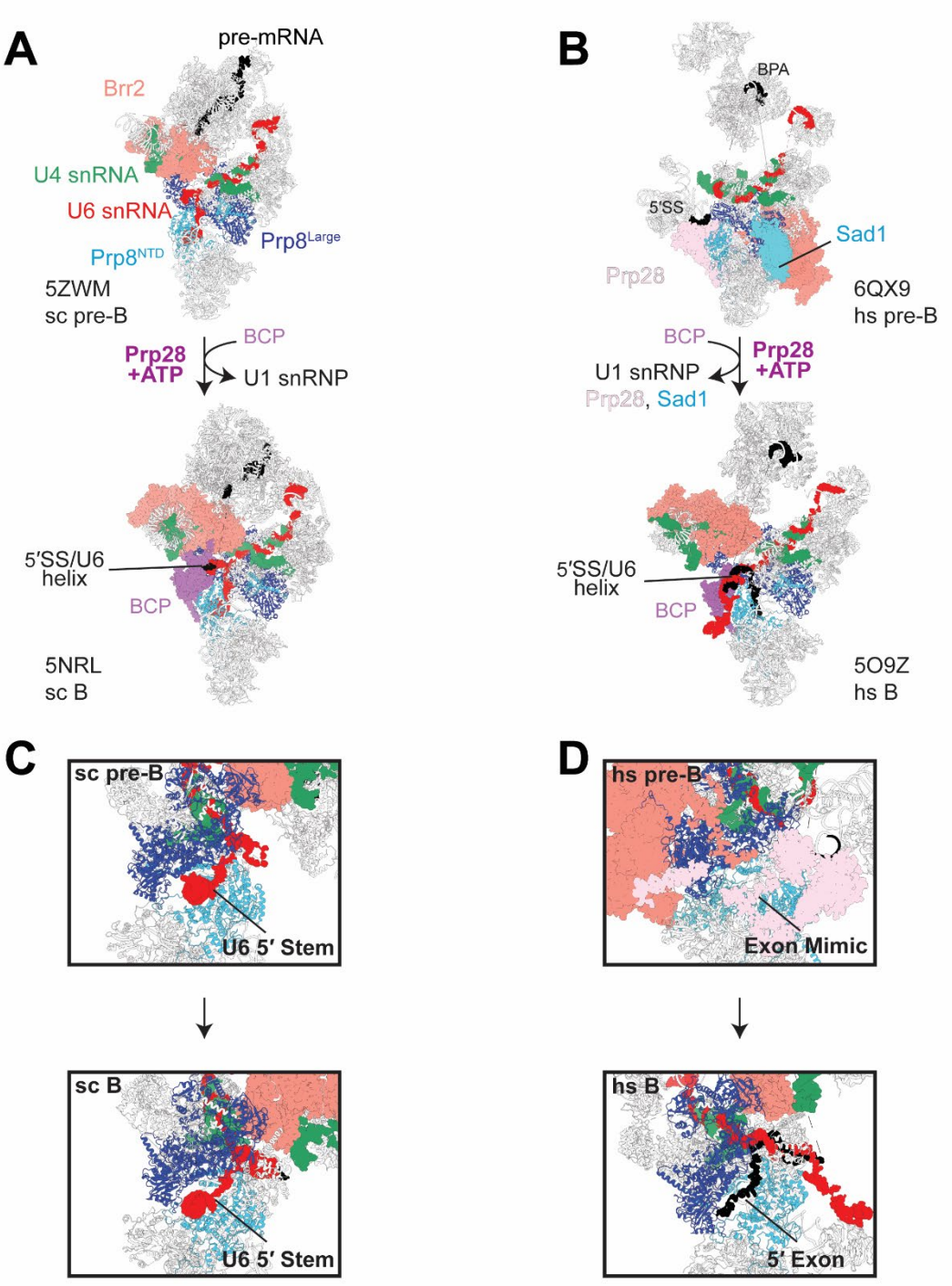


Figure 1.6 Similarities and Differences between Yeast and Human Cryo-EM Structures During the Early Activation Transition of pre-B to B Complex.

(A-B) In both models, Prp28 mediates the transfer of 5'SS from U1 snRNA to U6 snRNA. BCP is recruited and U1 snRNP is released. However, Prp28 and Sad1 are present in human pre-B, but

not in yeast pre-B. Brr2 rearranges from inactive to active location in the human model, while it is already present in the active location in both yeast pre-B and B complexes. **(C-D)** Boxes show specific regions of back views to the intermediates shown in (**A, B**). These boxes highlight changes of the content in exon exit channel, which is formed between $Prp8^{NTD}$ and $Prp8^{Large}$. The U6 5' stem remains in the channel for the yeast model, while the exon mimic of Prp28 is replaced by 5' exon for the human model. PDB numbers are indicated for each intermediate. U1 snRNPs are omitted from pre-B intermediates for simplicity.

The Middle Activation Stage – Brr2-mediated Transition from the B to B^{ACT} Complex

In the next transition, the ATP-dependent Brr2 helicase likely initiates the unwinding of U4/U6 stem I, followed by the exchange of ~50 factors in yeast (more in humans). The released factors include di-snRNP components (except for the U6 snRNA), the BCP, tri-snRNP specific factors, and Dib1 (**Fig. 1.5**). The elimination of these factors could make space for the folding of the U6 ISL into the spliceosome active site and for bringing the U2 snRNA/branch site duplex closer to the catalytic center (**Fig. 1.4C-D, 1.7A**). The recruited factors (NTC, NTR and B^{ACT}-specific factors, Prp2, RES) can stabilize the nascent catalytic center in this new conformation or facilitate the next set of structural rearrangements (**Fig. 1.7C**).

Since the BCP are needed to form B complex intermediates and are released in the transition to the B^{ACT} complex, they are likely to play important roles during this transition. Among the many human BCP components, only Prp38, Snu23 and MFAP (yeast Spp381) are conserved in yeast. They colocalize with one another at the same location in the B complex and may be important for loading Brr2 onto its U4 substrate and stabilizing the nucleotides near the 5' SS/U6 duplex (**Fig. 1.7B**). The center of the BCP subunit localizes in a pocket created by Brr2, $Prp8^{En}$ (a subdomain of $Prp8^{Large}$) and $Prp8^{NTD}$, with alpha helical regions extending in different directions on the spliceosome (**Fig. 1.7C**). For instance, the N-terminus of Snu23 appears to reach the ATPase active site in Brr2 (**Fig. 1.7B**), suggesting a role in regulation of the Brr2 activity. Alpha helices at each terminus of the elongated Spp381 protein appear to make peripheral interactions with both Brr2 and $Prp8^{NTD}$ domain, potentially acting as a protein bridge connecting them together (**Fig. 1.7C**). Prp38 and Snu23 localize near the 5'SS/U6 duplex in the front of exon channel (**Fig. 1.7B**), suggesting a potential role for stabilizing this duplex during activation. Once the BCP is released, this particular interaction is replaced by Cwc24 and Cwc27 during the transition to B^{ACT}. The C-terminus of hsMFAP1 along with hsUBL5 (another BCP component in humans) appears to stabilize the 5' exon at the back of the channel while the C-terminus of Spp381 has not yet

been resolved in yeast spliceosomes. In the transition to B^{ACT}, MFAP1 and UBL5 are released, the size of exon channel is further restrained by the relative movement of $Prp8^{Large}$ and $Prp8^{NTD}$ (**Fig. 1.7C**). The interactions around the 5' exon are substituted by Cwc21 and Cwc22, which remain associated until mRNA product release.

While the U2/U6 catalytic center has formed in B^{ACT}, the scissile phosphate at the 5' exon/5' SS boundary and the nucleophilic adenosine (BPA) within $U2^{34-39}$/BPS branch helix have not docked properly near one another in the active site for 1st step catalysis (**Fig. 1.7D**). The U2/U6 catalytic triplex, coordinating two Mg^{2+}, has already formed in the B^{ACT} spliceosome on the $Prp8^{NTD}$. While the 5'SS/ $A^{47}C^{48}A^{49}$ is placed close to the catalytic site, the scissile phosphate is protected by Prp11 (U2 snRNP protein) and Cwc24 (B^{ACT}-specific protein). Even though U2 snRNP makes a dramatic conformational change as it fills vacant space previously occupied by the U4 snRNP and other factors, the nucleophilic BPA is still placed ~50 Å away from the 5'SS and surrounded by Hsh155 (a U2 snRNP protein).

Finally, several important structural features are established along the U6 snRNA in the B to B^{ACT} transition. The catalytic triplex region ($U6^{54-85}$, including Ia, Ib and ISL) is positioned over the $Prp8^{NTD}$ and encircled by several NTC/NTR proteins which remain throughout catalysis (**Fig. 1.7E**; factors around the triplex not shown for clarity). Downstream of the catalytic region near the 3'-end of U6, U2 and U6 also pair with one another to form $U2^{3-13}$/$U6^{92-102}$ helix II. From the B to B^{ACT} transition, helix II is repositioned together with U2 snRNP towards $Prp8^{NTD}$ and stabilized in that position by NTC/NTR proteins in the B^{ACT} intermediate (**Fig. 1.7E**). Single stranded $U6^{26-46}$ in the docks onto $Prp8^{NTD}$ and is stabilized by Bud31/Ecm2/Cwc2 (or hRBM22/hCWC2). The 5' stem of $U6^{1-25}$, which is located in the exon channel in structures of the yeast spliceosome, extends toward the periphery of the spliceosome.

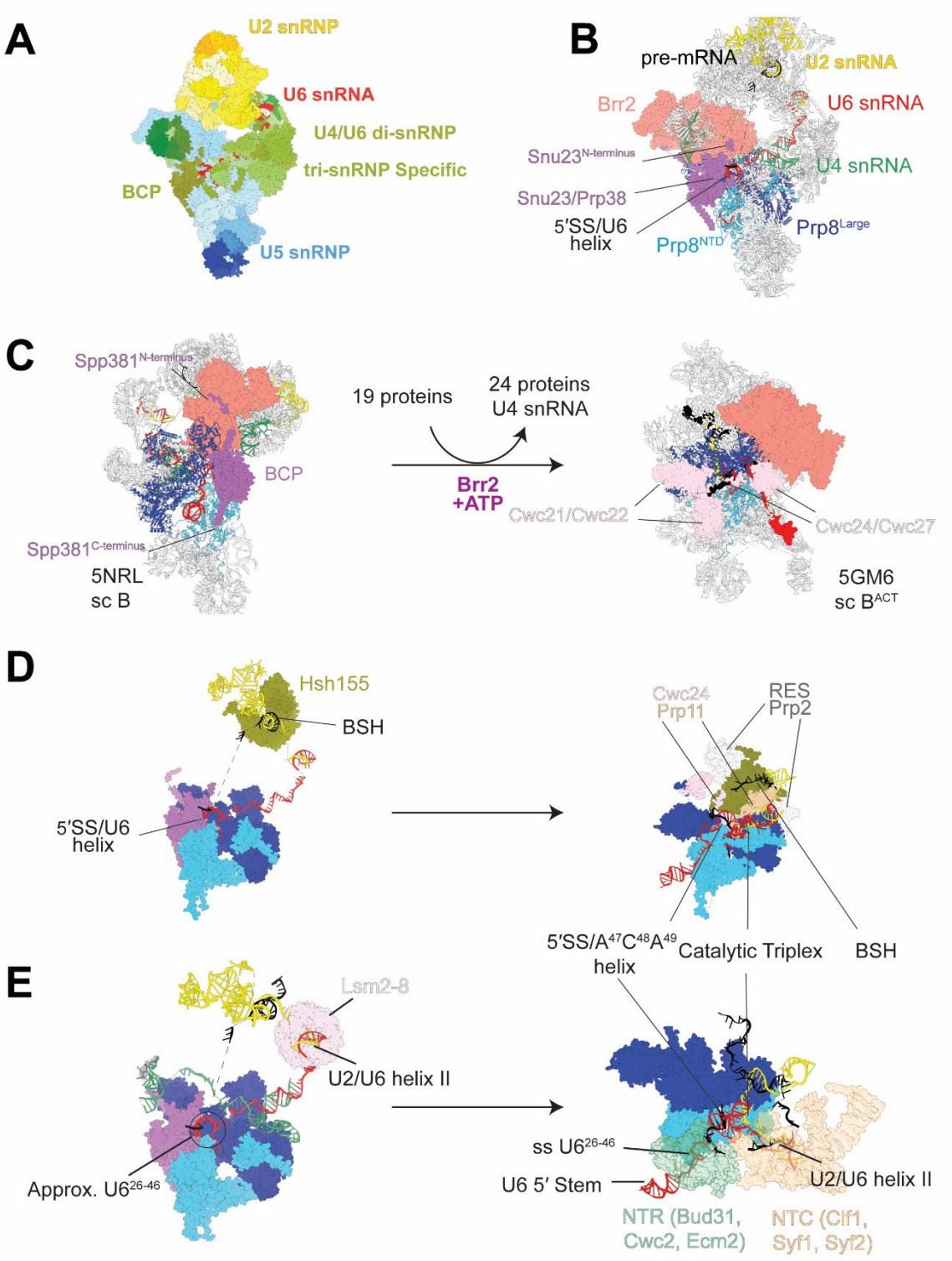


Figure 1.7 Factors That Contribute to the Formation of the RNA network During the Middle Activation Transition from B to B^{ACT} Complex.

(A) Overall spatial distribution of classes of factors in yeast B spliceosome. U2 snRNP, U5 snRNP and U6 snRNA are colored in shades of yellow, blue and red, respectively. Factors that are released are indicated in shades of green, including the di-snRNP components (except for U6 snRNA), tri-snRNP specific factors and BCP. **(B-C)** Extensive interactions made by the BCP in the B complex are replaced by Cwc21/Cwc22 and Cwc24/Cwc27 during this transition. Front view (**B**) of the B complex intermediate shows the BCP interacting with the active site of Brr2 (via Snu23$^{N\text{-terminus}}$) and with the 5'SS/U6 duplex (by Snu23/Prp38). Back view (**C;** before transition) of the B complex shows BCP interacting with the gate of the exon exit channel (between Prp8^{NTD} and Prp8^{Large}) and Brr2. C-terminus of human MFAP1 (yeast Spp381) interacts with RNA within the exon exit channel (not shown). After the transition to B^{ACT}, exon channel interactions with Spp381 are replaced by Cwc21/Cwc22 and interactions around the gate are replaced by B^{ACT} complex proteins (Cwc24/Cwc27). **(D)** Interactions around the splice sites and branch site from B to B^{ACT}. Prior to the transition, the scissile phosphate is positioned closed to the 5'SS/U6 helix, likely stabilized by the BCP. After the transition, the phosphate group is positioned closed to the catalytic center and protected by Prp11 and Cwc24. During this transition, the branch site helix (BSH) held by Hsh155 moves towards the catalytic center. (**E**) Interactions along the U6 snRNA from B to B^{ACT}. During the transition, U2/U6 helix II moves toward the Prp8^{NTD} and interactions around U2/U6 helix II are switched from Lsm2-8 to NTC (Clf1, Syf1, Syf2). The U6 nucleotides containing the catalytic regions are stabilized by NTC and NTR proteins after the transition (not shown for clarity). Upstream to U6 A^{47}CAGAGA sequences, the U6$^{26\text{-}46}$ region was was stabilized by BCP in B intermediate. After the transition, it is stabilized by NTR proteins (including Bud31, Cwc2 and Ecm2). The U6 5' stem, positioned in the exon exit channel in the yeast B intermediate, is removed from this location and now extends towards the periphery of Prp8^{NTD} after the formation of catalytic center.

The Late Activation Stage – Prp2-mediated Transition from the B^{ACT} to B Complex*

The last step of activation mediated by the Prp2 ATPase. Prp2's activity triggers a structural rearrangement that frees the scissile phosphate at the 5'SS and delivers the BPA to the active site. In the presence of coactivator Spp2 and ATP, Prp2 likely translocate on the intron in a 3' to 5' direction to release SF3a (including Prp11) and SF3b (including Hsh155) subcomplexes of the U2 snRNP, thus freeing the BSA for docking at the catalytic center. Meanwhile, structural rearrangements also lead to the release of B^{ACT}-specific proteins (Cwc24, Cwc27) allowing the scissile phosphate to dock at the catalytic site. Formation of B^{*} is accompanied with the recruitment of Prp17, Isy1, and Yju2. In the transition, the distance between the scissile phosphate and BPA is shortened from ~50 Å to ~4 Å. The distance is hypothesized to be further shortened upon recruiting another step1 factor (Cwc25) concomitant with the 1^{st} step of splicing (branching, B^{*} to C transition) occurring at the catalytic center.

Thesis Overview

My thesis work aims to understand the pathways of spliceosome activation using state-of-the-art single molecule techniques and the roles of splicing factors involved in activation using biochemical and genetic assays. Even with state-of-the-art cryo-EM techniques, the dynamic exchange of ~50 factors during activation remains a challenging area to study. Aiming to place the static snapshots of cryo-EM structures in a kinetic framework, I chose to take advantage of a multi-color single molecule technique (colocalization single molecule spectroscopy, CoSMoS) to monitor *in vitro* splicing in real time, carefully analyze the binding patterns of different proteins, and compile these patterns into a kinetic map of spliceosome activation. With use of a spatiotemporal map of spliceosome activation, I characterized the previously unknown roles of different factors involved in activation (Lsm2-8 and Spp381) and catalysis (Ecm2).

Chapter 2 focuses on understanding the roles of Lsm2-8 (a U6 snRNP protein) in spliceosome activation. Monitoring the dynamics of Lsm2-8 relative to landmark events of activation stages (U4 snRNP release and NTC recruitment) reveals the delayed release of Lsm2-8 after NTC recruitment. We then used *in vitro* EMSA assays to show that Lsm2-8 can potentially form a complex with U2/U6 helix II, implicating a previously unknown role for Lsm2-8 during spliceosome activation in stabilizing weak RNA duplexes.

Chapter 3 focuses on understanding the roles of poorly understood yeast B complex proteins during activation. Monitoring the dynamics of B complex proteins relative to each other and landmark activation events of activation stages reveals that the B complex proteins have temporally coupled dynamics suggesting they form their own subunit of the splicing machinery. This subunit is recruited after tri-snRNP integration and released after NTC binding under splicing conditions. However, if activation is inhibited by low [ATP], the proteins are recruited as part of the tri-snRNP. I propose a model for linking these two observations together based on ATP-dependent changes in tri-snRNP protein composition.

Within the same chapter, I described a parallel study on a tri-snRNP specific factor Snu66. Monitoring the dynamics of Snu66 relative to landmark events of activation and B complex proteins reveals that Snu66 has temporally coupled dynamics to U4 snRNP proteins, also cross validating the observed dynamics of B complex proteins in the activation process. Compiling the kinetics of spliceosomal intermediates that contain characteristic compositions allows me to build a map for showing a dynamic and ordered process of spliceosome activation.

As a continuation for understanding the roles of B complex proteins (Spp381, Snu23, Prp38) in activation, Chapter 4 focuses on the functional characterization of Spp381, which consists of several structurally extended alpha helices. My mutational analysis on Spp381 allowed me to identify helical regions and residues that are essential in a context-dependent manner. Functional characterization on the mutants reveals different extents of *in vivo* splicing defects and

growth defects across a temperature range. Whether expression and localization of *spp381* mutants are affected and what specific interactions made by Spp381 with the spliceosome are disrupted in these mutant strains remain to be characterized.

Chapter 5 focuses on the roles of Ecm2 (a non-essential NTR protein) in catalysis stage. This work was led by a former lab member, Clarisse van der Feltz. I was fortunate enough to help with carrying out some experiments. A genetic approach was taken to assay the synthetic effects of combining *ecm2* mutants with a variety of previously well-characterized catalysis-related mutants in splicing factors. Our results support a model that Ecm2 facilitates the formation and stabilization of the catalytic site in the 1st step as well as promotes the catalysis in the 2nd step.

Chapter 6 summarizes major contributions from this work and discusses my thoughts on near-term and far future directions. Future directions include understanding the molecular mechanism of forming Lsm2-8/U2/U6 complex, understanding roles of B-specific factors in activation, characterization of the yeast tri-snRNP composition and structure at high [ATP], forming a better understanding of spliceosome activation, and the significance of building a kinetic map for this process.

REFERENCES

Agafonov, D.E., Kastner, B., Dybkov, O., Hofele, R.V., Liu, W.T., Urlaub, H., Lührmann, R., and Stark, H. (2016). Molecular architecture of the human U4/U6.U5 tri-snRNP. Science *351*, 1416-1420. 10.1126/science.aad2085.

Bai, R., Wan, R., Yan, C., Lei, J., and Shi, Y. (2018). Structures of the fully assembled. Science *360*, 1423-1429. 10.1126/science.aau0325.

Baker, K.E., and Parker, R. (2004). Nonsense-mediated mRNA decay: terminating erroneous gene expression. Curr Opin Cell Biol *16*, 293-299. 10.1016/j.ceb.2004.03.003.

Berget, S.M. (1995). Exon recognition in vertebrate splicing. J Biol Chem *270*, 2411-2414. 10.1074/jbc.270.6.2411.

Bertram, K., Agafonov, D.E., Dybkov, O., Haselbach, D., Leelaram, M.N., Will, C.L., Urlaub, H., Kastner, B., Lührmann, R., and Stark, H. (2017). Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation. Cell *170*, 701-713.e711. 10.1016/j.cell.2017.07.011.

Charenton, C., Wilkinson, M.E., and Nagai, K. (2019). Mechanism of 5' splice site transfer for human spliceosome activation. Science *364*, 362-367. 10.1126/science.aax3289.

Coelho, M.B., and Smith, C.W. (2014). Regulation of alternative pre-mRNA splicing. Methods Mol Biol *1126*, 55-82. 10.1007/978-1-62703-980-2_5.

Fica, S.M., Mefford, M.A., Piccirilli, J.A., and Staley, J.P. (2014). Evidence for a group II intron-like catalytic triplex in the spliceosome. Nat Struct Mol Biol *21*, 464-471. 10.1038/nsmb.2815.

Galej, W.P., Wilkinson, M.E., Fica, S.M., Oubridge, C., Newman, A.J., and Nagai, K. (2016). Cryo-EM structure of the spliceosome immediately after branching. Nature *537*, 197-201. 10.1038/nature19316.

Hogg, R., McGrail, J.C., and O'Keefe, R.T. (2010). The function of the NineTeen Complex (NTC) in regulating spliceosome conformations and fidelity during pre-mRNA splicing. Biochem Soc Trans *38*, 1110-1115. 10.1042/BST0381110.

Kwan, T., Benovoy, D., Dias, C., Gurd, S., Provencher, C., Beaulieu, P., Hudson, T.J., Sladek, R., and Majewski, J. (2008). Genome-wide analysis of transcript isoform variation in humans. Nat Genet *40*, 225-231. 10.1038/ng.2007.57.

Lambowitz, A.M., and Zimmerly, S. (2011). Group II introns: mobile ribozymes that invade DNA. Cold Spring Harb Perspect Biol *3*, a003616. 10.1101/cshperspect.a003616.

Louhichi, A., Fourati, A., and Rebaï, A. (2011). IGD: a resource for intronless genes in the human genome. Gene *488*, 35-40. 10.1016/j.gene.2011.08.013.

López-Bigas, N., Audit, B., Ouzounis, C., Parra, G., and Guigó, R. (2005). Are splicing mutations the most frequent cause of hereditary disease? FEBS Lett *579*, 1900-1903. 10.1016/j.febslet.2005.02.047.

Moore, M.J., and Sharp, P.A. (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. Nature *365*, 364-368. 10.1038/365364a0.

Nguyen, T.H.D., Galej, W.P., Bai, X.C., Oubridge, C., Newman, A.J., Scheres, S.H.W., and Nagai, K. (2016). Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. Nature *530*, 298-302. 10.1038/nature16940.

Ohi, M.D., Link, A.J., Ren, L., Jennings, J.L., McDonald, W.H., and Gould, K.L. (2002). Proteomics analysis reveals stable multiprotein complexes in both fission and budding yeasts containing Myb-related Cdc5p/Cef1p, novel pre-mRNA splicing factors, and snRNAs. Mol Cell Biol *22*, 2011-2024. 10.1128/MCB.22.7.2011-2024.2002.

Orphanides, G., and Reinberg, D. (2002). A unified theory of gene expression. Cell *108*, 439-451. 10.1016/s0092-8674(02)00655-4.

Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet *40*, 1413-1415. 10.1038/ng.259.

Plaschka, C., Newman, A.J., and Nagai, K. (2019). Structural Basis of Nuclear pre-mRNA Splicing: Lessons from Yeast. Cold Spring Harb Perspect Biol *11*. 10.1101/cshperspect.a032391.

Proudfoot, N.J., Furger, A., and Dye, M.J. (2002). Integrating mRNA processing with transcription. Cell *108*, 501-512. 10.1016/s0092-8674(02)00617-7.

Roy, S.W., and Irimia, M. (2014). Diversity and evolution of spliceosomal systems. Methods Mol Biol *1126*, 13-33. 10.1007/978-1-62703-980-2_2.

Sakharkar, M.K., Chow, V.T., and Kangueane, P. (2004). Distributions of exons and introns in the human genome. In Silico Biol *4*, 387-393.

Schneider, M., Will, C.L., Anokhina, M., Tazi, J., Urlaub, H., and Lührmann, R. (2010). Exon definition complexes contain the tri-snRNP and can be directly converted into B-like precatalytic splicing complexes. Mol Cell *38*, 223-235. 10.1016/j.molcel.2010.02.027.

Singh, R.K., and Cooper, T.A. (2012). Pre-mRNA splicing in disease and therapeutics. Trends Mol Med *18*, 472-482. 10.1016/j.molmed.2012.06.006.

Ulrich, A.K., and Wahl, M.C. (2017). Human MFAP1 is a cryptic ortholog of the Saccharomyces cerevisiae Spp381 splicing factor. BMC Evol Biol *17*, 91. 10.1186/s12862-017-0923-1.

Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701-718. 10.1016/j.cell.2009.02.009.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature *456*, 470-476. 10.1038/nature07509.

Will, C.L., and Lührmann, R. (2011). Spliceosome structure and function. Cold Spring Harb Perspect Biol *3*. 10.1101/cshperspect.a003707.

Wongpalee, S.P., and Sharma, S. (2014). The pre-mRNA splicing reaction. Methods Mol Biol *1126*, 3-12. 10.1007/978-1-62703-980-2_1.

Chapter Two

Identification of Transient Intermediates During Spliceosome Activation by Single Molecule Fluorescence Microscopy

This chapter is published in the following form:

Fu, X., Kaur, H., Rodgers, M.L., Montemayor, E.J., Butcher, S.E., Hoskins, A.A. (2022) Identification of Transient Intermediates During Spliceosome Activation by Single Molecule Fluorescence Microscopy. Under Review.

https://biorxiv.org/cgi/content/short/2022.04.23.488636v1

Author contributions:

XF, MLR, and AAH conceived the project. XF, HK, and MLR carried out experiments. ECM and SEB provided purified Lsm proteins. XF, HK, MLR, and AAH analyzed data. XF and AAH wrote the manuscript with input from all authors.

ABSTRACT

Spliceosome activation is the process of creating the catalytic site for RNA splicing and occurs *de novo* on each intron following spliceosome assembly. Dozens of factors bind to or are released from the activating spliceosome including the Lsm2-8 heteroheptameric ring that binds the U6 small nuclear RNA (snRNA) 3'-end. Lsm2-8 must be released to permit active site stabilization by the Prp19-containing complex (NineTeen Complex, NTC); however, little is known about the temporal order of events and dynamic interactions that lead up to and follow Lsm2-8 release. We have used colocalization single molecule spectroscopy (CoSMoS) to visualize Lsm2-8 dynamics during activation of yeast spliceosomes. Lsm2-8 is recruited as a component of the tri-snRNP and is released after integration of the Prp19-containing complex (NineTeen Complex, NTC). Despite Lsm2-8 and the NTC being mutually exclusive in existing cryo-EM structures of yeast B complex spliceosomes, we identify a transient intermediate containing both ($\mathrm{B}^{\Delta U4}_{\mathrm{Lsm+NTC}}$) and provide a kinetic framework for its formation and transformation during activation. Prior to $\mathrm{B}^{\Delta U4}_{\mathrm{Lsm+NTC}}$ assembly, the NTC rapidly and reversibly samples the spliceosome suggesting a mechanism for preventing NTC sequestration by defective spliceosomes that fail to properly activate. In complementary ensemble assays, we show that a base-pairing dependent ternary complex can form between Lsm2-8 and U2 and U6 helix II RNAs. Together our data suggest a Hfq-like function for Lsm2-8 in maintaining U2/U6 helix II integrity before it can be transferred to the NTC by transient formation of the $\mathrm{B}^{\Delta U4}_{\mathrm{Lsm+NTC}}$ spliceosome.

INTRODUCTION

Pre-mRNA splicing is an essential step in eukaryotic gene expression. Disruption of the splicing process on a molecular level can lead to human diseases including retinitis pigmentosa, cancers, and amyotrophic lateral sclerosis (1, 2). Molecular mechanisms of splicing have been studied for decades *in vitro* and *in vivo* by a combination of techniques. The overall splicing process is evolutionarily conserved and involves formation of numerous intermediate spliceosome complexes (**Figs. 2.1A, 2.2**) (3-6). The building blocks of spliceosomes include five small nuclear ribonucleoproteins (U1, U2, U4, U5, and U6 snRNPs), each made of one snRNA and several protein factors. These snRNPs undergo assembly on the substrate, activation to form the catalytic site, catalysis of two sequential transesterification reactions, and disassembly/recycling. The NTC and NTC-related factors (NTR) are protein-only subunits that are necessary for catalysis and join the spliceosome during activation (7, 8). A successful splicing process is achieved by assembly and disassembly of intricate interaction networks involving 5 snRNAs and over 170 protein factors in humans and ~90 in *S. cerevisiae* (yeast) (5, 9).

The Sm and Lsm2-8 protein complexes are two important factors that contribute to pre-mRNA splicing by chaperoning snRNAs during snRNP biogenesis and forming interactions with other splicing factors (10-13). Sm and Lsm2-8 proteins in the spliceosome form heteroheptameric rings with small central channels for RNA binding (14). Individual Sm or Lsm proteins share a common fold with the Hfq protein, which forms homohexamers and regulates sRNA/mRNA annealing in bacteria (15). The Lsm2-8 complex preferentially binds to the post-transcriptionally processed 3' end of U6 snRNA (16) and is present in the U6 snRNP and U4/U6.U5 tri-snRNP. Lsm2-8 remains bound to U6 during spliceosome assembly but is released from yeast spliceosomes during activation in a process that requires the NTC (17). For the yeast splicing machinery, it is not known if Lsm2-8 release occurs before, after, or concertedly with NTC association; however, recent cryo-EM structures of human spliceosomes show that spliceosome

complexes containing both Lsm2-8 and a subset of NTC proteins can form (18). Formation of the spliceosome active site has been proposed to be directed by a series of mutually exclusive interactions in which Lsm2-8 and the NTC play critical roles (18). Therefore, knowledge of the kinetic pathways for Lsm2-8 release and NTC recruitment is essential for understanding activation of spliceosomes.

Spliceosome activation is complex and involves dozens of coordinated compositional and conformational changes. For example, the transition from the yeast B to B^{ACT} spliceosome involves the exchange of around 50 different factors (5) (**Figs. 2.1A, 2.2**), making it intrinsically difficult to study. While cryo-EM has captured snapshots of human spliceosome activation intermediates (18) and has provided key insights on the mechanism of splicing (13), it is difficult to derive kinetic pathways from these data. Previously, our lab has used Colocalization Single Molecule Spectroscopy (CoSMoS) to study the activation pathways of single spliceosomes by monitoring tri-snRNP binding, U4 snRNP protein release, and NTC recruitment (19). Single molecule techniques like CoSMoS can resolve heterogenous subpopulations, reveal new intermediates, and identify the temporal order of events occurring in complex, unsynchronized assembly processes (19-21). Prior CoSMoS studies of activation showed that while tri-snRNP binding was reversible, U4 snRNP release during activation is an irreversible step and that NTC association occurs predominantly after loss of U4 proteins (19). Whether or not Lsm2-8 release represents another irreversible step and how this is kinetically coupled with NTC recruitment are not known.

Here, we used CoSMoS to study Lsm2-8 dynamics during activation. Our results show that Lsm2-8, as expected, associates with the spliceosome as part of the tri-snRNP. Unexpectedly, Lsm2-8 release occurs after association of the NTC suggesting formation of B complex spliceosome lacking the U4 snRNP but containing both the NTC and Lsm2-8 ring ($B^{\Delta U4}_{Lsm+NTC}$). Formation of this complex is often preceded by the NTC reversibly sampling the spliceosome,

and stable NTC binding is followed by essentially irreversible release of Lsm2-8. Combined with *in vitro* biochemical data for ternary complex formation between Lsm2-8, U2, and U6 RNAs, we propose a kinetic model for spliceosome activation involving formation of at least two transient intermediates of unknown structure including the $B^{\Delta U4}_{Lsm+NTC}$ complex. We propose that release of Lsm2-8 from the $B^{\Delta U4}_{Lsm+NTC}$ complex may help to maintain U2/U6 helix II integrity during activation until it can be bound by NTC proteins.

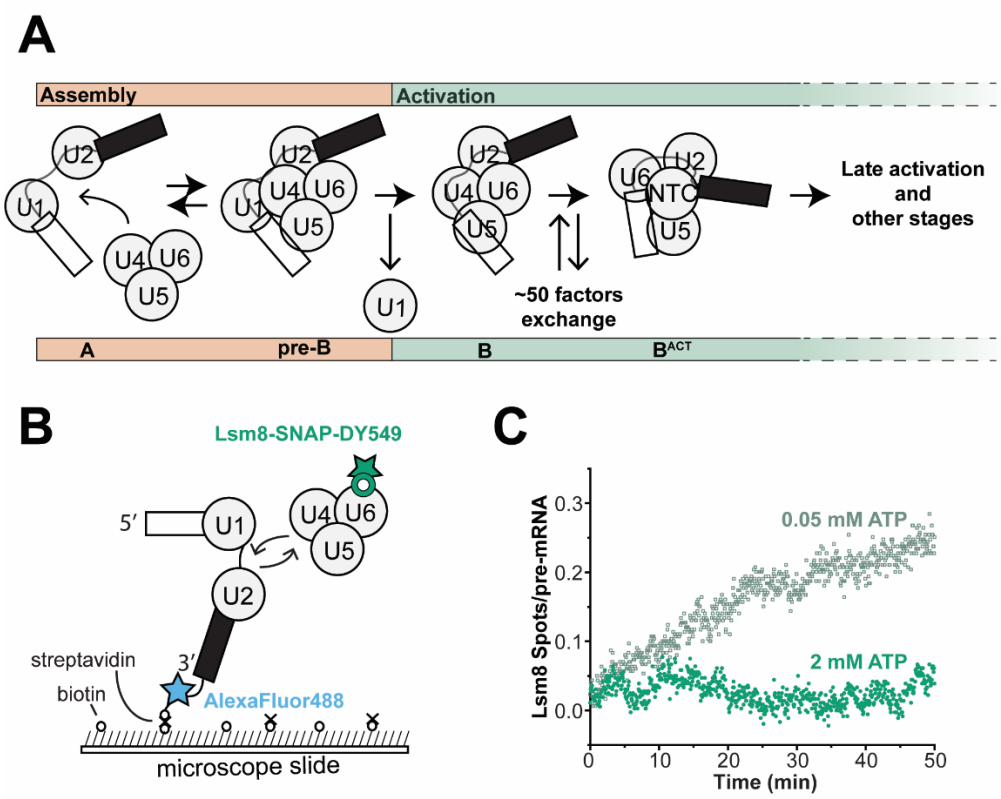


Figure 2.1 2-Color CoSMoS Assay to Study Lsm8 Dynamics During Spliceosome Activation.

(**A**) The U4/U6.U5 tri-snRNP reversibly interacts with the spliceosome A complex consisting of U1 and U2 snRNPs associated with pre-mRNA. Once the pre-B complex is formed, ATP-dependent activation can proceed by first ejecting the U1 snRNP to form the B complex. This is then followed by exchange of ~50 protein and snRNA factors to form the activated B complex (B^{ACT} spliceosome). During these steps, the U4 snRNP and Lsm2-8 proteins are released, and the NTC joins the spliceosome. Subsequent steps then lead to formation of the spliceosome active site and splicing. (**B**) Schematic of a 2-color CoSMoS experiment for observing Lsm ring binding dynamics. The U4/U6.U5 tri-snRNP contains a DY-549 fluorophore attached to the U6 Lsm ring component Lsm8 with a SNAP tag. The pre-mRNA was immobilized to the slide surface and contained an Alexa 488 fluorophore. (**C**) Time record of the number of Lsm8 fluorescent spots

that appeared on the surface relative to the number of surface-tethered pre-mRNA molecules at 0.05 mM (grey) and 2 mM ATP (green). At 0.05 mM ATP, spliceosome pre-B complexes can form but activation cannot proceed. At 2 mM ATP, spliceosome activation and splicing can proceed.

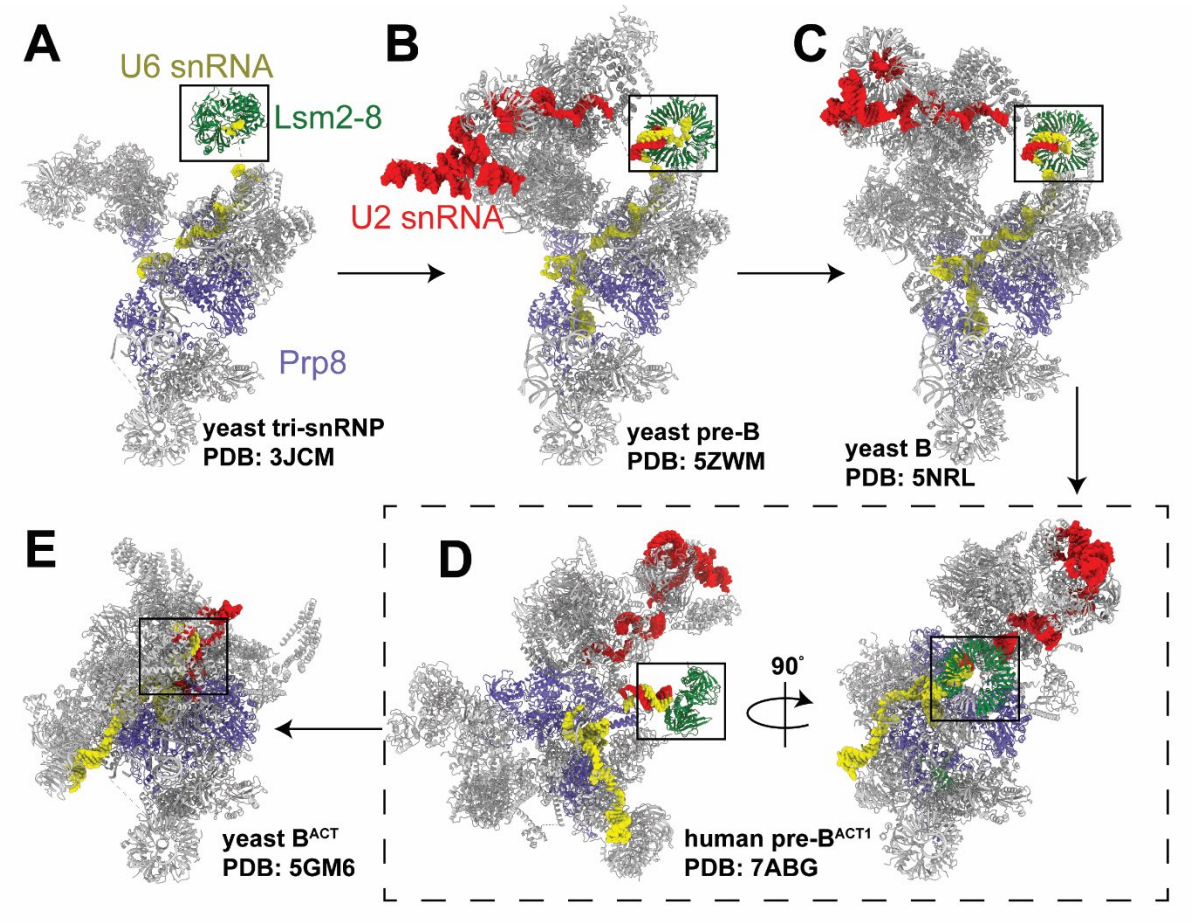


Figure 2.2 Structural Scheme for Spliceosome Activation.

Conformational and compositional changes occurring during spliceosome activation beginning with the yeast U4/U6.U5 tri-snRNP (**A**) and ending with the yeast B^{ACT} complex (**E**). The box in each structure is centered on the 3' end of the U6 snRNA while structures have been roughly aligned based on the position of the U5 snRNP Sm ring (grey ring at the base of each structure). The structures in panel (**D**) represent the human, not yeast, pre-B^{ACT1} spliceosome since no intermediate structures have yet been obtained for the yeast B to B^{ACT} transition.

MATERIAL AND METHODS

RNA Oligonucleotides

RNA oligos (**Table 2.1**) were purchased from Integrated DNA Technologies. Stocks (100 µM) were prepared by resuspending the oligos in nuclease-free water (Ambion). Initial stock concentrations were calculated from their absorbance values and the predicted extinction coefficients at 260nm using a NanoDrop.

Yeast Strains

Yeast strains (**Table 2.2**) were derived from the protease-deficient strain BJ2168 and splicing factors were c-terminally tagged by integrating fast SNAP or DHFR tags at the appropriate genomic locations by homologous recombination as previously described (19, 20).

Recombinant Lsm2-8 Expression and Purification

Lsm2-8 proteins were expressed in *E. coli* from a single plasmid and the heteroheptamer was purified as previously described (16).

5'-end Labeling of RNAs

Oligos (10 pmols for U6) were 5'-end labeled with [γ-^{32}P]-ATP (PerkinElmer), using T4 PNK and 10X reaction buffer A (Thermo Fisher Scientific) at 37^{o}C for 1 h, then heat inactivated at 75^{o}C for 10 min. The labelled oligos were then gel purified on a 19-20% PAGE denaturing gel (AccuGel 19:1, National Diagnostics), followed by extraction, ethanol precipitation, resuspension in buffer and quantification by liquid scintillation counting.

Electrophoretic Mobility Shift Assays (EMSAs)

Complexes between U2 and U6 RNAs and the Lsm2-8 protein were formed in a volume of 10 µL by mixing U2 and U6 oligos in RNA dilution buffer (100mM KCl, 20% w/v sucrose, 20mM

HEPES (pH 7), 1mM EDTA (pH 8), 1mM TCEP·HCl, 0.01% v/v Triton X-100, 0.2 mg/mL yeast tRNA [Ambion, Catalog No. AM7119], 0.2 mg/mL sodium heparin [Sigma Aldrich, Catalog No. H4784]) together with purified Lsm2-8 in protein dilution buffer (100mM KCl, 20% w/v sucrose, 20mM HEPES (pH 7), 1mM EDTA (pH 8), 1mM TCEP·HCl, 0.01% v/v Triton X-100, 0.2mg/mL BSA [Pierce, Catalog No. 23209]) in equal volume. Final concentrations for U6, U2 oligos and purified Lsm2-8 are described in the corresponding figures. All reactions were incubated at room temperature (22-24ºC) for 30 min prior to gel electrophoresis.

After 30min, the reactions were placed on ice and then loaded onto a native 6% polyacrylamide gel (AccuGel 29:1, National Diagnostics), which had been pre-run with 1x TBE buffer at 300 V for 30 min at 4ºC. Electrophoresis was then carried out at 300 V for 2-4 h. For detection of radioactive bands, the gels were dried and then exposed to a phosphor screen. Phosphor images of the screens were taken in the phosphorescence mode of a Typhoon scanner (Cytiva). Results were analyzed with ImageQuantTL (Cytiva) software.

Preparation of Yeast WCE

Yeast WCE for splicing was prepared according to published protocols using the ball method (43). SNAP-tagged proteins were fluorophore labeled as previously described (44). Briefly, SNAP-Surface® 549 dye (S9112S, New England BioLabs; abbreviated elsewhere as SNAP-DY-549) in DMSO was added to 1.2 mL yWCE to a final concentration of 1 µM. The reaction tube was rotated in the dark for 30 min at room temperature. The reaction was then loaded onto a pre-equilibrated G-25 Sephadex column (Kontes Flex Column) in SEC buffer (25 mM HEPES-KOH pH 7.9, 50 mM KCl, 1 mM DTT, 10% v/v glycerol) at 4ºC to remove excess dye. A Peristaltic Pump P-1 (Cytiva) was used for pre-equilibrating the column and eluting the labeled extract at a flow rate of ~0.25 mL/min with the SEC buffer. Fluorophore labeling of the proteins was confirmed by SDS-PAGE and fluorescence using a Typhoon FLA 9000 scanner (Cytiva) at 532nm. Results were analyzed with ImageQuantTL (Cytiva) software.

In Vitro Splicing Assays

[α-^{32}P] UTP radio-labeled (PerkinElmer) and m^{7}G(5')ppp(5')G capped (New England Biolabs) RP51A pre-mRNA substrates were made by *in vitro* transcription of a linear DNA template with T7 RNA polymerase (Agilent or purified in the laboratory). The DNA template was produced from a PCR reaction of pBS117 plasmid (45) using Taq DNA polymerase (M7122, Promega), followed by gel purification of the products with a Wizard SV Gel and PCR Clean-Up System kit (Promega). Transcription products were separated on a 6% denaturing polyacrylamide gel (AccuGel 19:1, National Diagnostics), followed by ethanol precipitation of the extracted RP51A transcripts and quantitation of the resuspended transcripts in nuclease-free water (Ambion, Fischer Scientific) with a liquid scintillation counter (LSC, tri-carb 2900TR, Packard).

A typical *in vitro* splicing reaction included 40% v/v WCE and 0.2 nM radio-labeled RP51A substrate in a splicing buffer [final concentrations: 100mM KPi pH 7.3, 3% w/v PEG-8000, 1mM DTT, 2.5mM $MgCl_2$, 0.2U/μL RNasin Plus (Promega)]. The reaction was incubated at room temperature for 45 min. The reaction was quenched in a splicing dilution buffer (100mM Tris base pH 7.5, 10mM EDTA pH 8.0, 1% w/v SDS, 150mM NaCl, 300 mM NaOAc pH 5.3). RNAs from the reaction were extracted using phenol-chloroform, ethanol precipitated, resuspended in deionized formamide, and separated on a 12% denaturing polyacrylamide gel. Gels were then dried and exposed to a phosphor screen overnight. The screen was imaged with a Typhoon FLA 9000 scanner (Cytiva). Results were analyzed with ImageQuantTL (Cytiva) software. The intensities (*I*) for RP51A pre-mRNAs and splicing products were determined by integrating the signals within same-sized rectangles around the band. The background corrected intensities for bands were then used for calculating splicing efficiencies, with Equations 1 and 2.

$$1st\ step\ efficiency\ = \frac{I(mRNA) + I\left(intron\ lariat - 3^{'}exon\right)}{I(pre-mRNA) + I(mRNA) + I(intron\ lariat - 3^{'}exon)} \qquad (Eq.1)$$

$$2nd\ step\ efficiency\ = \frac{I(mRNA)}{I(pre-mRNA) + I(mRNA) + I(intron\ lariat - 3'exon)} \quad (Eq.2)$$

Preparation of Fluorescently-Labeled RP51A pre-mRNAs

Unlabeled biotin handles (**Table 2.1**, Oligo 8) from IDT were dissolved in 0.091 M $Na_2B_4O_7$ buffer (NaB buffer, pH 8.50, adjusted with concentrated HCl) at a concentration of 1 mM. Alexa Fluor 488 NHS ester (A20000, Thermo Fisher Scientific) was prepared by dissolving to a concentration of 1 mg in 60 µL of anhydrous DMSO. For fluorophore labeling of the biotin handle, 15 µL of the fluorophore was mixed with 49 µL of the biotin handle and 36 µL NaB buffer. The reaction was kept in the dark and rotated overnight at room temperature. The labeling reaction mixture was then applied to a G-25 MicroSpin column (45-001-397, Fisher Scientific) to remove most of the unreacted dye. The eluate from the spin column was then loaded onto a 19% denaturing polyacrylamide gel (AccuGel 19:1, National Diagnostics). Fluorescent bands containing the oligo were excised from the gel, the oligo eluted from gel, and then ethanol precipitated. The products were resuspended in water and concentrations determined by UV-Vis.

Capped RP51A transcripts were made by *in vitro* transcription using T7 RNA polymerase and the biotin handle attached by splinted RNA ligation. The fluorescently labeled oligo (56 pmol) was 5'-phosphorylated using T4 PNK (5U, M0201S, New England Biolabs) according to the manufacturer's directions with the addition of 2 mM ATP. The PNK enzyme was then heat inactivated and DNA bridge oligos (50 pmol; **Table 2.1**, Oligo 9) and radio-labeled RP51A transcript (14 pmol) was added and then allowed to anneal by incubating at 90°C for 3 min, followed by cooling using a gradient of 0.1°C/s over 10 min. T4 RNA Ligase 2 (20U; M0239L, New England Biolabs) was then added to ligate the transcript to the biotinylated oligo. Ligation products were separated on a 5% denaturing polyacrylamide gel (AccuGel 19:1, National Diagnostics), followed by ethanol precipitation of the extracted products from the gel. Precipitated

RNAs were resuspended in nuclease-free water (Ambion, Fischer Scientific) and quantified with a liquid scintillation counter (LSC, tri-carb 2900TR, Packard).

CoSMoS Assays

Microscope slides (100490-396, VWR) and cover glasses (12-553-455, Fischer) were cleaned by sonication while immersing in the following solutions/solvents in the order listed: f 2% v/v micro-90, 100% ethanol, 1M KOH, and MiliQ water. Each sonication step took 1 h, followed by rinsing with MiliQ water between steps. After drying with high purity nitrogen (NI UHP300, Airgas), the slides were aminosilanized with VECTABOND (NC9280699, Fisher Scientific). Reaction chambers were created by drawing multiple parallel straight lines on the slides with vacuum grease, followed by covering with the cover glasses to create individual lanes for reaction chambers. Typically, four lanes can be made on a single slide. Lastly, the slides were passivated by addition of a mixture of mPEG-SVA (MPEG-SVA-5K, Laysan Bio) and mPEG-biotin-SVA (BIO-PEG-SVA-5K, Laysan Bio) at a ratio of 1:100 w/w in 100mM $NaHCO_3$ (pH 8) buffer and incubating overnight.

Prior to each CoSMoS assay, the mPEG mixture in the lane (~20 µL) was washed off with 1x PBS (200 µL) twice. Then, streptavidin (50 µL, 0.2 mg/mL, SA10, Kelowna International Scientific) was flowed in the lane, allowing it to bind with the biotin on the slide surface. After washing the lane off with 200 µL 1x PBS, 50 µL 50-150 pM fluorescently labeled and biotinylated RNA was added and allowed to bind with the streptavidin. The slide was washed again with PBS and the density of RNA was determined by exciting the fluorophore with the 488 nm laser. Finally, splicing assay buffer containing 40% v/v WCE, 20 nM Cy5-TMP (20), oxygen scavenger (5 mM PCA and 0.96 U/mL PCD), 1 mM Trolox, and triplet quenchers (0.5 mM propyl gallate, 1 mM cyclooctatetraene, 1 mM 4-nitrobenzyl alcohol) was added (20). The triplet quenchers were added as a mixture made as a 100x stock in DMSO.

CoSMoS Data Acquisition

A custom-built objective-type TIRF microscope (23, 46) together with Glimpse software (written in LabVIEW programming language, https://github.com/gelles-brandeis/Glimpse) was used for collecting single molecule movies. Briefly, a 60X 1.45(NA) PlanApo objective (Olympus) was used for the TIRF excitation and collecting the emission light. Different wavelengths of incoming and exit excitation light were directed to and away from the objective by two separate micromirrors. The emission light was further filtered and separated by a dichroic mirror at the cutoff wavelength of 635 nm. The spectrally separated light was then focused onto two different areas (FOV: field of view) of the same EMCCD camera. Four different lasers were used in this study, including 488 nm (blue), 532 nm (green), 633 nm (red) and 785 nm (infrared), the powers of which were set in the ranges of 1.2 – 1.5 mW, 500 – 600 μW, 400-440 μW, and 2.5 mW, respectively. Cycles of time-lapse imaging were used according to the following excitation scheme with each frame lasting 1s: in each cycle, the 785 nm was first used to illuminate the sample and correct sample positioning using an auto-focus system. Then, the 488 nm blue laser was turned on to collect two consecutive frames to image the immobilized RNAs. The 532 and 633 nm lasers were then turned on to simultaneously collect 14 frames with a 3 s delay between adjacent frames. The total cycle time was ~1 min and this cycle was repeated 50x to collect videos lasting for ~50 min. To avoid photo bleaching of DY-549 and Cy5 fluorophores by the 488 nm laser, the path of the laser was physically blocked to prevent the beam from reaching the sample after collection of 10 frames of blue images.

CoSMoS Data Analysis

Single molecule data were analyzed by the same method as described previously (20, 28) with slight modifications, using a custom program imscroll (https://github.com/gelles-brandeis/CoSMoS_Analysis) written in MATLAB (The Mathworks). A general analysis procedure includes, 1) creating a mapping file for correlating the locations in the <635 nm FOV to those in

the >635 nm FOV, using a fluorescent beads (T10711, ThermoFisher); 2) creating a driftlist file for correcting the drift of the fluorescent spots over the recording; 3) creating a AOI (area of interest; typically either 3×3 or 5×5 pixels) file listing the positions of the immobilized pre-mRNA molecules; 4) combining the AOI, mapping, and driftlist files to generate drift-corrected AOIs corresponding to the pre-mRNA locations in the fields of view used for DY-549 and Cy5-TMP imaging; and 5) integrating the measured intensities in the AOIs in those fields of view over the experimental time course (47). Binding events were identified as signals centered within the AOI that appeared with intensities greater than 3.6x the standard deviation above the mean of the background noise. Loss of signals were identified as points in time at which the signal fell below 1x standard deviation above the mean background.

Analysis of the measured dwell times and fits to kinetic equations were carried out using MATLAB and AGATHA software (https://github.com/hoskinslab/AGATHA) using maximum likelihood methods and fitting to equations containing one (Eq. 3), two (Eq. 4), or three (Eq. 5) exponential terms as described (28).

$$\frac{1}{\left(e^{-tm/\tau} - e^{-tx/\tau}\right)} * \left(\frac{1}{\tau} e^{-\frac{intervals}{\tau}}\right) \qquad (Eq. 3)$$

$$\frac{1}{a\left(e^{-tm/\tau 1} - e^{-tx/\tau 1}\right) + (1-a)\left(e^{-tm/\tau 2} - e^{-tx/\tau 2}\right)} * \left(\frac{a}{\tau 1} e^{-\frac{intervals}{\tau 1}} + \frac{1-a}{\tau 2} e^{-\frac{intervals}{\tau 2}}\right) \qquad (Eq. 4)$$

$$\frac{1}{a1\left(e^{-tm/\tau1}-e^{-tx/\tau1}\right)+a2\left(e^{-tm/\tau2}-e^{-tx/\tau2}\right)+(1-a1-a2)\left(e^{-tm/\tau3}-e^{-tx/\tau3}\right)}$$

$$*\left(\frac{a1}{\tau1}e^{-\frac{intervals}{\tau1}}+\frac{a2}{\tau2}e^{-\frac{intervals}{\tau2}}+\frac{1-a1-a2}{\tau3}e^{-\frac{intervals}{\tau3}}\right) \qquad (Eq.5)$$

Bootstrapping was used to calculate standard errors for all fitted parameters. Fit parameters are included in **Table 2.3**, and models that best fit the distributions were identified by using the log-likelihood ratio test (28).

Histograms for the distribution of events were generated in MATLAB with empty bins removed. The error (standard deviation) for each bin were calculated using Eq. 6, with the assumption that the number of events within a bin follows a binomial distribution.

$$\sigma = \sqrt{NP(1-P)} \qquad (Eq.6)$$

Table 2.1 Oligonucleotide Sequences

Oligo #	ID	Sequence (5′-3′)	Length (nts)	Notes
1	U6_WT	/5/UA CAA AGA GAU UUA UUU CGU UUU /3Phos/	23	WT U6 snRNA (90 - 112nt)
2	U6_Mut	/5/UA CAA AGA GAU UUA UUU CG**A AAA** /3Phos/	23	Mutate U6 3′ end to polyA
3	U6_WT_4U_XF	/5IRD800CWN/UA CAA AGA GAU UUA UUU CGU UUU /3Phos/	23	WT U6 snRNA (90 - 112nt), IR dye labeled
4	U2_WT	/5/ACG AAU CUC UUU GCC UUU UGG CUU AGA U/3/	28	WT U2 snRNA (1 - 28nt)
5	U2_Mut	/5/AC**U UUA GAG AAA C**CC UUU UGG CUU AGA U/3/	28	Mutate basepairing region to U6 sequence.
6	U6_Mut1	/5IRD800CWN/UA CAA AGA GAU UUA UUU CG**A AAA** /3Phos/	23	Mutate 3′ end to polyA, IR dye labeled
7	U6_Mut2_4U_XF	/5IRD800CWN/UA **GUU UCU CUA AG**A UUU CGU UUU /3Phos/	23	Mutate basepairing region to U2 sequence, IR dye labeled
8	Ah_08_biotin handle	/5/mAmUmC mCmGmG mAmGmC mGmAmG /iAmMC6T/mAmG mA/3Bio/	16	biotin handle oligo used for coupling to a fluorophore through the internal amino modifier C6T (iAmMC6T) and ligating to RP51A transcripts.
9	CF13_RP51a_ bridge_rnl2	/5/CTC GCT CCG GAT CGA CCC TTT TGG ATT CTC TTC ATC /3/	36	DNA bridge used for splint ligation of biotin handles to RP51A transcripts, with RNA Ligase II.

Table 2.2 Yeast Strains

Strain #	ID	Genotype	Notes
1	yAAH0001	MATa prc1-407 prb1-1122 pep4-3 leu2 trp1 ura3-52 gal2	BJ2168, parental strain
2	yAAH0018	yAAH0001 + cef1::cef1-DHFR-HPH + ntc90::ntc90-DHFR-BLE	contains two DHFR-tagged NTC proteins with hygromycin and phleomycin resistance markers
3	yAAH0329	yAAH0001 + prp3::prp3-DHFR-HPH + prp4::prp4-DHFR-BLE	contains two DHFR-tagged U4 snRNP proteins with hygromycin and phleomycin resistance markers
4	yAAH0362	yAAH0018 + lsm8::lsm8-SNAP$_f$-NAT	contains two DHFR-tagged NTC proteins and fast-SNAP-tagged lsm8 with nourseothricin resistance marker
5	yAAH1709	yAAH0329 + lsm8::lsm8-SNAP$_f$-NAT	contains two DHFR-tagged U4 snRNP proteins and fast-SNAP-tagged lsm8 with nourseothricin resistance marker

Table 2.3 Fitted Kinetic Parameters

Tagged protein	Tagged Subcomplex	Complex	Strain	[ATP] mM	a_1	τ_1 (s)	a_2	τ_2 (s)	a_3	τ_3 (s)	Number of Events	Related Figure
Lsm8-SNAP	Lsm2-8	--	yAAH1709	2	0.34±0.03	7.5±0.5	0.60±0.03	69.0±5.0	0.06±0.04	394.6±116.0	700	Fig. 2.5
Prp3/Prp4-DHFR	U4	--	yAAH1709	2	0.49±0.10	13.8±1.9	0.46±0.10	41.2±6.7	0.06±0.14	329.9±72.6	481	Fig. 2.5
Lsm8-SNAP	Lsm2-8	--	yAAH1709	0.05	0.71±0.05	14.2±2.8	0.29±0.05	543.5±121.3	NA	NA	145	Fig. 2.5
Prp3/Prp4-DHFR	U4	--	yAAH1709	0.05	0.48±0.05	9.0±1.7	0.52±0.05	413.5±54.8	NA	NA	145	Fig. 2.5
--	--	B_{Lsm+U4}	yAAH1709	2	NA	41.7±5.3	NA	NA	NA	NA	144	Fig. 2.5E
--	--	$B_{Lsm}^{\Delta U4}$	yAAH1709	2	NA	64.7±6.9	NA	NA	NA	NA	144	Fig. 2.5F
--	--	$B_{Lsm}^{\Delta NTC}$	yAAH0362	2	NA	68.6±8.9	NA	NA	NA	NA	83	Fig. 2.13B
--	--	$B_{Lsm+NTC}^{\Delta U4}$	yAAH0362	2	NA	21.1±2.7	NA	NA	NA	NA	83	Fig. 2.13C
--	--	$B_{NTC}^{\Delta Lsm}$	yAAH0362	2	NA	363.6±38.1	NA	NA	NA	NA	83	Fig. 2.13D

RESULTS

ATP-Dependent Accumulation of Lsm8 Complexes on Single Pre-mRNAs

To watch single molecules of Lsm8 assemble with and release from spliceosomes, we first genetically encoded a c-terminal fast SNAP tag on the Lsm8 protein. Lsm8 is essential for yeast growth in the absence of U6 snRNA overexpression (22), and it is the only member of the Lsm2-8 complex not found in Lsm1-7 complexes involved in mRNA decay. Lsm8-SNAP strains are viable, indicating that the tagged protein is functional for splicing *in vivo*. Lsm8-SNAP can be readily labeled with benzylguanine-fluorophores in yeast whole cell extracts (WCEs; **Fig. 2.3**) and these WCEs possessed high *in vitro* splicing activities (**Fig. 2.4**).

To observe Lsm8 dynamics, we labeled Lsm8 with SNAP-DY-549 [a 532nm (green) laser-excitable fluorophore] in WCE and added the WCE to passivated glass slides containing RP51A pre-mRNAs labeled with Alexa Fluor 488 [a 488nm (blue) laser-excitable fluorophore] and immobilized with biotin-streptavidin (**Fig. 2.1B**). Experiments were carried out at two concentrations of ATP: 0.05 mM ATP, which permits spliceosome assembly but not activation or Lsm release, and 2 mM ATP, which permits activation, Lsm release, and splicing (7). Using a custom-built CoSMoS microscope (23), we observed fixed fluorescent RNA spots upon 488 nm laser excitation and dynamic spots originating from Lsm8-SNAP upon 532 nm laser excitation.

The Lsm8-SNAP binding and unbinding events were dependent on ATP concentration. At 0.05 mM ATP, Lsm8-SNAP fluorescent spots persisted, and this resulted in their surface accumulation over time (**Fig. 2.1C**). In contrast, at 2 mM ATP, Lsm8-SNAP fluorescent spots appeared more briefly, and this resulted in little surface accumulation (**Fig. 2.1C**). The surface accumulation data strongly resemble those previously obtained for the U4 snRNP under these ATP conditions (19) and are consistent with assembly of Lsm8-SNAP-containing spliceosomes that cannot undergo activation and retain Lsm8-SNAP at 0.05 mM ATP. At 2 mM ATP, Lsm8-

SNAP-containing spliceosomes can both assemble and activate, resulting in Lsm8-SNAP release and transient association with the surface-tethered RNAs.

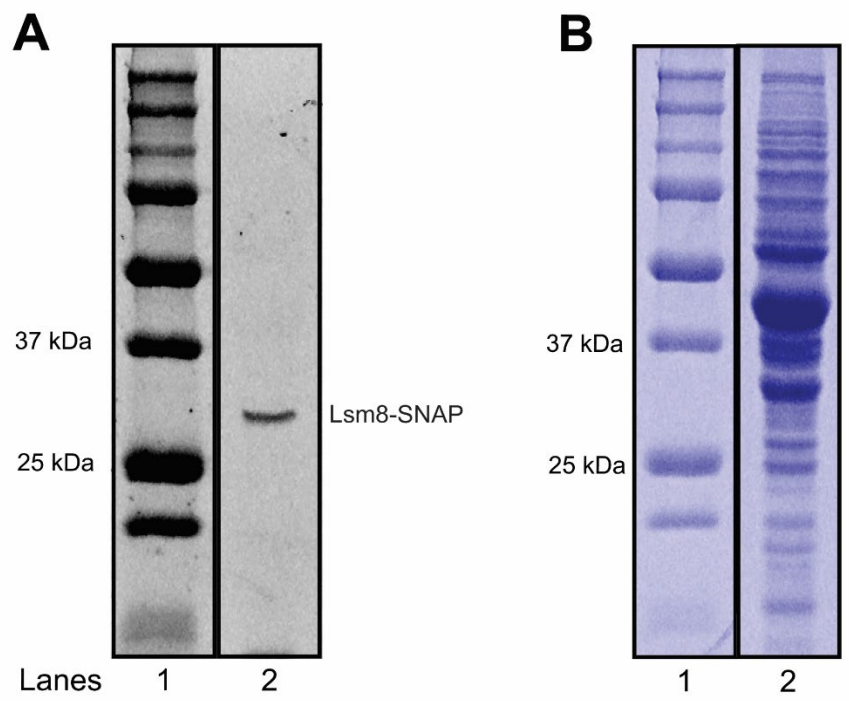


Figure 2.3 Fluorophore Labeling of Lsm8-SNAP.

Lsm8-SNAP proteins can be specifically labeled in yeast WCE with SNAP-DY-549. Shown is fluorescence image (**A**) and Coomassie stain (**B**) of the same SDS-PAGE gel used to analyze Lsm8-SNAP-containing extract after fluorophore labeling. The gel images were cropped (black boxes) to remove intervening lanes not relevant to this figure.

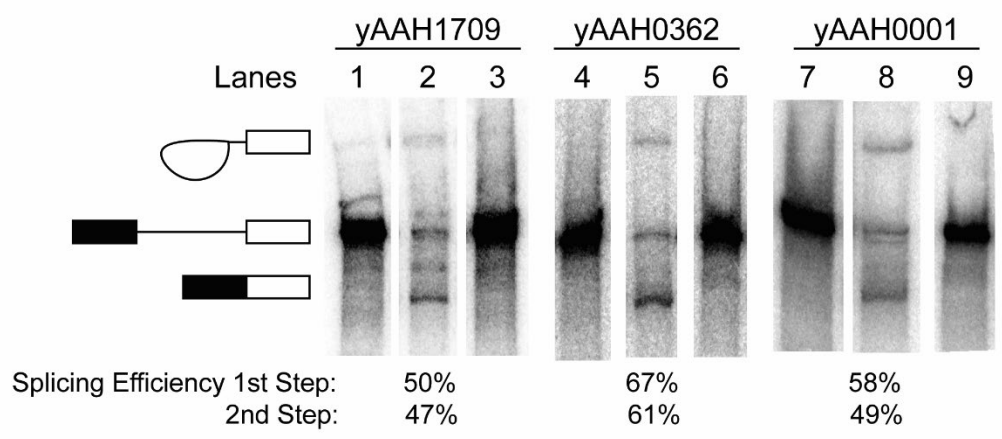


Figure 2.4 Splicing Activity of Lsm8-Labeled WCE.

Representative splicing assays and efficiencies for WCE containing Lsm8-SNAP labeled proteins (lanes 1-6) or a WT control (lanes 7-9). In lanes 2, 5, and 8, ATP was added to a concentration of 2 mM to promote splicing. In lanes 3, 6, and 9 ATP was added to a concentration of 0.05 mM to permit spliceosome assembly but prevent activation and splicing. Lanes 1, 4, and 7, represent samples at t=0 min while the other lanes represent reactions quenched at 45 min. The gel images were cropped (vertical white stripes) and combined into a single figure to remove intervening lanes not relevant for this figure and for clarity.

The Lsm Ring is Released after U4 snRNP Dissociation

To identify Lsm8-SNAP proteins associated with tri-snRNPs and correlate Lsm8 release with spliceosome activation, we carried out 3-color CoSMoS assays using extracts containing Lsm8-SNAP labeled with DY-549 fluorophores, the U4 snRNP proteins Prp3 and Prp4 both labeled with the DHFR tag and Cy5-TMP [632nm (red) laser-excitable] fluorophores, and Alexa Fluor 488 labeled pre-mRNAs (**Fig. 2.5A**). We have previously used CoSMoS to show that Prp3 and Prp4 are both lost during spliceosome activation and that their release is irreversible (19). In agreement with these previous experiments, Prp3/Prp4 fluorescence spots accumulated to a greater extent on a surface containing immobilized pre-mRNAs at 0.05 mM ATP than at 2 mM ATP (**Fig. 2.6**). In these 3-color experiments, the apparent trends of surface accumulation of Lsm8-SNAP and Prp3/Prp4-DHFR were quite similar at both ATP concentrations (**Figs. 2.6C, D**). Since Lsm8, Prp3, and Prp4 are all components of the tri-snRNP, the similarity originates from tri-snRNP binding events.

We analyzed individual fluorescence time trajectories to identify events at pre-mRNA locations in which Lsm8-SNAP and Prp3/Prp4-DHFR fluorescence colocalized (**Fig. 2.5B, 2.7**). Approximately 28% of Lsm8-SNAP or Prp3/Prp4-DHFR events at 2 mM ATP colocalized with one another at some point during the event lifetime (201 out of 723 Lsm8-SNAP events or 685 Prp3/Prp4-DHFR events). Since these proteins are expected to be stoichiometric with one another in the tri-snRNP, this suggests either incomplete fluorophore labeling or potential tri-snRNP heterogeneity. We restricted our analysis just to the subset of colocalized events. Of these, the majority (88%) showed simultaneous arrival of Lsm8-SNAP and Prp3/Prp4-DHFR fluorescent spots. This indicates that the predominant recruitment pathway for Lsm2-8 is concurrent with the U4 snRNP for the colocalized signals. This agrees with biochemical and structural data for the tri-snRNP showing simultaneous presence of these factors (24-26).

For this subset of Lsm8-SNAP and Prp3/Prp4-DHFR events, we then analyzed the order in which their signals disappeared from the immobilized RNAs. At 2 mM ATP, the predominant release pathway (82%, 144/176 events) showed loss of the Lsm8-SNAP signal after loss of the Prp3/Prp4-DHFR signals (**Fig. 2.5B**, **C**). The losses of the Lsm8-SNAP signals at 2 mM ATP were unlikely to be due entirely to photobleaching since we readily observed much longer-lived signals at 0.05 mM ATP (~8-fold longer fitted longest-lived lifetime, **Table 2.3**). Thus, we interpret these observations as U4 snRNP dissociation occurring prior to Lsm2-8 release during yeast spliceosome activation. This is in agreement with a proposed activation pathway for the human spliceosome deduced from cryo-EM structures of pre-B^{ACT} complexes (18).

The single molecule data support the existence of at least two spliceosomes with distinct compositions: a B_{Lsm+U4} complex that contains Lsm8, Prp3, and Prp4 and a $B_{Lsm}^{\Delta U4}$ complex that contains Lsm8 but neither Prp3 nor Prp4. B_{Lsm+U4} represents the yeast pre-B or B complex spliceosome that has been previously biochemically and structurally characterized (25, 27). In these complexes, the tri-snRNP has associated with the U1- and U2-containing A complex but activation has not yet proceeded to the point of U4 snRNP release. The $B_{Lsm}^{\Delta U4}$ intermediate is characterized by its formation from the B_{LSM+U4} complex, the absence of U4 snRNP proteins, and the presence of Lsm8. The $B_{Lsm}^{\Delta U4}$ complex, therefore, represents an activation intermediate formed after release of the U4 snRNP but from which the Lsm2-8 ring has not yet been released. It should be noted that in these experiments we cannot discriminate between $B_{Lsm}^{\Delta U4}$ complexes that contain the NTC and those that do not (see the following section). Though $B_{Lsm}^{\Delta U4}$ has not been previously identified in yeast, a cryo-EM structure of a human spliceosome containing a similar composition (pre-B^{ACT1}) has been obtained in the presence of an activation inhibitor (18).

To provide further insight into the B_{Lsm+U4} and $B_{Lsm}^{\Delta U4}$ complexes, we analyzed the arrival and departure times for each colocalized Lsm8-SNAP and Prp3/Prp4-DHFR event. The lifetimes of the B_{Lsm+U4} complexes were determined by subtracting the arrival times of the colocalized

signals from the U4 release times ($t_{U4}^{release} - t_{U4/Lsm8}^{arrival}$), and the lifetimes of the $B_{Lsm}^{\Delta U4}$ complexes were found by subtracting the Prp3/Prp4-DHFR release times from the release times of the Lsm8-SNAP signals ($t_{Lsm8}^{release} - t_{U4}^{release}$) (**Fig. 2.5D**). The resulting unbinned distributions of bound dwell times were then fit using maximum likelihood methods to exponential-based functions (28).

In both cases, the distributions could be best fit to functions containing single exponential terms with characteristic lifetimes of 41.7±5.3 and 64.7±6.9 s for B_{Lsm+U4} and $B_{Lsm}^{\Delta U4}$ complexes, respectively (**Fig. 2.5E, F; Table 2.3**). The lifetime of the B_{Lsm+U4} complex is very similar to that previously measured for U4 snRNP proteins (~ 34 s) under activation conditions (19) and represents the average residence time for the U4 snRNP in pre-B/B complex spliceosomes. Lsm8 release does not happen immediately after loss of U4 proteins. Instead, Lsm8 remains bound to the spliceosomes for ~ 1 min as given by the lifetime of the $B_{Lsm}^{\Delta U4}$ complex. The lifetimes for both B_{Lsm+U4} and $B_{Lsm}^{\Delta U4}$ complexes are smaller relative to the *in vitro* timescale of RP51A intron loss, which occurs over tens of minutes (29). This indicates that both complexes can be kinetically competent intermediates for activation. Finally, since we could measure the lifetimes of complexes formed in succession from the same molecules, we analyzed whether the lifetimes of the two complexes correlated with one another. In other words, we asked if rapid release of U4 was predictive of subsequent slow or rapid release of Lsm8 from the same molecular complex. Analysis of the paired B_{Lsm+U4} and $B_{Lsm}^{\Delta U4}$ lifetimes did not reveal a strong correlation between the two (**Fig. 2.8A**). This suggests that U4 snRNP release kinetics are to some degree decoupled from the release of the Lsm ring with the rate of the former not predictive of the latter.

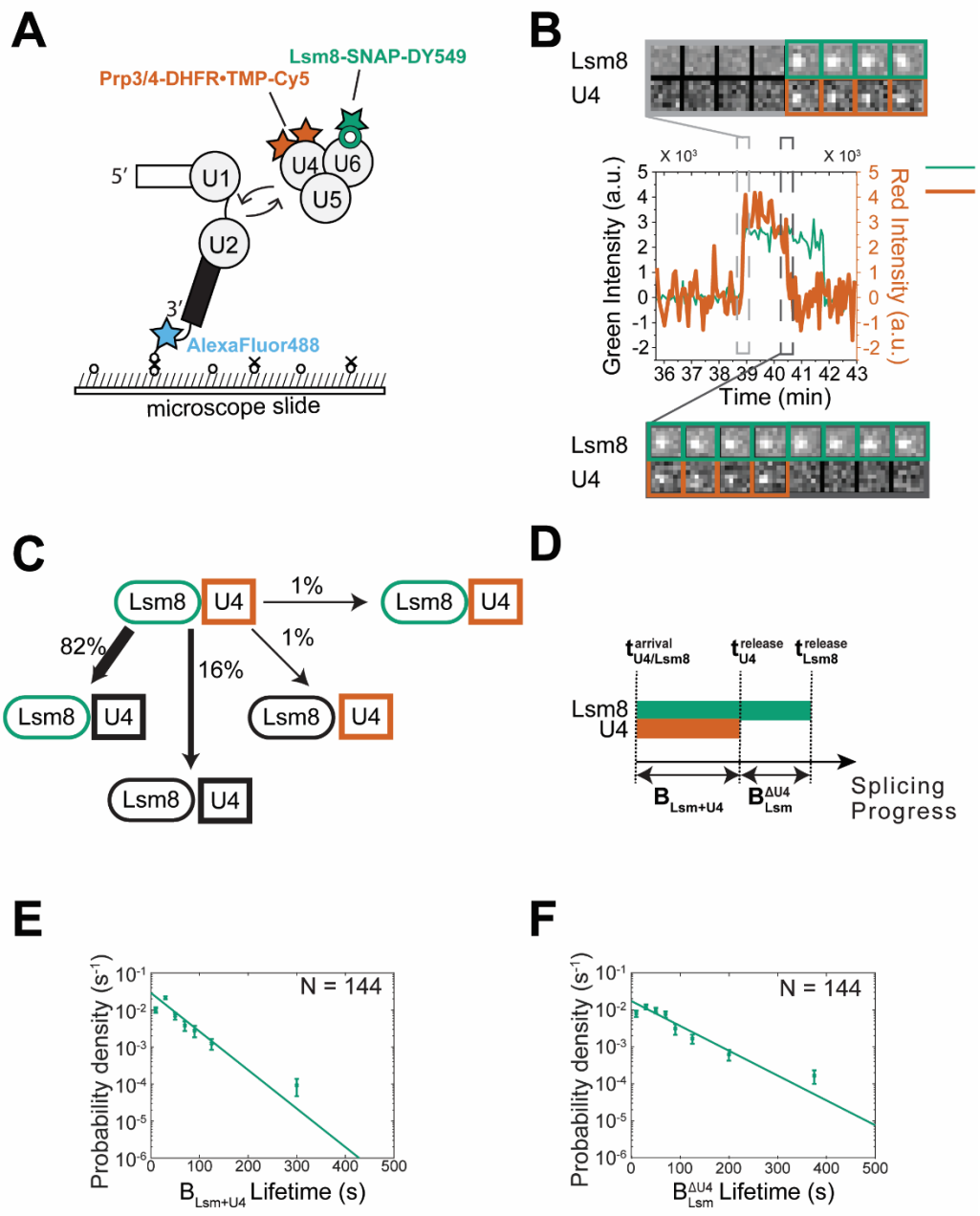


Figure 2.5 3-color CoSMoS observation of U4 snRNP and Lsm ring binding dynamics during activation.

(**A**) Schematic of a 3-color experiment in which U4 was labeled with Cy5-TMP fluorophores, the Lsm ring was labeled with a DY-549 fluorophore, and the surface-tethered pre-mRNA was labeled with an Alex Fluor 488 fluorophore. (**B**) Segment of a representative time record showing peaks in fluorescence intensity corresponding to colocalization of U4 (red, thick line) and Lsm8 proteins

(green, thin line) with the same individual pre-mRNA molecule. The light grey dashed rectangle marks an example of the simultaneous appearance of U4 and Lsm8 spots; galleries show consecutive images taken from the indicated part of the recording showing that spot appearance is simultaneous. The darker grey dashed rectangle marks an example of the ordered disappearance of U4 then Lsm spots; galleries show consecutive images taken from the indicated part of the recording showing that loss of the U4 signal precedes loss of the Lsm8 signal. (**C**) Routes for loss of either the U4 or Lsm fluorescent spots at 2 mM ATP for *N*=176 pairs of overlapping events. Green and red shapes represent observation of fluorescence from the corresponding DY-549 or Cy5 fluorophores on Lsm or U4, respectively; grey shapes represent the absence of fluorescence. Percentages represent the fraction of U4/Lsm complexes in which fluorescence disappeared by the indicated pathway; more prevalent pathways are emphasized with thicker arrows. (**D**) Schematic showing the relationship between the lifetimes of the $\mathrm{B}_{\mathrm{Lsm+U4}}$ and $\mathrm{B}_{\mathrm{Lsm}}^{\Delta \mathrm{U4}}$ complexes to the measured arrival and release times for U4 snRNP and Lsm8 proteins. (**E, F**) Probability density histograms of $\mathrm{B}_{\mathrm{Lsm+U4}}$ (panel E, $\mathrm{t}_{\mathrm{U4}}^{\mathrm{release}} - \mathrm{t}_{\mathrm{U4/Lsm8}}^{\mathrm{arrival}}$) and $\mathrm{B}_{\mathrm{Lsm}}^{\Delta \mathrm{U4}}$ (panel F, $\mathrm{t}_{\mathrm{Lsm8}}^{\mathrm{release}} - \mathrm{t}_{\mathrm{U4}}^{\mathrm{release}}$) complex lifetimes obtained from the subset of events (*N*) showing simultaneous arrival of U4 and Lsm spots followed by ordered loss of the U4 and then Lsm signals. Lines represent fits of the lifetime distributions with equations containing single exponential terms that yielded the parameters reported in **Table 2.3**. Error bars (SD) were calculated for each point as described in the Methods.

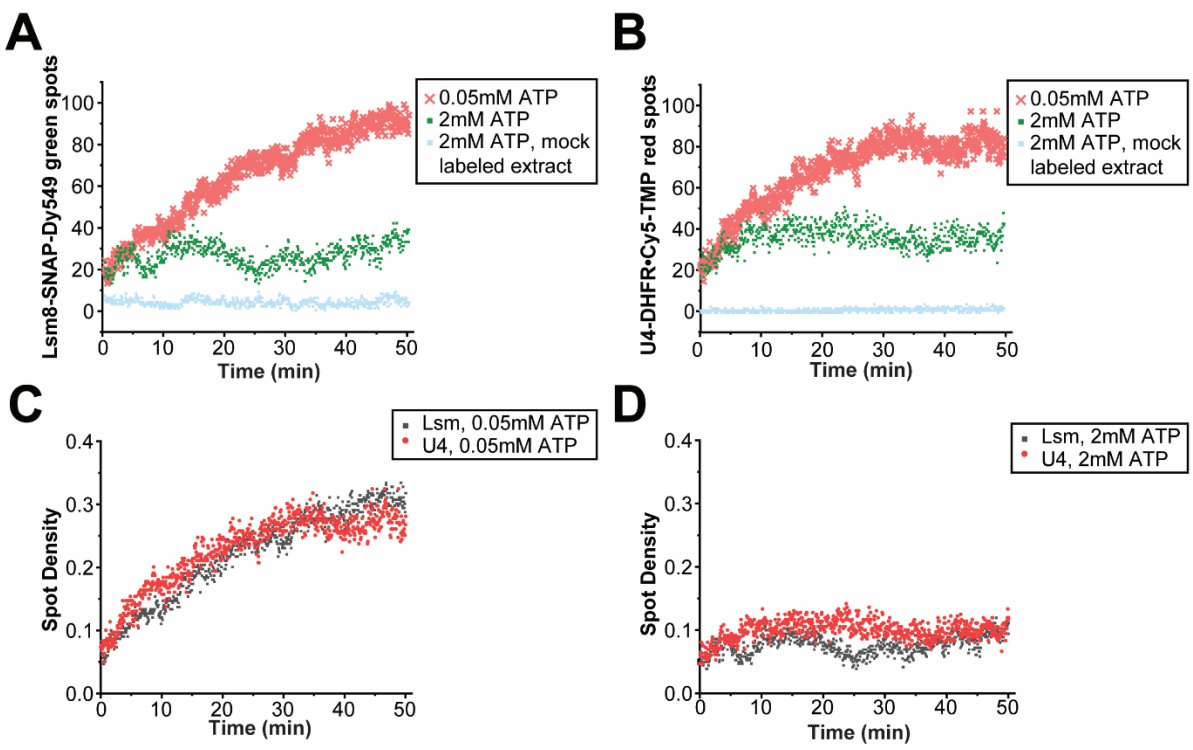


Figure 2.6 ATP-Dependent Accumulation of Lsm8 and U4 Protein Fluorescent Spots.

Shown are spot accumulation trajectories for CoSMoS experiments containing green-labeled Lsm8 and red-labeled U4proteins. In panels (**A** and **B**), spot accumulation trajectories are compared for the same experiment from both the green/Lsm8 (**A**) and red/U4 (**B**) channels under conditions that inhibit (0.05 mM ATP) or permit activation and splicing (2 mM ATP). As a control, the spot accumulation trajectory is shown for a WT extract labeled with SNAP-DY-549 and Cy5-TMP fluorophores but not containing any SNAP- or DHFR-tagged proteins (blue). In panels (**C** and **D**), normalized spot counts from the data shown in panels (**A** and **B**) were calculated by dividing the number of observed fluorescent spots by the number of surface-immobilized premRNAs in the field-of-view. Data for Lsm and U4 were then overlaid at low (**C**) and high (**D**) ATP to highlight the similarities in trajectories for the splicing factors.

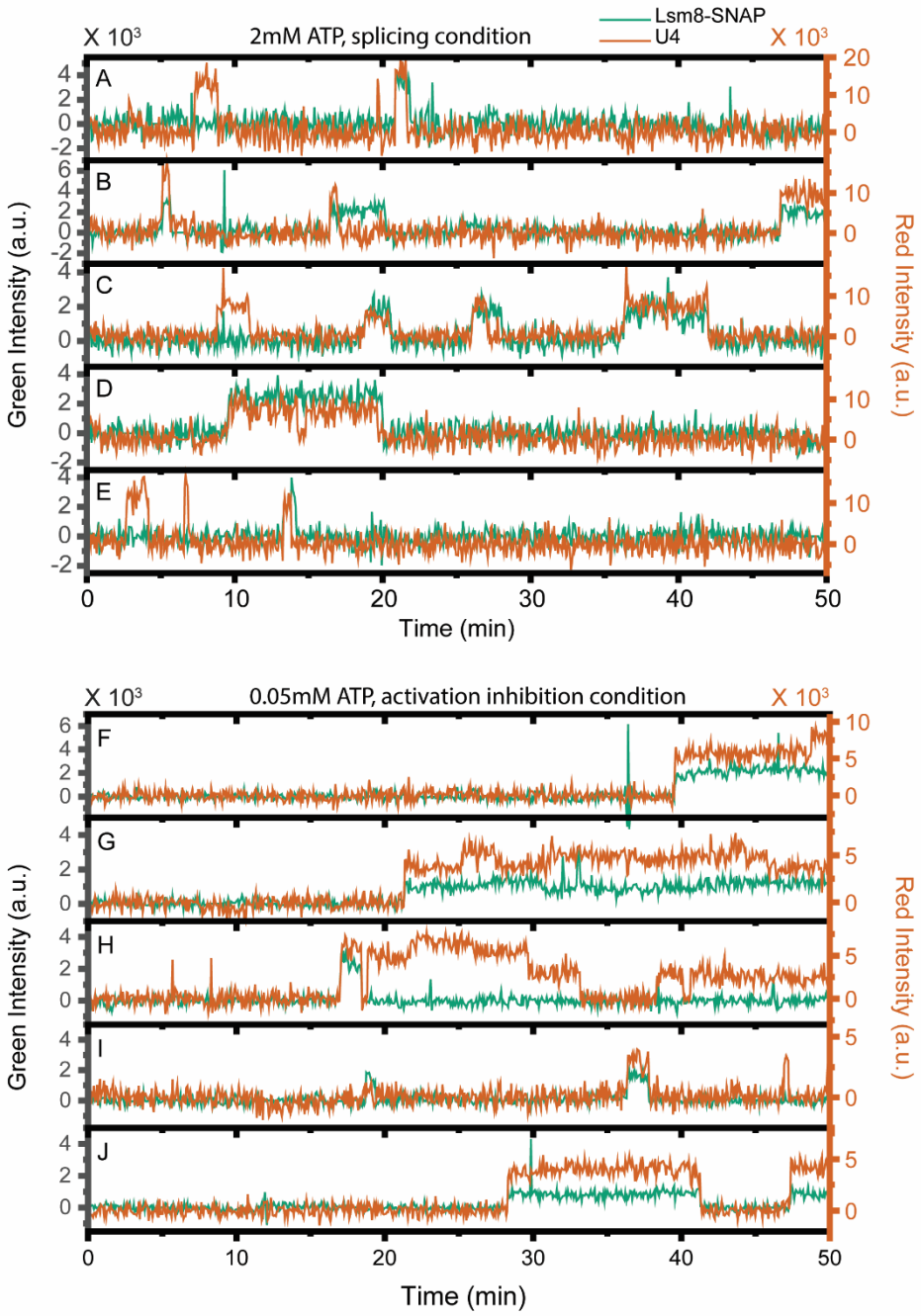


Figure 2.7 Additional Sample Fluorescence Trajectories from 3-color CoSMoS Experiments Monitoring Lsm8-SNAP and U4-DHFR Proteins.

Shown are super-imposed fluorescence intensities for Lsm8-SNAP (green traces) and U4-DHFR proteins (red traces) at 2 mM ATP (**A-E**) and 0.05 mM ATP (**F-J**).

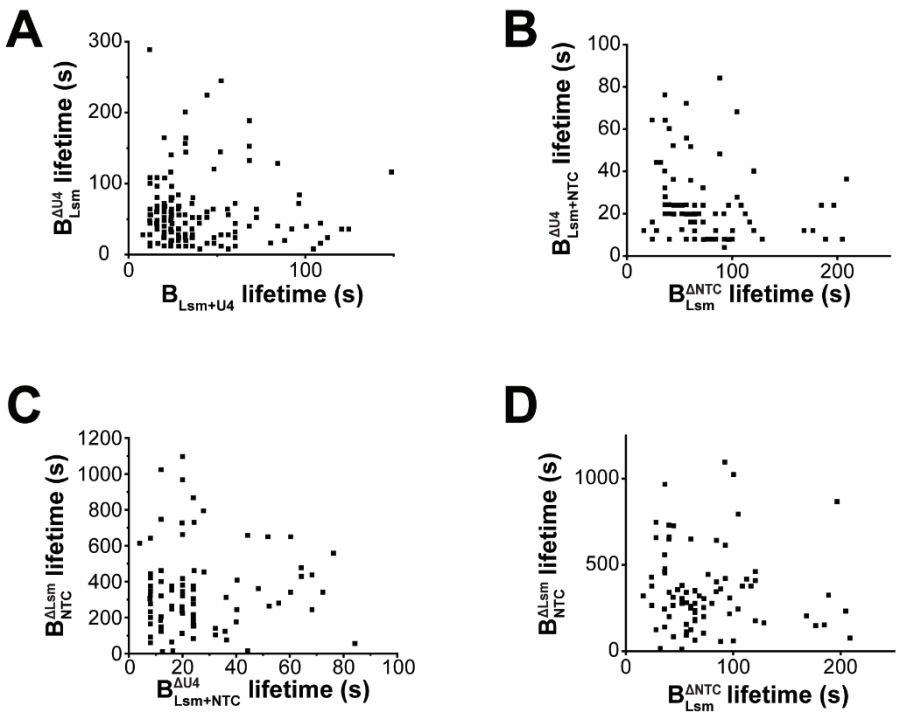


Figure 2.8 Lifetimes of Spliceosome Complexes Exhibit Little Correlation with One Another.

The lifetimes for paired events for the indicated spliceosome complexes obtained from 3-color CoSMoS experiments were plotted against one another. (**A**) Correlation of *N*=144 pairs of lifetimes for the indicated complexes. For this dataset, a Pearson's correlation coefficient (r) of 0.17 was observed. (**B**) Correlation for *N*=84 pairs of lifetimes yielded r = -0.09. (**C**) Correlation for *N*=83 pairs of lifetimes yielded r = 0.02. For this data set one outlier was removed with a NTC lifetime of >2000 s. (**D**) Correlation for *N*=84 pairs of lifetimes yielded r = -0.15.

The NTC Joins the Spliceosome Before Lsm2-8 Release

We next investigated the timing of Lsm8-SNAP release relative to NTC recruitment during activation. Previous work from our lab showed that the NTC associates after U4 snRNP release (19). Since Lsm2-8 release also occurs after U4 snRNP release, it is possible that the NTC associates while the Lsm2-8 ring is still present, concertedly with Lsm2-8 departure, or afterwards. To discriminate between these possibilities, we carried out 3-color CoSMoS assays using extracts containing Lsm8-SNAP labeled with DY-549, NTC proteins Cef1 and Syf1 both labeled with the DHFR tag and Cy5-TMP, and Alexa Fluor 488-labeled pre-mRNAs (**Figs. 2.9A, 2.4**).

Analysis of the fluorescence trajectories shows that Lsm8 signals frequently appeared before those from the NTC on single pre-mRNA molecules (**Figs. 2.9B, 2.10**). This agrees with the ordered addition of the tri-snRNP and the NTC to spliceosomes (17, 20). Most often the Lsm8 and Cef1/Syf1 signals were observed simultaneously on the same pre-mRNA. This suggests that the NTC associates with the spliceosome during activation before Lsm2-8 release. To determine the frequency of these type of events, we identified all pairs of Lsm8 and Cef1/Syf1 signals and determined the distribution of binding behaviors by calculating the difference between the Lsm8 release and NTC recruitment times ($t_{Lsm}^{release} - t_{NTC}^{recruitment}$) (**Fig. 2.9C**). Within the distribution, events in which the NTC binds before Lsm8 release result in positive values and events in which the NTC binds after Lsm8 release result in negative values. The order of Lsm8 release and NTC binding cannot be determined (indeterminate, I.D.) for events in which the last frame with a Lsm8 signal is the same as the first frame with a NTC signal ($t_{Lsm}^{release} - t_{NTC}^{recruitment} = 0$) or for cases where the last frame with a Lsm8 signal is immediately followed by the first frame with a NTC. In this latter case, colocalization cannot be judged during the time lapse interval between frames. Of the analyzed event pairs, the majority (~69%) have $t_{Lsm}^{release} - t_{NTC}^{recruitment} > 0$, indicating the predominant pathway for NTC protein binding is while Lsm8 is still present. In only ~17% of event pairs did loss of the Lsm8 signal precede NTC binding. Among the 157 pairs of events showing

simultaneous presence of Lsm8 and NTC proteins on the pre-mRNA molecules, 87% showed that Lsm2-8 is recruited prior to NTC arrival (**Fig. 2.9D**). This is consistent with tri-snRNP recruitment occurring before the integration of the NTC complex (19) and with the NTC arriving before Lsm2-8 release.

It has been proposed that for human spliceosomes undergoing activation, NTC proteins bind sequentially (18). A NTC subcomplex containing CDC5L (the human homolog of Cef1) binds first to form the pre-B^{ACT1} spliceosome. A different subcomplex containing Syf1 (the intron binding complex, IBC) then binds in a second step to form the pre-B^{ACT2} spliceosome. Since in our experiments, Cef1 and Syf1 were both DHFR labeled, we would expect to see a stepwise increase in Cy5-TMP fluorescence intensity during activation if the factors bound sequentially. While we could observe stepwise decreases fluorescence intensity consistent with the presence of both Cef1 and Syf1 followed by loss of the signal from one of the proteins (14 out of 83 NTC events analyzed, **Table 2.4**; likely due to photobleaching of one fluorophore), we never observed a stepwise increase in fluorescence consistent with sequential binding. This is unlikely due to sub-stoichiometric occupancy of Cy5-TMP on the labeled proteins, since nearly identical results were previously reported for single molecule studies of Cef1 and Syf1 labeled with the SNAP tag (20). These data suggest that in yeast, Cef1 and Syf1 are recruited to the spliceosome either very rapidly one after another or as part of the same NTC complex.

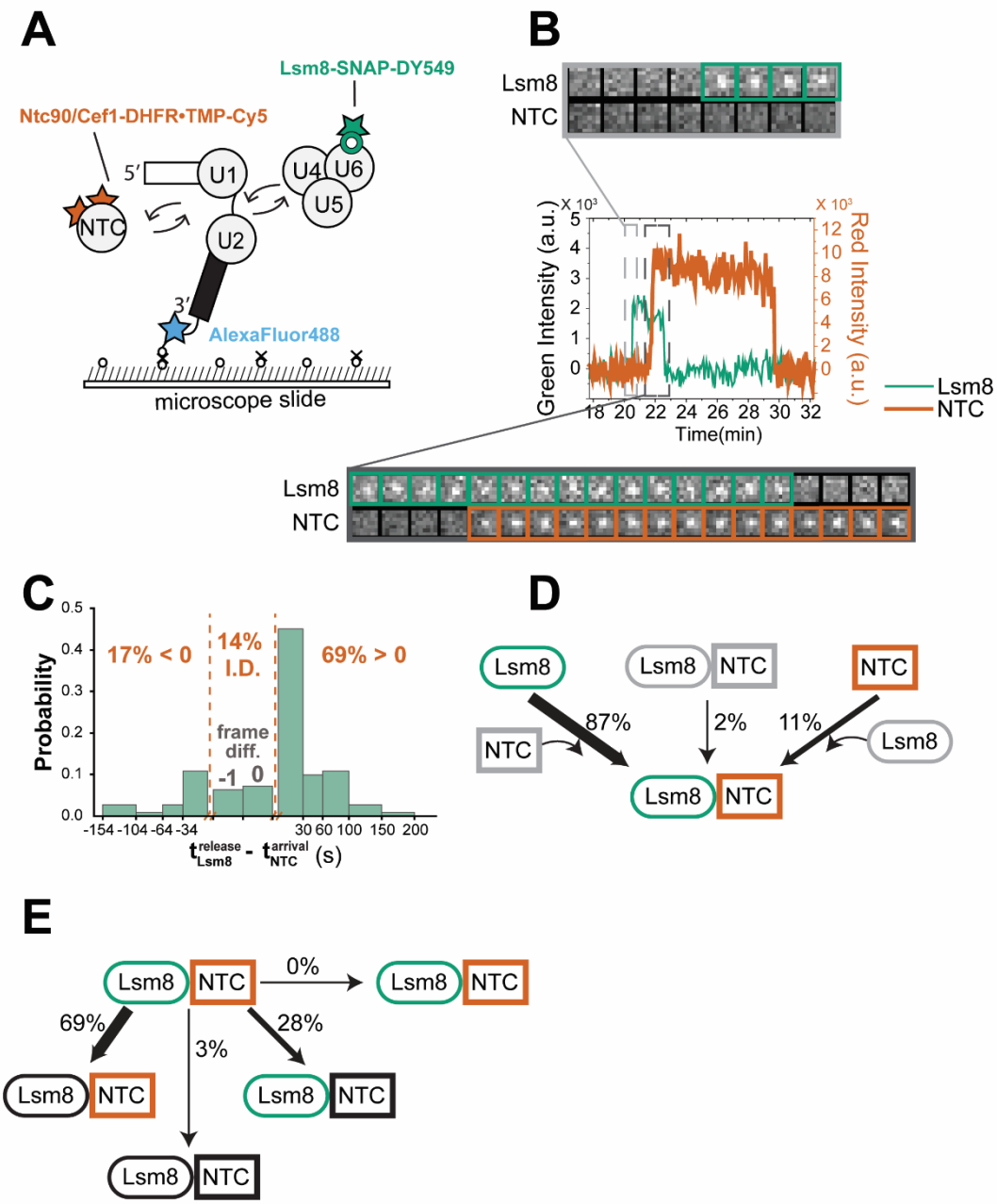


Figure 2.9 Three-color CoSMoS observation of NTC and Lsm ring binding dynamics during activation.

(**A**) Schematic of a 3-color experiment in which the NTC was labeled with Cy5, the Lsm ring was labeled with a DY-549, and the surface-tethered pre-mRNA was labeled with Alexa Fluor 488. (**B**) Segment of a representative time record showing peaks in fluorescence intensity corresponding

to colocalization of NTC (red, thick line) and Lsm proteins (green, thin line) with the same individual pre-mRNA molecule. The light grey dashed rectangle marks an example of the ordered appearance of NTC and Lsm8 spots; galleries show consecutive images taken from that part of the recording showing that spot appearance is not simultaneous. The dark grey dashed rectangle marks an example of simultaneous occupancy of the pre-mRNA by the NTC and Lsm ring followed by the ordered disappearance of Lsm8 and then NTC spots; galleries show consecutive images taken from the indicated part of the recording showing overlap of the Lsm8 and NTC signals. (**C**) Probability density histogram showing the delay between NTC arrival and Lsm8 release. Most often (69% of *N*= 293 total events), the NTC arrived before release of the Lsm ring ($t_{Lsm}^{release} - t_{NTC}^{arrival} > 0$). In 14% of cases, the exact order of events was indeterminate (I.D.) as described in the text. (**D**) Routes for the appearance of Lsm8 and NTC fluorescent spots at 2 mM ATP for *N*=157 pairs of overlapping events. (**E**) Routes for loss of either Lsm or NTC fluorescent spots at 2 mM ATP for *N*=136 pairs of overlapping events in which the Lsm8 spot appearance preceded arrival of the NTC.

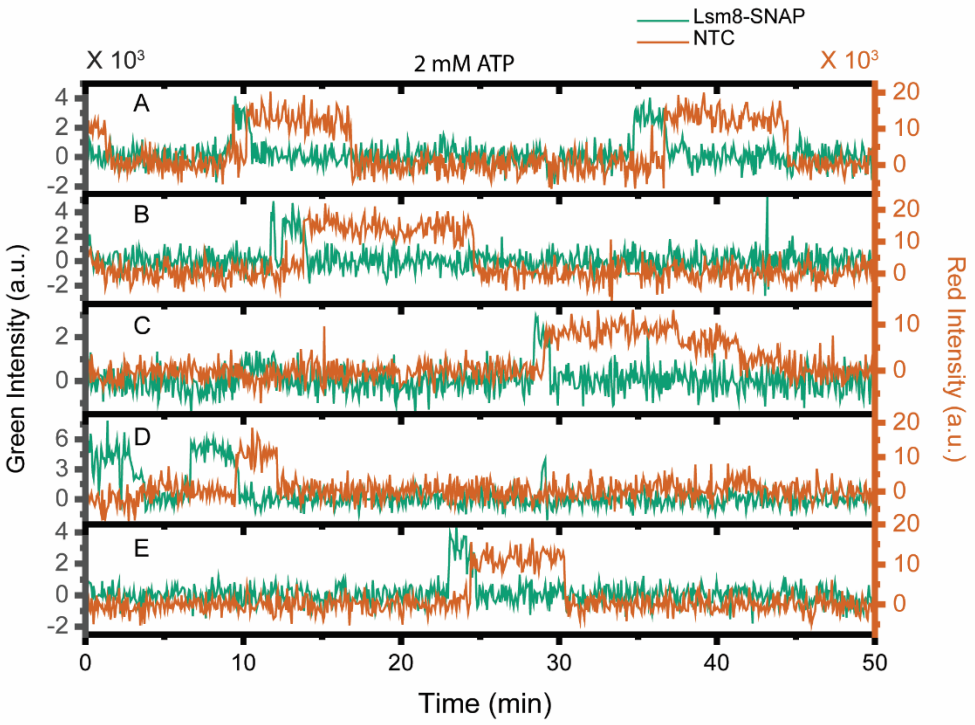


Figure 2.10 Additional Sample Fluorescence Trajectories from 3-color CoSMoS Experiments Monitoring Lsm8-SNAP and NTC-DHFR Proteins.

Shown are super-imposed fluorescence intensities for Lsm8-SNAP (green traces) and NTC-DHFR proteins (red traces) at 2 mM ATP.

Table 2.4 Number of Stepping Events Observed During NTC Binding

Strain/Subcomplex	Total Events	Step Down Events	Step Up Events
yAAH0020 (Syf1/Cef1-SNAP)[a]	84	23 (27%)	1 (1%)
yAAH0362 (Syf1/Cef1-DHFR) (this work)	83	14 (17%)	0 (0%)

[a]Data taken from Table S10 (20).

NTC Transiently Samples the Spliceosome during Activation

For the subset of molecules with simultaneous occupancy by Lsm8 and Cef1/Syf1, we classified each pair of events according to one of fourteen predicted patterns (**Fig. 2.11**). While most events showed release of the NTC proteins occurring after Lsm8 release (seq-13 events; **Fig. 2.11C**), we were surprised to see many events (24% of events), in which the NTC bound and released while the Lsm8 was still present (seq-15 events; **Figs. 2.12A, 2.11C**). These NTC events were often short-lived and lasted only a few frames, much shorter than other event types (**Fig. 2.12B**).

It is possible that the short-lived NTC events represent transient interactions with the spliceosome during activation in which stable integration of the NTC is not yet possible (sampling). We predicted that sampling is more likely to occur early in activation since the conformational changes needed for stable NTC binding may not yet be completed. To test this, we aligned the time of each NTC molecule's recruitment to the normalized lifetime of the colocalized Lsm8 binding event. A plot of these synchronized NTC recruitment times versus their corresponding NTC dwell times shows that short-lived NTC events are more commonly observed soon after the Lsm8 signal appears (**Fig. 2.12C**). Longer-lived NTC binding events appear only after ~50% of the Lsm8 dwell time on the spliceosome has elapsed. These observations are confirmed when we calculate the cumulative sum for NTC binding as a function of the normalized Lsm dwell time (**Fig. 2.12D**). In these plots, a lag phase is apparent for stable NTC accumulation for events in which sampling was observed as well as for those in which sampling was not detected. This shows that stable NTC binding requires a spliceosome remodeling event to take place during activation after arrival of the tri-snRNP and that this requirement is present on molecules for which we do and do not observe any sampling events.

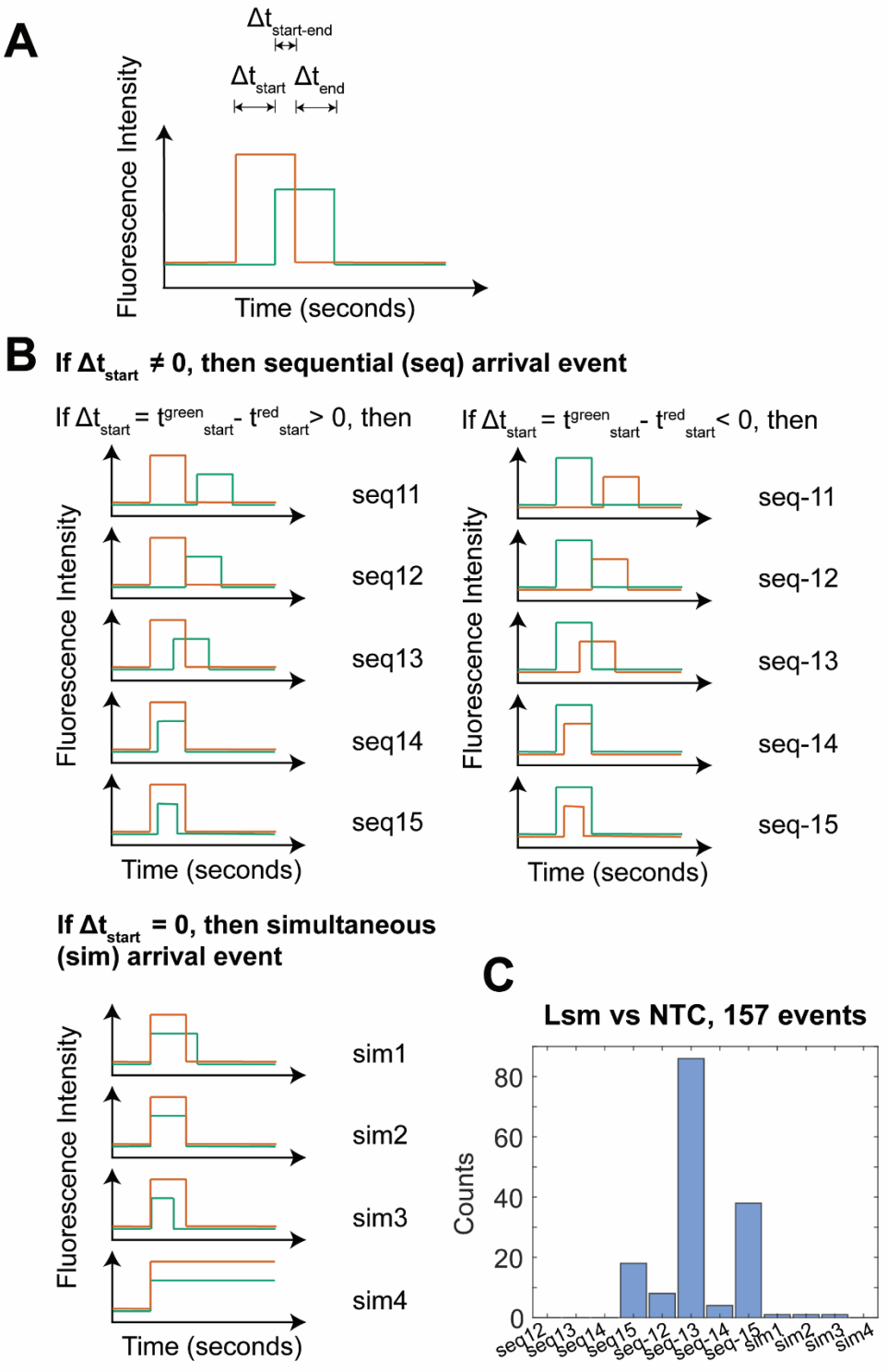

A
Δt start-end
Δt start
Δt end
Fluorescence Intensity
Time (seconds)
B
If Δt start ≠ 0, then sequential (seq) arrival event
If Δt start = t green start - t red start > 0, then
If Δt start = t green start - t red start < 0, then
Fluorescence Intensity
seq11
seq12
seq13
seq14
seq15
Time (seconds)
seq-11
seq-12
seq-13
seq-14
seq-15
If Δt start = 0, then simultaneous (sim) arrival event
sim1
sim2
sim3
sim4
C
Lsm vs NTC, 157 events
Counts
0
20
40
60
80
seq12
seq13
seq14
seq15
seq-12
seq-13
seq-14
seq-15
sim1
sim2
sim3
sim4

Figure 2.11 Scheme for assignment of observed binding and dissociation patterns in CoSMoS assays.

(**A**) For each set of paired events, the beginning and end times of the individual events were recorded. For the analysis shown here, Δt_{start} values were first calculated by subtracting the start time of the green event from the start time of the red event. (**B**) Δt_{start} values were separated such that all positive values indicated that the red event arrived before the green event (seq11 to 15), negative values indicated that the green event arrived after the red event (seq-11 to -15), and zero values indicated simultaneous arrival (sim1 to 4). Values of Δt_{start} and $\Delta t_{start\text{-}end}$ (**A**) were then determined to subcategorize each arrival class based on the disappearance patterns of the green and red events. (**C**) Distribution of Lsm8 and NTC binding events from 3-color CoSMoS experiments. Events of the seq-15 type arise from sampling behaviors.

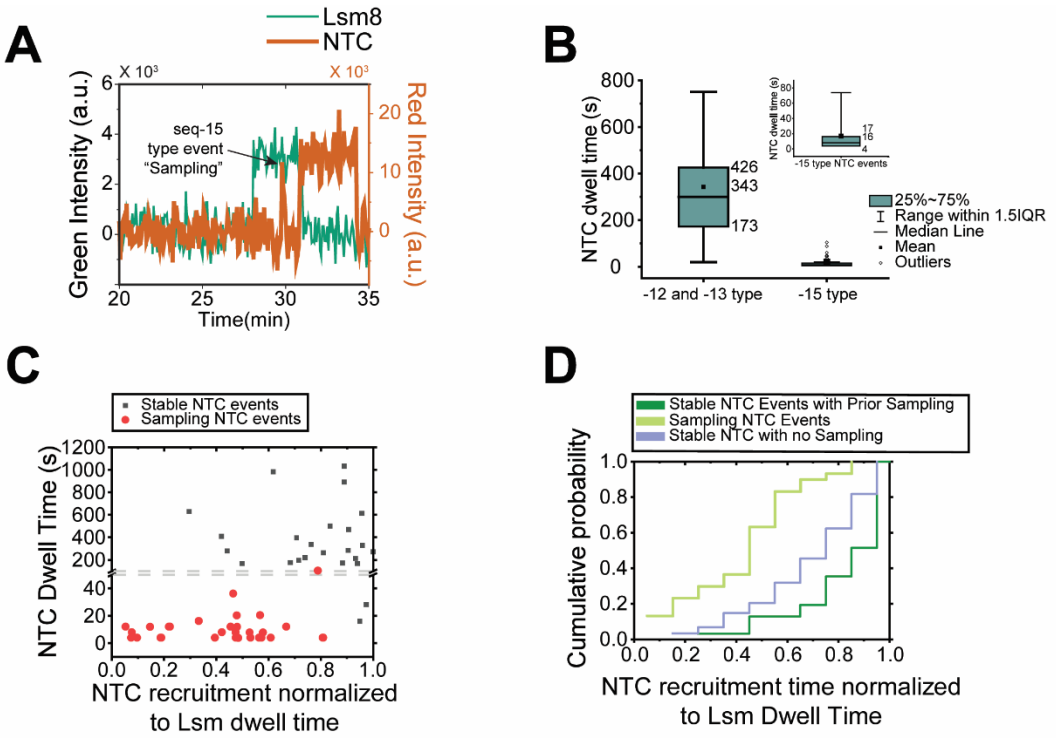


Figure 2.12 The NTC Frequently Samples the Spliceosome Prior to Stable Binding

(**A**) Example fluorescence trajectory showing a seq-15-type event/NTC sampling (arrow). (**B**) The observed dwell times for sampling events (-15 type) are much shorter than those observed for non-sampling events (-12 and -13 type). The inset shows the -15 type events with an expanded y-axis. Numbers adjacent to the boxes indicate the interquartile range (IQR, 25-75%) and mean. Outliers (hollow circles) are defined as those with dwell times >1.5xIQR. (**C**) Sampling events are more often observed soon after Lsm8 arrival and longer-lived NTC binding events are more often observed closer to the point of Lsm8 release. Fluorescence trajectories were synchronized post-experimentally to the Lsm8 dwell times which were then normalized (time = 0 represents the arrival of the Lsm8 signal and time = 1 represents loss of the Lsm8 signal). The dashed line indicates a break in the y-axis. (**D**) Normalized cumulative sum histogram for NTC binding normalized to Lsm8 dwell times for NTC binding events that do (light and dark green) and do not (blue) include sampling-type events. Long-lived, stable NTC binding events are more frequently observed towards the end of the Lsm8 binding time regardless if a sampling event was present.

Spliceosomes Containing Lsm8 and NTC are Short-Lived.

We then analyzed the kinetic features of spliceosomes containing Lsm8, Cef1, and Syf1 ($B^{\Delta U4}_{Lsm+NTC}$ complexes) by identifying the arrival and departure times for each colocalized Lsm8-SNAP and Cef1/Syf1-DHFR event. The lifetimes of $B^{\Delta NTC}_{Lsm}$ complexes were determined by subtracting the arrival times of Lsm8 signals from the arrival times of NTC signals ($t^{arrival}_{NTC} - t^{arrival}_{Lsm8}$), and the lifetimes of $B^{\Delta U4}_{Lsm+NTC}$ complexes were determined by subtracting the arrival times of NTC signals from the release times of Lsm8 ($t^{release}_{Lsm8} - t^{arrival}_{NTC}$) (**Fig. 2.13A**). Finally, the remaining lifetimes of complexes containing the NTC following Lsm8 release ($B^{\Delta Lsm}_{NTC}$ and later splicing complexes, L.C.) were found by subtracting the release time of Lsm8 from the release time of NTC ($t^{release}_{NTC} - t^{release}_{Lsm8}$). The distributions of dwell times were then fit as described above for the B_{Lsm+U4} and $B^{\Delta U4}_{Lsm}$ complexes.

In all cases, the distributions were best estimated using equations with a single exponential term as compared to more complex models using a loglikelihood ratio test (**Figs. 2.13B-D**, **Table 2.3**). The shortest-lived complex was $B^{\Delta U4}_{Lsm+NTC}$ at 21.1±2.7 s. This is three-fold shorter than the lifetime of the $B^{\Delta NTC}_{Lsm}$ complex (68.6±8.9 s) and represents only about 5% of the average total lifetime of the NTC on the spliceosome. We also tested whether the individual lifetimes of any of the $B^{\Delta NTC}_{Lsm}$, $B^{\Delta U4}_{Lsm+NTC}$ and $B^{\Delta Lsm}_{NTC}$ and subsequent complexes correlated with one another (**Figs. 2.8B-D**). As with U4 release, we could not identify any trends when the associated complex lifetimes were plotted against one another.

We previously determined that loss of U4 snRNP proteins during activation represented an irreversible step in splicing (19). Prp3/Prp4 almost never rejoined the pre-mRNA after their release. Instead, activated spliceosomes would either splice or disassemble to allow a different tri-snRNP molecule to associate and initiate a new activation process. In experiments with fluorescently labeled Lsm8 and Cef1/Syf1, Lsm8-SNAP rarely reappeared after its release while

the NTC remained bound (4 out of 83 events). Consequently, U4 snRNP and Lsm2-8 release appear to happen sequentially and nearly irreversibly during splicing.

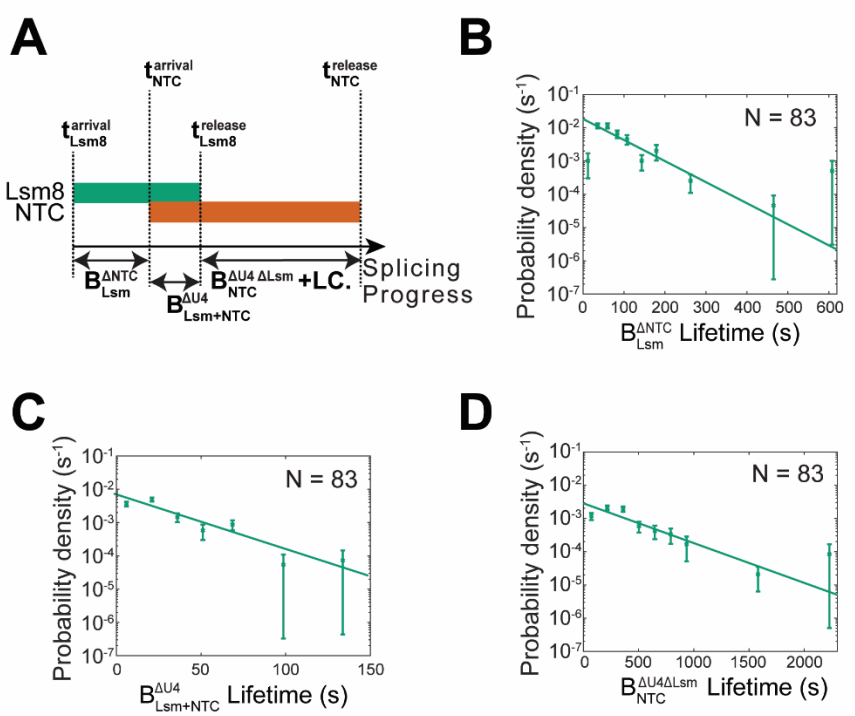


Figure 2.13 Lifetimes of Spliceosome Complexes Identified by CoSMoS

(**A**) Schematic showing the relationship between the lifetimes of the $B^{\Delta NTC}_{Lsm}$, $B^{\Delta U4}_{Lsm+NTC}$ and $B^{\Delta Lsm}_{NTC}$ and later complexes to the measured arrival and release times for Lsm8 and NTC proteins. (**B-D**) Probability density histograms of $B^{\Delta NTC}_{Lsm}$ (panel B, $t^{arrival}_{NTC} - t^{arrival}_{Lsm8}$), $B^{\Delta U4}_{Lsm+NTC}$ (panel C, $t^{release}_{Lsm8} - t^{arrival}_{NTC}$), and $B^{\Delta Lsm}_{NTC}$ and later complex lifetimes (panel D, $t^{release}_{NTC} - t^{release}_{Lsm8}$) obtained from the subset of events (N) showing ordered arrival of Lsm and then NTC spots followed by ordered loss of the Lsm and then NTC signals. Lines represent fits of the lifetime distributions with equations containing single exponential terms that yielded the parameters reported in **Table 2.3**. Error bars were calculated for each point as described in the Methods

Lsm2-8 Forms a Complex with U2/U6 helix II RNAs.

Since Lsm2-8 is not released until after the NTC joins, we speculated that this could indicate an important function for Lsm2-8 during activation that is then taken over by the NTC. One possible function could be to maintain U2/U6 helix II structural integrity since it consists of only ten base pairs and has a calculated melting temperature of near 35ºC in 1 M Na^+ (30). In activated spliceosomes, this helix is surrounded by the NTC proteins Syf1 and Syf3 and directly bound by Syf2 (31, 32). These interactions may help stabilize helix II during the catalytic steps of splicing. In contrast, helix II is not known to be bound by proteins in B complex spliceosomes and is located adjacent to Lsm2-8 (**Fig. 2.14A**).

We tested if Lsm2-8 could stabilize U2/U6 helix II *in vitro* with electrophoretic mobility shift assays (EMSAs) using purified recombinant Lsm2-8 and RNA oligo mimics of the 5' and 3' ends of the U2 and U6 snRNAs, respectively (**Fig. 2.14B**). As expected, Lsm2-8 tightly bound the $U6_{WT}$ oligo (**Fig. 2.14C**, lane 2). When the U2 oligo was added, we observed a supershift dependent on the U2 oligo concentration (lanes 6-8). No supershift was observed if the U2 oligos could not base pair to U6 (lanes 15-17) or if the U6 oligo was missing a Lsm2-8 binding site (lanes 21-23). Together, the data show that Lsm2-8 can form a ternary complex with the U2 and U6 RNAs dependent on base pairing and Lsm2-8 binding to U6. Lsm2-8 may function in spliceosomes to maintain U2/U6 helix II until it can be stabilized by NTC proteins, possibly by handover from Lsm2-8 to the NTC in the $B^{\Delta U4}_{Lsm+NTC}$ complex.

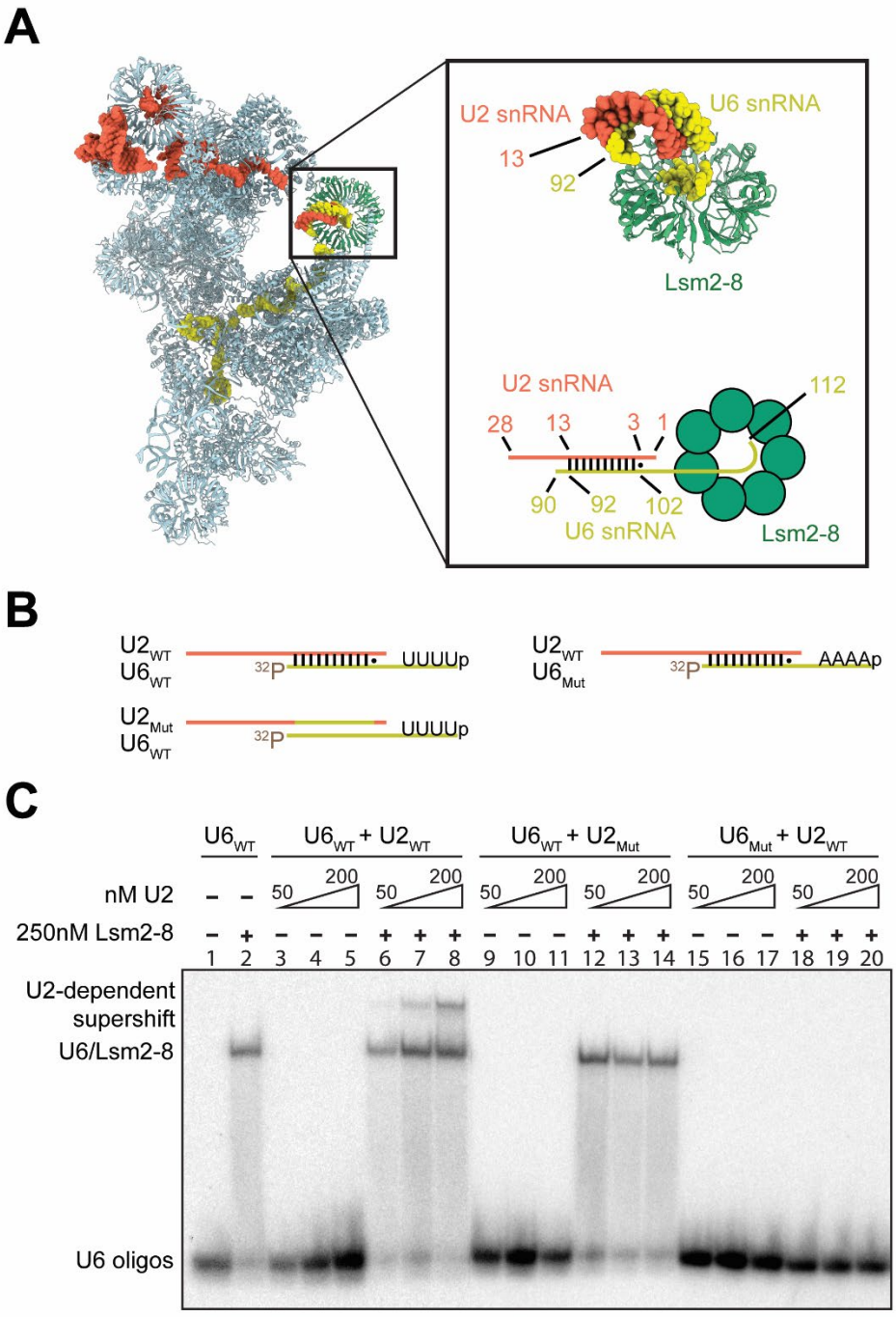


Figure 2.14 Formation of a Lsm2-8/U2/U6 ternary complex *in vitro*.

(**A**) Cryo-EM structure (left) showing the location of Lsm2-8 (green), U2 snRNA (red) and U6 snRNA (yellow) in the yeast B complex spliceosome (PDB: 5NRL). Inset shows a close-up view of U2/U6 helix II located adjacent to Lsm2-8. The numbers of the terminal pairing nucleotides for

U2 and U6 are indicated. Below the close-up view is a schematic of the structure with nucleotides that form helix II indicated. The $U2_{WT}$ oligo used in EMSA assays represents nulceotides 1-28 of the U2 snRNA, and the $U6_{WT}$ oligo represents nucleotides 90-112 of the U6 snRNA. (**B**) Cartoon illustating the U2 and U6 oligo pairs used for the *in vitro* binding assay (**Table 2.1**). All U6 oligos contain a 5'-[^{32}P] label and a 3'-end phosphate for mimicking the processed 3'-end of yeast U6. The $U6_{Mut}$ oligo does not contain a Lsm-binding site, and the $U2_{MUT}$ oligo can not base pair to U6 to form helix II. (**C**) EMSA analysis of Lsm2-8/U2/U6 complex formation. Phosphor image of a native PAGE gel showing the presence of a U2-dependent supershifted complex (lanes 6-8). The supershifted complex does not appear if the U2 oligo cannot base pair to U6 (lanes 15-17) or if the U6 oligo lacks a Lsm-binding site (lanes 21-23). In all cases, the final concentrations of [^{32}P]-labeled U6 oligos and Lsm2-8 were kept constant at 1nM and 250nM, respectively, while the U2 oligo concentration varied 50-200 nM if present.

DISCUSSION

A Kinetic Scheme for Spliceosome Activation

By integrating our experimental data, we can define a kinetic scheme for yeast spliceosome activation *in vitro* (**Fig. 2.15**). In the scheme, the tri-snRNP first joins the spliceosome A complex to form the pre-B and subsequent B complex. The average combined lifetime of these complexes is ~42 s. Multiple rearrangements are expected to occur during this time including release of the U1 snRNP, which has not yet been kinetically characterized. Following the loss of U4 snRNP, a $\mathrm{B}_{\mathrm{Lsm}}^{\Delta \mathrm{U4} \Delta \mathrm{NTC}}$ complex forms, which we define as containing the Lsm2-8 ring but lacking the U4 snRNP and NTC. To our knowledge, this complex has not previously been reported for either yeast or human spliceosomes. While we did not directly observe this intermediate in our experiments, we can infer its existence from previous data showing that the NTC is primarily recruited after U4 snRNP release (19) and experiments reported here showing that Lsm2-8 release occurs after NTC recruitment (**Fig. 2.9**). The lifetime of the $\mathrm{B}_{\mathrm{Lsm}}^{\Delta \mathrm{U4} \Delta \mathrm{NTC}}$ intermediate can be estimated by either subtracting the lifetime of $\mathrm{B}_{\mathrm{Lsm+NTC}}^{\Delta \mathrm{U4}}$ from the lifetime of $\mathrm{B}_{\mathrm{Lsm}}^{\Delta \mathrm{U4}}$ (65-21=44s) or by subtracting the lifetime of $\mathrm{B}_{\mathrm{Lsm+U4}}$ from the lifetime of $\mathrm{B}_{\mathrm{Lsm}}^{\Delta \mathrm{NTC}}$ (69-42=27s) (**Table 2.3**). This yields a range of 27-43s for the lifetime of $\mathrm{B}_{\mathrm{Lsm}}^{\Delta \mathrm{U4} \Delta \mathrm{NTC}}$.

Subsequently, the NTC is recruited to $\mathrm{B}_{\mathrm{Lsm}}^{\Delta \mathrm{U4} \Delta \mathrm{NTC}}$ to form the $\mathrm{B}_{\mathrm{Lsm+NTC}}^{\Delta \mathrm{U4}}$ spliceosome, which persists for only ~21 s before the release of Lsm2-8 ring. Like the preceding complex, $\mathrm{B}_{\mathrm{Lsm+NTC}}^{\Delta \mathrm{U4}}$ has not previously been observed for yeast spliceosomes but shares some similarities with the human pre-B^{ACT1}. Following this step, the NTC typically remains associated for ~364 s. This remaining NTC lifetime is a composite of subsequent steps in splicing and spliceosome discard. Our kinetic scheme also supports the hypothesis that the human pre-B^{ACT1} complex observed in the presence of splicing inhibitors (18) is a competent activation intermediate. We were able to detect spliceosomes compositionally similar to human pre-B^{ACT1} being formed and transformed at

rates rapid enough to be consistent with *in vitro* splicing. Overall kinetic efficiency of the activation process may be achieved, in part, by the presence of at least two irreversible steps: ordered release of the U4 snRNP and Lsm2-8.

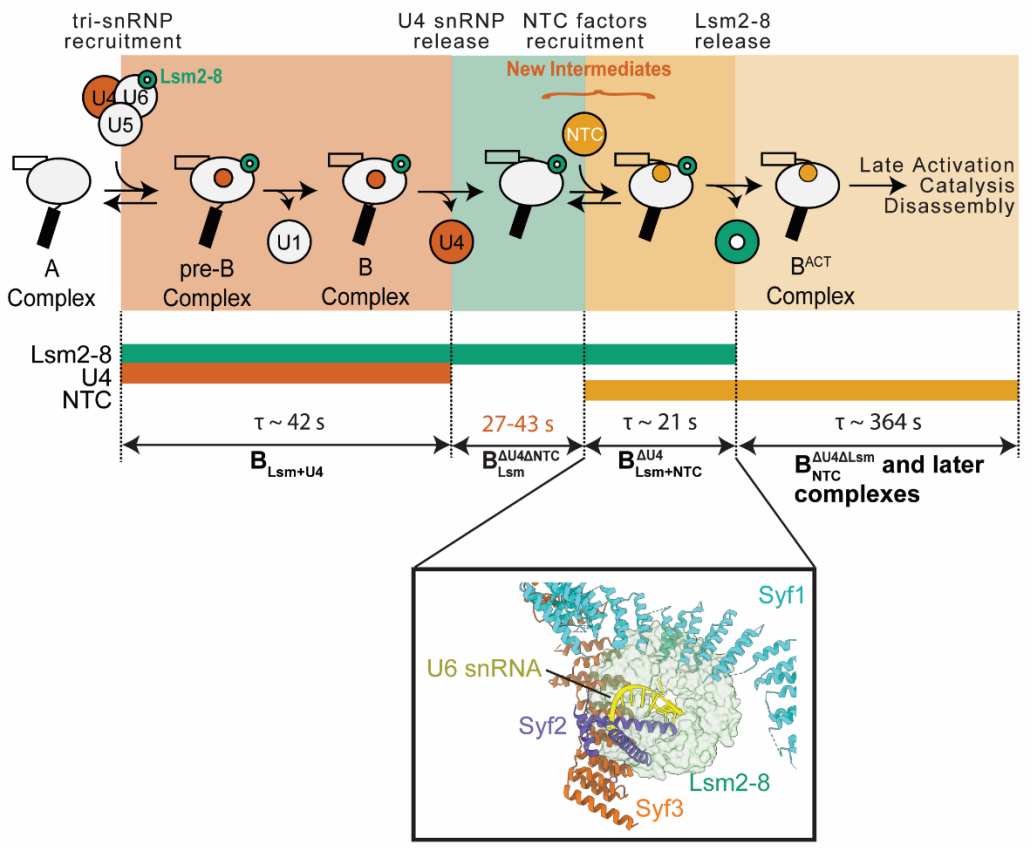


Figure 2.15 Transient Intermediates Formed During Spliceosome Activation.

In this kinetic scheme, U4 snRNP release precedes NTC recruitment, which in turn precedes Lsm2-8 release. This would involve formation of at least two activation intermediates, the $B^{\Delta U4}_{Lsm}$ and $B^{\Delta U4}_{Lsm+NTC}$ complexes, which have not been previously biochemically or structurally characterized. Single molecule data allows determination of the characteristic lifetimes (t) of these complexes. The $B^{\Delta U4}_{Lsm}$ complex lifetime (red) was not directly measured in our experiments but can be inferred as described in the text (19). Not pictured in this scheme are NTC sampling events that begin soon after tri-snRNP recruitment. (*Inset*) The structure of the $B^{\Delta U4}_{Lsm+NTC}$ complex represents a unique conformation of the spliceosome since the Lsm2-8 ring and the NTC proteins Syf1 and Syf3 are mutually exclusive with one another when the structures of yeast B and B^{ACT} complexes are superimposed (PDBs: 5NRL and 5GM6).

Activation Intermediates Suggest New Spliceosome Conformations

It is probable that the $B_{Lsm}^{\Delta U4\Delta NTC}$ and $B_{Lsm+NTC}^{\Delta U4}$ spliceosomes have unique conformations not yet observed in cryo-EM spliceosome structures. No structural information yet exists for spliceosomes lacking both U4 snRNP and NTC; however, biochemical evidence for such a complex has been obtained for the human splicing machinery in the presence of inhibitors (33). Lsm2-8 and NTC proteins must be uniquely positioned in the $B_{Lsm+NTC}^{\Delta U4}$ spliceosome since these factors are mutually exclusive when the structures of yeast B and B^{ACT} complexes are superimposed (**Fig. 2.15B**). The structure of $B_{Lsm+NTC}^{\Delta U4}$ is also distinct from that of human pre-B^{ACT1} complex since our data indicate that Syf1 is present in $B_{Lsm+NTC}$ but this protein is missing from pre-B^{ACT1}. Indeed, it was noted that human Syf1 cannot bind pre-B^{ACT1} complexes due to a mutually exclusive binding site with Lsm2-8 (18). Whether or not a transient human spliceosome containing both Syf1 and Lsm2-8 (analogous to yeast $B_{Lsm+NTC}^{\Delta U4}$) is formed during the pre-B^{ACT1} to pre-B^{ACT2} transition is not yet known.

NTC Samples Spliceosomes during Activation

By watching NTC recruitment in real time, we were able to observe transient NTC binding events to B complex spliceosomes that precede stable NTC integration. Since these transient events are more likely to occur early in activation and stable events more likely to occur later, it is consistent with transition of the spliceosome from a low to high affinity NTC-binding state. We have not identified the transition required for high affinity binding, but it could involve structural remodeling permitted by U4 snRNP release to allow the spliceosome to accommodate both the NTC and Lsm2-8.

The sampling behavior itself may allow the NTC to surveil spliceosomes for efficient and correct progress through activation. Rapid and reversible NTC binding may prevent NTC complexes from being sequestered by malfunctioning spliceosomes such as those which fail to

release U4. Since the NTC is also involved in nuclear processes other than splicing (34-36), these features of NTC binding dynamics might be important for preventing depletion of the available NTC pool by limiting formation of stable but unproductive complexes.

Divergent Pathways for NTC Binding in Yeast and Human Spliceosomes

Recently, it has been proposed that the human NTC is composed of discrete subunits that associate with the spliceosome during activation in a stepwise mechanism (18). Initially, the hNTC subunit containing CDC5L (yeast Cef1) and five other factors associates to form the pre-B^{ACT1} complex. In a second step, the intron binding complex (IBC) that contains SYF1 along with the hNTC-related (hNTR) proteins and other factors associate to form the pre-B^{ACT2} complex. In contrast, biochemical and mass spectrometry data support the existence of a single yeast NTC complex containing Cef1, Syf1, Prp19 along with IBC and NTR components (37, 38).

Our single molecule data also support the existence of a NTC complex containing Cef1 and Syf1. While it was previously speculated that the smaller size of the hNTC relative to its yeast counterpart permits the simultaneous presence of hNTC with Lsm2-8 during activation (18), our results show that this is still possible even with the larger yeast NTC, possibly due to structural rearrangements in the NTC or Lsm2-8 binding region. Thus, stable NTC association with the spliceosome in yeast appears to involve both fewer steps and individual subcomplexes than in humans. This may contribute to greater efficiency of yeast splicing but at the expense of fewer potential points at which splicing can be regulated during activation.

A Function for Lsm2-8 in Maintaining U2/U6 Helix II during Activation

One activity that could occur within the $B^{\Delta U4}_{Lsm+NTC}$ complex is transfer of the relatively unstable U2/U6 helix II to the NTC. Since Lsm2-8 is located adjacent to U2/U6 helix II, it could facilitate formation of this RNA duplex and/or help to stabilize it once formed. Avoiding release of Lsm2-8 in the absence of NTC binding may therefore be important for maintaining helix II integrity

during activation and preventing spurious unwinding which could lead to spliceosome disassembly. Consistent with this hypothesis, we observed that purified Lsm2-8 can form a ternary complex with U2/U6 helix II RNAs dependent on base pairing (**Fig. 2.14**). How important this function is for the yeast splicing machinery is unclear since neither helix II nor the helix II-interacting NTC protein Syf2 are essential for yeast viability (39, 40). Lsm2-8 stabilization of helix II during activation may only become essential under certain conditions in which the nascent spliceosome catalytic center is destabilized. Indeed, helix II becomes essential when the spliceosome catalytic center is mutated (39). Interestingly, Lsm2-8 also facilitates annealing of the U4 and U6 snRNAs by Prp24 (41). In this case, it was proposed to occur through Lsm2-8-facilitated folding of the intramolecular U6 telestem RNA duplex. The 3' half of the telestem ultimately pairs with U2 to form U2/U6 helix II. It is possible that similar properties of Lsm2-8 contribute to either U6 telestem or U2/U6 helix II stabilization, both of which involve duplex formation to the same region of U6 upstream of the Lsm2-8 binding site.

Formation of a complex between two complementary RNAs (U2 and U6) and Lsm2-8 is also intriguing due to similarities between eukaryotic Lsm proteins and bacterial Hfq. Even though Hfq proteins assemble into homohexamers and not heteroheptamers like Lsm2-8, both Hfq and Lsm proteins share a common fold (10). Hfq has been extensively studied for its ability to anneal bacterial small RNAs to mRNAs to regulate their expression (42). Our data suggest that in addition to a common fold, interaction with unstable RNA duplexes has also been evolutionarily conserved between the Lsm and Hfq proteins. The spliceosome may use these interactions to help form and/or stabilize U2/U6 helix II through assembly and the conformational and compositional changes occurring during activation. Whether or not Lsm1-7 or Sm complexes also possess similar duplex stabilizing activities is not yet known but could yield additional functional parallels to Hfq.

REFERENCES

1. Y. Lee, D. C. Rio, Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu Rev Biochem* **84**, 291-323 (2015).
2. R. K. Singh, T. A. Cooper, Pre-mRNA splicing in disease and therapeutics. *Trends Mol Med* **18**, 472-482 (2012).
3. R. Wan, R. Bai, Y. Shi, Molecular choreography of pre-mRNA splicing by the spliceosome. *Curr Opin Struct Biol* **59**, 124-133 (2019).
4. J. P. Staley, C. Guthrie, Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* **92**, 315-326 (1998).
5. C. L. Will, R. Lührmann, Spliceosome structure and function. *Cold Spring Harb Perspect Biol* **3** (2011).
6. C. Yan, R. Wan, Y. Shi, Molecular mechanisms of pre-mRNA splicing through structural biology of the spliceosome. *Cold Spring Harbor perspectives in biology* **11** (2019).
7. W. Y. Tarn, K. R. Lee, S. C. Cheng, Yeast precursor mRNA processing protein PRP19 associates with the spliceosome concomitant with or just after dissociation of U4 small nuclear RNA. *Proc Natl Acad Sci U S A* **90**, 10821-10825 (1993).
8. W. Y. Tarn, K. R. Lee, S. C. Cheng, The yeast PRP19 protein is not tightly associated with small nuclear RNAs, but appears to associate with the spliceosome after binding of U2 to the pre-mRNA and prior to formation of the functional spliceosome. *Mol Cell Biol* **13**, 1883-1891 (1993).
9. H. Kaur, C. van der Feltz, Y. Sun, A. A. Hoskins, Network theory reveals principles of spliceosome structure and dynamics. *Structure* **30**, 190-200.e192 (2022).
10. J. D. Beggs, Lsm proteins and RNA processing. *Biochem Soc Trans* **33**, 433-438 (2005).
11. H. Urlaub, V. A. Raker, S. Kostka, R. Lührmann, Sm protein-Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure. *EMBO J* **20**, 187-196 (2001).
12. M. E. Wilkinson, C. Charenton, K. Nagai, RNA Splicing by the Spliceosome. *Annu Rev Biochem* **89**, 359-388 (2020).
13. Y. Shi, Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat Rev Mol Cell Biol* **18**, 655-670 (2017).
14. P. Khusial, R. Plaag, G. W. Zieve, LSm proteins form heptameric rings that bind to RNA via repeating motifs. *Trends Biochem Sci* **30**, 522-528 (2005).
15. T. Møller *et al.*, Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. *Mol Cell* **9**, 23-30 (2002).
16. E. J. Montemayor *et al.*, Architecture of the U6 snRNP reveals specific recognition of 3'-end processed U6 snRNA. *Nat Commun* **9**, 1749 (2018).
17. S. P. Chan, D. I. Kao, W. Y. Tsai, S. C. Cheng, The Prp19p-associated complex in spliceosome activation. *Science* **302**, 279-282 (2003).
18. C. Townsend *et al.*, Mechanism of protein-guided folding of the active site U2/U6 RNA during spliceosome activation. *Science* **370** (2020).
19. A. A. Hoskins, M. L. Rodgers, L. J. Friedman, J. Gelles, M. J. Moore, Single molecule analysis reveals reversible and irreversible steps during spliceosome activation. *eLife* **5** (2016).

20. A. A. Hoskins *et al.*, Ordered and dynamic assembly of single spliceosomes. *Science* **331**, 1289-1295 (2011).

21. J. D. Larson, M. L. Rodgers, A. A. Hoskins, Visualizing cellular machines with colocalization single molecule microscopy. *Chem Soc Rev* **43**, 1189-1200 (2014).

22. C. F. Fernandez, B. K. Pannone, X. Chen, G. Fuchs, S. L. Wolin, An Lsm2-Lsm7 complex in Saccharomyces cerevisiae associates with the small nucleolar RNA snR5. *Mol Biol Cell* **15**, 2842-2852 (2004).

23. J. Larson *et al.*, Design and construction of a multiwavelength, micromirror total internal reflectance fluorescence microscope. *Nat Protoc* **9**, 2317-2328 (2014).

24. S. W. Stevens *et al.*, Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Mol Cell* **9**, 31-44 (2002).

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

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

27. R. Bai, R. Wan, C. Yan, J. Lei, Y. Shi, Structures of the fully assembled. *Science* **360**, 1423-1429 (2018).

28. H. Kaur, F. Jamalidinan, S. G. F. Condon, A. Senes, A. A. Hoskins, Analysis of spliceosome dynamics by maximum likelihood fitting of dwell time distributions. *Methods* **153**, 13-21 (2019).

29. I. Shcherbakova *et al.*, Alternative spliceosome assembly pathways revealed by single-molecule fluorescence microscopy. *Cell Rep* **5**, 151-165 (2013).

30. N. R. Markham, M. Zuker, DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res* **33**, W577-581 (2005).

31. R. Rauhut *et al.*, Molecular architecture of the Saccharomyces cerevisiae activated spliceosome. *Science* **353**, 1399-1405 (2016).

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

33. A. Sidarovich *et al.*, Identification of a small molecule inhibitor that stalls splicing at an early step of spliceosome activation. *Elife* **6** (2017).

34. S. Chanarat, M. Seizl, K. Strässer, The Prp19 complex is a novel transcription elongation factor required for TREX occupancy at transcribed genes. *Genes Dev* **25**, 1147-1158 (2011).

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

36. J. A. Henriques, E. Moustacchi, Isolation and characterization of pso mutants sensitive to photo-addition of psoralen derivatives in Saccharomyces cerevisiae. *Genetics* **95**, 273-288 (1980).

37. W. Y. Tarn *et al.*, Functional association of essential splicing factor(s) with PRP19 in a protein complex. *EMBO J* **13**, 2421-2431 (1994).

38. M. D. Ohi *et al.*, Proteomics analysis reveals stable multiprotein complexes in both fission and budding yeasts containing Myb-related Cdc5p/Cef1p, novel pre-mRNA splicing factors, and snRNAs. *Mol Cell Biol* **22**, 2011-2024 (2002).

39. D. J. Field, J. D. Friesen, Functionally redundant interactions between U2 and U6 spliceosomal snRNAs. *Genes Dev* **10**, 489-501 (1996).

40. S. Ben-Yehuda *et al.*, Genetic and physical interactions between factors involved in both cell cycle progression and pre-mRNA splicing in Saccharomyces cerevisiae. *Genetics* **156**, 1503-1517 (2000).

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

42. J. Vogel, B. F. Luisi, Hfq and its constellation of RNA. *Nat Rev Microbiol* **9**, 578-589 (2011).

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

44. E. G. Anderson, A. A. Hoskins, Single molecule approaches for studying spliceosome assembly and catalysis. *Methods Mol Biol* **1126**, 217-241 (2014).

45. B. Séraphin, M. Rosbash, The yeast branchpoint sequence is not required for the formation of a stable U1 snRNA-pre-mRNA complex and is recognized in the absence of U2 snRNA. *EMBO J* **10**, 1209-1216 (1991).

46. L. J. Friedman, J. Chung, J. Gelles, Viewing dynamic assembly of molecular complexes by multi-wavelength single-molecule fluorescence. *Biophys J* **91**, 1023-1031 (2006).

47. L. J. Friedman, J. Gelles, Multi-wavelength single-molecule fluorescence analysis of transcription mechanisms. *Methods* **86**, 27-36 (2015).

Chapter Three

Transient Interaction of a Prp38, Snu23, and Spp381 Protein Subcomplex with the Spliceosome During Active Site Assembly

This chapter is not yet published.

Fu, X., Hoskins, A.A.

Author contributions:

XF and AAH conceived the project. XF carried out experiments. XF analyzed data. XF and AAH wrote the manuscript.

ABSTRACT

Spliceosome activation is the process of creating the catalytic site for pre-mRNA splicing. Yeast B complex proteins (BCP; Spp381, Prp38 and Snu23) are recruited to and released from the spliceosome during activation. Using Colocalization Single Molecular Spectroscopy (CoSMoS), we investigated the dynamics of BCP in real-time during spliceosome activation on single pre-mRNA molecules in real time. Kinetic analysis revealed that BCP components exhibited similar dynamic behaviors and that these were distinct from a different protein factor that is also released during activation, Snu66. To further understand activation dynamics, we monitored the dynamics of BCP components and Snu66 relative to two activation landmark events: release of U4 snRNP proteins and recruitment of NTC proteins. While Snu66 is recruited and released along with the U4 snRNP, BCP proteins are recruited after U4 binding and released after both U4 loss and arrival of the NineTeen Complex (NTC). These results indicate that the BCP components form a stable splicing factor subcomplex that is recruited to and released from the spliceosome at distinct times during splicing. Further, we show that ATP can regulate association of the BCP with the U4/U6.U5 tri-snRNP with it forming a spliceosome-independent complex with the tri-snRNP at low ATP but not under splicing conditions. In a sum, our results reveal the dynamics of the BCP and Snu66 during activation and allow us to propose a kinetic scheme for spliceosome active site formation *in vitro*.

INTRODUCTION

RNA splicing is the process of removing non-coding introns from precursor mRNAs (pre-mRNAs) concomitant with ligation of the flanking exons. The process is mediated by a megaDalton ribonucleoprotein (RNP) complex, the spliceosome. The components of spliceosomes can be broadly categorized as small nuclear ribonucleoprotein complexes (U1, U2, U4, U5, U6 snRNPs), protein only complexes (NineTeen Complex, NTC; NTC related complexes, NTR), and transiently interacting protein splicing factors (e.g., several DExD/H-box ATPases required for splicing). While spliceosome assembly is partially ordered, it is also a highly dynamic process (**Fig. 3.1A, 3.2**) (Hoskins et al., 2011; Shcherbakova et al., 2013; Wilkinson et al., 2020). In brief, spliceosome assembly begins by recruitment of the U1, U2, and U4/U6.U5 tri-snRNPs to pre-mRNA to form the pre-B complex spliceosome. This is then followed by spliceosome activation during which the nascent active site is formed, the U1 and U4 snRNPs are released, and the NTC and NTR complexes are recruited to form the B^{ACT} complex. Ultimately, large-scale conformational and compositional changes lead to the formation of a U2/U6 di-snRNA catalytic center associated with the intron and capable of catalyzing splicing chemistry. Upon completion of exon ligation, the spliceosome disassembles, and its components are recycled for processing of other pre-mRNAs.

During the activation of human spliceosomes, B-specific factors (PRPF38, MFAP1, ZMAT2, FBP21, SMU1, RED, NPW38, NPW38BP) are recruited during the transition from the pre-B to B complex spliceosome and released during the transition from the B to B^{ACT} complex (Bertram et al., 2017; Townsend et al., 2020; Zhang et al., 2018). The locations of these proteins are interesting in that they interact with the U6 snRNA/5' splice site (SS) duplex and a binding site previously occupied by the DEAD-box ATPase Prp28 which is involved in transfer of the 5' SS from the U1 to the U6 snRNA (Staley and Guthrie, 1999). Consequently, presence of the B-specific factors on the human spliceosome may reveal a great deal about the state of the

spliceosome including that activation has begun, the pre-B complex has been converted to the B complex, the U1 snRNP has dissociated, the 5' SS has been identified and transferred to U6, and Prp28 has departed.

Only PRPF38, ZMAT2 and MFAP1 are evolutionarily conserved between human and *S. cerevisiae* (hereafter yeast; Prp38, Snu23 and Spp381 respectively) (Ulrich and Wahl, 2017). However, the three yeast proteins have not been considered as step-specific factors that associate only during a certain stage of activation because they can also be isolated as part of the yeast U4/U6·U5 tri-snRNP in the absence of splicing (Gottschalk et al., 1999; Nguyen et al., 2016; Stevens and Abelson, 1999; Ulrich and Wahl, 2017; Wan et al., 2016). Given the evolutionarily conservation of the splicing machinery, the origins and consequences for this variation in association of Prp38, Snu23, and Spp381 with the tri-snRNP remains elusive as do the implications for 5' SS identification and Prp28 function in yeast (Ulrich and Wahl, 2017).

B complex proteins (BCPs; Prp38, Snu23 and Spp381) are all important for yeast viability with Snu23 and Prp38 being essential and loss of Spp381 affective vegetative growth in some strains and being lethal in others (Blanton et al., 1992; Lybarger et al., 1999; Stevens et al., 2001) (Fu and Hoskins, unpublished data). Though no structures of a yeast Prp38-Snu23-Spp381 co-complex have yet been solved, a crystal structure containing a ternary complex formed from fragments of *Chaetomium thermophilum* Prp38, Snu23, and MFAP1 has been obtained (Ulrich et al., 2016). This structure and accompanying biochemical data suggest that they can potentially function as a heterotrimeric subcomplex during splicing and associate with one another in the absence of the spliceosome. Further, the yeast proteins can form an isolable complex by gel filtration experiments (Ulrich and Wahl, 2017). Despite evidence for interactions between the three proteins, it is still unclear if they are recruited to and release from tri-snRNPs or spliceosomes individually or as a single subunit.

It is also unclear if other splicing factors known to be released from yeast spliceosomes during activation would coordinate their binding with the BCP. For example, Snu66 is considered a tri-snRNP specific factor in both human and yeast (Behrens and Lührmann, 1991; Gottschalk *et al.*, 1999; Stevens and Abelson, 1999). However, it is intrinsically disordered, and its structural interactions with the human and yeast tri-snRNPs are largely undefined; although, portions of Snu66 do directly contact two key regulators of the activation process: the Brr2 helicase that unwinds U4 from the U6 snRNA and the Prp8 protein that scaffolds the active site and regulate the activity of Brr2 and other factors (Agafonov et al., 2016; Nguyen *et al.*, 2016) (**Fig. 3.3**). Since both Snu66 and BCP have been predicted to be recruited to yeast spliceosomes as part of the tri-snRNP and released during the transition from the B to B^{ACT} complex, it is possible that BCPs could form a larger spliceosomal subcomplex together with Snu66 and coordinate their binding dynamics with one another during splicing.

Given that the transition from B to B^{ACT} spliceosome involves the exchange of about fifty factors (including the release of BCPs, Snu66, and the U4 snRNP and recruitment of the NTC and NTR), it is unsurprising that this conversion likely involves the formation of several intermediates (**Fig. 3.1A, 3.2**) (Plaschka et al., 2017). Two recent human pre-B^{ACT} spliceosome structures stalled during this transition using a small molecule splicing inhibitor revealed at least two of these potential intermediates (Townsend *et al.*, 2020). In the pre-B^{ACT1} spliceosome (predicted to form first from the B complex; **Fig. 3.2**), U4 snRNP and Snu66 have been released, BCPs are retained, and a subset of NTC proteins have associated and can be resolved in the structure. Similar activation intermediates have not yet been characterized by either biochemical or structural methods for the yeast spliceosome. Whether the activation pathways are conserved between human and yeast is unknown.

The highly compositional and conformational dynamic activation process renders the underlying mechanism intrinsically complicated and difficult to study. Even though cryo-EM

structures have provided a few snapshots of along the activation path, the number of kinetic steps and characteristic spliceosome complexes involved in each transition are challenging to predict. Previously, our lab has employed Colocalization Single Molecule Spectroscopy (CoSMoS) to study the dynamics of major factors (U4 snRNP proteins, NTC proteins and Lsm2-8 proteins) involved in the activation pathway (Hoskins et al., 2016) (see Chapter 2). Here, we measured the dynamics of yeast BCPs during activation and compared these to Snu66. We then ordered these binding and release events relative to two landmark events during activation: loss of the U4 snRNP and NTC recruitment. Our results show that a human-like activation pathway is conserved for the yeast splicing machinery with simultaneous binding and release of BCPs together as a subcomplex. Association of BCPs with the tri-snRNP is ATP-dependent with BCPs and tri-snRNP associating with pre-mRNA simultaneously only under low ATP conditions that inhibit splicing. This observation suggests that tri-snRNP composition may be regulated by ATP and explains differences in BCP association observed between yeast and human spliceosomes. Finally, we combine our single molecule data to propose a kinetic map for yeast spliceosome activation *in vitro*.

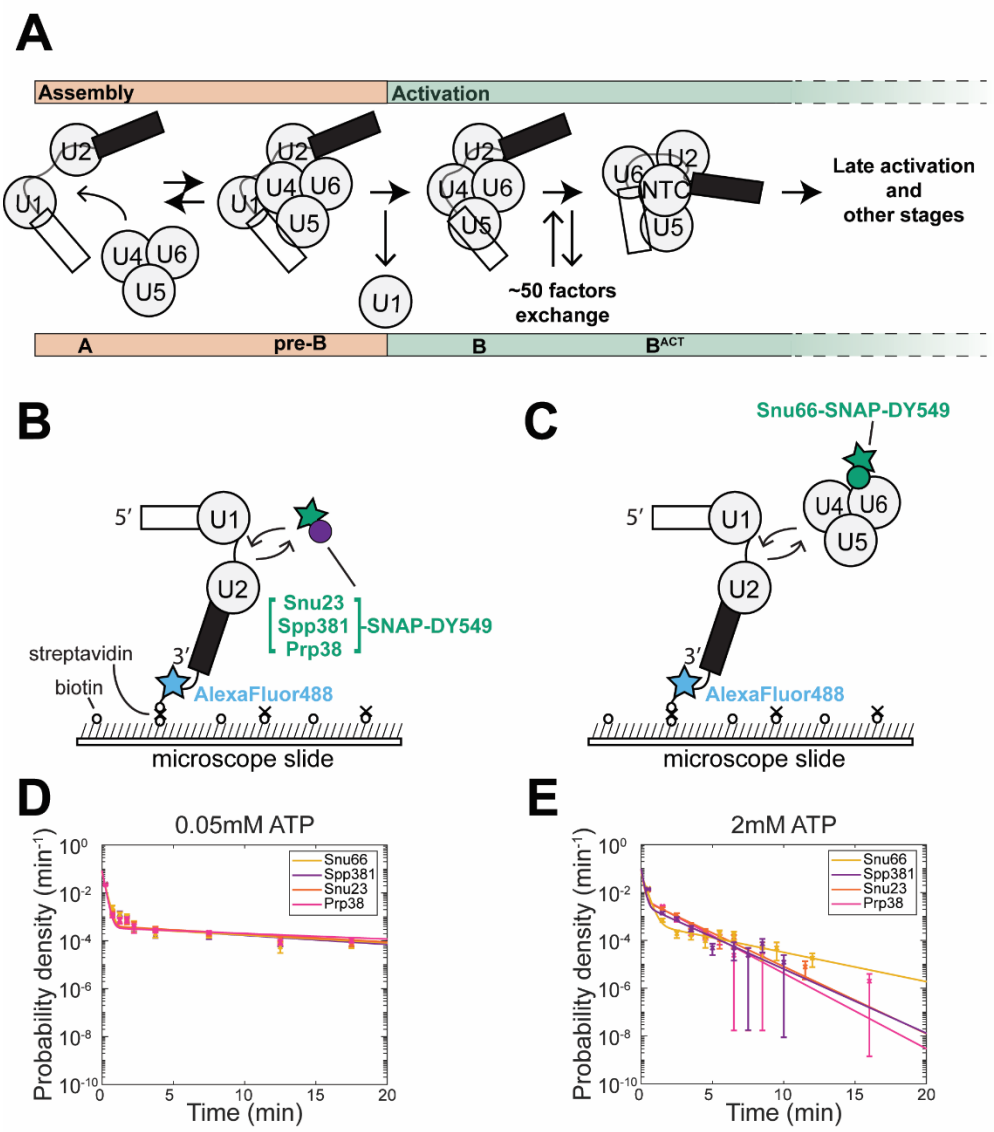


Figure 3.1 2-Color CoSMoS Assay to Study BCP and Snu66 Dynamics During Spliceosome Activation.

(A) The U4/U6.U5 tri-snRNP reversibly interacts with the spliceosome A complex consisting of U1 and U2 snRNPs associated with pre-mRNA. Once the pre-B complex is formed, ATP-dependent activation can proceed by first ejecting the U1 snRNP to form the B complex. This is then followed by exchange of ~50 protein and snRNA factors to form the activated B complex (B^{ACT} spliceosome). During these steps, the U4 snRNP and Lsm2-8 proteins are released and the

NTC joins the spliceosome. Subsequent steps then lead to formation of the spliceosome active site and splicing. (B-C) Schematic of 2-color CoSMoS experiments for observing a single BCP protein or Snu66 binding dynamics. A BCP or Snu66 was fluorescently labeled with a DY-549 fluorophore via a SNAP tag. The pre-mRNA was immobilized to the slide surface and contained an Alexa 488 fluorophore. (D-E) Probability density histogram for dwell times of BCP and Snu66 at 0.05mM ATP (D) or at 2mM ATP (E).

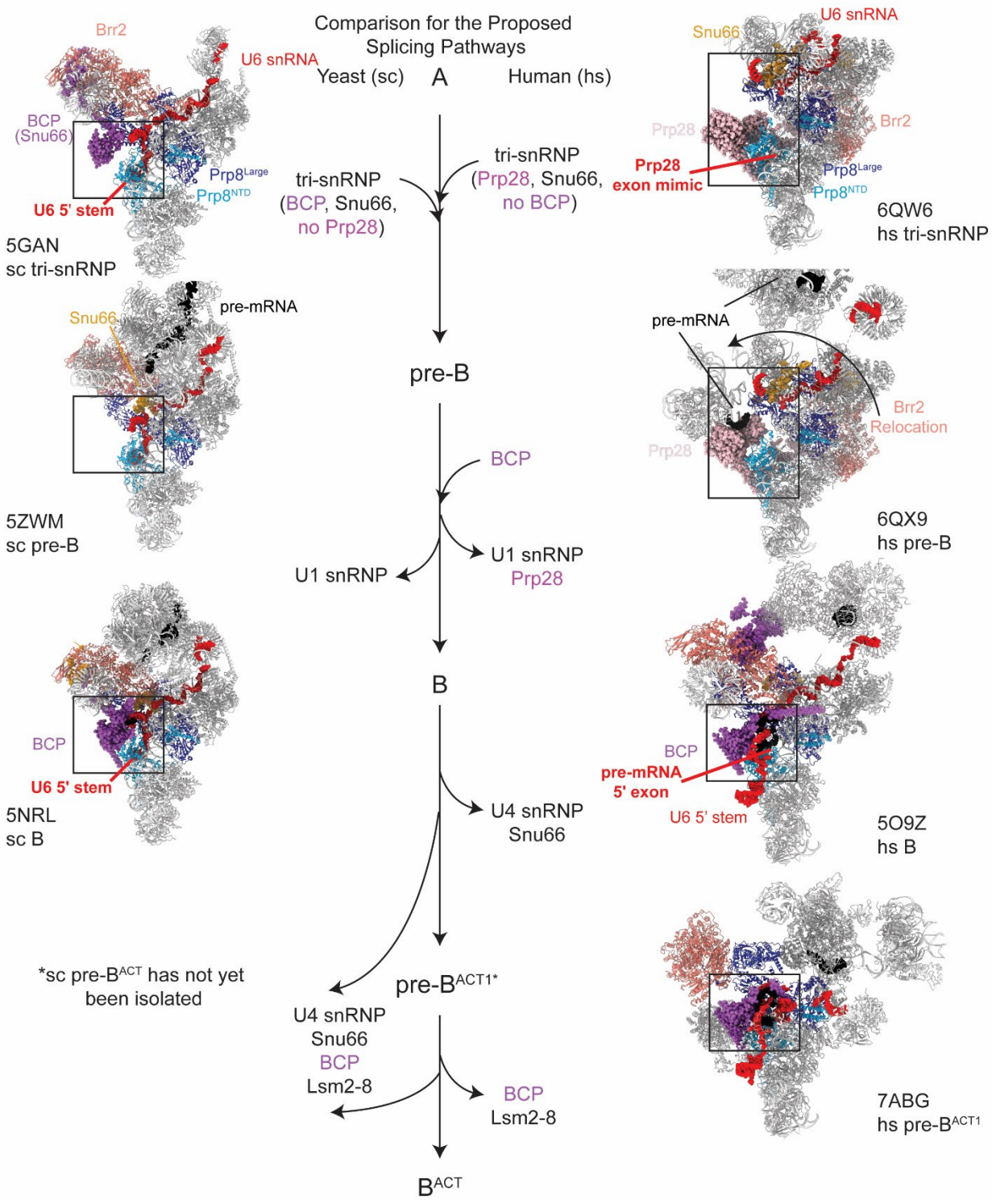

Comparison for the Proposed Splicing Pathways
Yeast (sc)
A
Human (hs)
tri-snRNP (BCP, Snu66, no Prp28)
tri-snRNP (Prp28, Snu66, no BCP)
pre-B
BCP
U1 snRNP
U1 snRNP Prp28
B
U4 snRNP Snu66
pre-B^ACT1*
U4 snRNP Snu66 BCP Lsm2-8
BCP Lsm2-8
B^ACT
*sc pre-B^ACT has not yet been isolated
Brr2
U6 snRNA
BCP (Snu66)
Prp8^Large
Prp8^NTD
U6 5' stem
5GAN
sc tri-snRNP
Snu66
pre-mRNA
5ZWM
sc pre-B
BCP
U6 5' stem
5NRL
sc B
Snu66
U6 snRNA
Prp28
Brr2
Prp28 exon mimic
Prp8^Large
Prp8^NTD
6QW6
hs tri-snRNP
pre-mRNA
Brr2 Relocation
Prp28
6QX9
hs pre-B
BCP
pre-mRNA 5' exon
U6 5' stem
5O9Z
hs B
7ABG
hs pre-B^ACT1

Figure 3.2 Proposed Schemes for Spliceosome Activation in Yeast and Humans based on cryo-EM Structures.

Conformational and compositional changes occurring during spliceosome activation beginning with the yeast (Left Column) and human U4/U6.U5 tri-snRNP (Right Column) structures on top and ending with the yeast B spliceosome and human pre-B^{ACT1} spliceosome structures at the bottom. Middle column represents a comparison of simplified linear model between yeast (to the left of the line) and human (to the right of the line) activation, with relevant compositions highlighted. The box in each structure is centered around the supposed binding sites of BCPs in spliceosomes. The binding site locates at the gate of the exon channel, formed between $Prp8^{NTD}$ and $Prp8^{Large}$. The occupancy of the exon channel and binding factors locatedat the gate change as activation progresses. All human and yeast structures are aligned in orientation to the $Prp8^{NTD}$ (residue 1 to 885 for yeast; residue 1 to 812 for human).

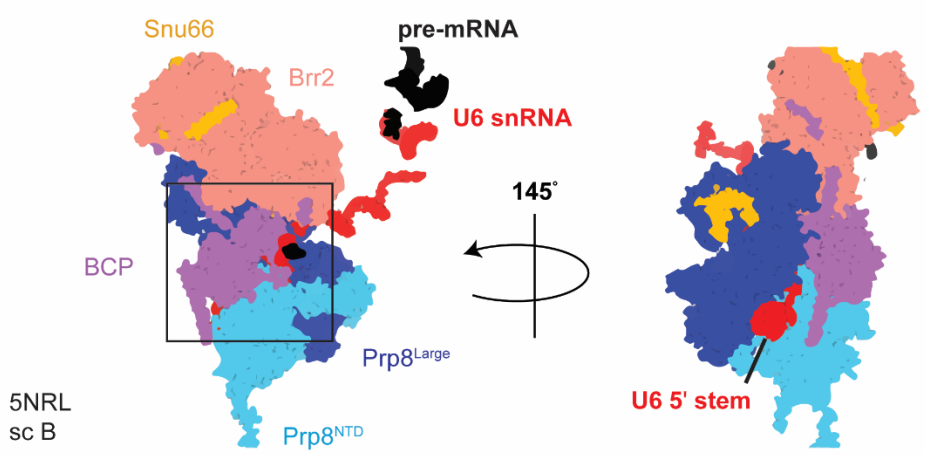


Figure 3.3 Observed Interactions between Yeast BCP and Snu66 in the B Complex Spliceosome.

The BCP is located at the gate of the exon channel. The view on the left shows BCP interacting with the U6/5' exon RNA duplex at the entrance of the channel. The view on the right is rotated by 145° to show occupancy of the channel opening by the U6 5' stem. Snu66 (yellow) is only partially resolved in this structure but contacts Brr2 and the large central domain of Prp8. This model was derived from PDB ID 5NRL.

MATERIAL AND METHODS

Yeast Strains

Yeast strains (**Table 3.1**) were derived from the protease-deficient strain BJ2168 and splicing factors were c-terminally tagged by integrating fast SNAP or DHFR tags at the appropriate genomic locations by homologous recombination as previously described (Hoskins *et al.*, 2011; Hoskins *et al.*, 2016).

Yeast Growth Assay

Selected yeast strains were grown in YPD media to saturation in a 30ºC shaking incubator. Concentrated yeast cultures were then diluted down to 0.03 with YPD media in wells of a Corning Costar 48-well Flat Bottom Cell Culture plate. The 48-well plate without lid was then covered with a Breathe-Easy plate sealing membrane to limit evaporation during incubation. A Tecan Infinite 200Pro plate reader was used to monitor the yeast growth for 30 h at 30 ºC.

Preparation of Yeast WCE

Yeast WCE for splicing was prepared according to published protocols using the ball method (Mayas et al., 2006). SNAP-tagged proteins were fluorophore labeled as previously described (Anderson and Hoskins, 2014). Briefly, SNAP-Surface® 549 dye (S9112S, New England BioLabs; abbreviated elsewhere as SNAP-DY-549) in DMSO was added to 1.2 mL yWCE to a final concentration of 1 µM. The reaction tube was rotated in the dark for 30 min at room temperature. The reaction was then loaded onto a pre-equilibrated G-25 Sephadex column (Kontes Flex Column) in SEC buffer (25 mM HEPES-KOH pH 7.9, 50 mM KCl, 1 mM DTT, 10% v/v glycerol) at 4ºC to remove excess dye. A Peristaltic Pump P-1 (Cytiva) was used for pre-equilibrating the column and eluting the labeled extract at a flow rate of ~0.25 mL/min with the SEC buffer. Fluorophore labeling of the proteins was confirmed by SDS-PAGE and fluorescence

using a Typhoon FLA 9000 scanner (Cytiva) at 532nm. Results were analyzed with ImageQuantTL (Cytiva) software.

In Vitro Splicing Assays

[α-^{32}P] UTP radio-labeled (PerkinElmer) and m^7G(5′)ppp(5′)G capped (New England Biolabs) RP51A pre-mRNA substrates were made by *in vitro* transcription of a linear DNA template with T7 RNA polymerase (Agilent or purified in the laboratory). The DNA template was produced from a PCR reaction of pBS117 plasmid (Séraphin and Rosbash, 1991) using Taq DNA polymerase (M7122, Promega), followed by gel purification of the products with a Wizard SV Gel and PCR Clean-Up System kit (Promega). Transcription products were separated on a 6% denaturing polyacrylamide gel (AccuGel 19:1, National Diagnostics), followed by ethanol precipitation of the extracted RP51A transcripts and quantitation of the resuspended transcripts in nuclease-free water (Ambion, Fischer Scientific) with a liquid scintillation counter (LSC, tri-carb 2900TR, Packard).

A typical *in vitro* splicing reaction included 40% v/v WCE and 0.2 nM radio-labeled RP51A substrate in a splicing buffer [final concentrations: 100mM KPi pH 7.3, 3% w/v PEG-8000, 1mM DTT, 2.5mM $MgCl_2$, 0.2U/μL RNasin Plus (Promega)]. The reaction was incubated at room temperature for 45 min. The reaction was quenched in a splicing dilution buffer (100mM Tris base pH 7.5, 10mM EDTA pH 8.0, 1% w/v SDS, 150mM NaCl, 300 mM NaOAc pH 5.3). RNAs from the reaction were extracted using phenol-chloroform, ethanol precipitated, resuspended in deionized formamide, and separated on a 12% denaturing polyacrylamide gel. Gels were then dried and exposed to a phosphor screen overnight. The screen was imaged with a Typhoon FLA 9000 scanner (Cytiva). Results were analyzed with ImageQuantTL (Cytiva) software. The intensities (I) for RP51A pre-mRNAs and splicing products were determined by integrating the signals within same-sized rectangles around the band. The background corrected intensities for bands were then used for calculating splicing efficiencies, with following equations:

$$1st\ step\ efficiency: \frac{I(mRNA) + I(intron\ lariat - 3'exon)}{I(pre-mRNA) + I(mRNA) + I(intron\ lariat - 3'exon)}$$

$$2nd\ step\ efficiency: \frac{I(mRNA)}{I(pre-mRNA) + I(mRNA) + I(intron\ lariat - 3'exon)}$$

Preparation of Fluorescently-Labeled RP51A pre-mRNAs

Unlabeled biotin handles (See Chapter 2, Oligo Table) from IDT were dissolved in 0.091M $Na_2B_4O_7$ buffer (NaB buffer, pH 8.50, adjusted with concentrated HCl) at a concentration of 1 mM. Alexa Fluor 488 NHS ester (A20000, Thermo Fisher Scientific) was prepared by dissolving to a concentration of 1 mg in 60 µL of anhydrous DMSO. For fluorophore labeling of the biotin handle, 15 µL of the fluorophore was mixed with 49 µL of the biotin handle and 36 µL NaB buffer. The reaction was kept in the dark and rotated overnight at room temperature. The labeling reaction mixture was then applied to a G-25 MicroSpin column (45-001-397, Fisher Scientific) to remove most of the unreacted dye. The eluate from the spin column was then loaded onto a 19% denaturing polyacrylamide gel (AccuGel 19:1, National Diagnostics). Fluorescent bands containing the oligo were excised from the gel, the oligo eluted from gel, and then ethanol precipitated. The products were resuspended in water and concentrations determined by UV-Vis.

Capped RP51A transcripts were made by *in vitro* transcription using T7 RNA polymerase and the biotin handle attached by splinted RNA ligation. The fluorescently labeled oligo (56 pmol) was 5'-phosphorylated using T4 PNK (5U, M0201S, New England Biolabs) according to the manufacturer's directions with the addition of 2 mM ATP. The PNK enzyme was then heat inactivated and DNA bridge oligos (50 pmol; See Chapter 2, Oligo Table) and radio-labeled RP51A transcript (14 pmol) was added and then allowed to anneal by incubating at 90°C for 3 min, followed by cooling using a gradient of 0.1°C/s over 10 min. T4 RNA Ligase 2 (20U; M0239L, New England Biolabs) was then added to ligate the transcript to the biotinylated oligo. Ligation products were separated on a 5% denaturing polyacrylamide gel (AccuGel 19:1, National

Diagnostics), followed by ethanol precipitation of the extracted products from the gel. Precipitated RNAs were resuspended in nuclease-free water (Ambion, Fischer Scientific) and quantified with a liquid scintillation counter (LSC, tri-carb 2900TR, Packard).

CoSMoS Assays

Microscope slides (100490-396, VWR) and cover glasses (12-553-455, Fischer) were cleaned by sonication while immersing in the following solutions/solvents in the order listed: f 2% v/v micro-90, 100% ethanol, 1M KOH, and MiliQ water. Each sonication step took 1 h, followed by rinsing with MiliQ water between steps. After drying with high purity nitrogen (NI UHP300, Airgas), the slides were aminosilanized with VECTABOND (NC9280699, Fisher Scientific). Reaction chambers were created by drawing multiple parallel straight lines on the slides with vacuum grease, followed by covering with the cover glasses to create individual lanes for reaction chambers. Typically, four lanes can be made on a single slide. Lastly, the slides were passivated by addition of a mixture of mPEG-SVA (MPEG-SVA-5K, Laysan Bio) and mPEG-biotin-SVA (BIO-PEG-SVA-5K, Laysan Bio) at a ratio of 1:100 w/w in 100mM $NaHCO_3$ (pH 8) buffer and incubating overnight.

Prior to each CoSMoS assay, the mPEG mixture in the lane (~20 µL) was washed off with 1x PBS (200 µL) twice. Then, streptavidin (50 µL, 0.2 mg/mL, SA10, Kelowna International Scientific) was flowed in the lane, allowing it to bind with the biotin on the slide surface. After washing the lane off with 200 µL 1x PBS, 50 µL 50-150 pM fluorescently labeled and biotinylated RNA was added and allowed to bind with the streptavidin. The slide was washed again with PBS and the density of RNA was determined by exciting the fluorophore with the 488 nm laser. Finally, splicing assay buffer containing 40% v/v WCE, 20 nM Cy5-TMP (Hoskins *et al.*, 2011), oxygen scavenger (5 mM PCA and 0.96 U/mL PCD), 1 mM Trolox, and triplet quenchers (0.5 mM propyl gallate, 1 mM cyclooctatetraene, 1 mM 4-nitrobenzyl alcohol) was added (Hoskins *et al.*, 2011). The triplet quenchers were added as a mixture made as a 100x stock in DMSO.

CoSMoS Data Acquisition

A custom-built objective-type TIRF microscope (Friedman et al., 2006; Larson *et al.*, 2014) together with Glimpse software (written in LabVIEW programming language, https://github.com/gelles-brandeis/Glimpse) was used for collecting single molecule movies. Briefly, a 60X 1.45(NA) PlanApo objective (Olympus) was used for the TIRF excitation and collecting the emission light. Different wavelengths of incoming and exit excitation light were directed to and away from the objective by two separate micromirrors. The emission light was further filtered and separated by a dichroic mirror at the cutoff wavelength of 635 nm. The spectrally separated light was then focused onto two different area (FOV: field of view) of the same EMCCD camera. Four different lasers were used in this study, including 488 nm (blue), 532 nm (green), 633 nm (red) and 785 nm (infrared), the powers of which were set in the ranges of 1.2 – 1.5 mW, 500 – 600 μW, 400-440 μW, and 2.5 mW, respectively. Cycles of time-lapse imaging were used according to the following excitation scheme with each frame lasting 1s: in each cycle, the 785 nm was first used to illuminate the sample and correct sample positioning using an auto-focus system. Then, the 488 nm blue laser was turned on to collect two consecutive frames to image the immobilized RNAs. The 532 and 633 nm lasers were then turned on to simultaneously collect 14 frames with a 3 s delay between adjacent frames. The total cycle time was ~1 min and this cycle was repeated 50x to collect videos lasting for ~50 min. To avoid photo bleaching of DY-549 and Cy5 fluorophores by the 488 nm laser, the path of the laser was physically blocked to prevent the beam from reaching the sample after collection of 10 frames of blue images.

CoSMoS Data Analysis

Single molecule data were analyzed by the same method as described previously (Hoskins *et al.*, 2011; Kaur et al., 2019) with slight modifications, using a custom program imscroll (https://github.com/gelles-brandeis/CoSMoS_Analysis) written in MATLAB (The Mathworks). A

general analysis procedure includes, 1) creating a mapping file for correlating the locations in the <635 nm FOV to those in the >635 nm FOV, using a fluorescent beads (T10711, ThermoFisher); 2) creating a driftlist file for correcting the drift of the fluorescent spots over the recording; 3) creating a AOI (area of interest) file listing the positions of the immobilized pre-mRNA molecules; 4) combining the AOI, mapping, and driftlist files to generate drift-corrected AOIs corresponding to the pre-mRNA locations in the fields of view used for DY-549 and Cy5-TMP imaging; and 5) integrating the measured intensities in the AOIs in those fields of view over the experimental time course (Friedman and Gelles, 2015). Binding events were identified as signals centered within the AOI that appeared with intensities greater than 3.6x the standard deviation above the mean of the background noise. Loss of signals were identified as points in time at which the signal fell below 1x standard deviation above the mean background.

Analysis of the measured dwell times and fits to kinetic equations were carried out using MATLAB and AGATHA software (https://github.com/hoskinslab/AGATHA) using maximum likelihood methods and fitting to equations containing one (Eq. 1), two (Eq. 2), or three (Eq. 3) exponential terms as described (Kaur *et al.*, 2019).

$$\frac{1}{\left(e^{-tm/\tau} - e^{-tx/\tau}\right)} * \left(\frac{1}{\tau} e^{-\frac{intervals}{\tau}}\right) \qquad (Eq.\,1)$$

$$\frac{1}{a\left(e^{-tm/\tau 1} - e^{-tx/\tau 1}\right) + (1-a)\left(e^{-tm/\tau 2} - e^{-tx/\tau 2}\right)} * \left(\frac{a}{\tau 1} e^{-\frac{intervals}{\tau 1}} + \frac{1-a}{\tau 2} e^{-\frac{intervals}{\tau 2}}\right) \qquad (Eq.\,2)$$

$$\frac{1}{a1\left(e^{-tm/\tau1}-e^{-tx/\tau1}\right)+a2\left(e^{-tm/\tau2}-e^{-tx/\tau2}\right)+(1-a1-a2)\left(e^{-tm/\tau3}-e^{-tx/\tau3}\right)}$$

$$*\left(\frac{a1}{\tau1}e^{-\frac{intervals}{\tau1}}+\frac{a2}{\tau2}e^{-\frac{intervals}{\tau2}}+\frac{1-a1-a2}{\tau3}e^{-\frac{intervals}{\tau3}}\right) \qquad (Eq.3)$$

Bootstrapping was used to calculate standard errors for all fitted parameters. Fit parameters are included in **Table 3.1** (Kaur *et al.*, 2019).

Histograms for the distribution of events were generated in MATLAB with empty bins removed. The error (standard deviation) for each bin were calculated using Eq. 4, with the assumption that the number of events within a bin follows a binomial distribution.

$$\sigma = \sqrt{NP(1-P)} \qquad (Eq.4)$$

Table 3.1 Yeast Strains

Strain #	ID	Genotype	Notes
1	yAAH0001	MATa prc1-407 prb1-1122 pep4-3 leu2 trp1 ura3-52 gal2	BJ2168, parental strain
2	yAAH0018	yAAH0001 + cef1::cef1-DHFR-HPH + ntc90::ntc90-DHFR-BLE	contains two DHFR-tagged NTC proteins with hygromycin and phleomycin resistance markers
3	yAAH0329	yAAH0001 + prp3::prp3-DHFR-HPH + prp4::prp4-DHFR-BLE	contains two DHFR-tagged U4 snRNP proteins with hygromycin and phleomycin resistance markers
4	yAAH1708	yAAH0329 + snu66::snu66-SNAP$_f$-NAT	contains two DHFR-tagged U4 snRNP proteins and fast-SNAP-tagged snu66 with nourseothricin resistance marker
5	yAAH0367	yAAH0329 + snu23::snu23-SNAP$_f$-KanMX	contains two DHFR-tagged U4 snRNP proteins and fast-SNAP-tagged snu23 with kanamycin resistance marker
6	yAAH0372	yAAH0329 + prp38::prp38-SNAP$_f$-NAT	contains two DHFR-tagged U4 snRNP proteins and fast-SNAP-tagged prp38 with nourseothricin resistance marker
7	yAAH0376	yAAH0329 + spp381::spp381-SNAP$_f$-NAT	contains two DHFR-tagged U4 snRNP proteins and fast-SNAP-tagged spp381 with nourseothricin resistance marker
8	yAAH2425	yAAH0001 + prp38::prp38-SNAP$_f$-NAT + spp381::spp381-DHFR-HPH + snu23::snu23-DHFR-BLE	contains two DHFR-tagged B complex proteins with hygromycin and phleomycin resistance markers and fast-SNAP-tagged prp38 with nourseothricin resistance marker
9	yAAH2424	yAAH0373 + spp381::spp381-DHFR-HPH + snu23::snu23-DHFR-BLE	contains two DHFR-tagged B complex proteins with hygromycin and phleomycin resistance markers and fast-SNAP-tagged snu66 with nourseothricin resistance marker
10	yAAH2381	yAAH0018 + snu23::snu23-SNAP$_f$-NAT	contains two DHFR-tagged NTC proteins and fast-SNAP-tagged snu23 with nourseothricin resistance marker
11	yAAH2380	yAAH0018 + prp38::prp38-SNAP$_f$-NAT	contains two DHFR-tagged NTC proteins and fast-SNAP-tagged prp38 with nourseothricin resistance marker
12	yAAH1710	yAAH0018 + snu66::snu66-SNAP$_f$-NAT	contains two DHFR-tagged NTC proteins and fast-SNAP-tagged snu66 with nourseothricin resistance marker
13	yAAH0373	yAAH0001 + snu66::snu66-SNAP$_f$-NAT	contains a fast-SNAP-tagged snu66 with nourseothricin resistance marker

RESULTS

BCPs and Snu66 Have Similar pre-mRNA Binding Kinetics at Low but not High ATP

To be able to measure the dynamics of different proteins under splicing conditions, we genetically encoded c-terminal fast SNAP tags on each of the three BCPs (Spp381, Snu23, and Prp38) and Snu66 in four different strains (**Table 3.1**)**.** We confirmed that these SNAP-tagged proteins can be readily labeled with benzylguanine-fluorophores (SNAP-DY549, 532nm laser (green) excitable) in yeast whole cell extracts (WCEs) and that the labeled extracts efficiently spliced RP51A substrate pre-mRNAs (**Fig. 3.4, 3.5**). We do note, however, that strains containing tagged Prp38 grew more slowly (**Fig. 3.5B**) and produced WCE with lower splicing activity (**Fig. 3.5A**) relative to others. It is possible that tagging of Prp38 disrupts expression of its neighboring gene, SMD1 (a component of the U snRNP Sm ring), and that this contributes to the differences in growth and splicing activity (Rymond, 1993).

To observe splicing factor dynamics, we flowed a pre-mixed solution consisting of a WCE containing DY549-labeled splicing factors and splicing buffer into a reaction chamber containing surface-immobilized, Alexa Fluor 488-labeled (488nm laser (blue) excitable) RP51A pre-mRNAs (**Fig. 3.1B, C**). We carried out the experiments at two ATP concentrations that either permit spliceosome activation and splicing (2mM ATP) or inhibit activation and splicing but permit assembly (0.05mM ATP) (Tarn et al., 1993). Using a custom-built CoSMoS microscope (Larson et al., 2014), we could observe binding and unbinding dynamics of fluorescently labeled proteins on the fixed fluorescent pre-mRNA molecules in real time.

Both BCPs and Snu66 showed ATP-dependent dynamics. At 0.05mM ATP, the dwell times of each protein to single pre-mRNA molecules often lasted for tens of minutes (**Fig. 3.6**). In comparison, the binding events at 2 mM ATP were much shorter and lasted from a few seconds to a few minutes. These results resemble single molecule data obtained from other proteins

involved in spliceosome activation, such as Prp3/4 from U4 snRNP complexes (Hoskins *et al.*, 2016) and Lsm8 from the Lsm2-8 ring complex (see Chapter 2). It is likely that the long dwell times observed at 0.05 mM ATP originated from stalled spliceosome complexes while the shorter times seen at 2 mM ATP originated from transient binding due to release of these factors during spliceosome activation.

We further analyzed the binding dynamics by fitting the collected, unbinned dwell times to exponential-based functions (**Table 3.2**) and generating probability density histograms. At low ATP, histograms and fits for all three BCP members and Snu66 are very similar to one another (**Fig. 3.1D**, **Table 3.2**). In contrast, under conditions that permit activation at 2mM ATP, the distributions and fits for the BCPs remained similar to one another but were distinct from those of Snu66 (**Fig. 3.1E, Table 3.2**). Together, these results indicate that BCPs interact with the splicing machinery with similar kinetic features. These features are shared with another factor released during the B to B^{ACT} transition (Snu66) only under low ATP conditions that inhibit spliceosome activation and splicing.

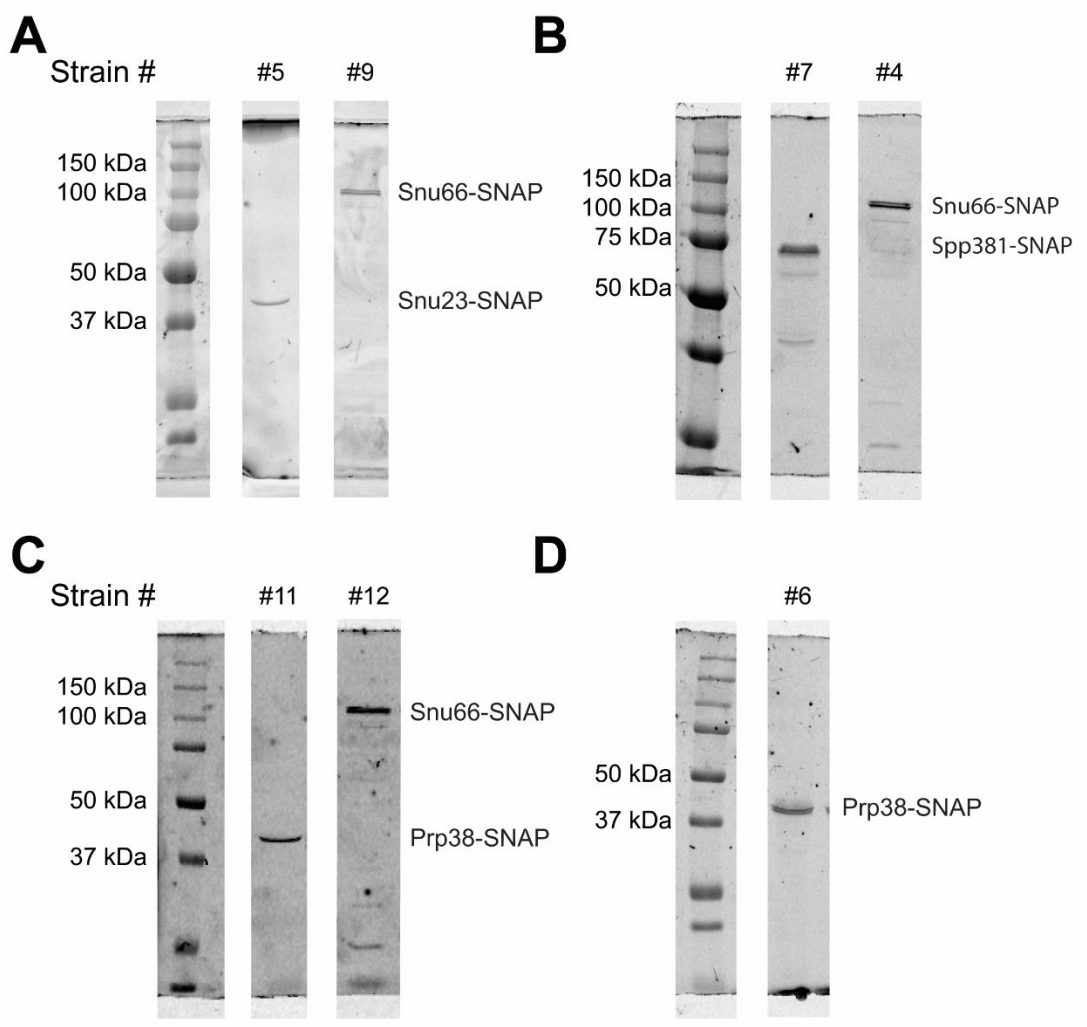


Figure 3.4 Fluorophore Labeling of Prp38, Snu23, Snu66, and Spp381 in Various SNAP-Tagged Proteins Can Be Specifically Labeled in Yeast WCE with SNAP-DY-549.

Shown are fluorescence images of SDS-PAGE gel for confirming successful labeling. The gel images were cropped (vertical white stripes) to remove intervening lanes not relevant to this figure. In our assays, Snu66 sometimes is observed as a doublet. Anomalous migration of Snu66 by SDS-PAGE has previously been reported (migrating as a ~80 kDa protein rather than its expected MW of 66 kDa) (Stevens and Abelson, 1999). In addition, we observe that Spp381 also migrates at a larger than expected MW. It was previously shown that a Spp381-HA tag fusion migrates as a ~51 kDa species compared to its expected MW of ~34 kDa (Lybarger *et al.*, 1999).

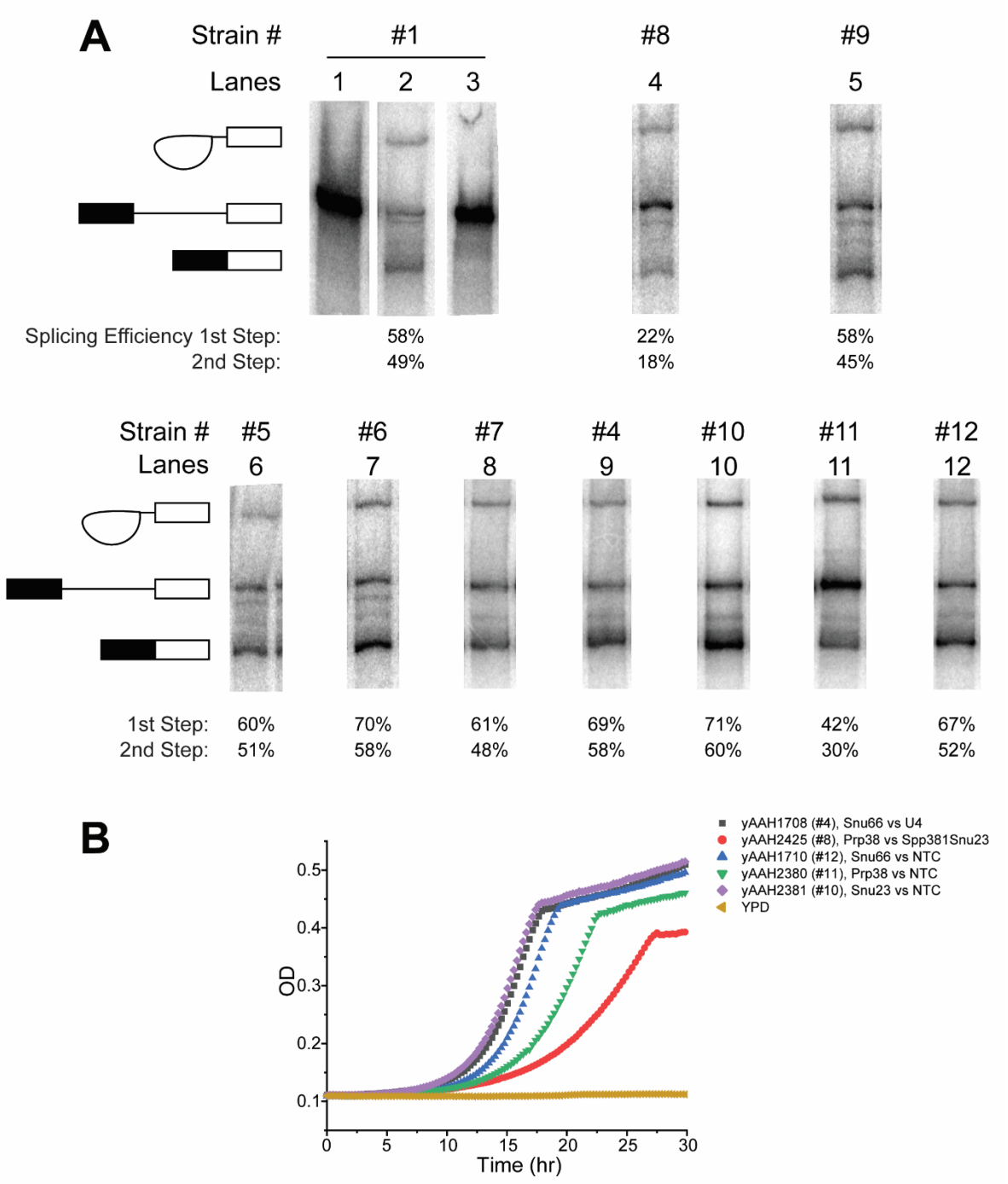
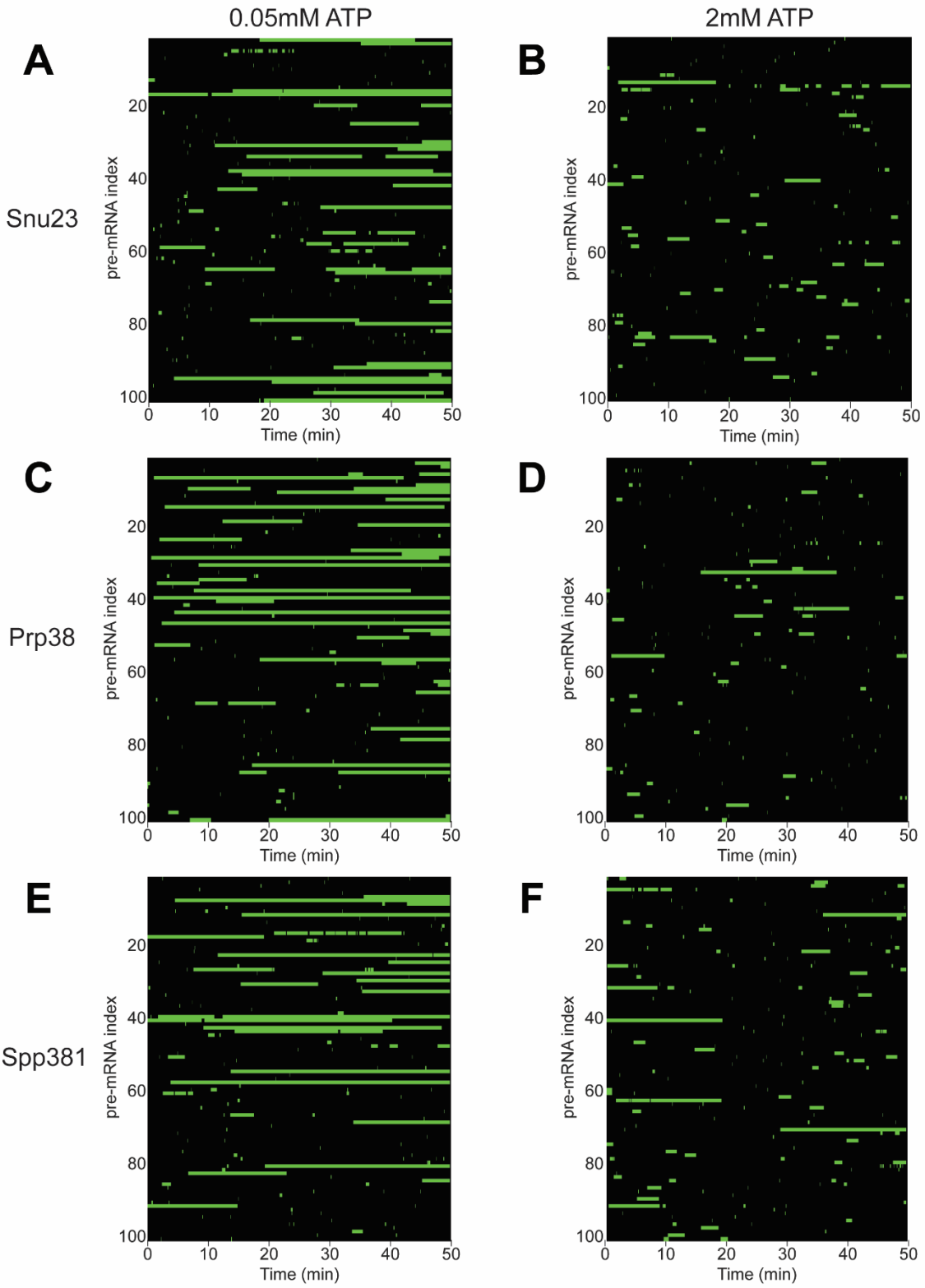


Figure 3.5 Splicing Activities of Dy549-Labeled WCEs.

(A) Representative splicing assays and efficiencies for WCE containing fluorophore-labeled proteins (lanes 4-12) or a WT control (lanes 1-3). In lanes 4-12, ATP was added to a concentration of 2 mM to promote splicing. In lane 3, ATP was added to a concentration of 0.05 mM to permit

spliceosome assembly but prevent activation and splicing. Lanes 1 represents a sample at t=0 min while the other lanes represent reactions quenched at 45 min. Cartoons for splicing products are indicated to the left of the gel. The gel images were cropped (vertical white stripes) and combined into a single figure to remove intervening lanes not relevant for this figure and for clarity. (B) Yeast growth assay results for the indicated strains. Strains containing labels on Spp381grow more slowly than others.


0.05mM ATP
2mM ATP
A
B
C
D
E
F
Snu23
Prp38
Spp381
pre-mRNA index
Time (min)
20
40
60
80
100
0
10
20
30
40
50

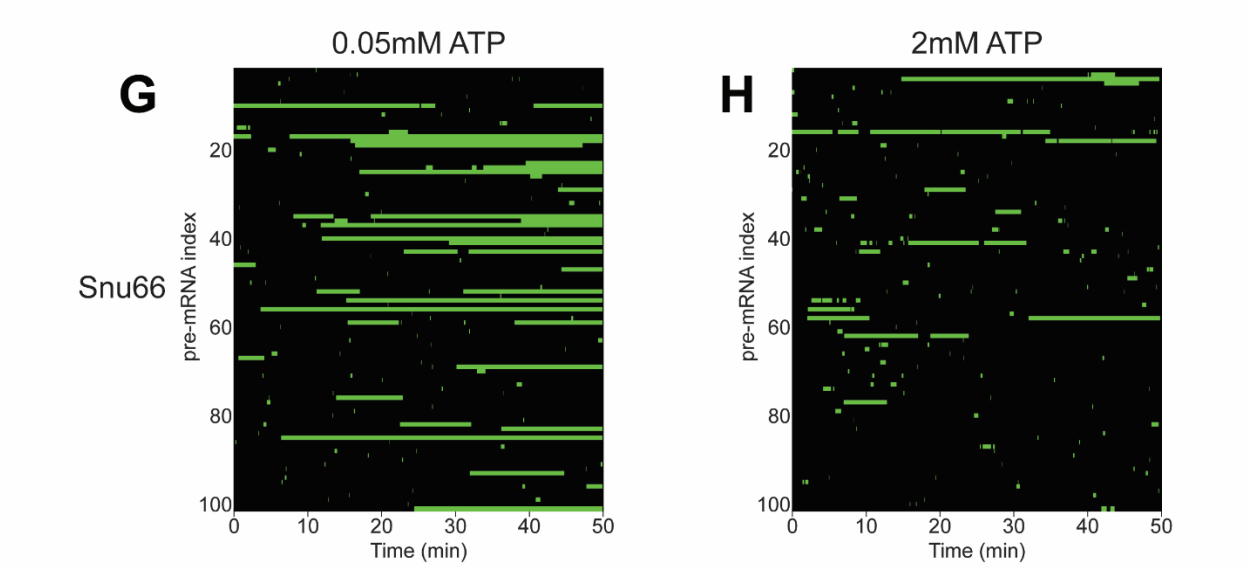


Figure 3.6 ATP-Dependent Binding Intervals of BCP and Snu66 on Single pre-mRNA Molecules.

Shown are binding intervals of fluorescently labeled BCP and Snu66 molecules on single pre-mRNAs under conditions that inhibit (0.05 mM ATP, left columns) or permit activation and splicing (2 mM ATP, right columns). Within each panel, each row represents association and dissociation of of fluorescent molecules on an individual pre-mRNA molecule over 50 min with green and black bars representing bound/fluorescent and unbound/nonfluorescent states, respectively.

Table 3.2 Fit parameters for binding dynamics of proteins from 2-color CoSMoS experiments

Tagged protein	Tagged Subcomplex	Strain	[ATP] mM	a_1	τ_1 (s)	a_2	τ_2 (s)	a_3	τ_3 (s)	Number of Events	Related Figure
Snu66-SNAP	tri-snRNP	yAAH1708	2	0.89±0.02	15.2±0.8	0.11±0.02	213.0±26.0			688	Fig. 3.1
Snu23-SNAP	BCP	yAAH0367	2	0.54±0.03	6.9±0.5	0.46±0.03	92.9±8.3			471	Fig. 3.1
Spp381-SNAP	BCP	yAAH0376	2	0.72±0.02	7.9±0.4	0.28±0.02	97.2±9.2			684	Fig. 3.1
Prp38-SNAP	BCP	yAAH0372	2	0.53±0.05	6.9±0.4	0.47±0.03	82.7±7.5			697	Fig. 3.1
Snu66-SNAP	tri-snRNP	yAAH1708	0.05	0.63±0.05	7.8±0.7	0.23±0.05	45.7±14.8	0.14±0.07	435.8±80.0	385	Fig. 3.1
Snu23-SNAP	BCP	yAAH0367	0.05	0.77±0.02	8.1±0.5	0.07±0.03	77.4±49.6	0.16±0.04	640.0±132.2	514	Fig. 3.1
Spp381-SNAP	BCP	yAAH0376	0.05	0.74±0.03	7.6±0.6	0.17±0.03	61.9±21.6	0.09±0.04	908.3±645.1	356	Fig. 3.1
Prp38-SNAP	BCP	yAAH0372	0.05	0.79±0.02	8.6±0.6	0.21±0.02	616.7±87.1			466	Fig. 3.1

BCPs are Recruited to and Released from Spliceosomes Simultaneously

To see if the similarity in BCP kinetics is coincidental or results from their presence together in a single spliceosomal subcomplex, we carried out experiments to visualize the recruitment and release of BCPs relative to one another during splicing. We carried out 3-color CoSMoS assays using WCE containing Prp38-SNAP labeled with SNAP-DY549 fluorophores and DHFR tags on Spp381 and Snu23 labeled with Cy5-TMP fluorophores (633nm (red) laser-excitable) and surface immobilized Alexa Fluor 488-labeled pre-mRNAs (**Fig. 3.7A**).

We identified events in which green signals from Prp38 and red signals from Spp381 and Snu23 were colocalized on the immobilized RNAs based on the assumption that at some point during spliceosome activation Prp38, Snu23, and Spp381 should all be simultaneously present (**Fig. 3.7B, 3.8**). We note that not all green signals showed a corresponding red signal and vice versa. This may be due to incomplete fluorophore labeling of BCP factors or partially assembled BCP subcomplexes. These uncolocalized signals were not analyzed further.

For the subset of signals showing colocalization, the binding patterns of the paired events were then classified based on their relative recruitment and release orders (**Fig. 3.9**). For example, we determined if the green signal appeared or disappeared before, after, or simultaneously with the red signal for each colocalized pair. We found that most of the event pairs (91%; 102/112) showed simultaneous arrival of Prp38-SNAP and Spp381/Snu23-DHFR signals (**Fig. 3.7C**). Among this subset, 77% showed simultaneous release of all three proteins (**Fig. 3.7D**). This suggests that the roles of BCPs in the spliceosome are temporally coupled. While it is possible that BCPs individually bind and dissociate too quickly one after the other for our measurements to detect, our data is most consistent with structural and gel filtration data (Bertram *et al.*, 2017; Plaschka *et al.*, 2017; Townsend *et al.*, 2020; Ulrich *et al.*, 2016) supporting formation of a BCP ternary complex in the absence of a spliceosome. This protein-only subcomplex binds to and releases from the spliceosome as a single unit.

We next used the measured dwell times for the colocalized event pairs to determine the lifetime of the spliceosome complex containing the BCP (the B_{BCP} complex, **Fig. 3.7E**). The distribution of lifetime measurements could be described by a function containing a single exponential term, resulting in a characteristic lifetime of ~41 s for B_{BCP} (**Fig. 3.7F, Table 3.3**). Since splicing *in vitro* occurs over tens of minutes, transformation of the B_{BCP} complex occurs rapidly enough to support splicing. The lifetime of the B_{BCP} complex is unlikely to be limited by photobleaching since we observe much longer (>10-fold) lifetimes under low ATP conditions which block spliceosome activation (**Table 3.2, 3.3**).

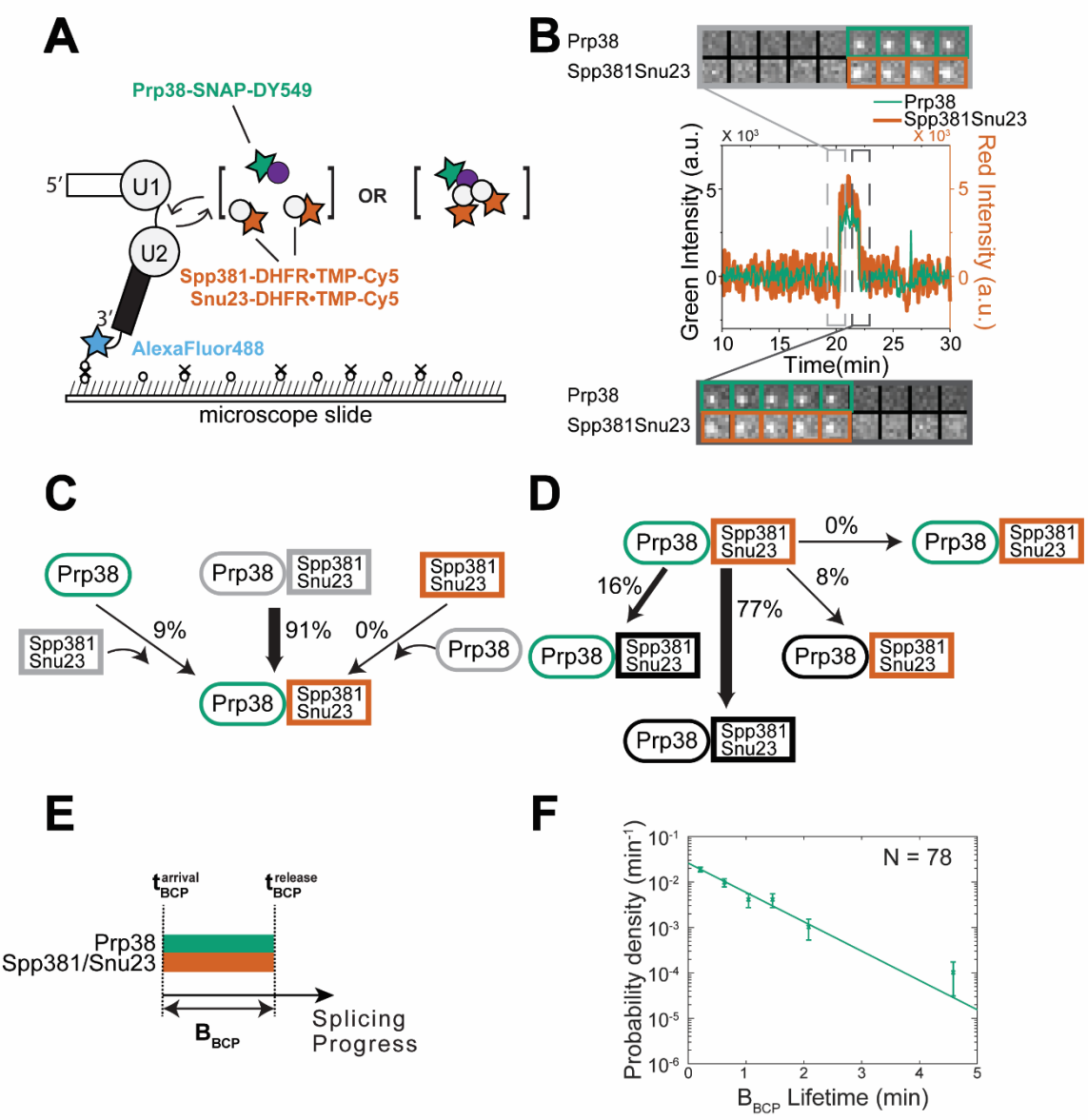


Figure 3.7 3-color CoSMoS Detection of a BCP Subcomplex.

(A) Schematic of a 3-color experiment in which Spp381 and Snu23 were labeled with Cy5-TMP fluorophores, Prp38 was labeled with a DY-549 fluorophore, and the surface-tethered pre-mRNA was labeled with an Alex Fluor 488 fluorophore. (B) Segment of a representative time record showing peaks in fluorescence intensity corresponding to colocalization of Spp381 and Snu23 (red, thick line) and Prp38 proteins (green, thin line) with the same individual pre-mRNA molecule. The dashed rectangles mark examples of the simultaneous appearance and disappearance of

Spp381/Snu23 and Prp38 spots; galleries show consecutive images taken from the indicated part of the recording showing that both spot appearance and disappearance are simultaneous. (C) Routes for recruitment of either the Prp38 or Spp381 and Snu23 fluorescent spots at 2 mM ATP for *N*=112 pairs of overlapping events. Green and red shapes represent observation of fluorescence from the corresponding DY-549 or Cy5 fluorophores on Prp38 or Spp381 and Snu23, respectively; grey shapes represent the absence of fluorescence. Percentages represent the fraction of complexes in which fluorescence appeared by the indicated pathway; more prevalent pathways are emphasized with thicker arrows. (D) Routes for loss of either the Prp38 or Spp381 and Snu23 fluorescent spots at 2 mM ATP for *N*=102 pairs of overlapping events in which the spots also appeared simultaneously on the pre-mRNAs. Black shapes represent the disappearance of fluorescence. (E) Schematic showing the relationship of the B_{BCP} complex lifetime to the arrival and departure times of BCP proteins. (F) Probability density histograms of $\mathbf{B_{BCP}}$ (panel E, $\mathbf{t_{BCP}^{release} - t_{BCP}^{arrival}}$) complex lifetimes obtained from the subset of events (*N*) showing the predominant recruitment and release times. The line represents the fit of the lifetime distribution with an equation containing a single exponential term as reported in Table 3.3. Error bars (SD) were calculated for each point as described in the Methods.

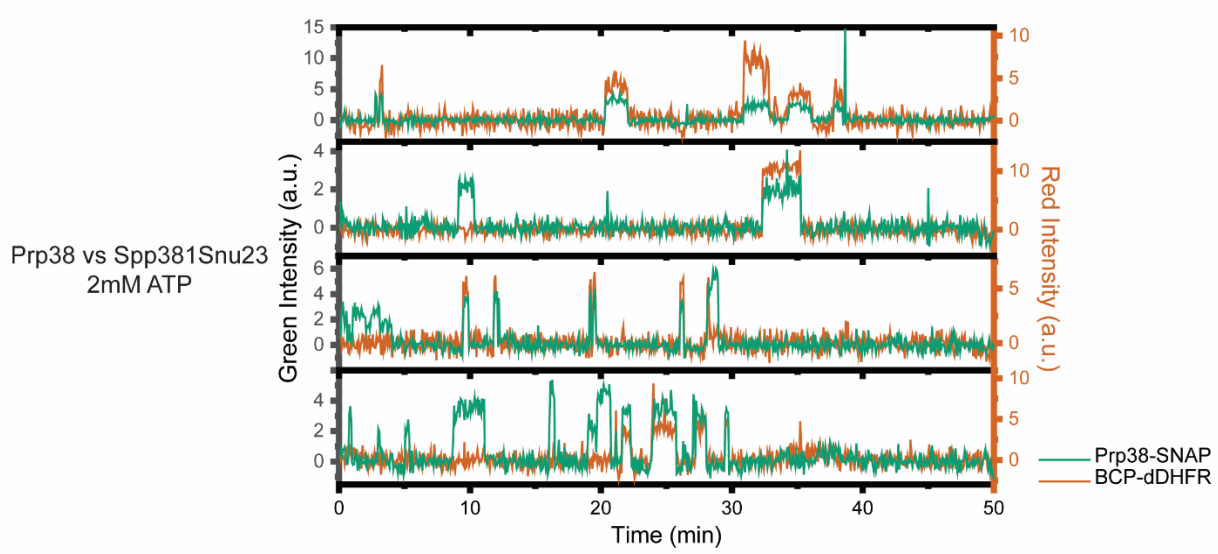


Figure 3.8 Additional Sample Fluorescence Trajectories from 3-color CoSMoS Experiments Monitoring Prp38-SNAP and Spp381- and Snu23-DHFR Proteins.

Shown are super-imposed fluorescence intensities for Prp38-SNAP (green traces) and Spp381-DHFR and Snu23-DHFR proteins (red traces) at 2 mM ATP.

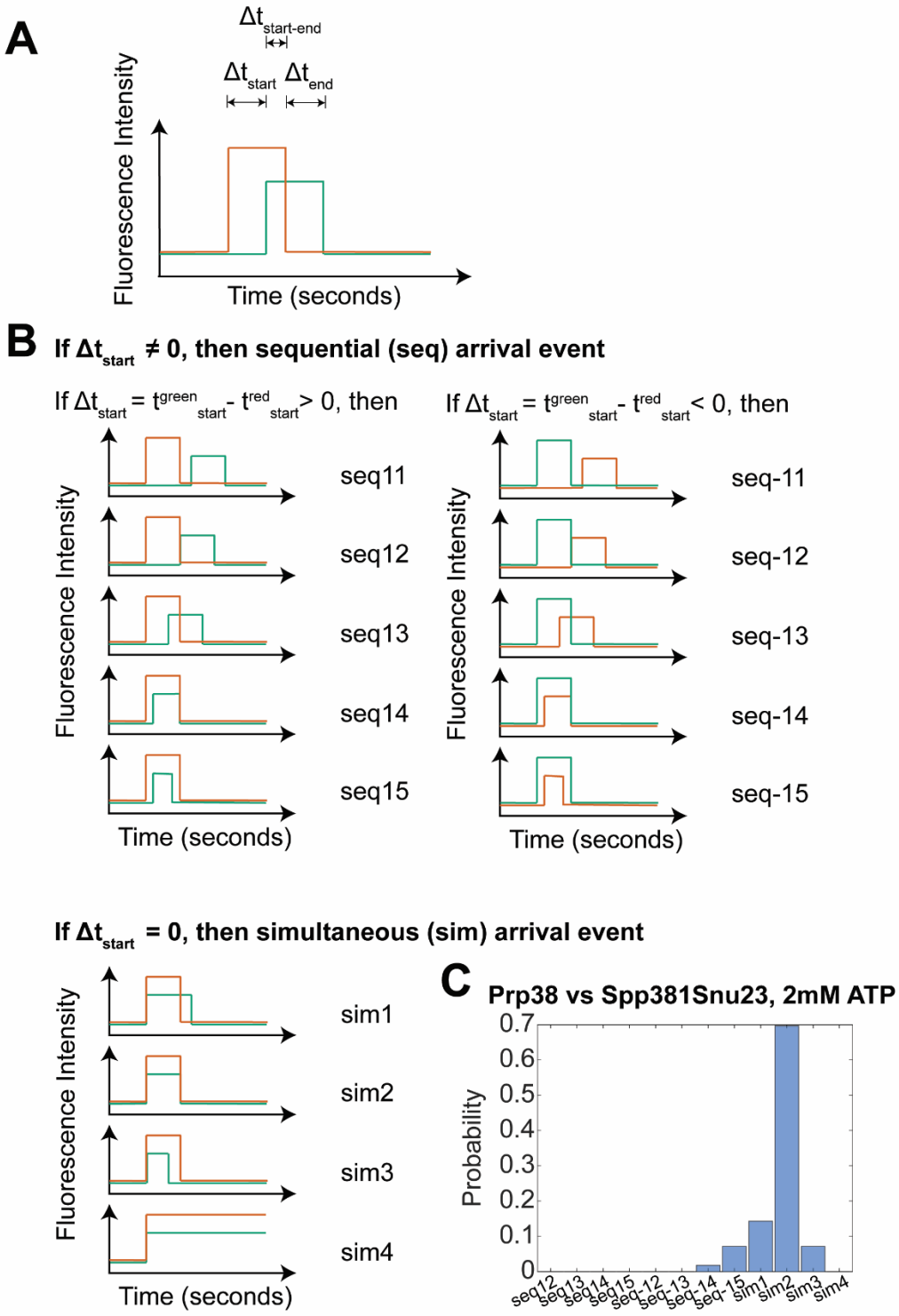

A
Δt start-end
Δt start
Δt end
Fluorescence Intensity
Time (seconds)
B
If Δt start ≠ 0, then sequential (seq) arrival event
If Δt start = t green start - t red start > 0, then
If Δt start = t green start - t red start < 0, then
Fluorescence Intensity
seq11
seq12
seq13
seq14
seq15
Time (seconds)
seq-11
seq-12
seq-13
seq-14
seq-15
Time (seconds)
If Δt start = 0, then simultaneous (sim) arrival event
Fluorescence Intensity
sim1
sim2
sim3
sim4
Time (seconds)
C
Prp38 vs Spp381Snu23, 2mM ATP
Probability
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
seq12
seq13
seq14
seq15
seq-12
seq-13
seq-14
seq-15
sim1
sim2
sim3
sim4

Figure 3.9 Scheme for assignment of observed binding and dissociation patterns in CoSMoS assays.

(A) For each set of paired events, the beginning and end times of the individual events were recorded. For the analysis shown here, Δt_{start} values were first calculated by subtracting the start time of the green event from the start time of the red event. (B) Δt_{start} values were separated such that all positive values indicated that the red event arrived before the green event (seq11 to 15), negative values indicated that the green event arrived after the red event (seq-11 to -15), and zero values indicated simultaneous arrival (sim1 to 4). Values of Δt_{end} and $\Delta t_{start-end}$ (A) were then determined to subcategorize each arrival class based on the disappearance patterns of the green and red events. (C) Distribution of Prp38-SNAP and Snu23-DHFR and Spp381-DHFR signal classes from overlapped binding events.

Table 3.3 Fitted Kinetic Parameters from 3-color CoSMoS experiments

Tagged protein	Tagged Subcomplex	Complex	Strain	[ATP] mM	a_1	τ_1 (s)	a_2	τ_2 (s)	a_3	τ_3 (s)	Number of Events	Related Figure
Snu66-SNAP	tri-snRNP	--	yAAH1708	2	0.88±0.03	18.1±1.2	0.11±0.03	133.2±23.4	NA	NA	446	Fig. 3.14
Prp3/Prp4-DHFR	U4 snRNP	--	yAAH1708	2	0.76±0.14	13.9±2.6	0.23±0.13	59.8±19.6	0.01	477.9±183.3	283	Fig. 3.14
Spp381-SNAP	BCP	--	yAAH0376	2	0.64±0.04	8.6±0.7	0.33±0.04	67.5±9.7	0.03	479.3±249.3	290	Fig. 3.11-12
Prp3/Prp4-DHFR	U4 snRNP	--	yAAH0376	2	0.85±0.08	16.7±2.2	0.15±0.08	104.4±32.8	NA	NA	164	Fig. 3.11-12
Snu23-SNAP	BCP	--	yAAH0367	2	0.53±0.06	7.0±0.8	0.47±0.06	100.6±18.2	NA	NA	135	Fig. 3.10
Prp3/Prp4-DHFR	U4 snRNP	--	yAAH0367	2	0.45±0.07	6.5±0.8	0.41±0.10	53.6±13.4	0.13	259.2±92.9	114	Fig. 3.10
Prp38-SNAP	BCP	--	yAAH0372	2	0.23±0.05	5.3±0.4	0.72±0.05	55.4±5.3	0.05	359.8±122.5	216	Fig. 3.11-12
Prp3/Prp4-DHFR	U4 snRNP	--	yAAH0372	2	0.56±0.24	20.9±9.1	0.44±0.24	70.1±56.3	NA	NA	209	Fig. 3.11-12
Snu66-SNAP	tri-snRNP	--	yAAH1708	0.05	0.67±0.03	10.1±1.1	0.33±0.03	381.5±50.6	NA	NA	281	Fig. 3.18
Prp3/Prp4-DHFR	U4 snRNP	--	yAAH1708	0.05	0.51±0.05	13.2±3.0	0.49±0.05	401.9±57.0	NA	NA	167	Fig. 3.18
Spp381-SNAP	BCP	--	yAAH0376	0.05	0.78±0.03	11.2±1.4	0.22±0.03	380.8±84.6	NA	NA	255	Fig. 3.18
Prp3/Prp4-DHFR	U4 snRNP	--	yAAH0376	0.05	0.56±0.05	8.0±1.6	0.44±0.05	331.2±56.3	NA	NA	133	Fig. 3.18
Snu23-SNAP	BCP	--	yAAH0367	0.05	0.54±0.04	10.1±1.6	0.46±0.04	395.2±52.6	NA	NA	249	Fig. 3.17
Prp3/Prp4-DHFR	U4 snRNP	--	yAAH0367	0.05	0.45±0.05	12.2±3.4	0.55±0.05	371.5±48.7	NA	NA	217	Fig. 3.17
Prp38-SNAP	BCP	--	yAAH0372	0.05	0.63±0.05	12.5±2.6	0.37±0.05	545.7±95.1	NA	NA	158	Fig. 3.18
Prp3/Prp4-DHFR	U4 snRNP	--	yAAH0372	0.05	0.51±0.04	9.4±1.0	0.49±0.04	516.1±65.4	NA	NA	186	Fig. 3.18
Prp38-SNAP	BCP	--	yAAH2425	2	0.44±0.05	8.5±0.9	0.51±0.05	49.0±4.6	0.05±0.07	435.8±11818	445	Fig. 3.7
Spp381/Snu23-DHFR	BCP	--	yAAH2425	2	0.42±0.08	8.8±1.6	0.52±0.08	41.4±6.2	0.06±0.11	332.9±86.2	250	Fig. 3.7
Snu66-SNAP	tri-snRNP	--	yAAH2424	2	0.59±0.03	12.2±1.1	0.41±0.09	190.0±29.4	NA	NA	708	Fig. 3.16
Spp381/Snu23-DHFR	BCP	--	yAAH2424	2	0.56±0.06	7.9±2.6	0.44±0.06	128.2±81.7	NA	NA	290	Fig. 3.16

Snu23-SNAP	BCP	--	yAAH2381	2	0.44±0.03	8.2±0.5	0.48±0.03	69.5±5.1	0.08±0.04	370.8±61.0	721	Fig. 3.19
Syf1/Cef1-DHFR	NTC	--	yAAH2381	2	0.65±0.02	7.9±0.5	0.35±0.02	338.7±22.2			485	Fig. 3.19
--	--	B_{BCP}	yAAH2425	2	NA	40.6±5.8	NA	NA	NA	NA	78	Fig. 3.7F
--	--	$B_{U4}^{\Delta Spp381}$	yAAH0376	2	NA	47.0±22.8	NA	NA	NA	NA	25	Fig. 3.13
--	--	$B_{Spp381+U4}$	yAAH0376	2	NA	10.8±2.5	NA	NA	NA	NA	25	Fig. 3.13
--	--	$B_{Spp381}^{\Delta U4}$	yAAH0376	2	NA	114.0±47.8	NA	NA	NA	NA	25	Fig. 3.13
--	--	$B_{U4}^{\Delta Snu23}$	yAAH0367	2	NA	80.5±29.1	NA	NA	NA	NA	25	Fig. 3.10
--	--	$^{*}B_{Snu23+U4}$	yAAH0367	2	NA	18.2	NA	NA	NA	NA	25	Fig. 3.10
--	--	$B_{Snu23}^{\Delta U4}$	yAAH0367	2	NA	108.7±31.2	NA	NA	NA	NA	25	Fig. 3.10
--	--	$B_{U4}^{\Delta Prp38}$	yAAH0362	2	NA	43.4±9.9	NA	NA	NA	NA	51	Fig. 3.13
--	--	$B_{Prp38+U4}$	yAAH0362	2	NA	8.5±1.4	NA	NA	NA	NA	51	Fig. 3.13
--	--	$B_{Prp38}^{\Delta U4}$	yAAH0362	2	NA	81.4±24.2	NA	NA	NA	NA	51	Fig. 3.13
--	--	$B_{Snu66}^{\Delta BCP}$	yAAH2424	2	NA	172.5±35.7	NA	NA	NA	NA	39	Fig. 3.16
--	--	$B_{Snu66+BCP}$	yAAH2424	2	NA	12.4±2.2	NA	NA	NA	NA	39	Fig. 3.16
--	--	$B_{BCP}^{\Delta Snu66}$	yAAH2424	2	NA	87.9±17.6	NA	NA	NA	NA	39	Fig. 3.16
--	--	$B_{Snu23}^{\Delta NTC}$	yAAH2381	2	NA	65.5±8.8	NA	NA	NA	NA	69	Fig. 3.19G
--	--	$B_{Snu23+NTC}$	yAAH2381	2	NA	52.7±7.2	NA	NA	NA	NA	69	Fig. 3.19H
--	--	$B_{NTC}^{\Delta Snu23}$	yAAH2381	2	NA	409.0±47.9	NA	NA	NA	NA	69	Fig. 3.19

*Lifetime of $B_{Snu23+U4}$ cannot be fitted by exponential-based functions and was estimated by averaging the observed data.

Conservation of BCP Recruitment During Activation between Yeast and Human Spliceosomes

We next studied when the BCP subcomplex is recruited to the spliceosome and when it is released during activation relative to the U4 snRNP. The U4 snRNP is a core component of the tri-snRNP and thus reports on tri-snRNP binding to the pre-mRNA. Release of U4 snRNP is a key event during the B to B^{ACT} complex transition since it permits formation of the U2/U6 di-snRNA catalytic site. We carried out 3-color CoSMoS assays using extracts containing SNAP-tagged Snu23 and DHFR-tagged U4 snRNP proteins (Prp3 and Prp4) along with surface-immobilized Alexa Fluor 488 labeled pre-mRNAs (**Fig. 3.10A, 3.11A**). We identified all paired binding events in which fluorescent signals from the U4 snRNP proteins were simultaneously present on the pre-mRNA with signals from Snu23. We found that most of the event pairs (89%; 31/35 events) showed sequential arrival. The U4 snRNP protein signals appeared first followed by appearance of Snu23, suggesting that Snu23 and the BCP are not recruited as part of the tri-snRNP (**Fig. 3.10B, C, 3.12A**). Consistently, the same recruitment pattern was also found when extracts containing SNAP-tagged Spp381 or Prp38 were used along with DHFR-tagged U4 snRNP proteins (**Fig. 3.11B-C, 3.12B-E**). Together these results indicate that a BCP subcomplex containing Prp38, Snu23, and Spp381 is recruited to the spliceosome after the tri-snRNP associates. These observations are consistent with schemes for activation involving ordered addition of the tri-snRNP and BCP complex to the spliceosome proposed for the human splicing machinery and by cryo-EM analysis of endogenous yeast spliceosome complexes isolated in the presence of a dominant negative mutant of the splicing factor Prp28 (Bai et al., 2018; Bertram *et al.*, 2017; Gottschalk *et al.*, 1999; Lybarger *et al.*, 1999; Stevens and Abelson, 1999; Xie et al., 1998).

We further analyzed the order of BCP and U4 snRNP release for pairs of colocalized events. In the case of the BCP protein Snu23, the majority of event pairs (83%) showed loss of

signals from the U4 snRNP proteins occurring before loss of signals from Snu23 (**Fig. 3.10D, 3.12A**). This was also the predominant pattern for signal loss from Spp381 and Prp38 when WCEs containing those labeled proteins were used (**Fig. 3.12A-C, F, G**). Together, the results indicate that during activation the BCP subcomplex is recruited to the spliceosome after the tri-snRNP associates and that the BCP dissociates as a single unit after release of the U4 snRNP. Ordered release of U4 and then the BCP is consistent with structural data for B complex spliceosomes showing that the BCP has no structural contacts with U4 snRNP proteins that could be modeled (Bertram *et al.*, 2017; Plaschka *et al.*, 2017). Thus, it is feasible that their release could be coordinated as separate events. This result is also consistent with a previously determined structure of the human pre-B^{ACT1} activation intermediate that contains human BCP but not the U4 snRNP (**Fig. 3.2E**) (Townsend *et al.*, 2020). The yeast spliceosome appears to share a conserved pathway with humans for recruitment and release of BCP proteins during activation.

The single molecule data also indicate the existence of at least three types of spliceosome B complexes: $\mathrm{B}_{\mathrm{U4}}^{\Delta\mathrm{BCP}}$ in which the tri-snRNP has associated with the pre-mRNA but the BCP has not yet joined, $\mathrm{B}_{\mathrm{U4+BCP}}$ in which the BCP has bound but U4 has not been released, and $\mathrm{B}_{\mathrm{BCP}}^{\Delta\mathrm{U4}}$ in which the U4 snRNP has been released and the BCP has not yet dissociated (**Fig 3.10E**). We were able to determine the characteristic lifetimes of each complex to be ~57s, ~13s, and ~101s for the $\mathrm{B}_{\mathrm{U4}}^{\Delta\mathrm{BCP}}$, $\mathrm{B}_{\mathrm{U4+BCP}}$,and $\mathrm{B}_{\mathrm{BCP}}^{\Delta\mathrm{U4}}$ complexes, respectively by averaging results obtained for each BCP component in independent CoSMoS experiments (**Fig 3.13, Table 3.3**). The $\mathrm{B}_{\mathrm{U4+BCP}}$ intermediate that we observed at 2 mM ATP compositionally resembles the pre-catalytic B complex spliceosome that was observed at 0.05 mM ATP by cryo-EM (Plaschka *et al.*, 2017). Our data suggest that the lifetime of this complex at higher ATP concentrations is very short—often only a few seconds—before U4 snRNP release is completed. To our knowledge, the yeast $\mathrm{B}_{\mathrm{BCP}}^{\Delta\mathrm{U4}}$ complex, while having a longer lifetime in our experiments than $\mathrm{B}_{\mathrm{U4+BCP}}$, has not yet been been biochemically or structurally characterized.

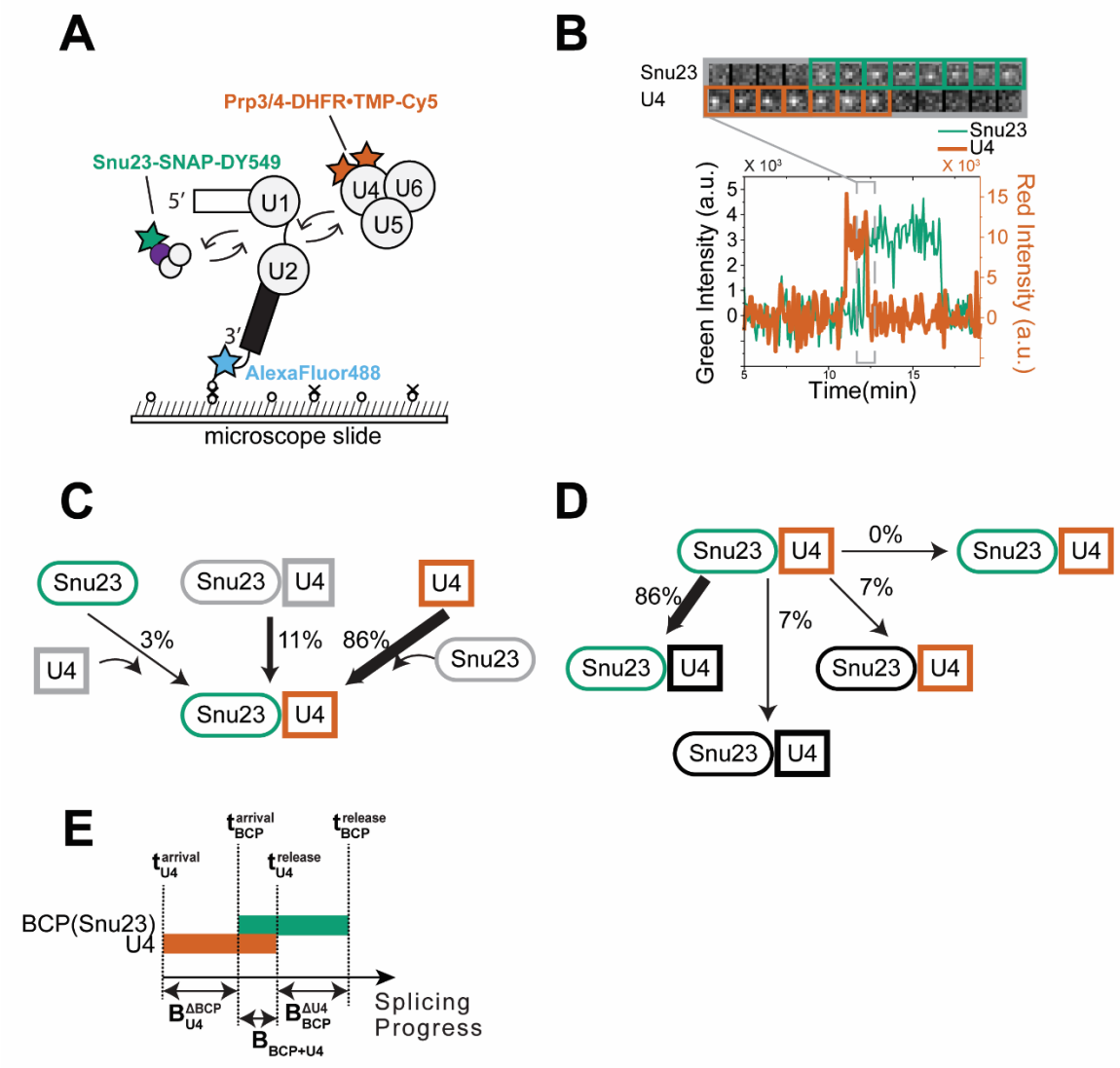


Figure 3.10 The BCP Binds After the tri-snRNP and Releases After U4 Dissociation.

(A) Schematic of a 3-color experiment in which U4 was labeled with Cy5-TMP fluorophores, Snu23 was labeled with a DY-549 fluorophore, and the surface-tethered pre-mRNA was labeled with an Alexa Fluor 488 fluorophore. (B) Segment of a representative time record showing peaks in fluorescence intensity corresponding to colocalization of U4 (red, thick line) and Snu23 proteins (green, thin line) with the same individual pre-mRNA molecule. The light grey dashed rectangle marks an example of the ordered recruitment of Snu23 followed by the release of U4 spots; galleries show consecutive images taken from that part of the recording showing the ordered spot

appearance and disapperance. (C) Routes for the appearance of Snu23 and U4 fluorescent spots at 2 mM ATP for *N*=35 pairs of overlapping events. Red and green shapes represent observation of fluorescence from the corresponding DY-549 or Cy5 fluorophores on Snu23 or U4, respectively. Percentages represent the fraction of Snu23/U4 complexes in which fluorescence appeared by the indicated pathway; more prevalent pathways are emphasized with thicker arrows. (D) Routes for loss of either Snu23 or U4 fluorescent spots at 2 mM ATP for *N*=30 pairs of overlapping events in which the U4 spot appearance preceded arrival of Snu23. (E) Schematic showing the definition for identified intermediate based on the relative recruitment and release times of Snu23 and U4 proteins. Data for Spp381 and Snu23 are included in **Fig. 3.12-3.13**.

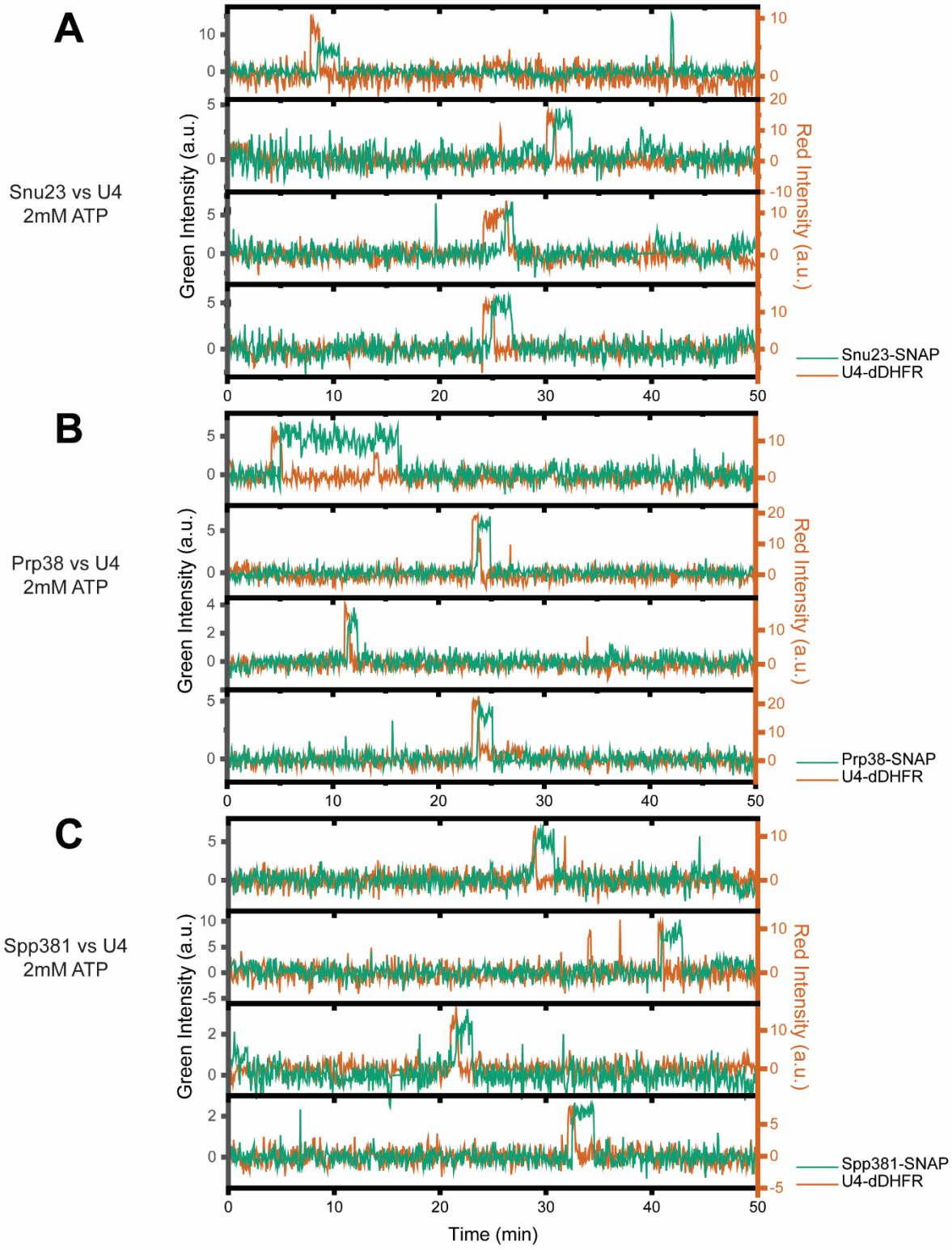

A
Snu23 vs U4
2mM ATP
B
Prp38 vs U4
2mM ATP
C
Spp381 vs U4
2mM ATP
Green Intensity (a.u.)
Red Intensity (a.u.)
Time (min)
Snu23-SNAP
U4-dDHFR
Prp38-SNAP
U4-dDHFR
Spp381-SNAP
U4-dDHFR

Figure 3.11 Additional Sample Fluorescence Trajectories from 3-color CoSMoS Experiments Monitoring BCP-SNAP and U4-DHFR Proteins.

Shown are super-imposed fluorescence intensities for Prp38-SNAP, Snu23-SNAP or Spp381-SNAP (green traces) and U4DHFR proteins (red traces, Prp3- and Prp4-DHFR) at 2 mM ATP.

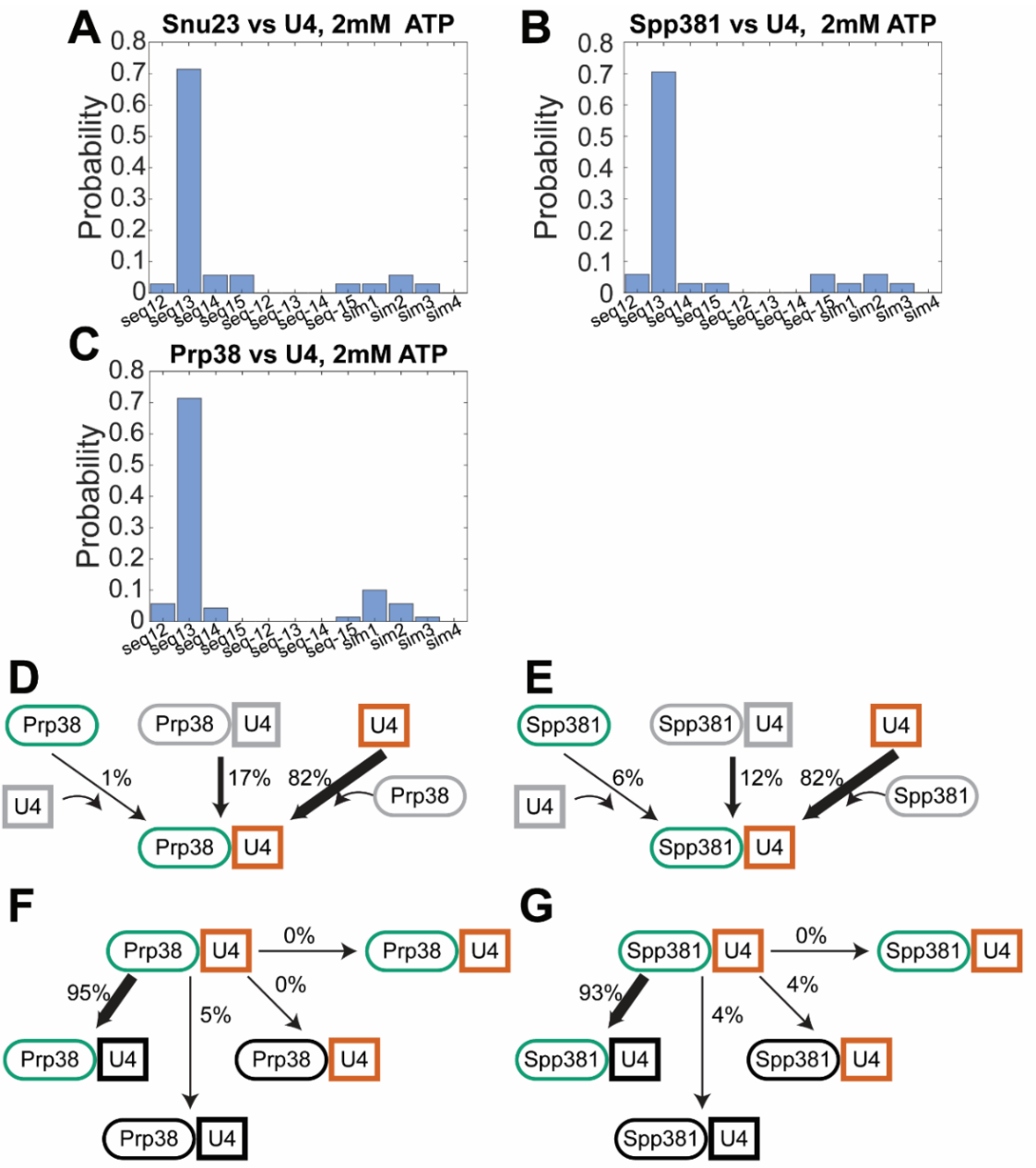


Figure 3.12 Additional Analysis of BCP and U4 Binding Dynamics During Activation.

(A-C) Assignment of overlapped BCP and U4 binding events from 3-color CoSMoS experiments according to the orders of signal arrival and disappearance. The majority of events in all cases are for sequential arrive of the U4 proteins followed by a BCP protein followed by loss of U4 preceding loss of the BCP protein (seq13-type events). (D-E) Routes for the appearance of Prp38 (D, *N=70*) or Spp381 (E, *N = 34*) and U4 fluorescent spots at 2 mM ATP. (F-G) Routes for loss of

Prp38 (F, *N=57*) or Spp381 (G, *N=28*) and U4 fluorescent spots at 2 mM ATP for event pairs in which the U4 spot appearance preceded arrival of Prp38 or Spp381.

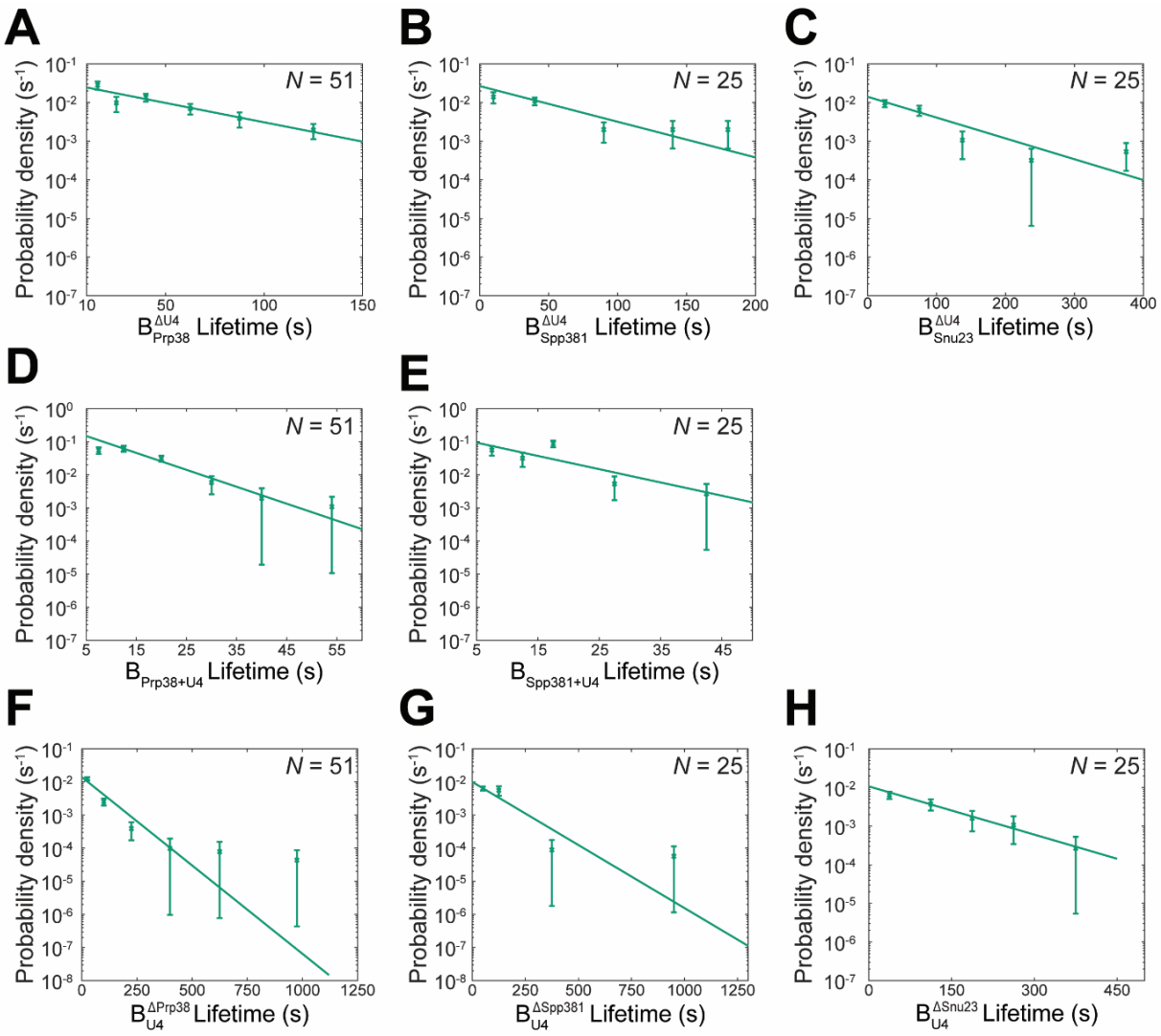


Figure 3.13 Probability Density Histogram of Lifetimes for Identified Activation Intermediates.

Probability density histograms of $\mathbf{B_{BCP}^{\Delta U4}}$ (panel A-C, $\mathbf{t_{BCP}^{arrival} - t_{U4}^{arrival}}$), $\mathbf{B_{BCP+U4}}$ (panel D-E, $\mathbf{t_{U4}^{release} - t_{BCP}^{arrival}}$) and $\mathbf{B_{BCP}^{\Delta U4}}$ (panel F-H, $\mathbf{t_{BCP}^{release} - t_{U4}^{release}}$) obtained from the subset of events (*N*) showing ordered arrival of U4 and then BCP spots (either Prp38, Spp381, or Snu23) followed by ordered loss of the BCP and then U4 signals. Distribution of $\mathbf{B_{Snu23+U4}}$ lifetimes cannot be fitted with the exponential-based distributions described in the methods, due to limited number of

binding intervals. Lines represent fits of the lifetime distributions with equations containing single exponential terms that yielded the parameters reported in **Table 3.3**. Error bars were calculated for each point as described in the Methods.

Snu66 is Released with the U4 snRNP and Before the BCP Subcomplex

We next tested if another protein released during the B to B^{ACT} transition but not part of the BCP also dissociates subsequent to U4 snRNP release. We carried out three-color CoSMoS assays using WCEs containing Snu66-SNAP labeled with SNAP-DY549 fluorophore (green), doubly DHFR-tagged U4 snRNP proteins (Prp3 and Prp4) labeled with Cy5-TMP fluorophores (red) and Alexa Fluor 488 labeled pre-mRNAs (blue) (**Fig. 3.14A**). As before, we identified colocalized pairs of events and then studied the orders in which the signals appeared or were lost. At 2 mM ATP, the predominant pathway for Snu66 association with the pre-mRNA is coincident with U4 snRNP proteins (80%; 77/96 event pairs, **Fig. 3.14B, C; 3.15**). Among the subset in which Snu66 and U4 snRNP protein signals appeared simultaneously, the majority also showed simultaneous loss of the signals (72%, 55/77 event pairs, **Fig. 3.14B, D**). While we cannot exclude the possibility that Snu66 binding or release occurs very swiftly (< 4 s) after tri-snRNP binding or separately from U4 dissociation, our data are most consistent with Snu66 being recruited to spliceosomes as part of the tri-snRNP and released along with the U4 snRNP. Association of Snu66 with the tri-snRNP is in agreement with mass spectrometry and cryo-EM data for both yeast and human spliceosomes (Bertram *et al.*, 2017; Gottschalk *et al.*, 1999; Plaschka *et al.*, 2017; Stevens and Abelson, 1999). Release of Snu66 along with U4 snRNP proteins is in agreement with direct interactions between Snu66 and Prp3 observed in human tri-snRNPs (Liu et al., 2006). We cannot determine if Snu66 may also be a U4 snRNP component from these data or if its interaction with U4 is transient and Snu66/U4 snRNP complexes are resolved prior to U4/U6 di-snRNP assembly during tri-snRNP regeneration.

The above data also indicate that not all proteins released during the B to B^{ACT} transition dissociate simultaneously. When combined with the data from **Fig. 3.10**, we can predict that Snu66 and the BCP are released as separate complexes. Specifically, since the BCP is released after the U4 snRNP and Snu66 along with the U4 snRNP, then the BCP should also be released

after Snu66. To verify the prediction, we carried out 3-color CoSMoS assay using extracts containing Snu66-SNAP labeled with SNAP-DY549 fluorophore (green), doubly DHFR-tagged B complex proteins (Spp381 and Snu23) labeled with Cy5-TMP fluorophores (red) and Alexa Fluor 488 labeled pre-mRNAs (blue) (**Fig. 3.16A**). As expected, 84% (66/79) of event pairs showed that Snu66 associates with the pre-mRNA prior to the BCP (**Fig. 3.16B, C**). Among this subset of event pairs, 67% (44/66) showed that BCP is released after the U4 snRNP proteins (**Fig. 3.16B, D**), in agreement with our prediction.

On many pre-mRNAs we also observed multiple rounds of tri-snRNP and BCP binding (**Fig. 3.16B**). The data are similar to those previously described for tri-snRNP binding in which pre-mRNAs can undergo multiple rounds of activation (Hoskins *et al.*, 2016). However, if activation or subsequent steps fails the spliceosome likely disassembles and a different tri-snRNP molecule is recruited. Here, our data suggest that failure also results in release of the BCP and that subsequent attempts at spliceosome assembly and activation originate from recruitment of different tri-snRNP and BCP molecules.

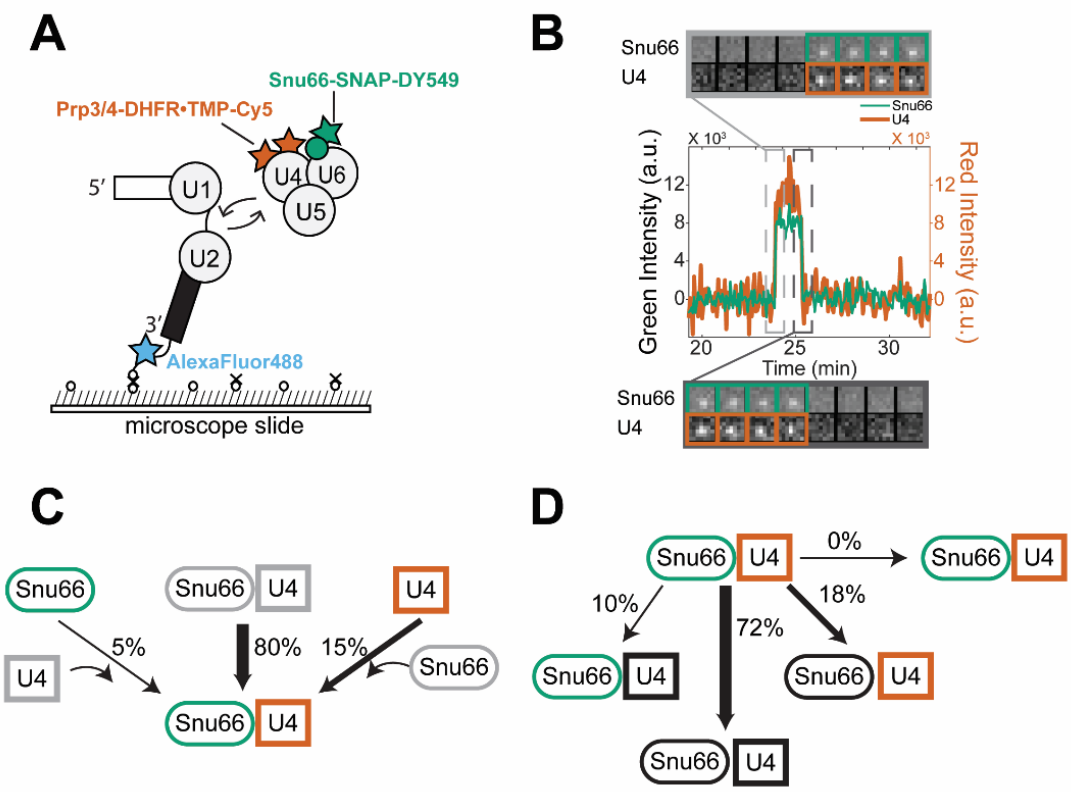


Figure 3.14 Snu66 Binds with the tri-snRNP and Releases with U4.

(A) Schematic of a 3-color experiment in which U4 was labeled with Cy5-TMP fluorophores, Snu66 was labeled with a DY-549 fluorophore, and the surface-tethered pre-mRNA was labeled with an Alexa Fluor 488 fluorophore. (B) Segment of a representative time record showing peaks in fluorescence intensity corresponding to colocalization of U4 (red, thick line) and Snu66 proteins (green, thin line) with the same individual pre-mRNA molecule. The dashed rectangles mark examples of the simultaneous appearance and disappearance of U4 and Snu66 spots; galleries show consecutive images taken from the indicated part of the recording showing that both spot appearance and disappearance are simultaneous. (C) Routes for the appearance of Snu66 and U4 fluorescent spots at 2 mM ATP for *N*=96 pairs of overlapping events. Red and green shapes represent observation of fluorescence from the corresponding DY-549 or Cy5 fluorophores on Snu66 or U4, respectively. Percentages represent the fraction of Snu66/U4 complexes in which

fluorescence appeared by the indicated pathway; more prevalent pathways are emphasized with thicker arrows. (D) Routes for loss of either Snu66 or U4 fluorescent spots at 2 mM ATP for *N*=77 pairs of overlapping events that show predominant recruitment patterns.

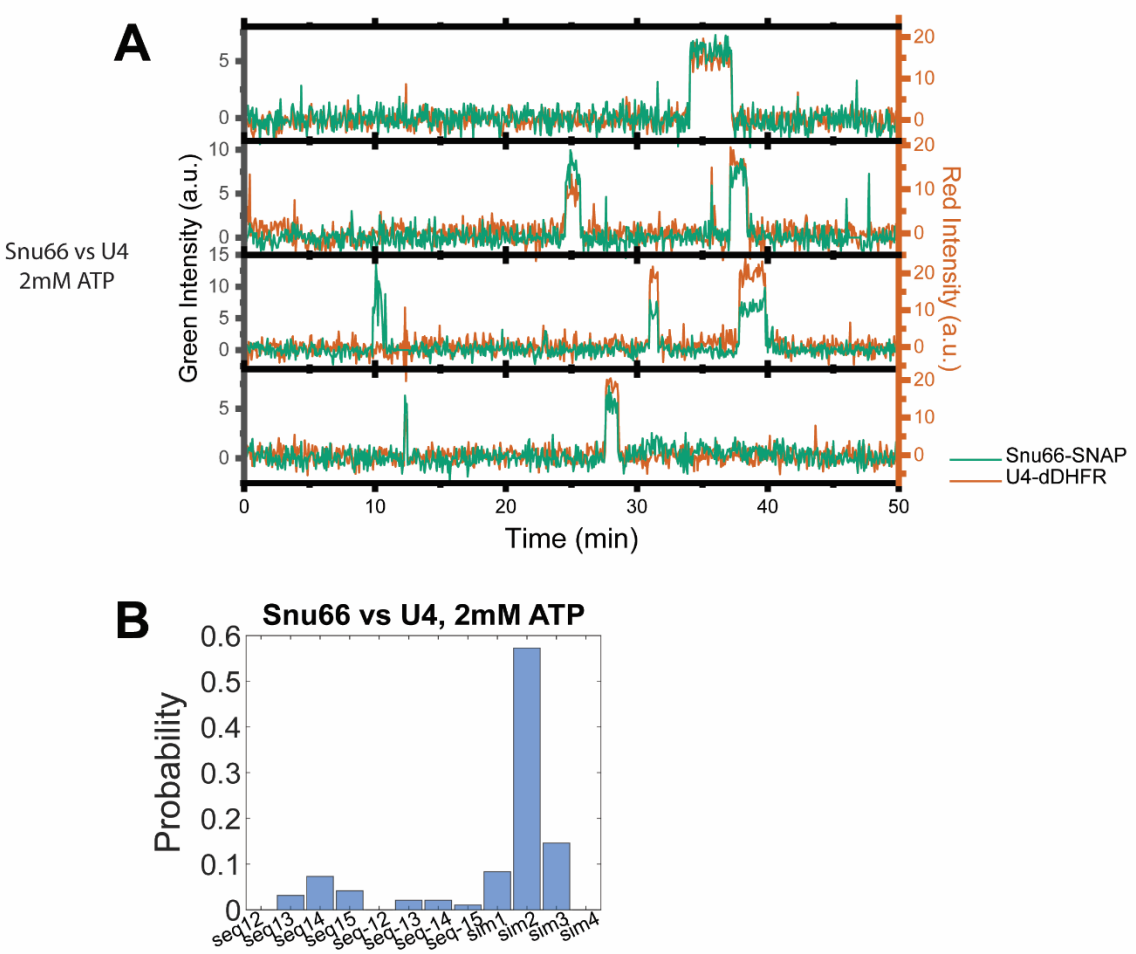


Figure 3.15 Additional Trajectories and Analysis of Snu66 and U4 Binding Dynamics.

(A) Shown are super-imposed fluorescence intensities for Snu66-SNAP (green traces) and U4-DHFR proteins (red traces) at 2 mM ATP. (B) Distribution of Snu66 and U4 overlapped binding events from 3-color CoSMoS experiments.

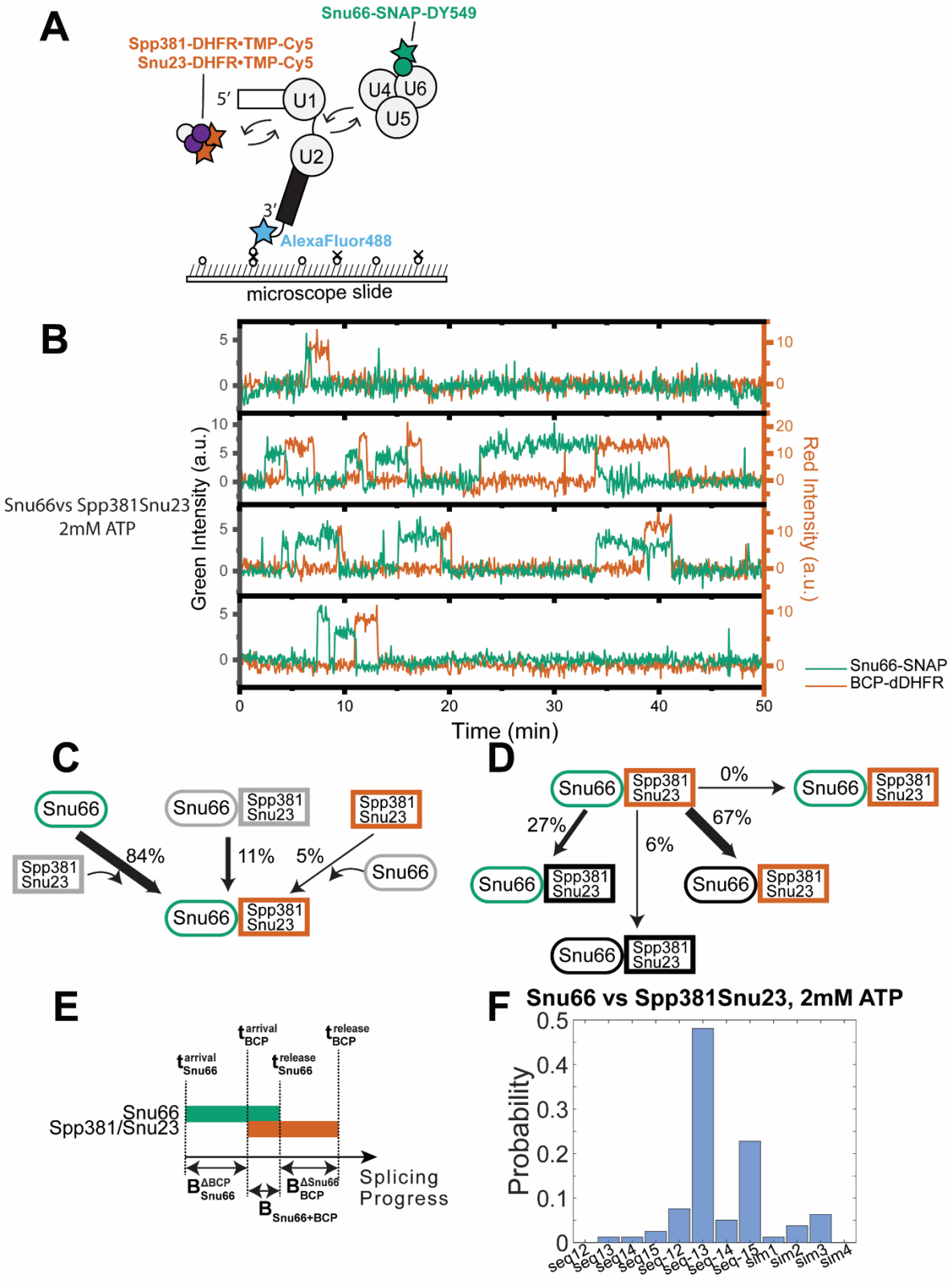

A
Snu66-SNAP-DY549
Spp381-DHFR•TMP-Cy5
Snu23-DHFR•TMP-Cy5
5′
U1
U2
U4
U5
U6
3′
AlexaFluor488
microscope slide
B
Snu66vs Spp381Snu23
2mM ATP
Green Intensity (a.u.)
Red Intensity (a.u.)
Time (min)
Snu66-SNAP
BCP-dDHFR
C
Snu66
Spp381
Snu23
84%
11%
5%
D
0%
27%
67%
6%
E
arrival BCP
release BCP
arrival Snu66
release Snu66
Snu66
Spp381/Snu23
Splicing
Progress
F
Snu66 vs Spp381Snu23, 2mM ATP
Probability
seq12
seq13
seq14
seq15
seq-12
seq-13
seq-14
seq-15
sim1
sim2
sim3
sim4

Figure 3.16 3-color CoSMoS observation of Snu66 and BCP binding dynamics during activation.

(A) Schematic of a 3-color experiment in which BCP was labeled with Cy5-TMP fluorophores, Snu66 was labeled with a DY-549 fluorophore, and the surface-tethered pre-mRNA was labeled with an Alexa Fluor 488 fluorophore. (B) Shown are super-imposed fluorescence intensities for Snu66-SNAP (green traces) and BCP-DHFR proteins (red traces) at 2 mM ATP. (C) Routes for the appearance of Snu66 and BCP fluorescent spots at 2 mM ATP for *N*=79 pairs of overlapping events. (D) Routes for loss of either Snu66 or BCP fluorescent spots at 2 mM ATP for *N*=66 pairs of overlapping events in which the Snu66 spot appearance preceded arrival of BCP. (E) Schematic showing the definition for the identified intermediate based on the relative recruitment and release times of Snu66 and BCP proteins. (F) Distribution of Snu66 and BCP overlapped binding events from 3-color CoSMoS experiments.

The BCP Subcomplex is a Stable Component of the tri-snRNP Only at Low ATP

The observation that the BCP is associating with the pre-mRNA after tri-snRNP binding was surprising since the BCP has previously been identified as a tri-snRNP component by mass spectrometry (Gottschalk *et al.*, 1999; Lybarger *et al.*, 1999; Stevens and Abelson, 1999; Xie *et al.*, 1998) and was present in samples used to determine the cryo-EM structure of the tri-snRNP, although not modeled (Nguyen et al., 2015). One difference between those experiments and ours is that we carried out our single molecule assays under conditions that permit spliceosome activation and splicing (2 mM ATP). Under these conditions, the yeast tri-snRNP is unstable, and it has instead been isolated in the absence of added ATP (Nguyen *et al.*, 2015; Nguyen *et al.*, 2016). We wondered if the presence or absence of ATP changes BCP association with the tri-snRNP. To test this, we repeated our three-color CoSMoS assays using labeled BCP and U4 snRNP but under conditions that permit spliceosome assembly but inhibit activation (0.05 mM ATP) (**Fig. 3.17A**).

At 0.05 mM ATP, Snu23 no longer predominantly associated with the pre-mRNA separately from the tri-snRNP. Instead, the most common pathway was simultaneous binding of Snu23 and U4 snRNP proteins (73%; 68/93 event pairs, **Fig. 3.17B, C, 3.18A**). Simultaneous binding with U4 snRNP proteins was also observed for Spp381, Prp38 and Snu66 when these proteins were labeled instead of Snu23 (**Fig. 3.18B-G**). This observation likely explains why the BCP and Snu66 showed similar binding kinetics in two-color CoSMoS assays at 0.05 mM but not 2 mM ATP (**Fig. 3.1D, E**). These results indicate that under conditions that stall spliceosome activation, the BCP is recruited to the pre-mRNA as a stable component of the tri-snRNP and remains stably bound to stalled spliceosome complexes.

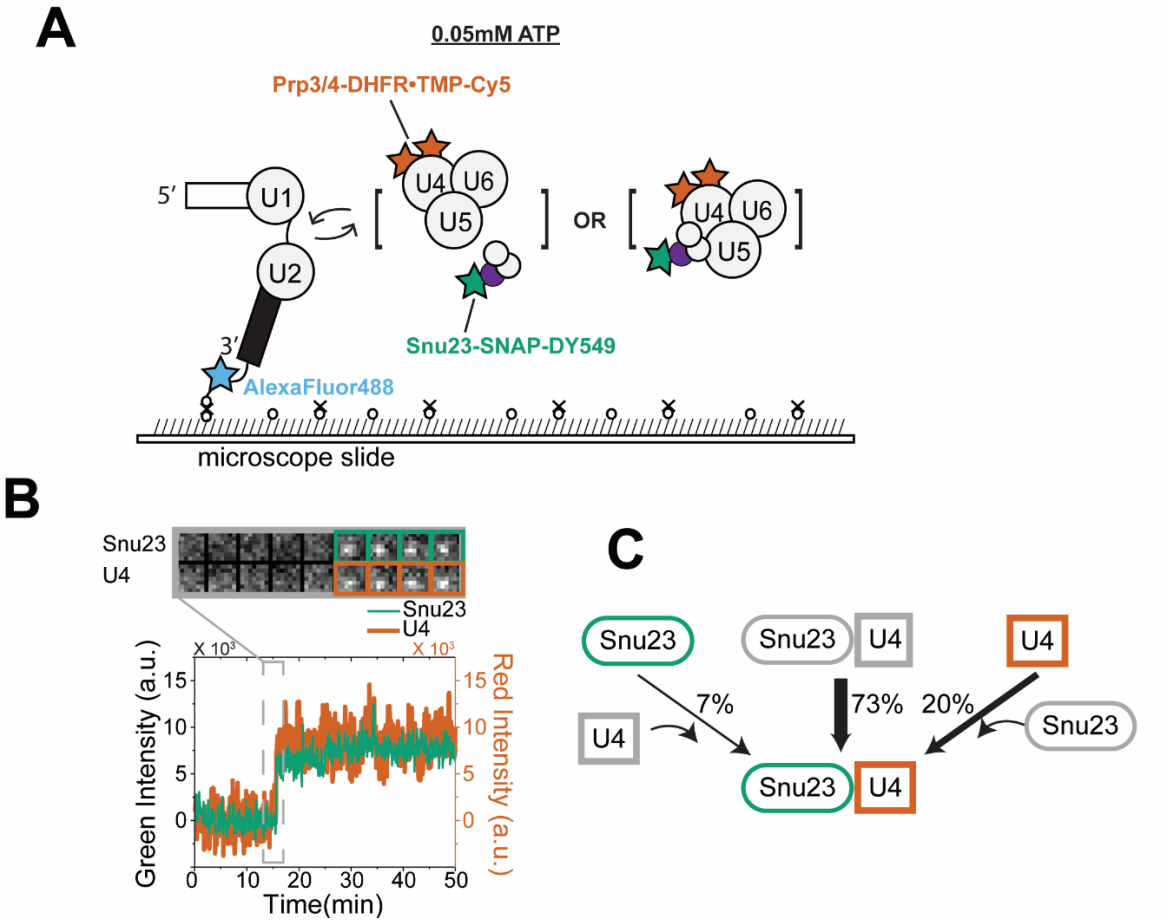


Figure 3.17 3-color CoSMoS observation of Snu23 and U4 binding dynamics under activation inhibition condition.

(A) Experimental schematic as in **Fig. 3.10A** except in the presence of 0.05 mM ATP (B) Segment of a representative time record showing peaks in fluorescence intensity corresponding to colocalization of U4 (red, thick line) and Snu23 proteins (green, thin line) with the same individual pre-mRNA molecule. The dashed rectangle mark an example of the simultaneous appearance of U4 and Snu23 spots; galleries show consecutive images taken from the indicated part of the recording showing that spot appearance is simultaneous. (C) Routes for the appearance of Snu23 and U4 fluorescent spots at 0.05 mM ATP for *N*=93 pairs of overlapping events. Red and green shapes represent observation of fluorescence from the corresponding DY-549 or Cy5 fluorophores on Snu23 or U4, respectively. Percentages represent the fraction of Snu23/U4

complexes in which fluorescence appeared by the indicated pathway; more prevalent pathways are emphasized with thicker arrows. Data for Spp381 and Snu23 is included in **Fig. 3.18**.

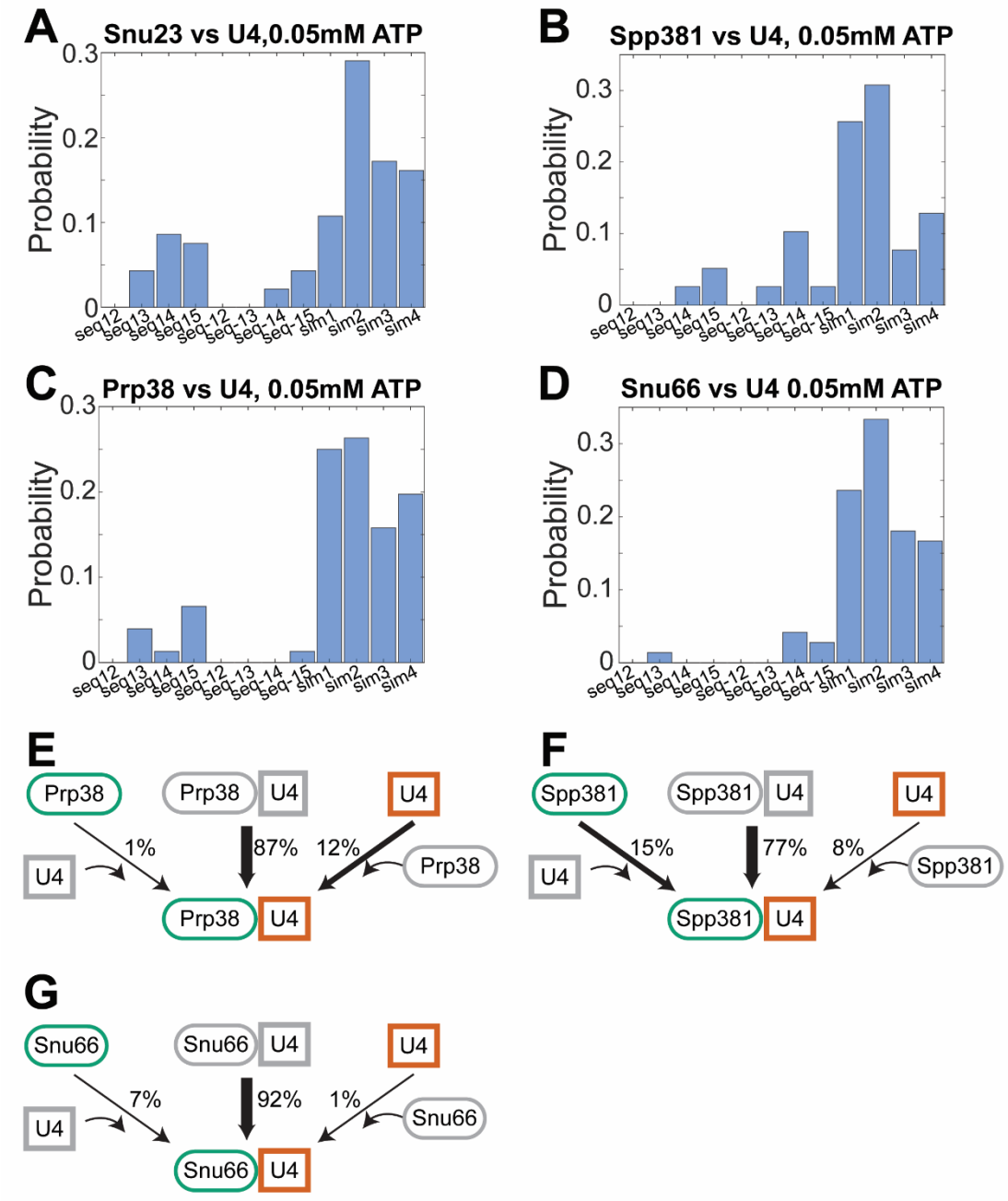


Figure 3.18 Additional Analysis on BCP and U4 binding dynamics at 0.05 mM ATP.

(A-D) Distribution of BCP and U4 binding events from 3-color CoSMoS experiments at 0.05 mM ATP. (E-G) Routes for the appearance of Prp38 (E), Spp381 (F) or Snu66 (G) and U4 fluorescent spots at 0.05 mM ATP for *N*=76, 39 or 72 pairs of overlapping events, respectively.

The BCP Subcomplex is Released after NTC Recruitment.

We previously determined that the NTC is predominantly recruited after U4 snRNP release (Hoskins *et al.*, 2016). Since the BCP is also released after U4 departure, we wondered if this occurred coincident, before, or after NTC recruitment. To answer this question, we carried out 3-color CoSMoS assays using WCE containing SNAP-tagged Snu23 labeled with SNAP-DY549 fluorophore and two DHFR tags on NTC components (Syf1 and Cef1) labeled with Cy5-TMP fluorophores (red) (**Fig. 3.19A**).

We first identified pairs of Snu23 and NTC binding events closest in time to one another with a requirement that binding events for both components should be at least 2 frames in duration ($\geq$ 8 s) to eliminate sampling interactions of the NTC (see Chapter 2). For each pair, we then subtracted the time of NTC binding ($t_{NTC}^{recruitment}$) from the time of Snu23 signal loss ($t_{Snu23}^{release}$) to yield a distribution of $t_{Snu23}^{release} - t_{NTC}^{recruitment}$ values. In this distribution, positive values would indicate that Snu23 (and the BCP) is released after NTC recruitment while negative values would indicate that BCP release occurs before NTC recruitment. The distribution of $t_{Snu23}^{release} - t_{NTC}^{recruitment}$ values for the selected pairs of events showed that 85% had positive value (97/114 events), suggesting that the NTC is recruited while the BCP is still present and that the BCP is released after the NTC binds (**Fig. 3.19C**). In comparison, the same analysis carried out for extracts containing a SNAP-tagged tri-snRNP protein (Snu66) and doubly DHFR-tagged NTC proteins showed predominantly negative values (**Fig. 3.20**).

The single molecule data also indicate the existence of several intermediate complexes: $B_{BCP}^{\Delta NTC}$ in which the BCP has associated with the pre-mRNA but the NTC has not yet joined, $B_{BCP+NTC}$ in which the NTC and BCP are both simultaneously present, and $B_{NTC}^{\Delta BCP}$ and later complexes in which the BCP has released but the NTC remains (**Fig 3.19F**). The characteristic lifetimes of the $B_{BCP}^{\Delta NTC}$ and $B_{BCP+NTC}$complexes are relatively short (~1 min; **Fig. 3.19G, H, Table**

3.3). The $\mathrm{B}_{\mathrm{BCP+NTC}}$ intermediate has not yet been described for yeast spliceosomes and likely possesses a similar composition and structure to human pre-B^{ACT1} complexes (Townsend *et al.*, 2020).

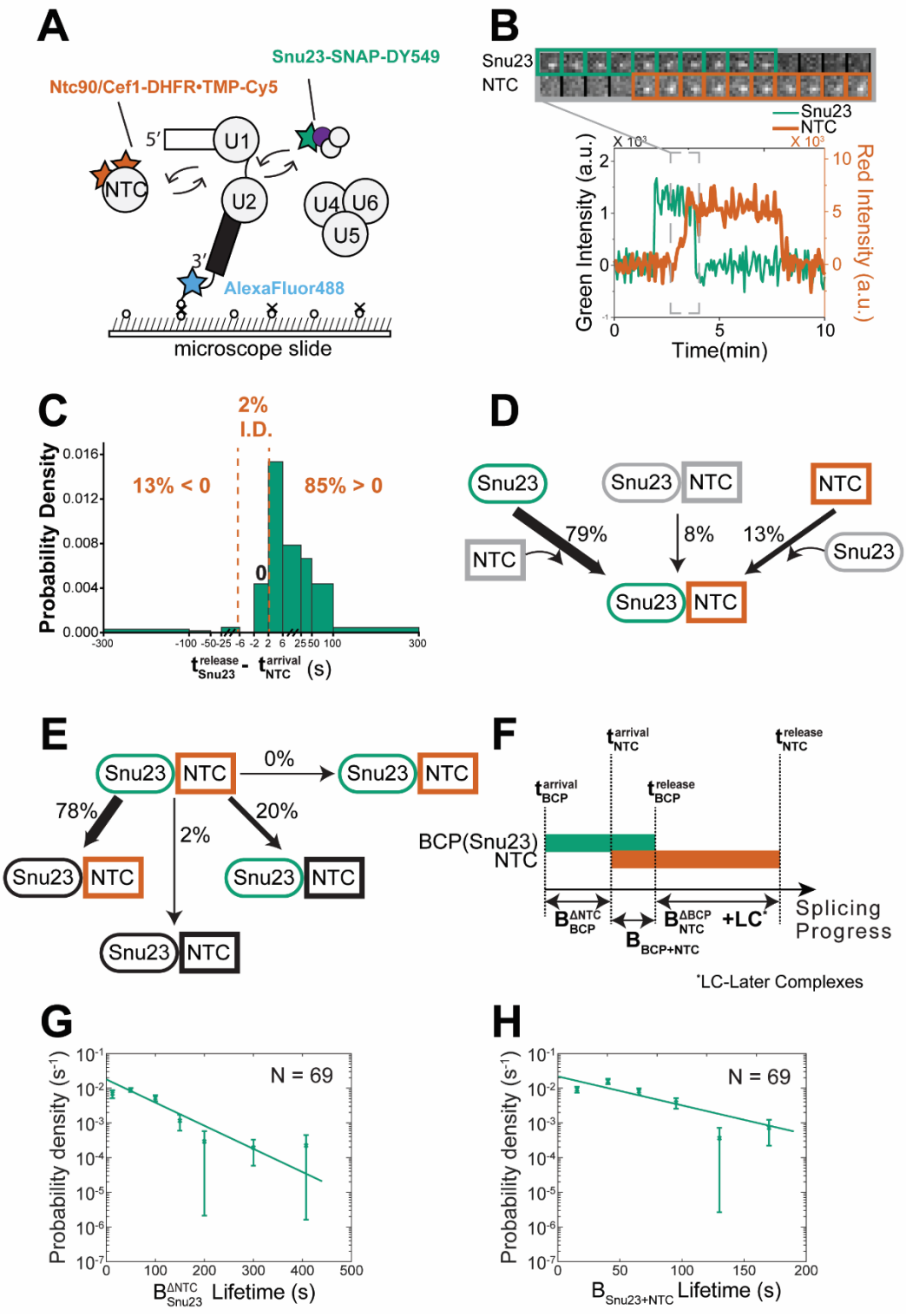


Figure 3.19 Three-color CoSMoS observation of NTC and Snu23 binding dynamics during activation.

(A) Experimental schematic as in **Fig. 3.10A** except that the NTC was labeled with Cy5-TMP fluorophores. (B) Segment of a representative time record showing peaks in fluorescence intensity corresponding to colocalization of the NTC (red, thick line) and Snu23 (green, thin line) with the same individual pre-mRNA molecule. The light grey dashed rectangle marks an example of the ordered recruitment of NTC followed by the release of Snu23 spots; galleries show consecutive images taken from that part of the recording showing the ordered spot appearance and disappearance. (C) Probability density histogram showing the delay between NTC arrival and Snu23 release. Most often (85% of *N*= 117 total events), the NTC arrived before release of the Snu23 ($\mathbf{t}_{\mathbf{Snu23}}^{\mathbf{release}} - \mathbf{t}_{\mathbf{NTC}}^{\mathbf{arrival}} > 0$). In 2% of cases, the exact order of events was indeterminate (I.D.) as described in the text. (D) Routes for the appearance of Snu23 and NTC fluorescent spots at 2 mM ATP for *N*=116 pairs of overlapping events. Red and green shapes represent observation of fluorescence from the corresponding DY-549 or Cy5 fluorophores on Snu23 or the NTC, respectively; grey shapes represent the absence of fluorescence. Percentages represent the fraction of Snu23/NTC complexes in which fluorescence appeared by the indicated pathway; more prevalent pathways are emphasized with thicker arrows. (E) Routes for loss of either Snu23 or NTC fluorescent spots at 2 mM ATP for *N*=92 pairs of overlapping events in which the Snu23 spot appearance preceded arrival of the NTC. (F) Schematic showing the definition for the identified intermediates based on the relative recruitment and release times of Snu23 and NTC proteins. (G-H) Probability density histograms of $\mathbf{B}_{\mathbf{Snu23}}^{\mathbf{\Delta NTC}}$ (panel G, $\mathbf{t}_{\mathbf{NTC}}^{\mathbf{arrival}} - \mathbf{t}_{\mathbf{Snu23}}^{\mathbf{arrival}}$) and $\mathbf{B}_{\mathbf{Snu23+NTC}}$ (panel H, $\mathbf{t}_{\mathbf{Snu23}}^{\mathbf{release}} - \mathbf{t}_{\mathbf{NTC}}^{\mathbf{arrival}}$) obtained from the subset of events (*N*) showing ordered arrival of Snu23 and then NTC spots followed by ordered loss of the Snu23 and then NTC signals. Lines represent fits of the lifetime distributions with equations containing single exponential terms that yielded the parameters reported in **Table 3.3**. Error bars were calculated for each point as described in the Methods.

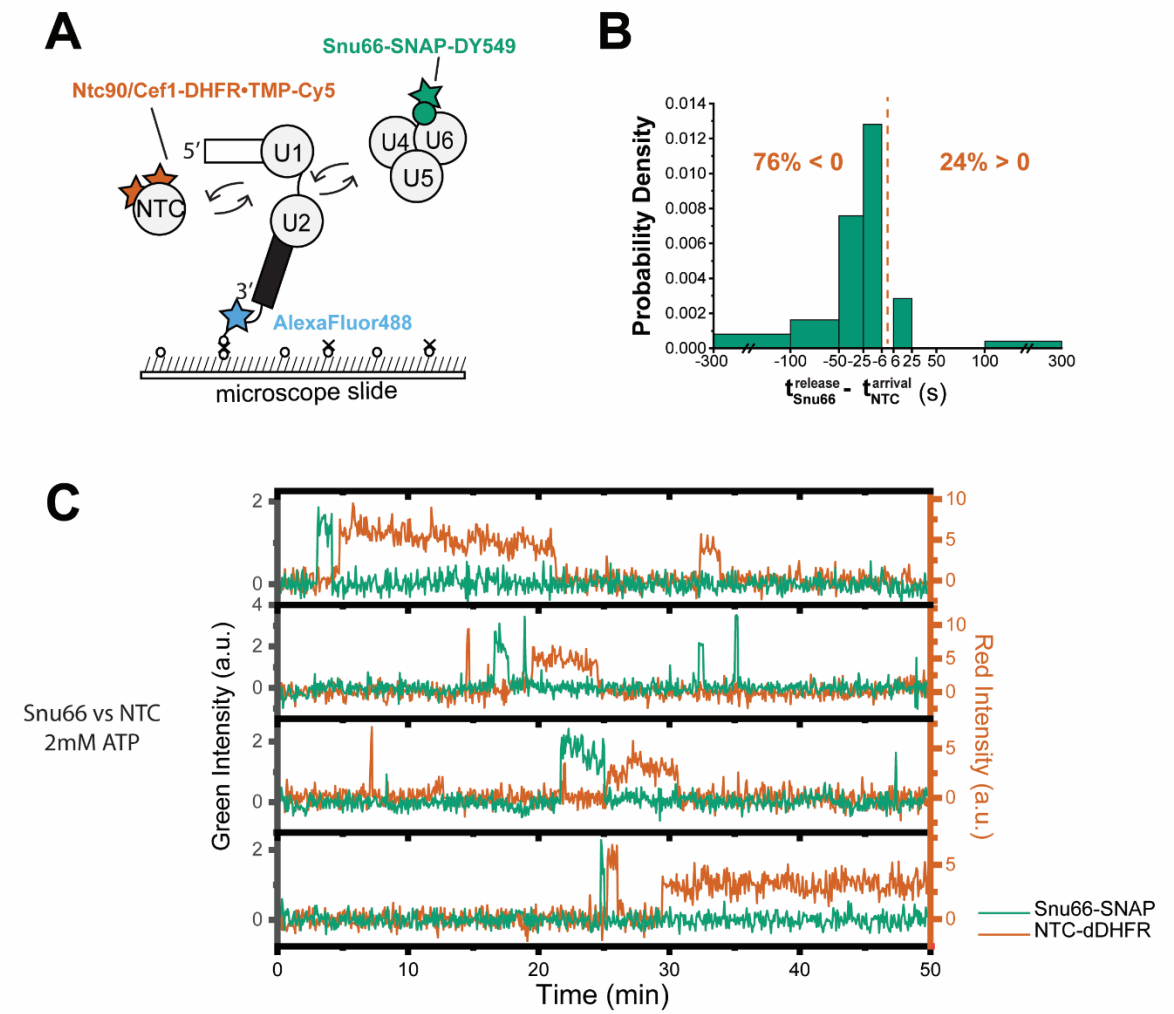


Figure 3.20 3-color CoSMoS observation of Snu66 and NTC binding dynamics during activation.

(A) Schematic of a 3-color experiment in which the NTC was labeled with Cy5-TMP fluorophores, Snu66 was labeled with a DY-549 fluorophore, and the surface-tethered pre-mRNA was labeled with an Alexa Fluor 488 fluorophore. (B) Probability density histogram showing the delay between NTC arrival and Snu66 release. Most often (76% of *N*= 39 total events), the NTC arrived after release of the Snu66 ($\mathbf{t^{release}_{Snu66} - t^{arrival}_{NTC} < 0}$). (C) Shown are super-imposed fluorescence intensities for Snu66-SNAP (green traces) and NTC-DHFR proteins (red traces) at 2 mM ATP.

DISCUSSION

Integrating previous (see Chapter 2) with single molecule data described here, we can derive a kinetic scheme for spliceosome activation *in vitro* (**Fig. 3.21A**). In our scheme, the tri-snRNP, lacking the BCP, joins the spliceosome A complex containing U1 and U2 to form the pre-B complex. In our single molecule data this is defined as a complex containing tri-snRNP proteins (U4 proteins, Lsm2-8, and Snu66; B_{Lsm+U4}) but not the BCP. After ~57s, the BCP joins as a single subunit to form the $B_{Lsm+U4+BCP}$ complex. Based on data obtained for activation of human spliceosomes and for yeast spliceosomes in the presence of a Prp28 mutant, the BCP associates near the U6 snRNA/5' SS duplex at the same site vacated by Prp28 (Bai *et al.*, 2018; Charenton et al., 2019; Townsend *et al.*, 2020). Thus, the ~57s time interval reports on both BCP association and transfer of the 5' SS to U6 by Prp28. We have not yet analyzed U1 snRNP dynamics during activation, but it is likely that it also dissociates during this time. The resulting $B_{Lsm+U4+BCP}$ complex may have the same composition as the biochemically and structurally characterized B complex.

The lifetime of $B_{Lsm+U4+BCP}$ is very short, ~13 s, before the U4 snRNP is lost along with at least a subset of tri-snRNP specific factors including Snu66. This forms a spliceosome complex lacking U4 and Snu66 but containing the Lsm2-8 and BCP complexes ($B_{Lsm+BCP}^{\Delta U4\Delta NTC}$). This complex then persists for 49-53 s before the NTC associates to form the $B_{Lsm+BCP+NTC}$ complex. This complex in turn persists for another ~52 s before dissociation of the BCP and Lsm2-8 complexes in an order that has not yet been determined. This forms a spliceosome with a composition similar to that of the characterized B^{ACT} complexes. Once B^{ACT} is formed, spliceosomes usually persist for ~407 s before loss of the NTC due to either successful splicing or disassembly of stalled complexes.

The kinetic scheme involves a series of intermediates several of which have not yet been biochemically or structurally characterized in yeast including the $B_{Lsm+BCP}^{\Delta U4\Delta NTC}$ and $B_{Lsm+BCP+NTC}$ complexes. By analysis of existing cyro-EM structures of yeast and human spliceosomes, we are

able to form a model for the types of conformational changes that are predicted to occur within these intermediates (**Fig. 3.21B**). Due to the disappearance of U4 snRNP proteins from B complex, U2 snRNP proteins and Lsm2-8 in complex with U2/U6 helix II are likely to move towards the void created by U4 release (**Fig. 3.21A**). The proper docking of these proteins onto Prp8 may then facilitate stable recruitment of NTC (see Chapter 2). Concertedly, the BCP may coordinate the stable docking of U2 snRNP proteins and the formation of the U2/U6 catalytic center (**Fig. 3.21B**). It is possible that Prp38 may serve as an interaction hub (Schütze et al., 2016) and facilitate U2 snRNP protein docking, especially for Hsh155. These conformational changes may be locked in place and made essentially irreversible through loss of Lsm2-8 and the BCP coordinated with binding of NTC components (Syf1, 2, and 3) and B^{ACT}-specific factors (Cwc24 and Cwc27).

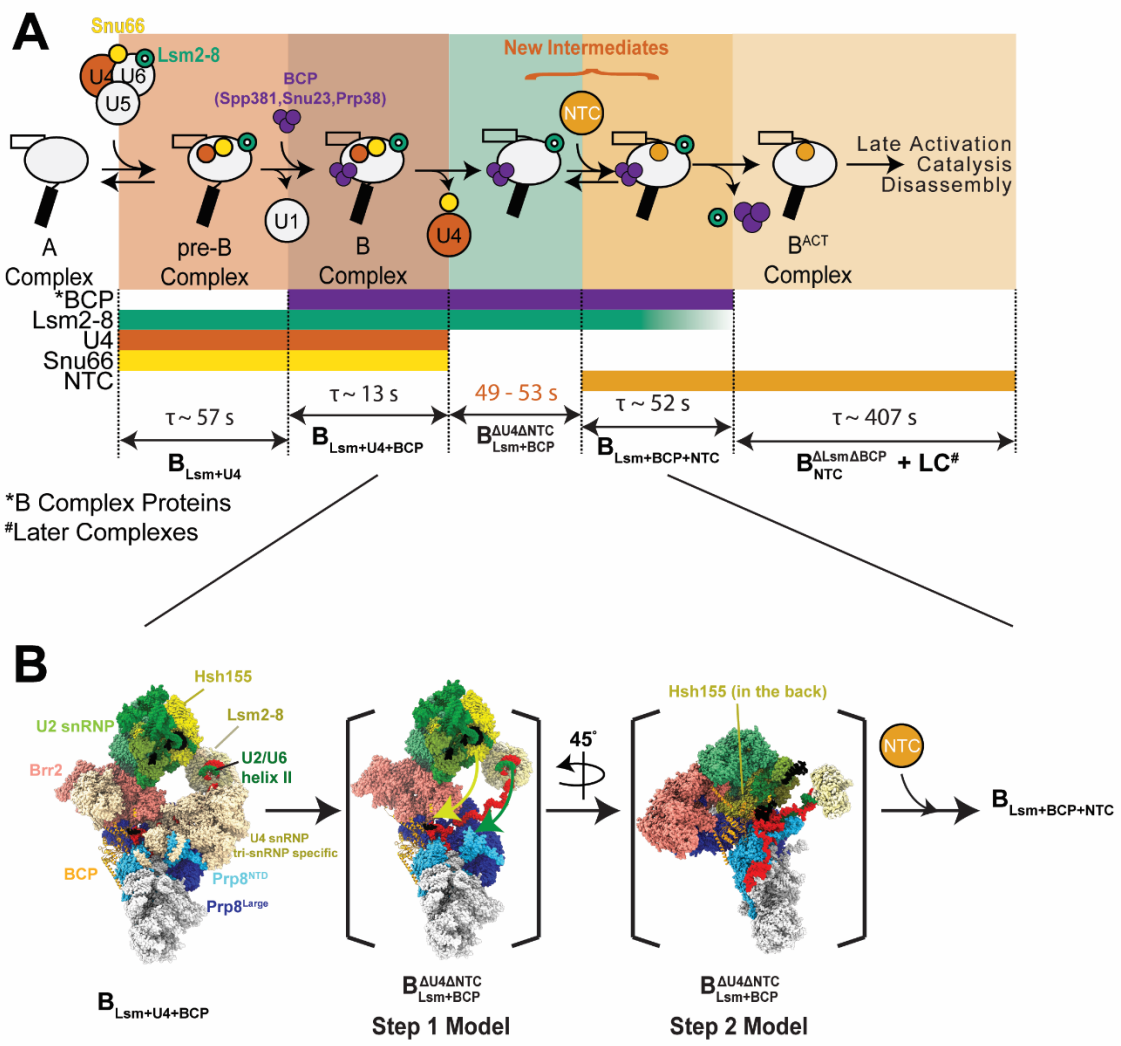


Figure 3.21 Transient Intermediates Formed During Spliceosome Activation.

(A) In this kinetic scheme for spliceosome activation, U4 snRNP release precedes a single NTC recruitment step, which in turn precedes BCP and Lsm2-8 (see Chapter 2) release. This would involve formation of at least two activation intermediates, the $\mathbf{B^{\Delta U4\Delta NTC}_{Lsm+BCP}}$ and $\mathbf{B^{\Delta U4}_{Lsm+BCP+\,NTC}}$ complexes, which have not been previously biochemically or structurally characterized. Single molecule data allows determination of the characteristic lifetimes (t) of these complexes during activation. The lifetime of the $\mathbf{B^{\Delta U4\Delta NTC}_{Lsm+BCP}}$ complex (red) was not directly measured in our experiments but can be inferred from our studies in combination with previous experiments (Hoskins *et al.*, 2016). (B) Two conformational states of $\mathbf{B^{\Delta U4\Delta NTC}_{Lsm+BCP}}$ in the transition from B to B^{ACT}

intermediate (PDBs: 5NRL and 5GM6; from $\mathbf{B_{Lsm+U4+\ BCP}}$ to $\mathbf{B_{NTC}^{\Delta LsmΔBCP}}$ in the nomenclature used in panel A) could be modeled, with the step 1 model ($\mathbf{B_{Lsm+BCP}^{\Delta U4\Delta NTC}}$) resembling the conformation of the yeast B complex except without the U4 snRNP. The step 2 model ($\mathbf{B_{Lsm+BCP}^{\Delta U4\Delta NTC}}$) resembles the conformation of yeast B^{ACT} except with Lsm2-8 and BCP proteins and without NTC proteins. The transition from step 1 to step 2 is likely to involve the conformational changes within the intermediate, possibly involving the movement of the U2 snRNP and Lsm2-8 towards the void created by the release of U4 snRNP proteins (yellow and green arrows). Retention of the BCP proteins may faciliate positioning of the U2 snRNP by docking of the U2 snRNP protein Hsh155 onto the BCP component Prp38.

Composition of the Yeast tri-snRNP is ATP Dependent

It has previously been proposed that BCP components are part of the tri-snRNP and recruited to the spliceosome with that complex. This was largely based on mass spectrometry data and biochemical assays of complexes isolated under no or limiting ATP conditions (Gottschalk *et al.*, 1999; Lybarger *et al.*, 1999; Nguyen *et al.*, 2016; Stevens and Abelson, 1999; Wan *et al.*, 2016; Xie *et al.*, 1998). However, data from the human splicing machinery has found no evidence that these proteins stably associate with the human tri-snRNP in the absence of spliceosome formation. Instead, they bind tri-snRNP factors only in the context of an activating spliceosome (Agafonov *et al.*, 2016; Charenton *et al.*, 2019). Our CoSMoS data resolve these observations. Under conditions that permit activation and splicing, the BCP is recruited to the spliceosome after incorporation of the tri-snRNP. This is in agreement with a recent activation mechanism for yeast based structural comparison of yeast and human spliceosomes and from structures of endogenous yeast pre-B and B complexes (Bai *et al.*, 2018; Plaschka *et al.*, 2017). This also suggests evolutionary conservation of these steps in activation between yeast and humans. However, under limiting ATP concentrations, the BCP is recruited to the spliceosome as part of the tri-snRNP. Thus, the BCP is an ATP-dependent yeast tri-snRNP component in the absence of the spliceosome.

These observations simplify models for yeast spliceosome activation by eliminating the need to reconcile BCP occupancy near the 5' SS with Prp28 activity. The BCP and Prp28 occupy overlapping binding sites (Bai *et al.*, 2018; Charenton *et al.*, 2019; Plaschka et al., 2018). Stepwise recruitment of the tri-snRNP and BCP to spliceosomes would facilitate Prp28 activity by allowing Prp28 to facilitate transfer the 5' SS from U1 to U6 and dissociate from the spliceosome before BCP binding.

Pre-mature Association of the BCP with the tri-snRNP may cause Spliceosome Stalling

It is unclear if tri-snRNPs formed at low ATP and containing the BCP are equally competent for splicing as those without the BCP. Previous single molecule data showed that the majority (68%) of spliceosomes formed at 0.05 mM ATP were disassembled at 2 mM ATP (Hoskins *et al.*, 2016). This would support the hypothesis that many of the tri-snRNP complexes that bind to the pre-mRNA pre-associated with the BCP are not competent for subsequent steps and may represent stalled states.

Formation of stalled yeast spliceosome complexes at low ATP is also supported by analysis of cryo-EM data for yeast and human complexes. When comparing the yeast to human pre-catalytic B complex, it is obvious that the 5' stem of the U6 snRNA and 5' exon of the pre-mRNA have exchanged locations (**Fig. 3.2, 3.23**). In catalytic spliceosomes from both yeast and humans, the 5' exon is accommodated in an “exon channel” formed by Prp8. In the human B spliceosome, the 5' exon is already located in the exon channel and the 5' SS base pairs with the U6 snRNA ACAGAGA sequence proximal to the channel (**Fig. 3.2**). These interactions are stabilized by the human BCP. However, in the yeast B spliceosome the 5' SS has not base paired with the U6 ACAGAGA sequence but is instead pairing at an adjacent upstream sequence (the ACAGAGA stem). Additionally, the 5' exon is located outside the channel, and the U6 5' stem is located in the exon channel, stabilized by the BCP. Previously, the differences in the 5' SS/U6 base-pairing and around the exon channel were interpreted as different activation mechanisms taken by yeast and human splicing machinery with yeast requiring additional steps to correctly pair U6 and position the 5' exon (Plaschka et al., 2019). Interestingly, the ability for the U6 snRNA to dock into the 5' exon channel may explain data seen with the nematode splicing machinery in which an exon could be ligated onto the 5' end of the U6 snRNA during *trans*-splicing (Yu et al., 1993). In this case, the protocols involved in depleting the nematode extract of RNA (including incubation at 30^{o}C in the absence of ATP) followed by its reconstitution with deproteinized total

RNA containing a synthetic, mutant U6 may have led to misassembly of the tri-snRNP and allowed U6 access to the exon exit channel.

We propose that yeast B complex structures formed at limiting ATP (Plaschka *et al.*, 2017) represent stalled intermediates due to pre-mature association of the BCP (**Fig. 3 22**). The BCP may structurally lock the U6 5' stem in the exon channel of the B complex as well as block correct transfer of the 5' SS to U6 by preventing Prp28 binding. This results in formation of spliceosomes not competent for carrying out the splicing reaction and that are substrates for disassembly at higher concentrations of ATP. It is interesting to note that the differences in the 5' SS region were also observed in spliceosome structures obtained from endogenous complexes found in lysates but from a strain expressing a dominant negative mutant of Prp28 (Bai *et al.*, 2018). This suggests that this region is labile and different conformations may give rise to spliceosomes with different kinetic or catalytic properties.

It is unclear if structures of the yeast tri-snRNP represent similarly stalled complexes due to presence of the BCP. When comparing yeast and human tri-snRNP cryo-EM structures, several differences are apparent and it has already been speculated that the purified yeast tri-snRNP is in a pre-maturely activated state (Agafonov *et al.*, 2016; Bertram *et al.*, 2017; Nguyen *et al.*, 2016). In yeast, the helicase Brr2 is already loaded onto the U4 snRNA binding site, resembling its location in the subsequent pre-B and pre-catalytic B spliceosome intermediates (**Fig. 3.2**). In contrast, the Brr2 has not been loaded onto the U4 snRNA binding site in the human tri-snRNP. It is instead held in a pre-activated conformation at a distant location by the Sad1.

While purified yeast tri-snRNPs do contain the BCP, the BCP has not yet been modeled within the cryo-EM density so it is unclear how they could facilitate stabilization of the yeast tri-snRNP structure. It was noted by Bertram et al., that cryo-EM density assigned to Snu66 in the purified yeast tri-snRNP (Nguyen *et al.*, 2016) could be better fit by the yeast BCP (Bertram *et al.*, 2017). It is possible that as in B complex spliceosomes, pre-mature recruitment of the BCP to the

yeast tri-snRNP leads to BCP association with the U6 snRNA and its stabilization in the wrong site, near the nascent 5' exon channel and that this is correlated with a partially closed structure of Prp8 between the N-terminal and central domains (Jia and Sun, 2018). Whether or not BCP-dependent formation of such stalled tri-snRNP or spliceosomes complexes at low ATP are used to regulate splicing *in vivo* is unknown. It is possible that switching of the tri-snRNP between catalytically competent and incompetent conformations depending on the presence or absence of the BCP could be used to globally regulate cellular splicing in response to ATP levels.

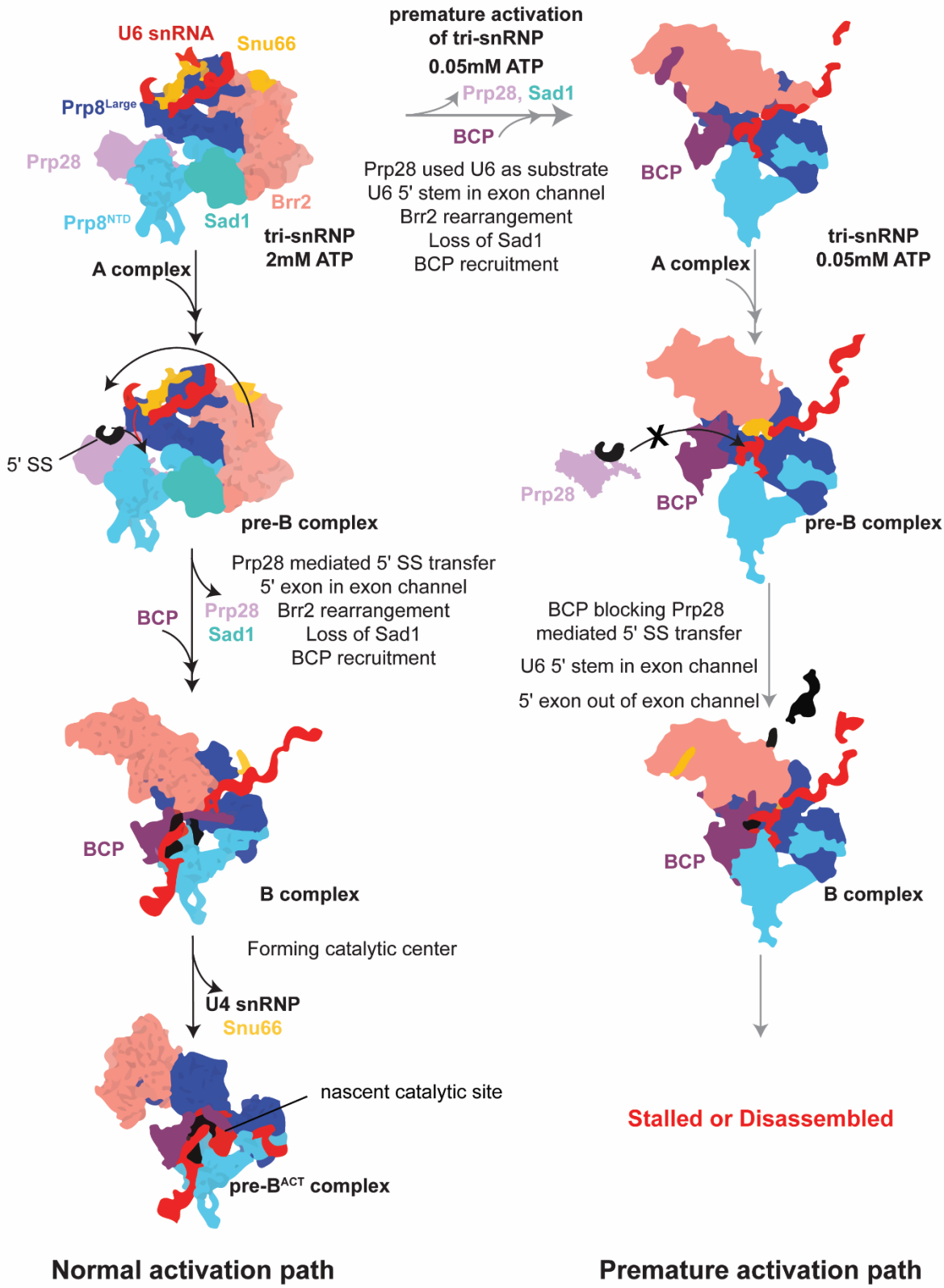

U6 snRNA
Snu66
Prp8Large
Prp28
Brr2
Prp8NTD
Sad1
tri-snRNP
2mM ATP
A complex
5' SS
pre-B complex
Prp28 mediated 5' SS transfer
5' exon in exon channel
Prp28 Brr2 rearrangement
Sad1 Loss of Sad1
BCP recruitment
BCP
B complex
Forming catalytic center
U4 snRNP
Snu66
nascent catalytic site
pre-BACT complex
Normal activation path
premature activation
of tri-snRNP
0.05mM ATP
Prp28, Sad1
BCP
Prp28 used U6 as substrate
U6 5' stem in exon channel
Brr2 rearrangement
Loss of Sad1
BCP recruitment
BCP
A complex
tri-snRNP
0.05mM ATP
Prp28
BCP
pre-B complex
BCP blocking Prp28
mediated 5' SS transfer
U6 5' stem in exon channel
5' exon out of exon channel
BCP
B complex
Stalled or Disassembled
Premature activation path

Figure 3.22 Proposed Structural Model for ATP-dependent Recruitment of BCP in Spliceosome Activation.

Integrating existing structural information and our single molecule data, the ATP-dependent recruitment of BCP could occur as follows. (*Left Pathway*) In the normal activation pathway for yeast and human spliceosomes, Prp28 is asscoiated with the tri-snRNP and Brr2 is held in an inactive state by Sad1. Association of the tri-snRNP with the spliceosome to form pre-B complex triggers release of Sad1, Brr2 reorientation and activation by binding of the U4 snRNA, and release of Prp28 after 5'SS transfer from U1 snRNA to U6 snRNA. The BCP is recruited to the now vacant binding site for Prp28 to complete the transition from the pre-B to the B complex. Finally, Brr2 helicase activity releases the U4 snRNP proteins and Snu66 and the now unbasepaired U6 snRNA can undergo self-folding to form nascent catalytic site. (*Right Pathway*) In the premature activation pathway, low ATP causes release of Prp28 and Sad1 from the tri-snRNP and recruitment of the BCP. Upon formation of the spliceosome pre-B complex, the BCP now interferes with Prp28 action by blocking its binding site. This disrupts transferring of the 5'SS to U6 snRNA, resulting in the 5' exon being located at the entrance to the exon channel instead of passing through the channel. The altered positions of the 5' exon and U6 snRNA originating from normal versus pre-mature activation pathways can potentially result in the formation of stalled or defective spliceosomes that are disassembled rather than being capable of completing activation and splicing.

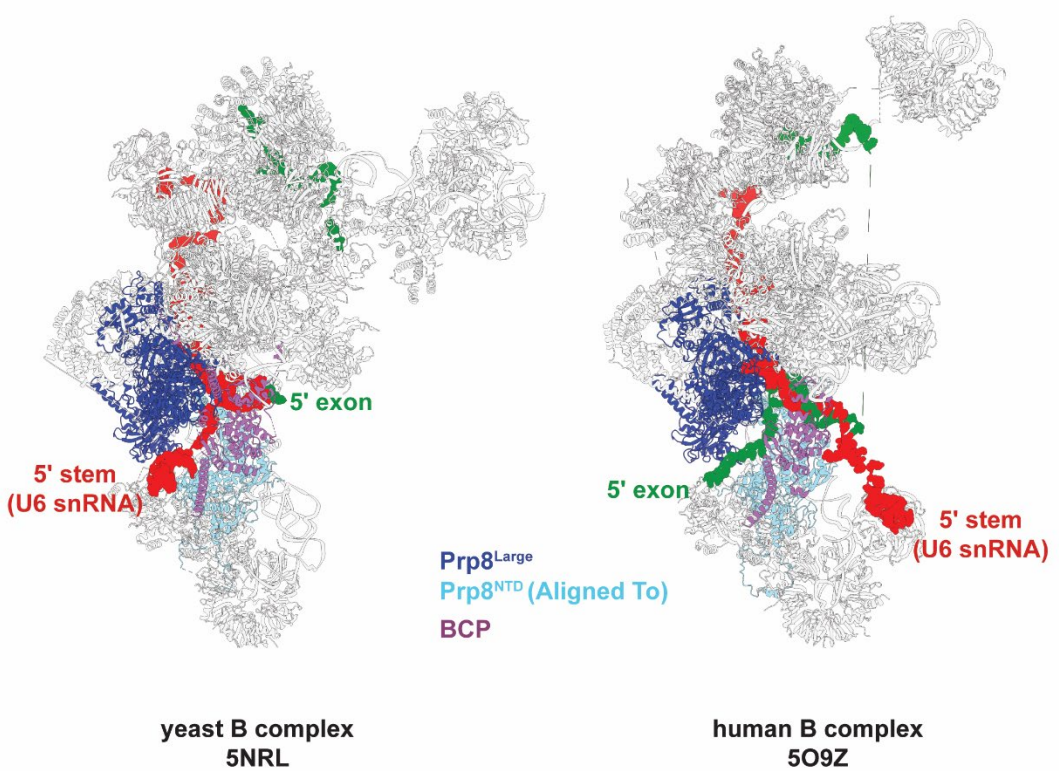


Figure 3.23 Swapping of the Positions of the 5' Exon and U5 5' Stem in Yeast and Human B Complex Spliceosomes.

In both yeast and human B complexes, the BCP is located at the gate of exon channel formed between $Prp8^{NTD}$ and $Prp8^{Large}$. The U6 5' stem and 5' exon appear to have switched locations when comparing the two structures. The structure of yeast B complex shown here was obtained at low ATP. Both human and yeast structures are aligned to the $Prp8^{NTD}$ (residue 1 to 885 for yeast; residue 1 to 812 for human).

REFERENCES

Agafonov, D.E., Kastner, B., Dybkov, O., Hofele, R.V., Liu, W.T., Urlaub, H., Lührmann, R., and Stark, H. (2016). Molecular architecture of the human U4/U6.U5 tri-snRNP. Science *351*, 1416-1420. 10.1126/science.aad2085.

Anderson, E.G., and Hoskins, A.A. (2014). Single molecule approaches for studying spliceosome assembly and catalysis. Methods Mol Biol *1126*, 217-241. 10.1007/978-1-62703-980-2_17.

Bai, R., Wan, R., Yan, C., Lei, J., and Shi, Y. (2018). Structures of the fully assembled. Science *360*, 1423-1429. 10.1126/science.aau0325.

Behrens, S.E., and Lührmann, R. (1991). Immunoaffinity purification of a [U4/U6.U5] tri-snRNP from human cells. Genes Dev *5*, 1439-1452. 10.1101/gad.5.8.1439.

Bertram, K., Agafonov, D.E., Dybkov, O., Haselbach, D., Leelaram, M.N., Will, C.L., Urlaub, H., Kastner, B., Lührmann, R., and Stark, H. (2017). Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation. Cell *170*, 701-713.e711. 10.1016/j.cell.2017.07.011.

Blanton, S., Srinivasan, A., and Rymond, B.C. (1992). PRP38 encodes a yeast protein required for pre-mRNA splicing and maintenance of stable U6 small nuclear RNA levels. Mol Cell Biol *12*, 3939-3947. 10.1128/mcb.12.9.3939-3947.1992.

Charenton, C., Wilkinson, M.E., and Nagai, K. (2019). Mechanism of 5' splice site transfer for human spliceosome activation. Science *364*, 362-367. 10.1126/science.aax3289.

Friedman, L.J., Chung, J., and Gelles, J. (2006). Viewing dynamic assembly of molecular complexes by multi-wavelength single-molecule fluorescence. Biophys J *91*, 1023-1031. 10.1529/biophysj.106.084004.

Friedman, L.J., and Gelles, J. (2015). Multi-wavelength single-molecule fluorescence analysis of transcription mechanisms. Methods *86*, 27-36. 10.1016/j.ymeth.2015.05.026.

Gottschalk, A., Neubauer, G., Banroques, J., Mann, M., Lührmann, R., and Fabrizio, P. (1999). Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6.U5] tri-snRNP. EMBO J *18*, 4535-4548. 10.1093/emboj/18.16.4535.

Hoskins, A.A., Friedman, L.J., Gallagher, S.S., Crawford, D.J., Anderson, E.G., Wombacher, R., Ramirez, N., Cornish, V.W., Gelles, J., and Moore, M.J. (2011). Ordered and dynamic assembly of single spliceosomes. Science *331*, 1289-1295. 10.1126/science.1198830.

Hoskins, A.A., Rodgers, M.L., Friedman, L.J., Gelles, J., and Moore, M.J. (2016). Single molecule analysis reveals reversible and irreversible steps during spliceosome activation. Elife *5*. 10.7554/eLife.14166.

Jia, X., and Sun, C. (2018). Structural dynamics of the N-terminal domain and the Switch loop of Prp8 during spliceosome assembly and activation. Nucleic Acids Res *46*, 3833-3840. 10.1093/nar/gky242.

Kaur, H., Jamalidinan, F., Condon, S.G.F., Senes, A., and Hoskins, A.A. (2019). Analysis of spliceosome dynamics by maximum likelihood fitting of dwell time distributions. Methods *153*, 13-21. 10.1016/j.ymeth.2018.11.014.

Larson, J., Kirk, M., Drier, E.A., O'Brien, W., MacKay, J.F., Friedman, L.J., and Hoskins, A.A. (2014). Design and construction of a multiwavelength, micromirror total internal reflectance fluorescence microscope. Nat Protoc *9*, 2317-2328. 10.1038/nprot.2014.155.

Liu, S., Rauhut, R., Vornlocher, H.P., and Lührmann, R. (2006). The network of protein-protein interactions within the human U4/U6.U5 tri-snRNP. RNA *12*, 1418-1430. 10.1261/rna.55406.

Lybarger, S., Beickman, K., Brown, V., Dembla-Rajpal, N., Morey, K., Seipelt, R., and Rymond, B.C. (1999). Elevated levels of a U4/U6.U5 snRNP-associated protein, Spp381p, rescue a mutant defective in spliceosome maturation. Mol Cell Biol *19*, 577-584. 10.1128/MCB.19.1.577.

Mayas, R.M., Maita, H., and Staley, J.P. (2006). Exon ligation is proofread by the DExD/H-box ATPase Prp22p. Nat Struct Mol Biol *13*, 482-490. 10.1038/nsmb1093.

Nguyen, T.H., Galej, W.P., Bai, X.C., Savva, C.G., Newman, A.J., Scheres, S.H., and Nagai, K. (2015). The architecture of the spliceosomal U4/U6.U5 tri-snRNP. Nature *523*, 47-52. 10.1038/nature14548.

Nguyen, T.H.D., Galej, W.P., Bai, X.C., Oubridge, C., Newman, A.J., Scheres, S.H.W., and Nagai, K. (2016). Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. Nature *530*, 298-302. 10.1038/nature16940.

Plaschka, C., Lin, P.C., Charenton, C., and Nagai, K. (2018). Prespliceosome structure provides insights into spliceosome assembly and regulation. Nature *559*, 419-422. 10.1038/s41586-018-0323-8.

Plaschka, C., Lin, P.C., and Nagai, K. (2017). Structure of a pre-catalytic spliceosome. Nature *546*, 617-621. 10.1038/nature22799.

Plaschka, C., Newman, A.J., and Nagai, K. (2019). Structural Basis of Nuclear pre-mRNA Splicing: Lessons from Yeast. Cold Spring Harb Perspect Biol *11*. 10.1101/cshperspect.a032391.

Rymond, B.C. (1993). Convergent transcripts of the yeast PRP38-SMD1 locus encode two essential splicing factors, including the D1 core polypeptide of small nuclear ribonucleoprotein particles. Proc Natl Acad Sci U S A *90*, 848-852. 10.1073/pnas.90.3.848.

Schütze, T., Ulrich, A.K., Apelt, L., Will, C.L., Bartlick, N., Seeger, M., Weber, G., Lührmann, R., Stelzl, U., and Wahl, M.C. (2016). Multiple protein-protein interactions converging on the Prp38 protein during activation of the human spliceosome. RNA *22*, 265-277. 10.1261/rna.054296.115.

Shcherbakova, I., Hoskins, A.A., Friedman, L.J., Serebrov, V., Corrêa, I.R., Xu, M.Q., Gelles, J., and Moore, M.J. (2013). Alternative spliceosome assembly pathways revealed by single-molecule fluorescence microscopy. Cell Rep *5*, 151-165. 10.1016/j.celrep.2013.08.026.

Staley, J.P., and Guthrie, C. (1999). An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. Mol Cell *3*, 55-64. 10.1016/s1097-2765(00)80174-4.

Stevens, S.W., and Abelson, J. (1999). Purification of the yeast U4/U6.U5 small nuclear ribonucleoprotein particle and identification of its proteins. Proc Natl Acad Sci U S A *96*, 7226-7231. 10.1073/pnas.96.13.7226.

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

Séraphin, B., and Rosbash, M. (1991). The yeast branchpoint sequence is not required for the formation of a stable U1 snRNA-pre-mRNA complex and is recognized in the absence of U2 snRNA. EMBO J *10*, 1209-1216.

Tarn, W.Y., Lee, K.R., and Cheng, S.C. (1993). Yeast precursor mRNA processing protein PRP19 associates with the spliceosome concomitant with or just after dissociation of U4 small nuclear RNA. Proc Natl Acad Sci U S A *90*, 10821-10825. 10.1073/pnas.90.22.10821.

Townsend, C., Leelaram, M.N., Agafonov, D.E., Dybkov, O., Will, C.L., Bertram, K., Urlaub, H., Kastner, B., Stark, H., and Lührmann, R. (2020). Mechanism of protein-guided folding of the active site U2/U6 RNA during spliceosome activation. Science *370*. 10.1126/science.abc3753.

Ulrich, A.K., and Wahl, M.C. (2017). Human MFAP1 is a cryptic ortholog of the Saccharomyces cerevisiae Spp381 splicing factor. BMC Evol Biol *17*, 91. 10.1186/s12862-017-0923-1.

Ulrich, A.K.C., Seeger, M., Schütze, T., Bartlick, N., and Wahl, M.C. (2016). Scaffolding in the Spliceosome via Single α Helices. Structure *24*, 1972-1983. 10.1016/j.str.2016.09.007.

Wan, R., Yan, C., Bai, R., Wang, L., Huang, M., Wong, C.C., and Shi, Y. (2016). The 3.8 Å structure of the U4/U6.U5 tri-snRNP: Insights into spliceosome assembly and catalysis. Science *351*, 466-475. 10.1126/science.aad6466.

Wilkinson, M.E., Charenton, C., and Nagai, K. (2020). RNA Splicing by the Spliceosome. Annu Rev Biochem *89*, 359-388. 10.1146/annurev-biochem-091719-064225.

Xie, J., Beickman, K., Otte, E., and Rymond, B.C. (1998). Progression through the spliceosome cycle requires Prp38p function for U4/U6 snRNA dissociation. EMBO J *17*, 2938-2946. 10.1093/emboj/17.10.2938.

Yu, Y.T., Maroney, P.A., and Nilsen, T.W. (1993). Functional reconstitution of U6 snRNA in nematode cis- and trans-splicing: U6 can serve as both a branch acceptor and a 5' exon. Cell *75*, 1049-1059. 10.1016/0092-8674(93)90315-h.

Zhang, X., Yan, C., Zhan, X., Li, L., Lei, J., and Shi, Y. (2018). Structure of the human activated spliceosome in three conformational states. Cell Res *28*, 307-322. 10.1038/cr.2018.14.

Chapter Four

Defining Essential Regions in the Protein Splicing Factor Spp381 and Isolation of Temperature Sensitive Mutants

This chapter is not yet published.

Fu, X., Voigts, L., Hoskins, A.A.

Author contributions:

XF and AAH conceived the project. XF and VL carried out experiments. XF and VL analyzed data. XF and AAH wrote the manuscript.

ABSTRACT

Many factors involved in pre-mRNA splicing have poorly characterized functions. About 50 splicing factors are exchanged at the stage of spliceosome activation. Spp381 is one of these factors contributing to the formation of active site for catalysis. Limited studies have been done on characterizing its cellular roles and its functions on splicing. Taking advantage of known Spp381 structures in spliceosome intermediates, we carried out structure-guided truncation analysis on Spp381 and studied the consequences of deleting characteristic regions on yeast growth. Using growth assays and ACT1-CUP1 assays, we identified multiple conditionally important regions in Spp381. By mapping these regions in spliceosome structures, we proposed these regions can potentially cause the observed phenotypes by disrupting the splicing processes via abolishing intermolecular interactions with adjacent interaction partners. More studies need to be done to confirm the expression level and cellular localization of Spp381 in mutant strains and to identify its interaction with other potential partners.

INTRODUCTION

Pre-mRNA splicing is the process of removing non-coding introns while simultaneously ligating the flanking exons. Splicing is carried out with nucleotide precision by a macromolecular machine called the spliceosome. The major building blocks for spliceosome are small nuclear ribonucleoproteins (snRNPs), namely the U1 snRNP, U2 snRNP, and U4/U6.U5 tri-snRNP. The splicing process undergoes four general stages, spliceosome assembly, activation, catalysis, and disassembly/recycling. Step-specific factors are required for transitioning the spliceosome between these stages. The B complex proteins subcomplex (BCP; Spp381, Snu23, and Prp38; see **Chapter 3**) is a group of three step-specific factors recruited to and released from the spliceosome during the spliceosome activation. How exactly its components contribute to activation is not yet clear.

The SPP381 (Suppressor of PrP38-1) gene encodes a protein 291 amino acid (aa) in length and was identified as a suppressor for the temperature sensitivity phenotype of *prp38-1* in *Saccharomyces cerevisiae* (hereafter, yeast) (Lybarger et al., 1999). Yeast two-hybrid analysis showed that the carboxyl half of Spp381 (aa 146-291), but not the amino terminal half, interacts with full length Prp38. Subsequently, structural data were obtained showing a direct interaction between *H. sapiens* or *C. thermophilum* Spp381 and Prp38 proteins in the absence of the spliceosome (Ulrich and Wahl, 2017) as well as similar complexes being present in the cryo-EM structures of yeast activation intermediates (pre-catalytic B complex) (Plaschka et al., 2017) (**Fig. 4.1A**). These genetic and structural data suggest that the physical interactions between Spp381 and Prp38 are important for *in vivo* splicing.

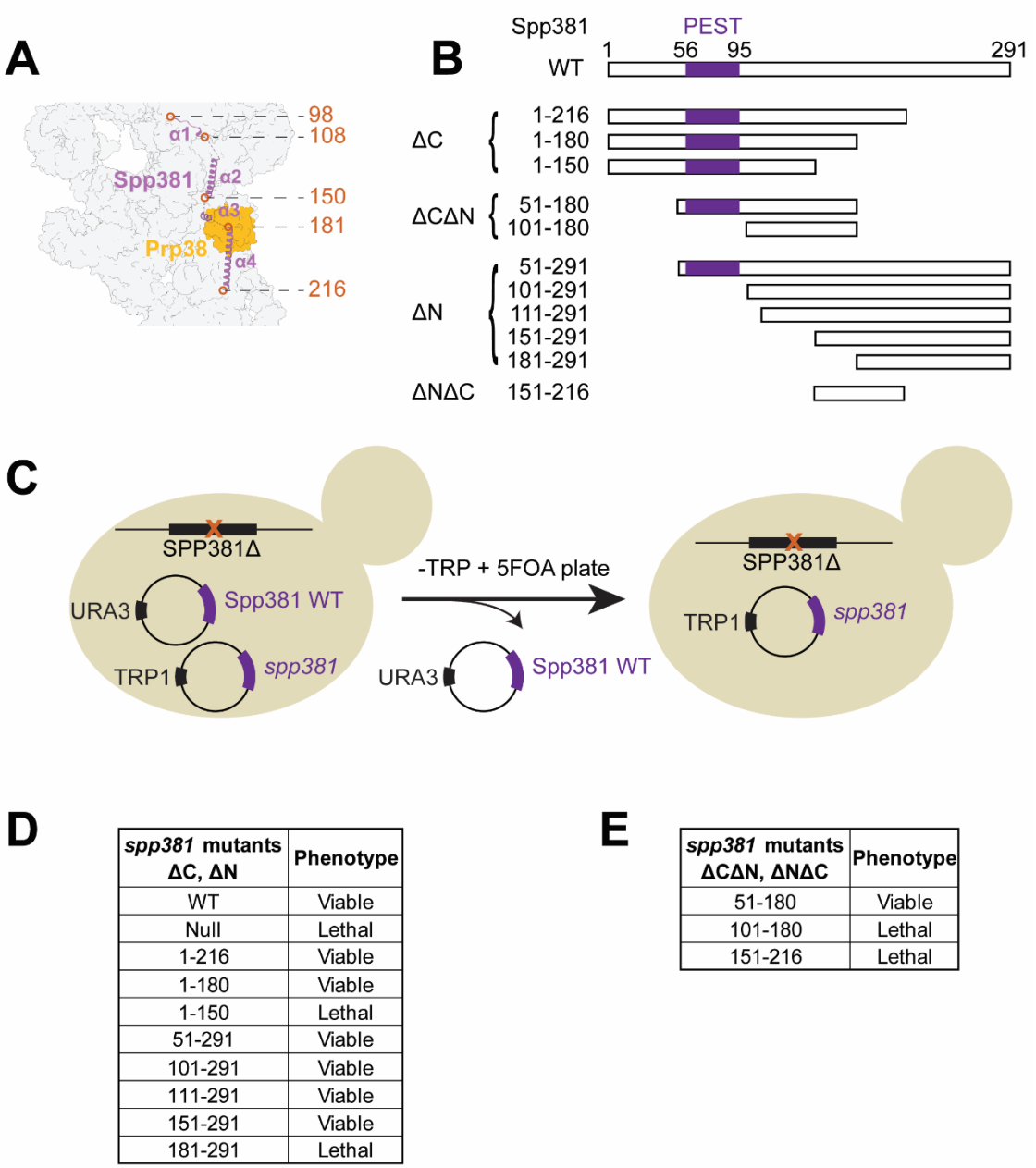
In addition to Prp38, Spp381 makes intermolecular interactions with many important splicing factors in the pre-catalytic B complex spliceosome via several extended helices (Plaschka *et al.*, 2017) (**Fig. 4.1A**). Starting at the Spp381 N-terminus, helix α1 interacts with Brr2 (the key helicase needed for unwinding the U4/U6 RNAs during activation), helix α2 runs along the groove

between Brr2 and Prp8$^{Jab1/MPN}$ (the interaction that regulates Brr2 activity (Boon et al., 2007; Maeder et al., 2009)), short helix α3 is positioned between Prp38 and Prp8^{Endo}, and the last helix α4 interacts with Prp38 and Prp8^{NTD} (near several Prp8 suppressors of U4 snRNA and Prp28 mutants that stall activation (Brow, 2019) (**Fig. 4.2**). No structural information is yet available for regions of Spp381 before α1 and after α4; however, a recent cryo-EM structure of the human pre-B^{ACT1} activation intermediate was able to model the C-terminus of the human Spp381 (MFAP1) (Townsend et al., 2020). In this structure, the C-terminus of MFAP1 interacts directly with the 5' exon of the pre-mRNA in the -6 to -8 region upstream of the exon's interaction with the U5 snRNA and the 5' splice site (SS)/U6 snRNA duplex. The multitude of intermolecular interactions between Spp381 and the pre-mRNA substrate and splicing factors has led us to ask how each region of the protein contributes to the splicing.

In addition to these regions, Spp381 and its orthologs also contain PEST (pro-asp-ser-thr) motifs associated with proteins with short half-lives (Rechsteiner and Rogers, 1996; Rogers et al., 1986). The Spp381/MFAP1 PEST motif score ranks highly among all PEST-containing proteins, implicating very short half lives in cells (Ulrich and Wahl, 2017). In Spp381, deletion of only the PEST motif (aa 56-95) has been shown to impair *in vivo* splicing (Lybarger *et al.*, 1999). In vivo measurements of protein half-life indicate that Spp381 is stable in yeast with half-live of 7.1 h, nearly the same as other BCP components and comparable to other splicing factors (Christiano et al., 2014). This indicates that the motif may contribute directly to the splicing reaction or only regulate Spp381 protein abundance under certain conditions.

Here, we assayed the effects of Spp381 truncations on yeast growth and splicing. Mutants containing truncations at either the N- or C-terminus, but not both, are viable if the protein also contains helix α3 that interacts with Prp38 and Prp8^{Endo}. Splicing assays carried out *in vitro* show that a Spp381 mutant that displays a growth defect at different temperatures also displays a

splicing defect. We propose that regions of Spp381 make redundant interactions with the spliceosome for ensuring the correct docking and anchoring of the BCP to the spliceosome.



D

spp381 mutants ΔC, ΔN	Phenotype
WT	Viable
Null	Lethal
1-216	Viable
1-180	Viable
1-150	Lethal
51-291	Viable
101-291	Viable
111-291	Viable
151-291	Viable
181-291	Lethal

E

spp381 mutants ΔCΔN, ΔNΔC	Phenotype
51-180	Viable
101-180	Lethal
151-216	Lethal

Figure 4.1 Schematic of Spp381 Protein Domains and Phenotypes of Truncation Mutants.

(A) Structure of Spp381 in the yeast pre-catalytic B complex spliceosome (PDB: 5NRL). Only a portion of the protein is visible, and Spp381 uses α1 to α4 helices to make intermolecular interactions with multiple factors. The previously investigated Prp38/Spp381$^{146\text{-}291}$ interaction consists of α3 and α4 regions (Lybarger *et al.*, 1999). (B) Domain representation of series of single-terminus and double-termini truncation mutants (ΔC, ΔCΔN, ΔN and ΔNΔC) of Spp381

used in this study. (C) Cartoon depiction of constructing *spp381* mutant strains by the plasmid shuffle method. (D-E) Summary tables for the growth phenotypes of *spp381* strains containing single terminus truncations (D) and double termini truncations (E) on 5FOA-containing plates at 30°C.

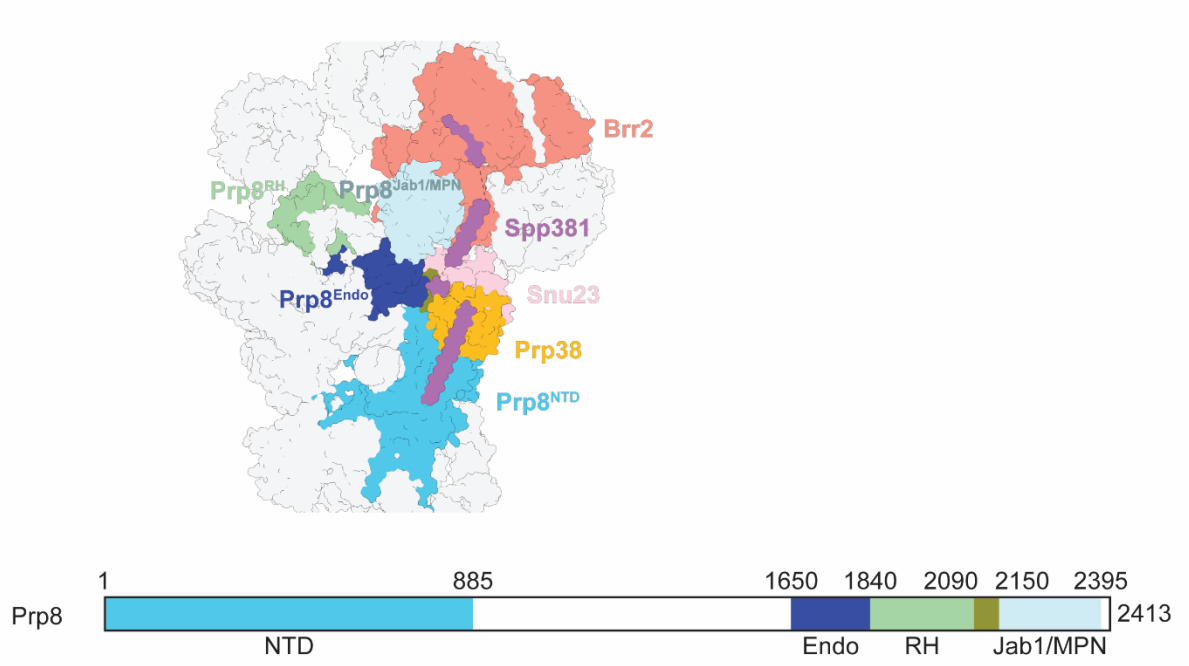


Figure 4.2 Spp381 Interacts with Essential Splicing Factors in the pre-catalytic B Complex Spliceosome.

Spp381 contacts multiple splicing factors Prp38, Brr2, Prp8^{NTD}, Prp8^{Endo}, and likely Prp8$^{2090\text{-}2150}$ and Snu23 (PDB: 5NRL). Also shown is the domain organization of Prp8.

MATERIALS AND METHODS

Strains and Plasmids

A shuffle strain (**Table 4.1, Strain #2**) used for selecting yeast containing *spp381* mutant plasmids was created by transforming a pRS416 plasmid (URA3/CEN) containing the wild type (WT) SPP381 gene into a parental strain (**Table 4.1, Strain #1**). The genomic SPP381 gene was then deleted using a hygromycin resistance cassette via homologous recombination (Goldstein and McCusker, 1999). Strains containing *spp381* mutants were created by first transforming the *spp381* mutant gene on a pRS414 plasmid (TRP1/CEN) into the shuffle strain, selecting for transformants containing the TRP1 marker on -TRP dropout plates, and streaking single colonies onto plates containing -TRP dropout medium and 5-fluoroorotic acid (5FOA; 1g/L) to select for loss of the URA3 plasmid. After incubating the plates at 30^{o}C for three days, viable colonies from the 5FOA plates were re-streaked and grown on fresh -TRP dropout plates. The shuffle strains for ACT1-CUP1 assays were constructed in a similar approach as described above except that a different parental strain was used (**Table 4.1, Strain #20**). After creating the *spp381* mutant strains, an ACT1-CUP1 splicing reporter on a pG1 plasmid (containing LEU2 selection marker) (Lesser and Guthrie, 1993) was transformed into the strain and yeast were grown on -TRP-LEU dropout plates.

Plasmids containing WT Spp381 sequences were created by cloning the WT Spp381 gene along with its flanking genomic sequences (±250bp) between the XhoI and NotI restriction sites in empty plasmid vectors. Plasmids containing *spp381* mutant sequences were created by using PCR along with the corresponding primer sets (**Table 4.2, 4.3**).

Growth Assays for *spp381* Mutants

spp381 mutant strains were grown in -TRP dropout liquid medium to saturation in a 30^{o}C shaking incubator. The yeast were then diluted with 10% (v/v) glycerol to an OD_{600} value of 0.50.

The diluted culture was then used to create 10-fold serial dilutions in a 96-well plate. Yeasts were then spotted onto -TRP dropout plates using a 48-pin inoculator. Plates were incubated at 16^{o}C for 10 days or for 3 days at 23^{o}C, 30^{o}C and 37^{o}C before growth was assessed and images recorded.

ACT1-CUP1 Assays for *spp381* Mutants

spp381 mutant strains were grown in -LEU-TRP dropout liquid medium to saturation in a 30^{o}C shaking incubator. The yeasts were then diluted with 10% (v/v) glycerol to an OD_{600} value of 0.50. The diluted *spp381* mutant cultures were then added into a 96-well plate and transferred using a 48-pin inoculator onto 30 -LEU-TRP dropout plates which contain different concentrations of copper sulfate (0, 0.025, 0.05, 0.075, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.25, or 2.5mM). The plates were incubated at 30^{o}C for 3 days before imaging and the maximum [Cu^{2+}] at which yeast colony was observed was recorded.

Table 4.1 Yeast strains used in this study

Strain #	ID	Genotype	Notes
1	yAAH0001	MATa prc1-407 prb1-1122 pep4-3 leu2 trp1 ura3-52 gal2	BJ2168, parental strain
2	yAAH2200	yAAH0001 + spp381Δ::HygR + pAAH1081	WT Spp381 on URA plasmid
3	yAAH2850	yAAH2200 + pAAH1082	WT spp381 on TRP plasmid
4	yAAH2341	yAAH2200 + pAAH1088	1-216aa spp381 on TRP plasmid
5	yAAH2344	yAAH2200 + pAAH1111	1-180aa spp381 on TRP plasmid
6	yAAH2255	yAAH2200 + pAAH1102	51-291aa spp381 on TRP plasmid
7	yAAH2256	yAAH2200 + pAAH1103	101-291aa spp381 on TRP plasmid
8	yAAH2342	yAAH2200 + pAAH1106	111-291aa spp381 on TRP plasmid
9	yAAH2343	yAAH2200 + pAAH1107	151-291aa spp381 on TRP plasmid
10	yAAH2345	yAAH2200 + pAAH1162	51-180aa spp381 on TRP plasmid
11	yAAH2349	yAAH2200 + pAAH1168	82-180aa spp381 on TRP plasmid
12	yAAH2346	yAAH2200 + pAAH1169	51-180aa, 63-81N spp381 on TRP plasmid
13	yAAH2347	yAAH22200 + pAAH1166	51-180aa, 74-81N spp381 on TRP plasmid
14	yAAH2348	yAAH22200 + pAAH1171	51-180aa, 90-93N spp381 on TRP plasmid
15	yAAH2352	yAAH22200 + pAAH1170	51-180aa, 84-93N spp381 on TRP plasmid
16	yAAH2354	yAAH22200 + pAAH1177	1-180aa, 84-93N spp381 on TRP plasmid
17	yAAH2353	yAAH22200 + pAAH1176	1-180aa, 74-93N spp381 on TRP plasmid
18	yAAH2350	yAAH22200 + pAAH1167	WT, 74-81N spp381 on TRP plasmid
19	yAAH2351	yAAH2200 + pXF0020	WT, 74-93N spp381 on TRP plasmid
20	yAAH0433	MATa cup1Δ ura3 his3 trp1 lys2 ade2 leu2	Parental strain for creating ACT1CUP1 strains
21	yAAH2356	yAAH0433 + spp381Δ::HygR + pAAH1081	WT Spp381 on URA plasmid, cup1Δ strain
22	yAAH2363	yAAH2356 + pAAH1082	WT Spp381 on TRP plasmid, cup1Δ strain
23	yAAH2362	yAAH2356 + pAAH1088	1-216aa spp381 on TRP plasmid, cup1Δ strain
24	yAAH2361	yAAH2356 + pAAH1111	1-180aa spp381 on TRP plasmid, cup1Δ strain
25	yAAH2357	yAAH2356 + pAAH1102	51-291aa spp381 on TRP plasmid, cup1Δ strain
26	yAAH2358	yAAH2356 + pAAH1103	101-291aa spp381 on TRP plasmid, cup1Δ strain
27	yAAH2359	yAAH2356 + pAAH1106	111-291aa spp381 on TRP plasmid, cup1Δ strain
28	yAAH2360	yAAH2356 + pAAH1107	151-291aa spp381 on TRP plasmid, cup1Δ strain
29	yAAH2376	yAAH2363 + pAAH0470	WT Spp381 on TRP plasmid, WT ACT1-CUP1 on LEU plasmid, cup1Δ strain
30	yAAH2369	yAAH2362 + pAAH0470	1-216aa spp381 on TRP plasmid, WT ACT1-CUP1 on LEU plasmid, cup1Δ strain
31	yAAH2368	yAAH2361 + pAAH0470	1-180aa spp381 on TRP plasmid, WT ACT1-CUP1 on LEU plasmid, cup1Δ strain
32	yAAH2364	yAAH2357 + pAAH0470	51-291aa spp381 on TRP plasmid, WT ACT1-CUP1 on LEU plasmid, cup1Δ strain

33	yAAH2365	yAAH2358 + pAAH0470	101-291aa spp381 on TRP plasmid, WT ACT1-CUP1 on LEU plasmid, cup1Δ strain
34	yAAH2366	yAAH2359 + pAAH0470	111-291aa spp381 on TRP plasmid, WT ACT1-CUP1 on LEU plasmid, cup1Δ strain
35	yAAH2367	yAAH2360 + pAAH0470	151-291aa spp381 on TRP plasmid, WT ACT1-CUP1 on LEU plasmid, cup1Δ strain
36	yAAH2377	yAAH2363 + pAAH0514	WT Spp381 on TRP plasmid, A3U ACT1-CUP1 on LEU plasmid, cup1Δ strain
37	yAAH2375	yAAH2362 + pAAH0514	1-216aa spp381 on TRP plasmid, A3U ACT1-CUP1 on LEU plasmid, cup1Δ strain
38	yAAH2374	yAAH2361 + pAAH0514	1-180aa spp381 on TRP plasmid, A3U ACT1-CUP1 on LEU plasmid, cup1Δ strain
39	yAAH2370	yAAH2357 + pAAH0514	51-291aa spp381 on TRP plasmid, A3U ACT1-CUP1 on LEU plasmid, cup1Δ strain
40	yAAH2371	yAAH2358 + pAAH0514	101-291aa spp381 on TRP plasmid, A3U ACT1-CUP1 on LEU plasmid, cup1Δ strain
41	yAAH2372	yAAH2359 + pAAH0514	111-291aa spp381 on TRP plasmid, A3U ACT1-CUP1 on LEU plasmid, cup1Δ strain
42	yAAH2373	yAAH2360 + pAAH0514	151-291aa spp381 on TRP plasmid, A3U ACT1-CUP1 on LEU plasmid, cup1Δ strain

Table 4.2 Primers used in this study

Primer #	ID	Sequence (5’ to 3’)
F1	CH-Spp381_Truncate_Reverse	5’- CAT CGT ATG TGA TTC AAA TCA -3’
F2	CH-Spp381_Truncate50_Forward	5’- GGT ACT GGA AAG AAC CGC -3’
F3	CH-Spp381_Truncate100_Forward	5’- TTA TTT ATG AAG AAA AAG GCC AAT AAT -3’
F4	XF6_spp381_trunc110F	5’- CAG AAA GCT ACC AAG ATA GAT C-3’
F5	XF7_spp381_trunc150F	5'- GAA ACG ATG AAG CTG AGA C -3'
F6	XF8_spp381_trunc180F	5'- GAT TCA GAA AAG GAG AGA CAG -3'
F7	XF10_spp381_stop151R	5'- TTA CTA TTA GTA GTT CTT TGC CAC -3'
F8	XF11_spp381_stop181R	5'- TTA CTA TTA GAC TTC ATC GTT ATC ATC -3'
F9	XF12_Spp381_74to81_N_f	5’- AAC AGT AAT AAT AGT AGC GAC AAC AG -3’
F10	XF13_Spp381_74to81_N_r	5’- GGG ACT ATT ATT ACT TTC GGG ACT CC -3’
F11	XF15_Spp381_63to81_N_f	5’- AAT GGG AGT CCC AAT AGT AAT AAT AGT C -3’
F12	XF17_Spp381_63to81_N_r	5’- ACT ATT ATT TGA ATT TTG GCC ATC ATT GG -3’
F13	XF18_Spp381_63to93_N_f	5’- AGC AAT AGT AAT AAT ATG AGA CCT TTA CC -3’
F14	XF19_Spp381_63to93_N_r	5’- CGA ATT ACT GTT ATT GCT ACT ATT ATT AC -3’
F15	XF21_Spp381_2to81del_f	5’- AGT AGC GAC AAC AGT GAT TCG -3’
F16	XF22_Spp381_84to93only_N_r	5’- CGA ATT ACT GTT ATT GCT ACT TTC -3’
F17	XF23_Spp381_84to88_r	5’- CGA ATC ACT GTT GTC GCT AC-3’

Table 4.3 Plasmids used in this study

Plasmid #	ID	Notes
1	pAAH1081	pRS416, WT Spp381 (cloned between XhoI and NotI restriction sites), URA3 marker
2	pAAH1082	pRS414, WT Spp381 (cloned between XhoI and NotI restriction sites), TRP1 marker
3	pAAH1088	pAAH1082 + XX primers, 1-216aa Spp381
4	pAAH1111	pAAH1082 + F6/F8 primers, 1-180aa Spp381
5	pAAH1110	pAAH1082 + F5/F7 primers, 1-150aa Spp381
6	pAAH1102	pAAH1082 + F1/F2 primers, 51-291aa Spp381
7	pAAH1103	pAAH1082 + F1/F3 primers, 101-291aa Spp381
8	pAAH1106	pAAH1082 + F1/F4 primers, 111-291aa Spp381
9	pAAH1107	pAAH1082 + F1/F5 primers, 151-291aa Spp381
10	pAAH1108	pAAH1082 + F1/F6 primers, 181-291aa Spp381
11	pAAH1164	pAAH1088 + F1/F5 primers, 151-216aa Spp381
12	pAAH1162	pAAH1111 + F1/F2 primers, 51-180aa Spp381
13	pAAH1163	pAAH1111 + F1/F3 primers, 101-180aa Spp381
14	pAAH1168	pAAH1111 + F1/F15 primers, 82-180aa Spp381
15	pAAH1169	pAAH1166 + F11/F12 primers, 51-180aa, 63-81N Spp381
16	pAAH1165	pAAH1166 + F13/F14 primers, 51-180aa, 74-93N Spp381
17	pAAH1166	pAAH1162 + F9/F10 primers, 51-180aa, 74-81N Spp381
18	pAAH1171	pAAH1162 + F13/F17 primers, 51-180aa, 90-93N Spp381
19	pAAH1170	pAAH1162 + F13/F16 primers, 51-180aa, 84-93N Spp381
20	pAAH1177	pAAH1111 + F13/F16 primers, 1-180aa, 84-93N Spp381
21	pAAH1175	pAAH1111 + F9/F10 primers, 1-180aa, 74-81N Spp381
22	pAAH1176	pAAH1175 + F13/F14 primers, 1-180aa, 74-93N Spp381
23	pAAH1167	pAAH1082 + F9/F10 primers, WT, 74-81N Spp381
24	pXF0020	pAAH1167 + F13/F14 primers, WT, 74-93N Spp381
25	pAAH0470	WT ACT1-CUP1 plasmid, LEU2 marker
26	pAAH0514	A3U ACT1-CUP1 plasmid, LEU2 marker

RESULTS

N- and C-terminal Truncation Mutants Reveal Regions of Spp381 Important for Viability

To systematically understand how each intermolecular interaction region of Spp381 affects yeast growth, we divided Spp381 into the following regions based on existing structural features or characteristic motifs: aa 1-50 (disordered), 51-100 (PEST motif), 101-110 (α1), 111-150 (α2), 151-180 (α3), 181-216 (α4), and 217-291 (disordered) (**Fig. 4.1A, B**). We then created N- and C-terminal truncation mutants that remove one or more of these regions in a TRP1/CEN plasmid containing the SPP381 gene (**Fig. 4.1B**). The *spp381* mutants were then transformed into a yeast strain harboring a deletion of the chromosomal copy of the SPP381 gene and a URA3/CEN-marked plasmid expressing WT Spp381. We select for loss of the plasmid containing WT Spp381 on plates containing -TRP dropout medium and 5FOA at 30^{o}C (**Fig. 4.1C**).

Yeast expressing a N-terminal deletion of the N-terminal disordered region, PEST motif, α1, and α2 were viable (*spp381*$^{151\text{-}291}$); however, a gene in which α3 was deleted and only coding α4 and the C-terminal disordered region was lethal (*spp381*$^{181\text{-}291}$) (**Fig. 4.1D**). This seems to suggest that the C-terminal region, but not the N-terminal region, is essential. However, yeast expressing a C-terminal deletion of α4 and the C-terminal disordered region were also viable (*spp381*$^{1\text{-}180}$). As before, additional deletion of α3 was lethal (*spp381*$^{1\text{-}150}$). While we have not confirmed expression or nuclear localization of the lethal mutants, this suggests that both the N- and C-terminal regions can independently contribute to yeast viability but must also contain α3 to do so. We hypothesize that this could be due to α3's interactions with Prp38 (**Fig. 4.1A, 4.2**) and Prp8^{Endo} that could be important for both formation of the BCP and binding of that complex to the spliceosome during activation. Consistently, this region matches the previously described Prp38 interaction site (Spp381 aa 146-291) obtained from yeast two hybrid analysis (Lybarger *et al.*, 1999).

A Minimal Spp381 Protein Contains the PEST motif and α1- α3

We then combined N- and C-terminal truncation mutants to define a minimal Spp381 polypeptide capable of supporting yeast viability. Yeasts that contain a *spp381* gene encoding for only the PEST motif and α1- α3 was viable (*spp381*$^{51-180}$ **Fig. 4.1E**). Further truncation of this gene to remove the PEST motif did not support yeast growth (*spp381*$^{101-180}$) nor does removal of the C-terminal disordered region if only α3 and α4 are present (*spp381*$^{151-216}$). These results indicate that the PEST motif and C-terminal disordered regions do contribute to yeast viability in the context of truncated Spp381 proteins. Further these results suggest that α3 cannot by itself support yeast growth since several mutants containing this region are not viable (*spp381*$^{101-180}$ and *spp381*$^{151-216}$). This implies that Spp381 has a function beyond providing a binding site for Prp38 and the BCP to Prp8.

To investigate the impact of truncation on yeast growth in further details, we assayed several viable mutants for growth either cold sensitivity (*cs* at 16 or 23°C) or heat sensitivity (*ts* at 37°C) (**Fig. 4.3**). We could not detect a significant *cs* or *ts* growth phenotype for truncations that remove either the N- or C-terminal disordered regions (*spp381*$^{51-291}$ or *spp381*$^{1-216}$) or when the deletions were extended from the N-terminus to include the PEST motif (*spp381*$^{101-291}$) and α2 (*spp381*$^{151-291}$) or from the C-terminus to include α3 (*spp381*$^{1-180}$). Interestingly, C-terminal deletions of the C-terminal disordered region include the region that may correspond to the 5' exon binding domain of Spp381 based on interactions observed for MFAP1 in human spliceosomes (Townsend *et al.*, 2020). This suggests that if such an interaction is present in yeast spliceosomes between Spp381 and pre-mRNA it is not essential, and its deletion does not significantly impact growth at low or high temperatures.

We did observe a *ts* phenotype in the double truncation mutants which either exhibited slow (*spp381*$^{51-180}$) or no growth at 37°C (*spp381*$^{82-180}$) (**Fig. 4.3**). This suggests a stronger

destabilizing effect of these mutants at higher temperatures possibly related to the fewer intermolecular interactions they are expected to make with the spliceosome.

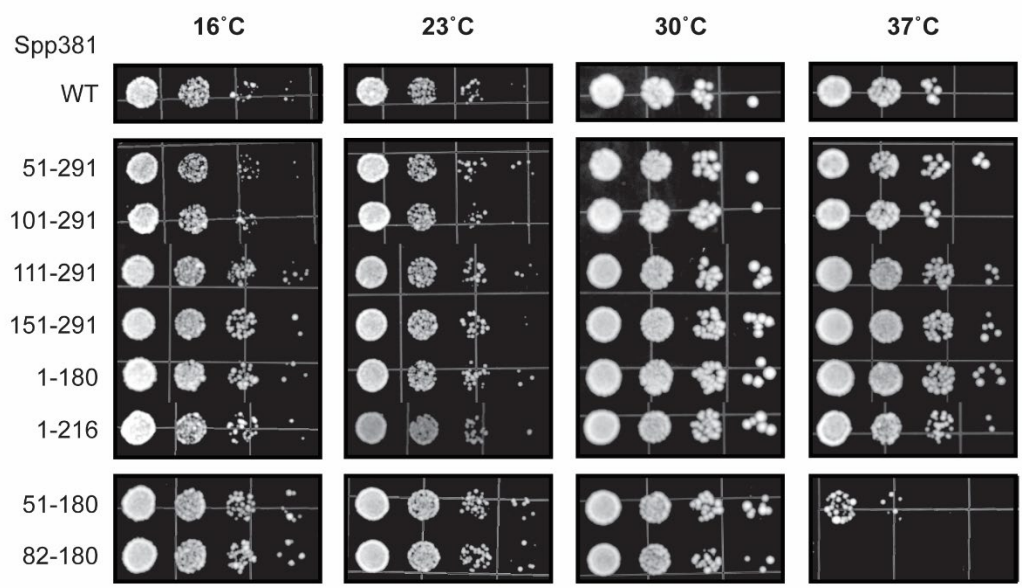


Figure 4.3 Effects of Spp381 Truncation Mutants on Yeast Growth at Various Temperatures.

The indicated *spp381* mutant strains were spotted onto -TRP dropout plates and incubated for either 10 (16ºC) or 3 days (23, 30, and 37ºC) before imaging.

Spp381 Truncation Mutants Impact Splicing In Vivo at a Weak 5' SS

While the *spp381* truncation mutants may not impact yeast viability, they may still perturb splicing. To test this, we carried out ACT1-CUP1 assays in which growth of a Cu^{2+}-sensitive yeast strain is dependent on the amount of spliced mRNA products generated from a ACT1-CUP1 reporter pre-mRNA (**Fig. 4.4A-B**) (Lesser and Guthrie, 1993). We tested either a WT reporter containing consensus splice sites or one containing a suboptimal 5' SS (G/GUAUGU → G/GUUUGU; A3U) (**Fig. 4.4C**) and quantified the *in vivo* splicing level by determining the maximum Cu^{2+} tolerance of each strain at 30°C.

As expected based on the growth assays (**Fig. 4.3**), all of the strains grew similarly to the highest tested Cu^{2+} concentrations (2.5mM) in the presence of the WT ACT1-CUP1 reporter (**Fig. 4.5**). This suggests that the single terminus Spp381 mutants have little effect on the *in vivo* splicing of genes that contain canonical splice sites. However, the strains that contain A3U splicing reporters exhibit a range of copper tolerances (**Fig. 4.5**). Truncation mutants of either the N- or C-terminal disordered regions (*spp381*$^{51-291}$ or *spp381*$^{1-216}$) have no significantly different copper tolerance from WT Spp381 (*spp381*$^{1-291}$, **Fig. 4.4D**). Further truncation from either terminus causes lower copper tolerances and likely a lower splicing level. This indicates that interactions at either terminus of Spp381 become important when introns contain non-canonical splice sites.

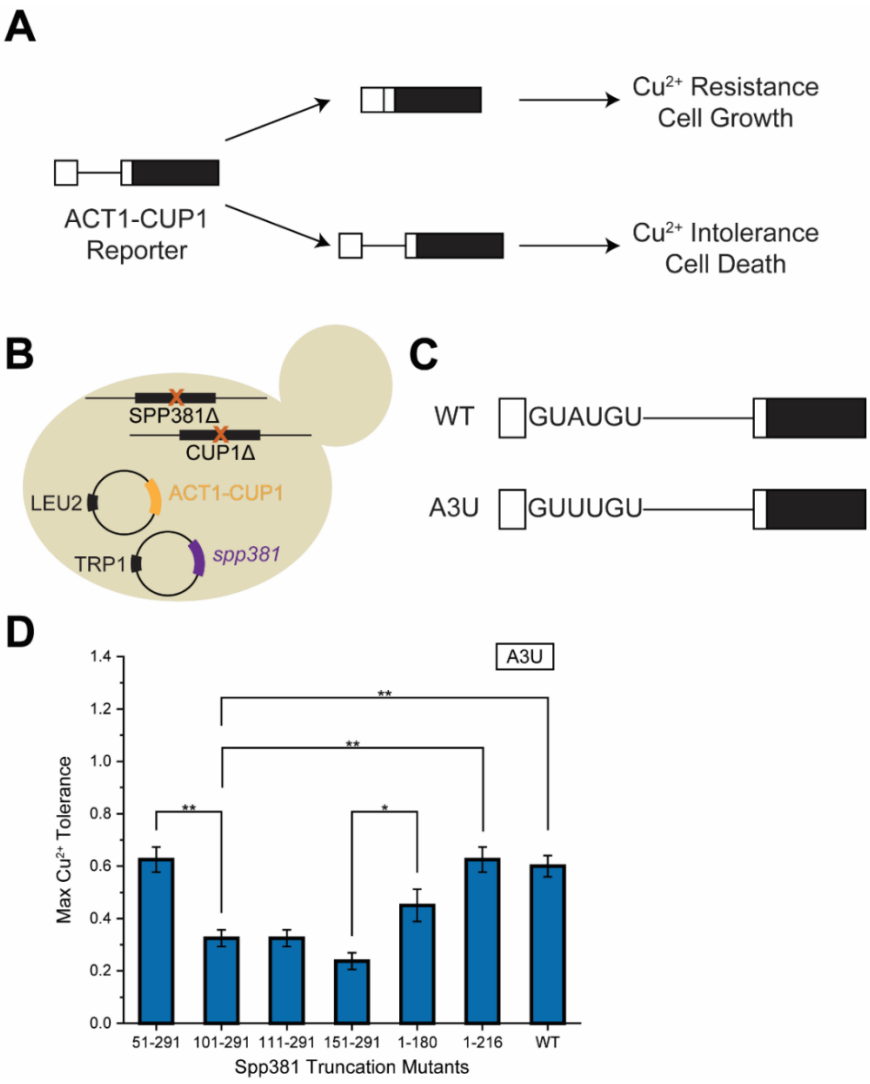


Figure 4.4 Analysis of *in vivo* Splicing for Yeast Containing Truncated Spp381 Proteins Using the ACT1-CUP1 Assay.

(A) Cartoon depiction of the ACT1-CUP1 assay. (B) A generic ACT1-CUP1 assay strain. (C) Schematic of ACT1-CUP1 reporters used for the experiment, with indicated 5′ splice site sequences. WT reporter has canonical splice sites. A3U reporter has the third canonical A mutated to U at 5′ splice site. (D) Bar chart showing the paired mean-comparison plot for the *in vivo* splicing levels of A3U splicing reporter in the indicated *spp381* strains (n = 4). Mean comparison is done by Tukey's test. Only comparisons showing significant differences are marked by brackets and asterisks. *p <= 0.05, **p <= 0.01.

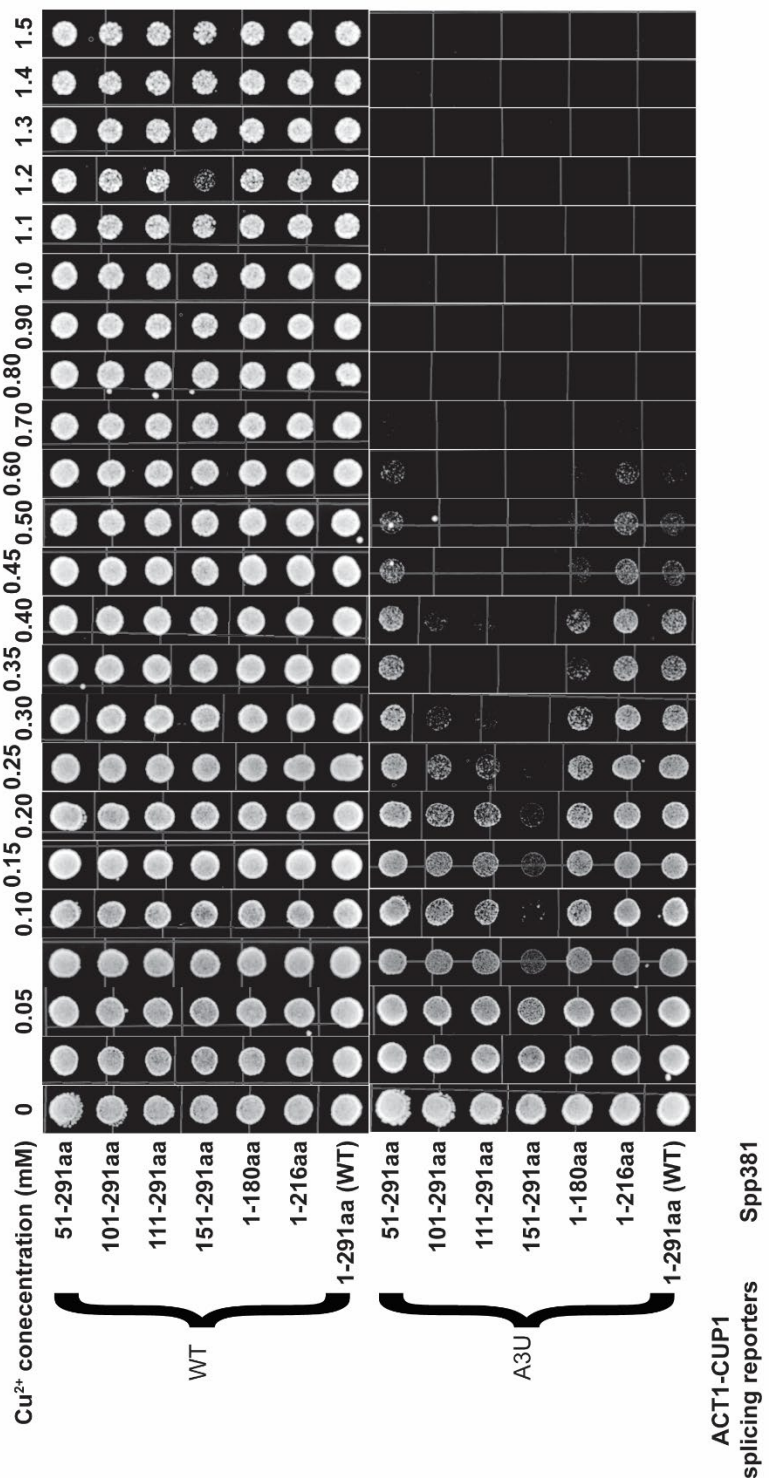


Figure 4.5 Results of ACT1-CUP1 Assays for Spp381 Truncation Mutants.

Each row shows the growth of a particular strain on growth media containing different copper concentrations.

Specific Negatively-Charged Residues in the PEST Motif Affect Yeast Growth in the Minimal Spp381

Deletion of the PEST motif region from the minimal Spp381 is lethal to yeast (**Fig. 4.1E;** *spp381*$^{51\text{-}180}$ vs. *spp381*$^{101\text{-}180}$). To determine if residues that are part of the PEST motif contribute to viability in this mutant, we first aligned PEST motif sequences of Spp381 from different model organisms and identified the most conserved residues within the sequences as those with negative charges (E, D) (**Fig. 4.6A**). To determine if these charged residues are important for yeast growth in the context of *spp381*$^{51\text{-}180}$, we introduced charge neutralizing asparagine point mutations into three regions of the PEST domain (63-73N, 74-81N, 84-93N) (**Fig. 4.6B**).

We first simultaneously introduced the asparagine point mutations into two adjacent regions (63-81 and 74-93N) and found that the latter, but not former, mutations are lethal to yeast at 30ºC (**Fig. 4.6C**). This suggests that negatively charged amino acids within the 84-93 region are important for yeast growth with the minimal Spp381. To test this, we mutated just this region within the minimal Spp381 and detected a slow growth phenotype at 30ºC. The lethality and growth defect phenotypes caused by 74-93N and 84-93N cannot be observed in longer Spp381 proteins (*spp381*$^{1\text{-}180}$ and *spp381*^{WT}) suggesting that other parts of Spp381 (or at least the 1-50aa region) can rescue potential disrupted intermolecular interactions made by the negative-to-neutral charge mutations within the PEST motif.

To understand how yeast growth is impacted by the asparagine point mutations across temperature gradients, we carried out spotting assays at 16°C, 23°C, 30°C, 37°C. *spp381*$^{51\text{-}180}$ mutants containing 63-81N, 74-81N point mutations do not affect yeast growth compared to *spp381*$^{51\text{-}180}$, suggesting that these negatively charged E/D residues have little effect on *in vivo* splicing. In contrast, the asparagine point mutations introduced into aa 84-93 (84-93N, 90-93N) showed more severe growth defects (**Fig. 4.6D**). The 90-93N mutant only showed a *ts* phenotype at 37°C while the 84-93N mutant showed growth defects at multiple temperatures. The full length

Spp381 containing asparagine point mutations in the same regions *(spp381^{WT},* 74-93N) does not exhibit any growth defect. This suggests that these mutations only result in temperature sensitivities in the context of the minimal Spp381.

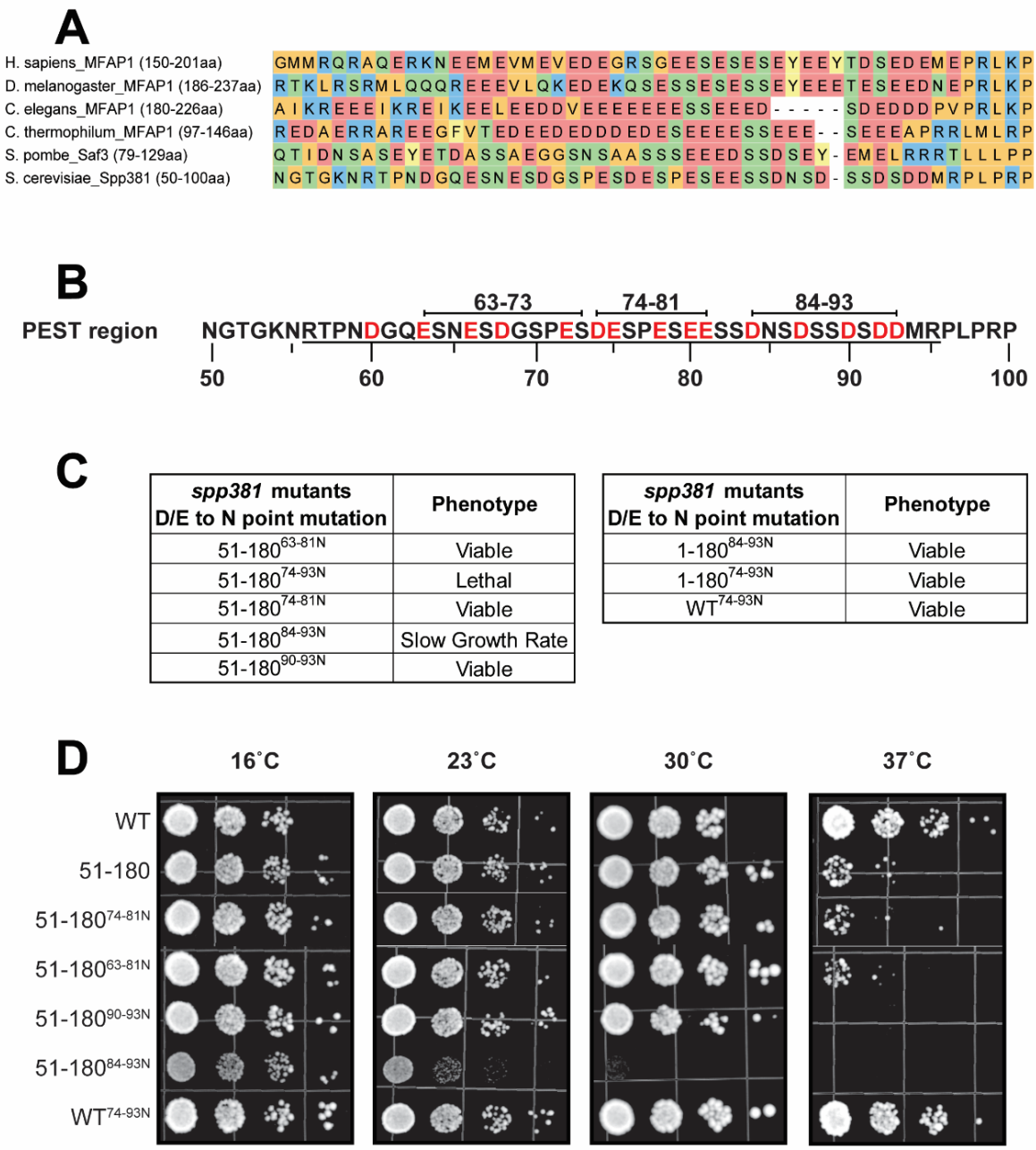


Figure 4.6 A Subset of Negatively-Charged Residues from the PEST region are Essential in Yeast Containing a Minimal Spp381 Protein.

(A) Aligned PEST regions of different Spp381 orthologues across model organisms by MUSCLE (Edgar, 2004). (B) Negatively charged D/E residues in the PEST motif (56-95aa) of Spp381 were manually divided into three regions (63-73aa, 74-81aa, 84-93aa). (C) Summary tables for the

growth of *spp381* strains containing asparagine point mutants on 5FOA selection plates at 30°C. (D) Spotting assay for the yeast strains containing different asparagine mutants at different temperatures as in **Figure 4.3**.

DISCUSSION

While the Prp38 and Spp381 (146-291aa) interaction was shown to be important for *in vivo* splicing some time ago (Lybarger *et al.*, 1999), the molecular details of this interaction remained elusive until only recently due to a lack of structural data. Here, we used a recent cryo-EM structure of Spp381 in the yeast pre-catalytic B spliceosome (Plaschka *et al.*, 2017) as a guide for performing truncation analysis of the protein. Our data suggest that the most essential region of Spp381 is aa151-180 residues (**Fig. 4.1D**), which appear to structurally interact with Prp38 and potentially with different domains of Prp8 ($Prp8^{Endo}$ and $Prp8^{2090\text{-}2150}$) in the spliceosome (**Fig. 4.2**).

Truncation mutants missing either a majority of the N- or C-teriminal domains (but not both) can be viable if they also contain the aa 150-181 region (**Fig. 4.3, 4.5**). We interpret this as Spp381 using functionally redundant N- and C-termini to make stable intermolecular interactions at corresponding sites in spliceosome for ensuring that the $Spp381^{151\text{-}180}$/Prp38 containing BCP complex can correctly dock onto the $Prp8^{NTD}$ (**Fig. 4.2**). The fact that either terminus can be substituted by the other suggests that the identity of their physical interaction sites is not important for splicing but being able to provide strong interaction sites with the spliceosome from either terminus matters. This mechanism may have implications for the recruitment of B complex proteins to the spliceosome or for stabilizing the location of B complex proteins in spliceosome while performing their presumed function during activation.

Our results also suggest that the putative PEST domain likely only plays a significant role in yeast splicing in the context of a minimal Spp381 protein containing both N- and C-terminal truncations. How this functions during splicing is unclear since the identity of its interaction partners remain unknown. Some potential candidates include a positively charged patch on the surface of Brr2 or with RNA within the 5' exon exit channel or the channel itself (**Fig. 4.7**). Interestingly, the C-terminal region of the human Spp381 homolog, MFAP1 has been shown to interact with the 5’ exon in human pre-B^{ACT1}. The corresponding C-terminus of Spp381 is not

essential for yeast growth suggesting that such an interaction is not required for yeast splicing. However, Spp381 mutants do impact splicing of a pre-mRNA with a 5’ SS mutant suggesting that the protein could play a minor role in stabilizing the interaction of the 5’ SS region of the substrate with the spliceosome.

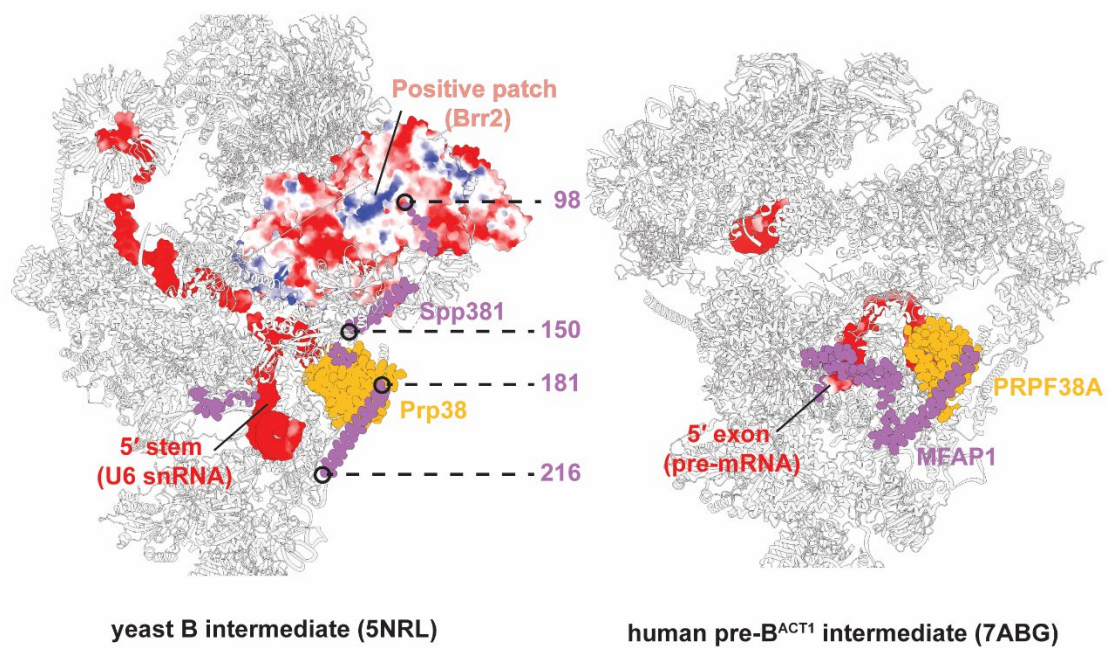


Figure 4.7 Potential Interaction Sites for Spp381 in spliceosomes.

Potential interaction sites for identified conditionally essential regions of yeast Spp381 are shown. Images showing Spp381 interaction partners in yeast B intermediates (Left: 5NRL) and human pre-B^{ACT1} intermediates (Right: 7ABG). The essential negatively charged residues (aa 84-93) from PEST region may interact with a positively charged patch on adjacent Brr2. The identified essential region (aa 151-180) from Spp381 truncation studies appears to interact with Prp38. The identified essential and disordered C-terminus of Spp381 may interact with the negatively charged RNAs within the exon exit channel (Left: 5' stem of U6 snRNA; Right: 5' exon of a model pre-mRNA). Coulombic electrostatic potential maps for Brr2, U6 snRNA and 5' exon were automatically calculated and displayed in red-white-blue color gradient in ChimeraX program, with red and blue being the negative and positive extreme values.

REFERENCES

Boon, K.L., Grainger, R.J., Ehsani, P., Barrass, J.D., Auchynnikava, T., Inglehearn, C.F., and Beggs, J.D. (2007). prp8 mutations that cause human retinitis pigmentosa lead to a U5 snRNP maturation defect in yeast. Nat Struct Mol Biol *14*, 1077-1083. 10.1038/nsmb1303.

Brow, D.A. (2019). An Allosteric Network for Spliceosome Activation Revealed by High-Throughput Suppressor Analysis in. Genetics *212*, 111-124. 10.1534/genetics.119.301922.

Christiano, R., Nagaraj, N., Fröhlich, F., and Walther, T.C. (2014). Global proteome turnover analyses of the Yeasts S. cerevisiae and S. pombe. Cell Rep *9*, 1959-1965. 10.1016/j.celrep.2014.10.065.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res *32*, 1792-1797. 10.1093/nar/gkh340.

Goldstein, A.L., and McCusker, J.H. (1999). Three new dominant drug resistance cassettes for gene disruption in Saccharomyces cerevisiae. Yeast *15*, 1541-1553. 10.1002/(SICI)1097-0061(199910)15:14<1541::AID-YEA476>3.0.CO;2-K.

Lesser, C.F., and Guthrie, C. (1993). Mutational analysis of pre-mRNA splicing in Saccharomyces cerevisiae using a sensitive new reporter gene, CUP1. Genetics *133*, 851-863. 10.1093/genetics/133.4.851.

Lybarger, S., Beickman, K., Brown, V., Dembla-Rajpal, N., Morey, K., Seipelt, R., and Rymond, B.C. (1999). Elevated levels of a U4/U6.U5 snRNP-associated protein, Spp381p, rescue a mutant defective in spliceosome maturation. Mol Cell Biol *19*, 577-584. 10.1128/MCB.19.1.577.

Maeder, C., Kutach, A.K., and Guthrie, C. (2009). ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C terminus of Prp8. Nat Struct Mol Biol *16*, 42-48. 10.1038/nsmb.1535.

Plaschka, C., Lin, P.C., and Nagai, K. (2017). Structure of a pre-catalytic spliceosome. Nature *546*, 617-621. 10.1038/nature22799.

Rechsteiner, M., and Rogers, S.W. (1996). PEST sequences and regulation by proteolysis. Trends Biochem Sci *21*, 267-271.

Rogers, S., Wells, R., and Rechsteiner, M. (1986). Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. Science *234*, 364-368. 10.1126/science.2876518.

Townsend, C., Leelaram, M.N., Agafonov, D.E., Dybkov, O., Will, C.L., Bertram, K., Urlaub, H., Kastner, B., Stark, H., and Lührmann, R. (2020). Mechanism of protein-guided folding of the active site U2/U6 RNA during spliceosome activation. Science *370*. 10.1126/science.abc3753.

Ulrich, A.K., and Wahl, M.C. (2017). Human MFAP1 is a cryptic ortholog of the Saccharomyces cerevisiae Spp381 splicing factor. BMC Evol Biol *17*, 91. 10.1186/s12862-017-0923-1.

Chapter Five

Saccharomyces cerevisiae Ecm2 Modulates the Catalytic Steps of pre-mRNA Splicing

This chapter is published in the following form:

van der Feltz, C., Nikolai, B., Schneider, C., Paulson, J.C., Fu, X., Hoskins, A.A. (2021) Saccharomyces cerevisiae Ecm2 Modulates the Catalytic Steps of pre-mRNA Splicing. *RNA,* 27(5), 591-603.

Author contributions:

This work was led by van der Feltz, C. Fu, X. contributed to the experiment in Figure 5.8B.

ABSTRACT

Genetic, biochemical, and structural studies have elucidated the molecular basis for spliceosome catalysis. Splicing is RNA catalyzed and the essential snRNA and protein factors are well-conserved. However, little is known about how non-essential components of the spliceosome contribute to the reaction and modulate the activities of the fundamental core machinery. Ecm2 is a non-essential yeast splicing factor that is a member of the Prp19-related complex of proteins. Cryo-electron microscopy (cryo-EM) structures have revealed that Ecm2 binds the U6 snRNA and is entangled with Cwc2, a factor previously found to promote a catalytically active conformation of the spliceosome. These structures also indicate that Ecm2 and the U2 snRNA likely form a transient interaction during 5' splice site (SS) cleavage. We have characterized genetic interactions between ECM2 and alleles of splicing factors that alter the catalytic steps in splicing. In addition, we have studied how loss of ECM2 impacts splicing of pre-mRNAs containing non-consensus or competing SS. Our results show that ECM2 functions during the catalytic stages of splicing. Our data are consistent with Ecm2 facilitating the formation and stabilization of the 1st-step catalytic site, promoting 2nd-step catalysis, and permiting alternate 5' SS usage. We propose that Cwc2 and Ecm2 can each fine-tune the spliceosome active site in unique ways. Their interaction network may act as a conduit through which splicing of certain pre-mRNAs, such as those containing weak or alternate splice sites, can be regulated.

INTRODUCTION

Spliceosomes are composed of small ribonucleoprotein particles (snRNPs), each containing proteins and a small nuclear RNA (U1, U2, U4, U5, or U6 snRNA), and dozens of additional protein splicing factors. Spliceosomes assemble from these factors before undergoing a number of conformational changes to form a catalytic center (activation) capable of carrying out the chemical steps of splicing (**Fig. 5.1A**): 5' splice site (SS) cleavage (1st step) and exon ligation (2nd step). Significant genetic, biochemical, and structural work over the past few decades has provided a wealth of information into how essential components of the splicing reaction such as the snRNAs, Prp8 protein, and DExD/H-box ATPases promote splicing (Kastner et al. ; Wahl et al. 2009; Mayerle et al. 2017; Kastner et al. 2019; Plaschka et al. 2019; Yan et al. 2019). In comparison, much less is known about how non-essential factors modulate the splicing reaction and interact with the core machinery.

Several of the non-essential splicing factors in yeast are associated with the Prp19-containing complex (NTC). Indeed, of 26 yeast proteins categorized as core-NTC components or NTC-associated, 12 can be considered as non-essential under certain growth conditions (*i.e.,* deletion strains may possess phenotypes but nonetheless proliferate in permissive environments) (Hogg et al. 2010; Hogg et al. 2014). Many of these proteins are well-conserved and have human splicing factor homologs. In general, the NTC is thought to stabilize catalytic conformations of the U6 snRNA and contribute to splicing fidelity (Hogg et al. 2010). Consistent with this model, cryo-EM structures and biochemical assays have shown that several non-essential NTC proteins directly interact with U6 including Cwc2, Ecm2, and Isy1 (Villa and Guthrie 2005; McGrail et al. 2009; Hogg et al. 2010; Rasche et al. 2012; Plaschka et al. 2019). Exactly how the NTC modulates RNA interactions within the spliceosome is not yet clear (Hogg et al. 2010), and it is difficult to infer potential mechanisms and the impact of mutations from cryo-EM structures alone (Mayerle and Guthrie 2017).

The NTC-associated protein Ecm2 was first isolated in a synthetic lethality screen for genetic interactors with U2 snRNA mutations (synthetic lethality with U2/*slt* screen) (Xu et al. 1998). This screen identified *slt11*/*ecm2* as well as other splicing factors including Prp8 and Brr2. Additional work demonstrated genetic interactions between *ecm2* and multiple components of the spliceosome but especially U2/U6 helix I and helix II mutants (Xu et al. 1998; Xu and Friesen 2001). Biochemical assays of splicing and spliceosome assembly showed that absence of Ecm2 results in loss of splicing activity at high temperatures and a block in spliceosome activation (Xu and Friesen 2001). Since it was known that U2/U6 form an intermolecular duplex during the early stages of activation (helix II) (Wassarman and Steitz 1992), it was proposed that Ecm2 functions during spliceosome activation to facilitate formation this duplex (Xu and Friesen 2001).

Structures of Ecm2 integrated into a number of spliceosome complexes have been determined by cryo-EM (Plaschka et al. 2019). Ecm2 contains two RNA binding domains separated by a linker: an N-terminal zinc finger motifs (ZNF) domain and a C-terminal RNA recognition motif (RRM) (**Fig. 5.1B**). Unexpectedly, Ecm2 does not directly bind U2/U6 helix II. Rather the ZNF domain interacts with the U6 nucleotides (nt) 29-32, which are located between the U6 5' stem loop and the ACAGAGA-box/5' SS pairing region (**Fig. 5.1C**). Ecm2 is also intertwined with another NTC-associated protein, Cwc2. Cwc2 contacts both the intronic RNA downstream of the 5' SS and the U6 snRNA at multiple locations.

Interestingly, the C-terminal RRM of Ecm2 closely approaches U2 snRNA stem IIb in the C complex spliceosome (**Fig. 5.1C**). Due to low resolution within this region, the exact molecular contacts between the RRM, U2 stem IIb, and nearby regions of Cwc2 are unclear. This interaction is likely transient: the Ecm2/U2 interaction has only been observed in structures captured just before (B* complex) and after (C complex) the 1st step of splicing (Galej et al. 2016; Wan et al. 2019). Large conformational changes place U2 stem IIb far away from any possible Ecm2 interaction in other structures (**Fig. 5.1A**, **C**) (Rauhut et al. 2016; Yan et al. 2016; Fica et al. 2017;

Yan et al. 2017). Toggling of Ecm2/U2 stem IIb contacts on-or-off in different complexes resembles structural toggle switches reported for other splicing factors. These include the RNaseH domain of Prp8, the U4/U6 di-snRNA, as well as interconversion of U2 stem II itself between two mutually exclusive structures: stem IIa/b and stem IIb/c (Hilliker et al. 2007; Perriman and Ares 2010; Rodgers et al. 2016a; Rodgers et al. 2016b; Abelson 2017; Mayerle et al. 2017). The significance of the Ecm2/stem IIb interaction has not been studied.

Human spliceosomes do not contain direct homologs of Cwc2 and Ecm2. Instead, a single protein, RBM22, binds the U6 snRNA at the corresponding positions. Based on limited sequence homology, it has previously been proposed that RBM22 represents a fusion of Cwc2 and Ecm2 (Rasche et al. 2012). This is supported by biochemical studies that show a similar function for Cwc2 and RBM22 in stabilizing the spliceosome active site (McGrail et al. 2009; Rasche et al. 2012; Hogg et al. 2014). In addition, both Cwc2 and RBM22 interact with intronic RNA at a location downstream of the 5'SS (Rasche et al. 2012; Kastner et al. 2019). Whether or not Ecm2 and RBM22 also share any conserved functions is unknown.

Here, we report genetic interactions between Ecm2 and splicing factors capable of modulating the 1st and 2nd steps of splicing including the Prp2 and Prp16 ATPases, Prp8, and the U2 and U6 snRNAs. Ecm2 also exhibits genetic interactions with mutations that disrupt U2 stem II toggling, consistent with a functional interaction between the protein and U2. Genetic deletion of ECM2 changes how pre-mRNAs containing non-consensus splice sites are processed, implicating Ecm2 in the catalytic steps of splicing in addition to its role in activation. Our results support a model in which Ecm2 has distinct functions for each catalytic step and are consistent with a proposal that several conditionally non-essential splicing factors (Ecm2, Cwc2, Isy1) function as a hub for regulating spliceosome catalysis (Hogg et al. 2010). These results have implications for the function of RBM22 in human spliceosomes as well as for how RBM22/intronic RNA interactions are formed.

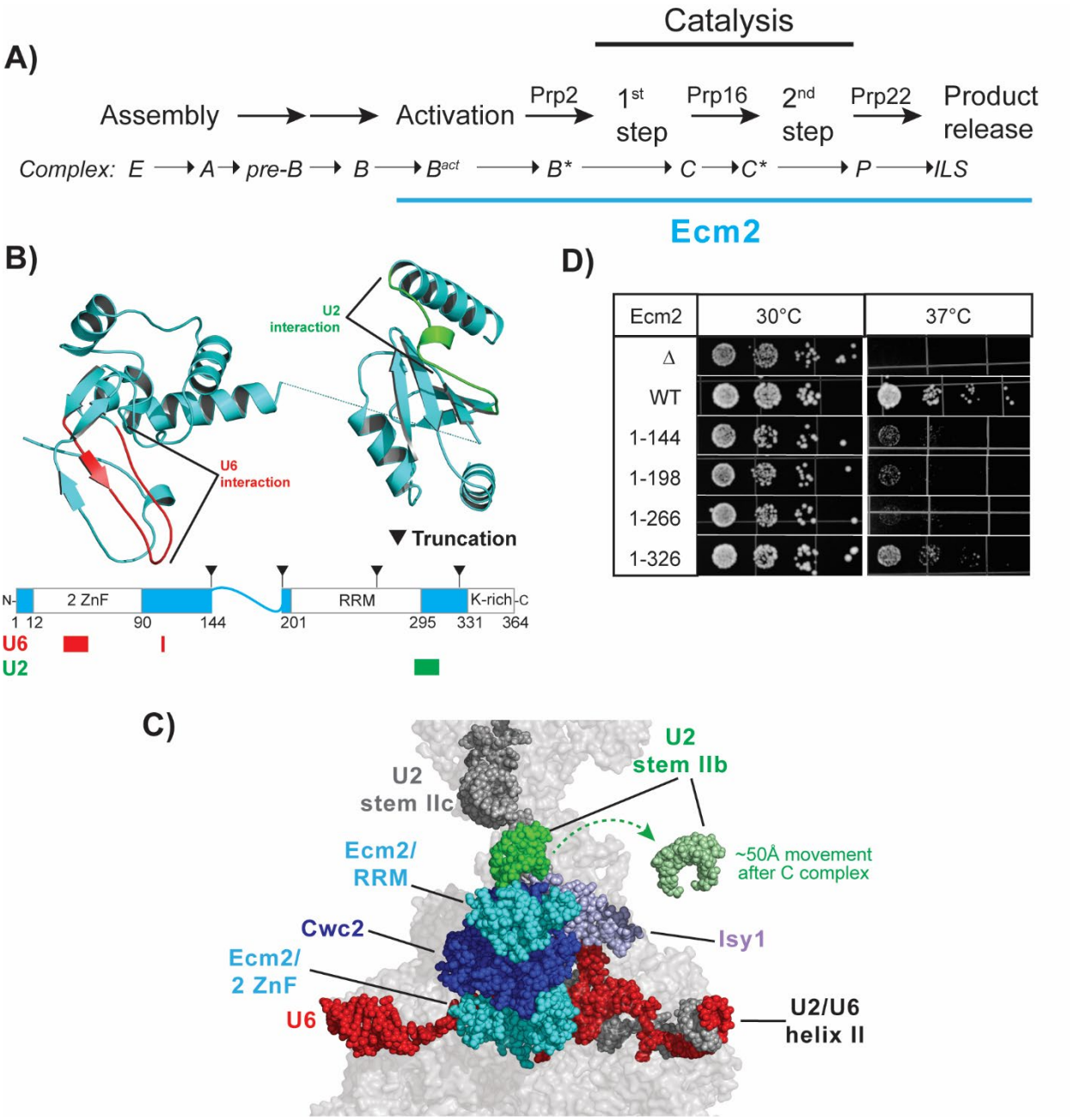


Figure 5.1 Structural Analysis of Ecm2 during Splicing.

(A) Schematic of the pre-mRNA splicing pathway. ATPases tested for genetic interactions with Ecm2 are shown above the arrows of the respective steps that they promote. Spliceosome complexes containing Ecm2 are underlined in blue. (B) Cryo-EM structure and domain organization of Ecm2. U6 and U2 snRNA interacting regions are marked and colored in red and

green, respectively. Locations of truncation mutants studied in panel (C) are indicated. Structure from 6EXN.pdb. (C) Structure of the Cwc2/Ecm2/Isy1 hub and Ecm2-RRM/U2 stem II interaction in C complex. The position of stem IIb after remodeling in C* complex has been superimposed on this structure. The U6 snRNA, Cwc2 and Ecm2 do not significantly change positions in C* complex. This figure was created using 5LJ5.pdb and 5MQ0.pdb. (D) Temperature sensitivity of *ecm2D* and truncation mutants on -ura dropout media after 3 days of growth.

MATERIALS AND METHODS

Yeast strains and plasmids used in these studies are described in **Tables 5.1** and **5.2**. Yeast transformation and growth were carried out using standard techniques and media.

Genetic Deletions of ECM2

Deletion of the ECM2 gene was carried out by replacement of the gene with an antibiotic resistance cassette (hygromycin or nourseothricin) by homologous recombination in the appropriate parental strain (**Tables 5.1**, (Goldstein and McCusker 1999)).

Cloning of ECM2 and Site-Directed Mutagenesis

ECM2 along with 300 base pairs of up- and downstream DNA was amplified from yeast genomic DNA by PCR. The resulting product was digested with NotI and SalI restriction enzymes and then ligated into pRS416 (URA3 CEN6) at those same restriction sites to create plasmid pAAH1056 containing the WT ECM2 gene. Novel mutants of Ecm2 were generated using inverse polymerase chain reaction (PCR) with Phusion DNA polymerase (New England Biolabs; Ipswich, MA). All plasmids were confirmed by sequencing.

Western Blot Analysis

A 3xFLAG epitope tag was genetically encoded at the C-terminus of Ecm2 or truncation mutants by synthesis of the corresponding genes (Genewiz). Strains expressing the epitope tagged proteins were grown in -ura dropout media and 10 mL of cells were collected at an OD_{600} of ~1.0. Cellular protein was precipitated with tricholoracetic acid (TCA) and resuspended in 2xSDS-PAGE loading buffer. Proteins were separated on a 12% SDS-PAGE gel before being transferred to a nitrocellulose membrane using a semi-dry transfer apparatus and according to the manufacturer's directions (Pierce Power Blotter, Thermo Fisher). The blot was then stained with 0.5% w/v Ponceau S (Sigma) in 2% v/v acetic acid and imaged. The stain was removed by washing with 200 µM NaOH followed by water. The gel was blocked, incubated with HRP-linked

anti-FLAG antibodies (Sigma), and washed using an iBind cassette (Invitrogen). Bands were detected using Clarity Western ECL substrate (Bio-Rad) and chemiluminescent imaging. Bands were quantified using ImageQuant TL v8.1 software (GE Healthcare) and normalized based on total protein content detected by Ponceau staining (Romero-Calvo et al. 2010).

ACT1-CUP1 Copper Tolerance Assays

Yeast strains expressing *ACT1-CUP1* reporters were grown to mid-log phase in -leu -trp dropout media to maintain selection for the plasmids, adjusted to OD_{600} = 0.5 and equal volumes were spotted onto plates containing 0-2.5 mM $CuSO_4$ (Lesser and Guthrie 1993a; Carrocci et al. 2018). Plates were scored and imaged after 3 days growth at 30°C.

Temperature Growth Assays

Yeast strains were grown to mid-log phase at permissive temperatures in YPD or -ura dropout liquid media. Cell growth was then quantified by measuring OD_{600}. Equal volumes of cells were diluted to an OD_{600} = 0.5 were stamped onto YPD-agar plates and incubated at 23°C, 30°C or 37°C for three days or at 16°C for ten days before imaging.

Primer Extension

Cells were grown at 30°C in 25 mL yeast -leu -trp dropout liquid media until OD_{600} reached 0.5–0.8, and 10 OD_{600} units were collected by centrifugation. Total yeast RNA was isolated following the MasterPure™ Yeast RNA Purification Kit (Epicenter, Madison, WI) protocol with minor changes as previously described (Carrocci et al. 2017). Primer extension was performed with IR dye conjugated probes yAC6: /5IRD700/GGCACTCATGACCTTC and yU6: /5IRD700/GAACTGCTGATCATCTCTG. purchased from Integrated DNA Technologies (Skokie, IL USA) (Carrocci et al. 2017; Kaur et al. 2020). Gels were imaged with the Amersham IR Typhoon 5 (GE Healthcare) excitation at 685 nm, emission filter 720BP20, PMT voltage of 700V, and 100 µm pixel size. Band intensities were quantified with ImageQuant TL v8.1 (GE Healthcare).

Table 5.1 Yeast strains used in this study.

Name	Genotype	Description	Source or Reference
yAAH1930	MATa ade2 cup1D::ura3 his3 leu2 lys2 trp1 ura3 GAL prp22D::loxP p360-22 (Prp22 URA CEN ARS)	Cu^{2+}-sensitive shuffle strain for Prp22. yMK-02	Charles Query/Magda Konarska
yAAH1931	yAAH1930 + pAAH1042	$Prp22^{WT}$	Charles Query/Magda Konarska
yAAH1932	yAAH1930 + pAAH1043	$Prp22^{T637A}$	Charles Query/Magda Konarska
yAAH1937	yAAH1930 + ECM2::hphMX	Cu^{2+}-sensitive, ecm2D shuffle strain for Prp22	This work.
yAAH1943	yAAH1937 + pAAH1042	$Prp22^{WT}$, ecm2D	This work.
yAAH2482	yAAH1937 + pAAH1043	$Prp22^{T637A}$, ecm2D	This work.
yAAH2044	yAAH1943 + pAAH0137	$Prp22^{WT}$, ecm2 D, empty URA3 plasmid	This work.
yAAH2045	yAAH1943 + pAAH1056	$Prp22^{WT}$, ecm2 D, ($ECM2^{WT}$ URA3 CEN ARS)	This work.
yAAH2047	yAAH1943 + pAAH1067	$Prp22^{WT}$, ecm2 D, ($ECM2^{1-143}$ URA3 CEN ARS)	This work.
yAAH2047	yAAH1943 + pAAH1068	$Prp22^{WT}$, ecm2 D, ($ECM2^{1-197}$ URA3 CEN ARS)	This work.
yAAH2047	yAAH1943 + pAAH1196	$Prp22^{WT}$, ecm2 D, ($ECM2^{1-265}$ URA3 CEN ARS)	This work.
yAAH2047	yAAH1943 + pAAH1197	$Prp22^{WT}$, ecm2 D, ($ECM2^{1-325}$ URA3 CEN ARS)	This work.
yAAH2257	MATa cup1D ura3 his3 trp1 lys2 ade2 leu2 hsh155::kanMX prp2::hphMX ecm2::natMX + (HSH155 TRP1 CEN ARS) + (PRP2 URA3 CEN ARS)	Cu^{2+}-sensitive, ecm2D shuffle strain for Prp2	This work.
yAAH2258	yAAH2257 + pAAH0778	$Prp2^{WT}$, ecm2D	This work.
yAAH2259	yAAH2257 + pAAH0790	$Prp2^{Q548N}$, ecm2D	This work.
yAAH1915	MATa cup1D ura3 his3 trp1 lys2 ade2 leu2 hsh155::kanMX prp2::hphMX + (HSH155 TRP1 CEN ARS) + ($PRP2^{WT}$ LEU2 CEN ARS)	$Prp2^{WT}$	(Carrocci et al. 2017)
yAAH1916	MATa cup1D ura3 his3 trp1 lys2 ade2 leu2 hsh155::kanMX prp2::hphMX + (HSH155 TRP1 CEN ARS) + ($PRP2^{Q548N}$ LEU2 CEN ARS)	$Prp2^{Q548N}$	(Carrocci et al. 2017)
yAAH0130	MATa trp1 ura3 his lys2 leu2 ade2 prp16D::lys2 pSB2-Prp16 (PRP16 URA)	shuffle strain for Prp16, yS78	Christine Guthrie (Wang and Guthrie 1998)
yAAH1936	yAAH0130 + ECM2::hphMX	ecm2D shuffle strain for Prp16	This work.
yAAH1946	yAAH0130 + pAAH1040	$Prp16^{WT}$	This work.
yAAH1947	yAAH0130 + pAAH1039	$Prp16^{R686I}$	This work.
yAAH1941	yAAH1936 + pAAH1040	$Prp16^{WT}$, ecm2D	This work.

yAAH1942	yAAH1936 + pAAH1039	$Prp16^{R686I}$, ecm2D	This work.
yAAH0434	MATa cup1D ura3 his3 trp1 lys2 ade2 leu2	Cu^{2+}-sensitive strain	David Brow
yAAH1691	yAAH0434 + ECM2::hphMX	Cu^{2+}-sensitive, ecm2D strain	This work.
yAAH2490	yAAH2435+pAAH0470	$Prp8^{WT}$ Cu^{2+}-sensitive strain with WT ACT-CUP1	This work.
yAAH2491	yAAH2446+pAAH0470	$Prp8^{WT}$ $Cu2^{+}$-sensitive, ecm2D strain with WT ACT-CUP1	This work.
yAAH2492	yAAH2435+pAAH1032	$Prp8^{WT}$ Cu^{2+}-sensitive strain with A3C ACT-CUP1	This work.
yAAH2493	yAAH2446+pAAH1032	$Prp8^{WT}$ Cu^{2+}-sensitive, ecm2D strain with A3C ACT-CUP1	This work.
yAAH2494	yAAH2435+pAAH0514	$Prp8^{WT}$ Cu^{2+}-sensitive strain with A3U ACT-CUP1	This work.
yAAH2495	yAAH2446+pAAH0514	$Prp8^{WT}$ Cu^{2+}-sensitive, ecm2D strain with A3U ACT-CUP1	This work.
yAAH2496	yAAH2435+pAAH0517	$Prp8^{WT}$ Cu^{2+}-sensitive strain with G5A ACT-CUP1	This work.
yAAH2497	yAAH2446+pAAH0517	$Prp8^{WT}$ Cu^{2+}-sensitive, ecm2D strain with G5A ACT-CUP1	This work.
yAAH2498	yAAH2435+pAAH1058	$Prp8^{WT}$ Cu^{2+}-sensitive strain with cyptic 5'SS ACT1-CUP1	This work.
yAAH2499	yAAH2446+pAAH1058	$Prp8^{WT}$ Cu^{2+}-sensitive, ecm2D strain with cryptic 5' SS ACT-CUP1	This work.
yAH2500	yAAH2435+pAAH0439	$Prp8^{WT}$ Cu^{2+}-sensitive strain with U257C ACT-CUP1	This work.
yAH2501	yAAH2446+pAAH0439	$Prp8^{WT}$ Cu^{2+}-sensitive, ecm2D strain with U257C ACT-CUP1	This work.
yAH2502	yAAH2435+pAAH0524	$Prp8^{WT}$ Cu^{2+}-sensitive strain with A258U ACT-CUP1	This work.
yAH2503	yAAH2446+pAAH0524	$Prp8^{WT}$ Cu^{2+}-sensitive, ecm2D strain with A258U ACT-CUP1	This work.
yAH2504	yAAH2435+pAAH0441	$Prp8^{WT}$ Cu^{2+}-sensitive strain with A259C, BS-C ACT-CUP1	This work.
yAH2505	yAAH2446+pAAH0441	$Prp8^{WT}$ Cu^{2+}-sensitive, ecm2D strain with A259C, BS-C ACT-CUP1	This work.
yAH2506	yAAH2435+pAAH0880	$Prp8^{WT}$ Cu^{2+}-sensitive strain with A259G, BS-G ACT-CUP1	This work.
yAH2507	yAAH2446+pAAH0880	$Prp8^{WT}$ Cu^{2+}-sensitive, ecm2D strain with A259G, BS-G ACT-CUP1	This work.
yAH2508	yAAH2435+pAAH0527	$Prp8^{WT}$ Cu^{2+}-sensitive strain with A302U ACT-CUP1	This work.

yAH2509	yAAH2446+pAAH0527	$Prp8^{WT}$ Cu^{2+}-sensitive, ecm2D strain with A302U ACT-CUP1	This work.
yAAH2449	yAAH1911 + pAAH1015	U2-WT	This work.
yAAH2455	yAAH1913 + pAAH1015	U2-WT, ecm2D	This work.
yAAH2453	yAAH1911 + pAAH1020	U2-IIc+	This work.
yAAH2460	yAAH1913 + pAAH1020	U2-IIc+, ecm2D	This work.
yAAH2452	yAAH1911 + pAAH1018	U2-2,4	This work.
yAAH2458	yAAH1913 + pAAH1018	U2-2,4, ecm2D	This work.
yAAH2451	yAAH1911 + pAAH1017	U2-G53A	This work.
yAAH2457	yAAH1913 + pAAH1017	U2-G53A, ecm2D	This work.
yAAH2453	yAAH1911 + pAAH1019	U2-DIIb	This work.
yAAH2459	yAAH1913 + pAAH1019	U2-DIIb, ecm2D	This work.
yAAH2450	yAAH1911 + pAAH1016	U2-DCC	This work.
yAAH2456	yAAH1913 + pAAH1016	U2-DCC, ecm2D	This work.
yAAH1911	U2 shuffle MATa ade2 cup1D::ura3 his3 leu2 lys2 prp5D::loxP trp1 snr20D::loxP (pRS314-PRP5 (PRP5 TRP1 CEN ARS) pRS316-U2 (SNR20 URA3 CEN ARS)	Cu^{2+}-sensitive shuffle strain for U2 and Prp5, yZX11.	(Xu and Query 2007)
yAAH1913	yAAH1911 + ECM2::hphMX	Cu^{2+}-sensitive shuffle strain for U2 and Prp5, ecm2D	This work.
yAAH0117	MATa, cup1D ura3 his3 leu2 lys2 prp8D:lys2 trp1 pJU169 (PRP8 URA3)	Cu^{2+}-sensitive shuffle strain for Prp8, yJU75	Christine Guthrie (Umen and Guthrie 1995)
yAAH1886	yAAH0117 + ECM2::hphMX	Cu^{2+}-sensitive shuffle strain for Prp8, ecm2D	This work.
yAAH2435	yAAH0117 + pAAH0997	$Prp8^{WT}$	This work.
yAAH2446	yAAH1886+ pAAH0997	$Prp8^{WT}$, ecm2D	This work.
yAAH2439	yAAH0117 + pAAH1045	$Prp8^{V1870N}$ (*prp8-162*)	This work.
yAAH2440	yAAH1886 + pAAH1045	$Prp8^{V1870N}$ (*prp8-162*), ecm2D	This work.
yAAH2441	yAAH0117 + pAAH1003	$Prp8^{E1960K}$ (*prp8-101*)	This work.
yAAH2442	yAAH1886 + pAAH1003	$Prp8^{E1960K}$ (*prp8-101)*, ecm2D	This work.
yAAH2437	yAAH0117 + pAAH1001	$Prp8^{P986T}$ (*prp8-161*)	This work.
yAAH2438	yAAH1886 + pAAH1001	$Prp8^{P986T}$ (*prp8-161)*, ecm2D	This work.
yAAH2443	yAAH0117 + pAAH1004	$Prp8^{R1753K}$	This work.
yAAH2444	yAAH1886 + pAAH1004	$Prp8^{R1753K}$, ecm2D	This work.
yAAH2445	yAAH0117 + pAAH1006	$Prp8^{P986T/R1753K}$	This work.
yAAH2446	yAAH1886 + pAAH1006	$Prp8^{P986T/R1753K}$, ecm2D	This work.
yAAH2447	yAAH0117 + pAAH1007	$Prp8^{E1960K/V1870N}$	This work.
yAAH2448	yAAH1886+ pAAH1007	$Prp8^{E1960K/V1870N}$, ecm2D	This work.
yAAH1908	yAAH0434 + snr6::natMX + pRS316-U4U6mini (SNR4 SNR6 URA3)	Cu^{2+}-sensitive shuffle strain for U6 snRNA	This work.
yAAH1910	yAAH1908 + ECM2::hphMX	Cu^{2+}-sensitive shuffle strain for U6 snRNA, ecm2D	This work.

yAAH2434	yAAH1910 + pAAH0137 + pAAH0412	WT U6 + Empty URA	This work.
yAAH2433	yAAH1910 + pAAH1056 + pAAH0412	WT U6 + $Ecm2^{WT}$	This work.
yAAH2432	yAAH1910 + pAAH1067 + pAAH0412	WT U6 + $Ecm2^{1-143}$	This work.
yAAH2476	yAAH1910 + pAAH0137 + pAAH1027	U57A U6 + Empty URA	This work.
yAAH2477	yAAH1910 + pAAH1056 + pAAH1027	U57A U6 + $Ecm2^{WT}$	This work.
yAAH2478	yAAH1910 + pAAH1067 + pAAH1027	U57A U6 + $Ecm2^{1-143}$	This work.
yAAH2479	yAAH1910 + pAAH0137 + pAAH1028	U57C U6 + Empty URA	This work.
yAAH2480	yAAH1910 + pAAH1056 + pAAH1028	U57C U6 + $Ecm2^{WT}$	This work.
yAAH2481	yAAH1910 + pAAH1067 + pAAH1028	U57C U6 + $Ecm2^{1-143}$	This work.

Table 5.2 Plasmids used in this study.

Plasmid ID	Plasmid Name	Description	Reference or Source
pAAH1056	WT Ecm2	ECM2 ±300 bp of up- and downstream sequence cloned into the NotI and SalI sites of pRS416	This work
pAAH1067	Ecm2 1-143	pAAH mutated to introduce stop codons after amino acid 143 of Ecm2	This work
pAAH1068	Ecm2 1-197	pAAH mutated to introduce stop codons after amino acid 197 of Ecm2	This work
pAAH1196	Ecm2 1-265	pAAH mutated to introduce stop codons after amino acid 265 of Ecm2	This work
pAAH1197	Ecm2 1-325	pAAH mutated to introduce stop codons after amino acid 325 of Ecm2	This work
pAAH0137	pRS416	URA3/CEN6 yeast centromeric plasmid	(Sikorski and Hieter 1989)
pAAH0470	ACT1-CUP1 WT	WT reporter used for ACT1-CUP1 assays. ACT1-CUP1 expressed from a GPD promoter on a Leu2-marked plasmid.	Gift from Charles Query
pAAH0524	ACT1-CUP1 A258U	BS A258U BS mutant reporter used for ACT1-CUP1 assays. ACT1-CUP1 expressed from a GPD promoter on a Leu2-marked plasmid.	Gift from Charles Query
pAAH0439	ACT1-CUP1 U257C	BS U257C BS mutant reporter used for ACT1-CUP1 assays. ACT1-CUP1 expressed from a GPD promoter on a Leu2-marked plasmid.	Gift from Charles Query
pAAH0880	ACT1-CUP1 A259G	BS A259G BS mutant reporter used for ACT1-CUP1 assays. ACT1-CUP1 expressed from a GPD promoter on a Leu2-marked plasmid.	Gift from Charles Query
pAAH1032	ACT1-CUP1 A3C	BS A3C 5'SS mutant reporter used for ACT1-CUP1 assays. ACT1-CUP1 expressed from a GPD promoter on a Leu2-marked plasmid.	Gift from Charles Query
pAAH0514	ACT1-CUP1 A3U	BS A3U 5'SS mutant reporter used for ACT1-CUP1 assays. ACT1-CUP1 expressed from a GPD promoter on a Leu2-marked plasmid.	Gift from Charles Query
pAAH0517	ACT1-CUP1 G5A	BS G5A 5'SS mutant reporter used for ACT1-CUP1 assays. ACT1-CUP1 expressed from a GPD promoter on a Leu2-marked plasmid.	Gift from Charles Query
pAAH0441	ACT1-CUP1 A259C	BS A259C BS mutant reporter used for ACT1-CUP1 assays. ACT1-CUP1 expressed from a GPD promoter on a Leu2-marked plasmid.	Gift from Charles Query
pAAH0527	ACT1-CUP1 A302U	BS A302U 3'SS mutant reporter used for ACT1-CUP1 assays. ACT1-CUP1 expressed from a GPD promoter on a Leu2-marked plasmid.	Gift from Charles Query
pAAH1058	ACT1-CUP1 G5A/U-5G	5'SS G5A mutant reporter with a competing, upstream 5'SS used for ACT1-CUP1 assays.	Gift from Charles Query
pAAH0778	$Prp2^{WT}$	Plasmid used to generate $Prp2^{WT}$ strains. Contains the Prp2 coding region in pRS415.	(Carrocci et al. 2017)
pAAH0790	$Prp2^{Q548N}$	Plasmid used to generate $Prp2^{Q548N}$ strains. Contains the Prp2-Q548N in pRS415.	(Carrocci et al. 2017)
pAAH1040	$Prp16^{WT}$	Plasmid used to generate $Prp16^{WT}$ strains (PRP16 TRP1 CEN ARS).	Gift from Charles Query
pAAH1039	$Prp16^{R686I}$	Plasmid used to generate $Prp16^{R686I}$ strains (PRP16-R686I TRP1 CEN ARS).	Gift from Charles Query

pAAH1042	$Prp22^{WT}$	Plasmid used to generate $Prp22^{WT}$ strains (PRP22 TRP1 CEN ARS).	Gift from Charles Query
pAAH1043	$Prp22^{T637A}$	Plasmid used to generate $Prp22^{T637A}$ strains (PRP22-T637A TRP1 CEN ARS).	Gift from Charles Query
pAAH1015	U2-WT	Plasmid used to generate U2-WT strains (SNR20 LEU2 CEN3 ARS1).	Gift from Manny Ares (Ares and Igel 1990)
pAAH1018	U2-2,4	Plasmid used to generate U2-2,4 strains (SNR20-U2-2,4 LEU2 CEN3 ARS1).	Gift from Manny Ares (Ares and Igel 1990)
pAAH1016	U2-DCC	Plasmid used to generate U2-DCC strains (SNR20-DCC LEU2 CEN3 ARS1).	Gift from Manny Ares (Ares and Igel 1990)
pAAH1017	U2-G53A	Plasmid used to generate U2-G53A strains (SNR20-G53A LEU2 CEN3 ARS1).	Gift from Manny Ares (Ares and Igel 1990)
pAAH1020	U2-IIc+	Plasmid used to generate U2-IIc+ strains (SNR20-IIc+ LEU2 CEN3 ARS1).	Gift from Manny Ares (Ares and Igel 1990)
pAAH1019	U2-DIIb	Plasmid used to generate U2-DIIb strains (SNR20-DIIb LEU2 CEN3 ARS1).	Gift from Manny Ares (Ares and Igel 1990)
pAAH0997	$Prp8^{WT}$	Plasmid used to generate $Prp8^{WT}$ strains (PRP8 2micron TRP1), pJU225-4	Gift from David Brow.
pAAH1004	$Prp8^{R1753K}$	Plasmid used to generate $Prp8^{R1753K}$ strains (PRP8-R1753K 2micron TRP1). pMK8-14	Gift from Magda Konarska (Liu et al. 2007)
pAAH1001	$Prp8^{P986T}$	Plasmid used to generate $Prp8^{P986T}$ strains (PRP8-161 2micron TRP1). *prp8-161* allele. pMK8-18	Gift from Magda Konarska (Liu et al. 2007)
pAAH1006	$Prp8^{R1753K/P986T}$	Plasmid used to generate $Prp8^{R1753K/P986T}$ strains (PRP8-R1753K/P986T 2micron TRP1). pMK8T-229	Gift from Magda Konarska (Liu et al. 2007)
pAAH1003	$Prp8^{E1960K}$	Plasmid used to generate $Prp8^{E1960K}$ strains (PRP8-101 2micron TRP1). *prp8-101* allele. pMK8-20.	Gift from Magda Konarska (Liu et al. 2007)
pAAH1045	$Prp8^{V1870N}$	Plasmid used to generate $Prp8^{V1870N}$ strains (PRP8-162 2micron TRP1). *prp8-162* allele. pMK8-19.	Gift from Magda Konarska (Liu et al. 2007)
pAAH1007	$Prp8^{E1960K/V1870N}$	Plasmid used to generate $Prp8^{WT}$ strains ($Prp8^{E1960K/V1870N}$ 2micron TRP1). pMK8T-230.	Gift from Magda Konarska (Liu et al. 2007)
pAAH0412	U6-WT	Plasmid used to generate U6 WT strains (snr6 TRP1 CEN).	(Rodgers et al. 2016a)
pAAH1027	U6-U57A	Plasmid used to generate U6 WT strains (snr6 TRP1 CEN).	This work. (McPheeters 1996)
pAAH1028	U6-U57C	Plasmid used to generate U6 WT strains (snr6 TRP1 CEN).	This work.
pAAH1239	WT Ecm2	pAAH1056 with 3x FLAG epitope tag at C-terminus	This work
pAAH1240	Ecm2 1-143	pAAH1067with 3x FLAG epitope tag at C-terminus	This work
pAAH1241	Ecm2 1-197	pAAH1068 with 3x FLAG epitope tag at C-terminus	This work
pAAH1242	Ecm2 1-265	pAAH1196 with 3x FLAG epitope tag at C-terminus	This work
pAAH1243	Ecm2 1-325	pAAH1197 with 3x FLAG epitope tag at C-terminus	This work

RESULTS

The Ecm2 U6-Binding Domain is Insufficient to Rescue Yeast Growth at 37°C

Previous studies of Ecm2 reported that *ecm2D* yeast exhibited a strong temperature-sensitive (*ts*) phenotype with significantly reduced or no growth at temperatures above 33°C (Xu and Friesen 2001). We replicated this result by deleting *ECM2* from a haploid strain of yeast and introducing plasmids containing *ecm2* variants under control of their endogenous promoters. As expected, *ecm2D* yeast containing an empty plasmid grew well at permissive temperatures (16-30°C) but possessed a severe *ts* phenotype at 37°C (**Fig. 5.1D**). When we included a plasmid containing the wild type (WT) *ECM2* gene, the *ts* phenotype was corrected, and growth was restored at 37°C.

To test if the N-terminal, U6-binding binding domain of Ecm2 alone was capable of rescuing the *ts* phenotype, we used recent cryo-EM structures of yeast spliceosomes to design truncation mutants of Ecm2. Nonsense mutations were incorporated at amino acids 144, 198, 266, and 326 to allow for expression of variants containing only the U6-binding ZNF domain ($Ecm2^{1-143}$), the ZNF domain plus the inter-domain linker ($Ecm2^{1-197}$), the ZNF and a partial U2-binding, RRM domain (lacking amino acids that come nearest to U2, $Ecm2^{1-265}$), or the complete ZNF and RRM domains truncated at the last amino acid modeled into cryo-EM density but missing the C-terminal lysine-rich region ($Ecm2^{1-325}$, **Fig. 5.1D**). Variants containing the U6-binding, ZNF domain but not the RRM were able to partially rescue the *ts* phenotype but still grew poorly at 37°C. Inclusion of the entire U2-binding, RRM ($Ecm2^{1-325}$) resulted in more significant suppression of the *ts* phenotype; although, cells still grew more slowly than those containing $Ecm2^{WT}$. Western blotting confirmed that all of the truncated proteins were expressed in cells; however, with the exception of $Ecm2^{1-325}$, they were expressed at lower levels than $Ecm2^{WT}$. Consequently, the inability to rescue the *ts* phenotype could be due to both truncation and decreased levels of the

proteins. Since Ecm2$^{1-325}$ was expressed at levels comparable to Ecm2^{WT}, the C-terminal lysine-rich region of the protein likely does play a role in suppression of the *ts* phenotype.

Genetic Interactions between Ecm2 and the Prp2 and Prp16 ATPases

Xu and Friesen provided ample evidence that Ecm2 plays a role in spliceosome activation (Xu and Friesen 2001). We and others have previously noted that key players in the activation process such as the U2 snRNP protein Hsh155/SF3B1 and U2/U6 helix I exhibit genetic interactions with a cold-sensitive (*cs*) mutant of the DEAH-box ATPase Prp2 (Prp2^{Q548N}) (Wlodaver and Staley 2014; Carrocci et al. 2017; Kaur et al. 2020). Prp2 binds the intronic RNA downstream of the branch site and uses ATP hydrolysis to trigger release of Hsh155/SF3B1 and other U2 snRNP proteins during activation (Lardelli et al. 2010; van der Feltz and Hoskins 2019). We tested if *ecm2D* would also show a genetic interaction with Prp2^{Q548N}. When we combined Prp2^{Q548N} with *ecm2D*, we observed no growth at low or high temperatures (16, 23, or 37ºC) and reduced growth at 30ºC (**Fig. 5.2A**). Prp2^{Q548N} is synthetic lethal with *ecm2D* at low temperatures and Prp2^{Q548N} does not rescue the *ts* phenotype of *ecm2D*. This genetic interaction is consistent with Ecm2's function in promoting spliceosome activation.

We next tested if other spliceosome DEAH-box ATPases would also show genetic interactions with *ecm2D* or if these results were specific to Prp2^{Q548N}. We combined *ecm2D* with a *cs* mutation of the ATPase Prp16 (Prp16^{R686I}) or a *cs* and *ts* mutation of the ATPase Prp22 (Prp22^{T637A}). Prp16 uses ATP hydrolysis to promote conformational changes of the spliceosome and splicing factor release during remodeling of the active site from the 1st to 2nd catalytic step (**Fig. 5.1A**) (Schwer and Guthrie 1992; Semlow et al. 2016; Plaschka et al. 2019). Prp16^{R686I} likely impedes this transition since this mutation is rescued by alleles of Prp8 that promote exon ligation (2nd-step alleles, discussed below) (Query and Konarska 2004). Prp22 also uses ATP hydrolysis to promote conformational change that enables release of the spliced mRNA product from the active site (**Fig. 5.1A**) (Schwer and Guthrie 1992; Wagner et al. 1998; Schwer 2008; Semlow et

al. 2016; Plaschka et al. 2019). In this case, Prp22^{T637A} likely impedes mRNA release and transition of the active site out of the exon ligation conformation since this mutation is exacerbated by 2nd-step alleles of Prp8 (Query and Konarska 2012).

When Prp16^{R686I} was combined with *ecm2D*, the *cs* phenotype of Prp16^{R686I} was suppressed and growth was restored at 16ºC (**Fig. 5.2B**). Yeast containing both Prp16^{R686I} and *ecm2D* also grew at 23 and 30ºC, albeit less well than when WT alleles were present. In addition, Prp16^{R686I} exacerbated the *ts* phenotype at 37ºC of *ecm2D* yeast. This is consistent with some degree of synthetic lethality between *ECM2* and *PRP16* at high temperatures and is consistent a previous report of synthetic lethality between the *slt11-1* and *prp16-1* alleles (Xu et al. 1998). We do note, however, that since the *ts* phenotype resulting from *ecm2D* in this yeast strain is less severe than in others (compare growth at 37ºC in **Fig. 5.1D** vs. **5.2B**), altered expression of Prp16 could also play a role appearance or suppression of phenotypes due to *ecm2D*.

While *ecm2D* suppressed the *cs* phenotype of Prp16^{R686I}, the *cs* phenotype of Prp22^{T637A} was not suppressed (**Fig. 5.2C**). Yeast containing both Prp22^{T637A} and *ecm2D* grew very poorly at 37ºC, and it was difficult to determine if Prp22^{T637A} was a weak suppressor of the *ts* phenotype of *ecm2D* yeast. In sum, these data strongly support genetic interactions between *ecm2D* and the Prp2 and Prp16 ATPases. Loss of Ecm2 exacerbates a *cs* defect in spliceosome activation caused by Prp2^{Q548N} and suppresses a *cs* defect in the 1st-to-2nd step conformational change caused by Prp16^{R686I}.

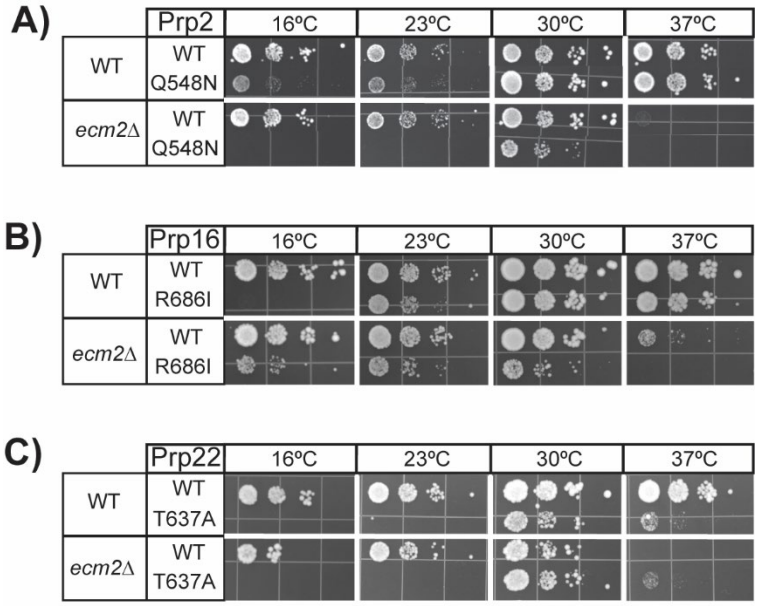


Figure 5.2 Genetic Interactions between Ecm2 and Spliceosomal ATPases.

(A-C) Mutations in Prp2 (A, *cs*), Prp16 (B, *cs*), and Prp22 (C, *cs* and *ts*) were combined with *ecm2D* and tested for suppression or exacerbation of temperature-dependent growth phenotypes. Yeasts were plated on YPD media and imaged after 3 (23, 30, or 37ºC) or 10 (16ºC) days of growth.

Genetic Interactions between Ecm2 and Mutations in U2 snRNA Stem II

The above results are consistent with a model in which Ecm2 stabilizes the 1st-step conformation of the spliceosome: aiding its formation during Prp2-initiated activation and inhibiting its remodeling by Prp16. To gain further insight into Ecm2's role during these steps, we combined *ecm2D* with mutations in the stem II region of the U2 snRNA which undergo a conformational change during activation. This region of U2 includes stem IIa/c as well as stem IIb—the RNA contacted by the C-terminal RRM of Ecm2 in cryo-EM structures of B* and C complex spliceosomes (**Fig. 5.1C**).

During activation, stem II undergoes a reversible conformational change from the stem IIa to the stem IIc structure, while stem IIb remains intact (**Fig. 5.3A**) (van der Feltz and Hoskins 2019). 5' SS cleavage is inhibited when formation of stem IIc is blocked by deletion of the 3' stem (DCC) or destabilized by mutation (Hilliker et al. 2007; Perriman and Ares 2007). In contrast, stabilization of stem IIc with additional base pairs (IIc+) promotes the 1st step of splicing (Perriman and Ares 2007). Like *ecm2D*, mutations that destabilize stem IIc or disrupt an interaction that is physically mutually exclusive with stem IIa also suppress Prp16 mutants defective in remodeling the 1st-step spliceosome active site (Hilliker et al. 2007; Perriman and Ares 2007). We predicted that if Ecm2 is facilitating activation by assisting stem IIc formation, then deletion of *ECM2* should exacerbate the phenotypes of mutants that antagonize stem IIc.

The U2-2,4 and DCC mutations both disrupt stem IIc formation: U2-2,4 stabilizes the competing stem IIa structure while DCC prevents stem IIc formation entirely by deletion of the nucleotides that comprise the 3' half of stem IIc (**Fig. 5.3A**) (Perriman and Ares 2007). These mutations have little phenotypic effect by themselves in our assay. However, when combined with *ecm2D* these mutations caused synthetic lethality at 30°C and *cs* phenotypes at 16 and 23°C (**Fig. 5.3B**). These results agree with our prediction that Ecm2 facilitates stem IIc formation.

This model also predicts that mutations in stem II that promote stem IIc formation may be able to suppress the *ts* phenotype of *ecm2D* yeast. The G53A and IIc+ mutants both favor stem IIc: G53A destabilizes the competing stem IIa structure while IIc+ extends base pairing of IIc (**Fig. 5.3A**) (Perriman and Ares 2007). These U2 mutants exhibit phenotypes even in the presence of Ecm2: both are *cs* while IIc+ also exhibits a modest growth defect at 30 and 37°C. Neither mutation suppressed the *ts* phenotype of *ecm2D* yeast, and *ecm2D* exacerbated the *cs* phenotypes of both mutations. These latter results could mean that Ecm2 has additional functions in the spliceosome while stem IIa is present or that snRNA structures containing these mutations are also disruptive for growth at lower temperatures in the absence of Ecm2.

If Ecm2 functions to assist stem IIc formation during activation, it is possible that this occurs through capture of stem IIb by the C-terminal RRM of Ecm2 during the B^{act} to B* complex transition. Stem IIb is non-essential in yeast (Ares and Igel 1990), and we tested if deletion of stem IIb (DIIb) resulted in a similar *ts* phenotype as *ecm2D*. The DIIb mutant yeast were not *ts* and exhibited minimal or no temperature-dependent phenotypes (**Fig. 5.3B**). The *ts* phenotype at 37°C of *ecm2D* was still observed when combined with the U2 DIIb mutation, and yeast containing both *ecm2D* and U2 DIIb grew similarly at other temperatures. This indicates that disruption of the Ecm2-RRM/stem IIb interaction is not solely responsible for the *ts* phenotype in *ecm2D* yeast.

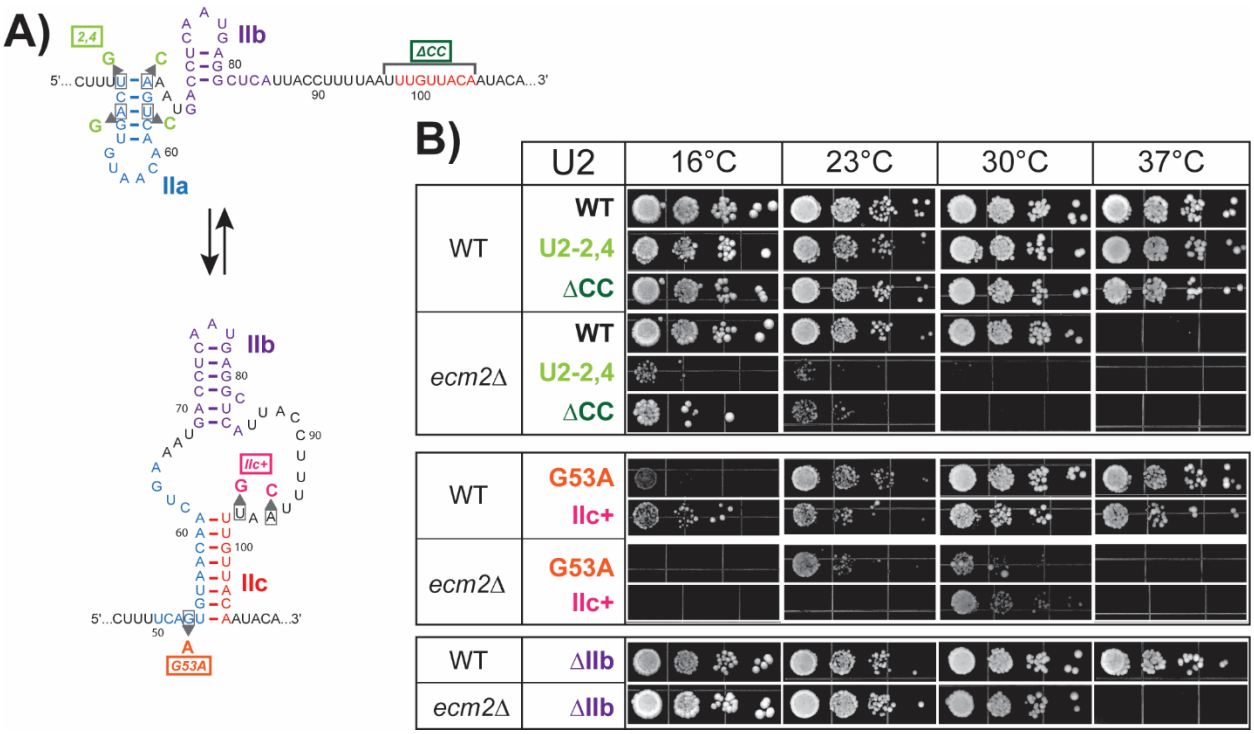


Figure 5.3 Genetic Interactions between Ecm2 and U2 stem II Mutations.

(A) Schematic of stem IIa/IIc toggling. Mutations which disfavor stem IIc (U2-2,4 and DCC; green-colored labels) are shown in the stem IIa structure. Mutations which disfavor stem IIa (G53A, IIc+; red-colored labels) are shown in the stem IIc structure. Nucleotides that are deleted in the DIIb mutant are colored in purple. (B) Mutations in stem II were combined with *ecm2D* and tested for suppression or exacerbation of temperature-dependent growth phenotypes. Yeasts were plated on YPD media and imaged after 3 (23, 30, or 37ºC) or 10 (16ºC) days of growth.

Ecm2 Impacts Splicing of Reporter pre-mRNAs Containing Non-consensus SS

We next studied how Ecm2 influences splicing catalysis *in vivo* with the ACT1-CUP1 splicing reporter (**Fig. 5.4A**). In this assay, splicing of the ACT1-CUP1 pre-mRNA is necessary for growth of a Cu^{2+} sensitive yeast strain on Cu^{2+}-containing media. The highest [Cu^{2+}] at which growth is observed is proportional to the amount of spliced mRNA in the cell (Lesser and Guthrie 1993a). In the presence of an ACT1-CUP1 reporter containing consensus SS, we observed no difference in Cu^{2+} tolerance between strains with or without *ECM2*. We used a primer extension assay to confirm that the similar Cu^{2+} tolerance results were correlated with high splicing efficiencies for both catalytic steps in the presence or absence of Ecm2 (**Fig. 5.5**). This indicates that splicing can still occur efficiently in the absence of Ecm2 and is consistent with lack of a significant growth phenotype in *ecm2D* strains at 30ºC in our assays.

We then assayed splicing in *ecm2D* yeast using ACT1-CUP1 reporters harboring substitutions at the SS. The most significant decreases in Cu^{2+} tolerance were observed using reporters containing the A3C substitution at the 5' SS, substitutions of the branch point adenosine (A259C or A259G), or substitutions flanking the branch point (U257C and C260G) (**Fig. 5.4B**). The large impact of *ecm2D* on A3C reporter splicing was intriguing since this substitution is only limiting for the 2nd catalytic step (Liu et al. 2007; Eysmont et al. 2019). Primer extension analysis of 1st- and 2nd-step splicing products confirmed a strong defect in exon ligation for the A3C reporter in the absence of Ecm2 (**Fig. 5.6**). Ecm2 can therefore have opposing effects on the 2nd-step active site: it can inhibit its formation but also promote 2nd-step catalysis on a substrate containing the A3C 5' SS substitution. It is possible that Ecm2 has distinct functions in both spliceosome structural transitions and in each catalytic reaction.

Deletion of ECM2 improved Cu^{2+} tolerance of yeast containing the A3U or G5A 5' SS reporters (**Fig. 5.4B**). However, analysis of 1st- and 2nd-step splicing products showed similar splicing efficiencies in the presence or absence of Ecm2 (**Fig. 5.6**). We did not study how decay

of unspliced RNAs or splicing intermediates influenced these primer extension results (Liu et al. 2007). Therefore, it is difficult to conclude from the primer extension assay if *ecm2D* truly changed the splicing efficiencies for the A3U and G5A substrates.

Nonetheless, the increase in Cu^{2+} tolerance observed with the G5A mutant in *ecm2D* yeast was noteworthy since this substitution can result in use of a cryptic, upstream 5' SS (Parker and Guthrie 1985; Kandels-Lewis and Séraphin 1993; Lesser and Guthrie 1993b). We used a modified ACT1-CUP1 reporter with competing 5' SS to test whether or not Ecm2 changes cryptic SS usage (**Fig. 5.4C**, **5.7**). When Ecm2 was present in the yeast, we detected usage of both the cryptic (21±1% of spliced products) and normal 5' SS. However, in the absence of Ecm2 use of the cryptic 5' SS was greatly reduced (7±4% of spliced products, **Fig. 5.4D**). This represents at least a 3-fold decrease based on our lower limit of detection. Combined, our results demonstrate that Ecm2 impacts the spliceosome active site to alter splicing of pre-mRNAs with non-consensus SS and permit the usage of an alternate, cryptic 5' SS.

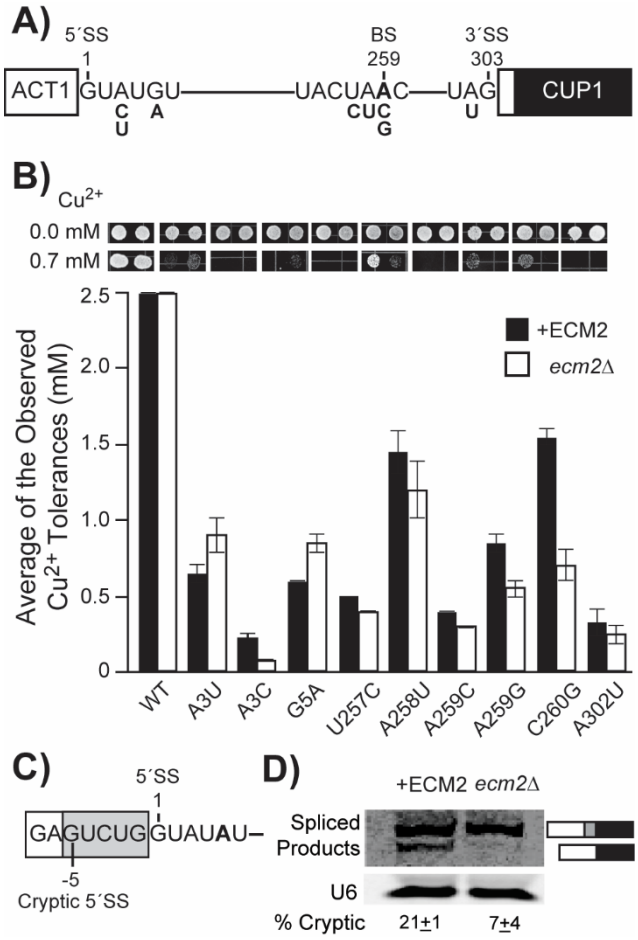


Figure 5.4 Impact of Ecm2 on Splicing of ACT1-CUP1 Reporter pre-mRNAs.

(A) Schematic of the ACT1-CUP1 reporter pre-mRNA with non-consensus substitutions noted. (B) ACT1-CUP1 assay results. Representative images of yeast growth after 3 days at 30ºC on agar plates made with -leu -trp dropout media containing 0 or 0.7 mM Cu^{2+} are shown above the bar graph. Each value in the graph represents the highest concentration of Cu^{2+} at which growth was observed on plates containing defined concentrations of Cu^{2+} and are representative of results obtained in replicate assays. (C) Schematic of the modified ACT1-CUP1 reporter containing a competing, cryptic 5' SS. (D) Primer extension assay of cryptic 5' SS usage using the reporter shown in panel (C). Primer extension of the U6 snRNA was included as a control. The percentages of cryptic products (ratios of cryptic products/total products) are shown below the gel and are the averages of three replicate experiments ± the standard deviation.

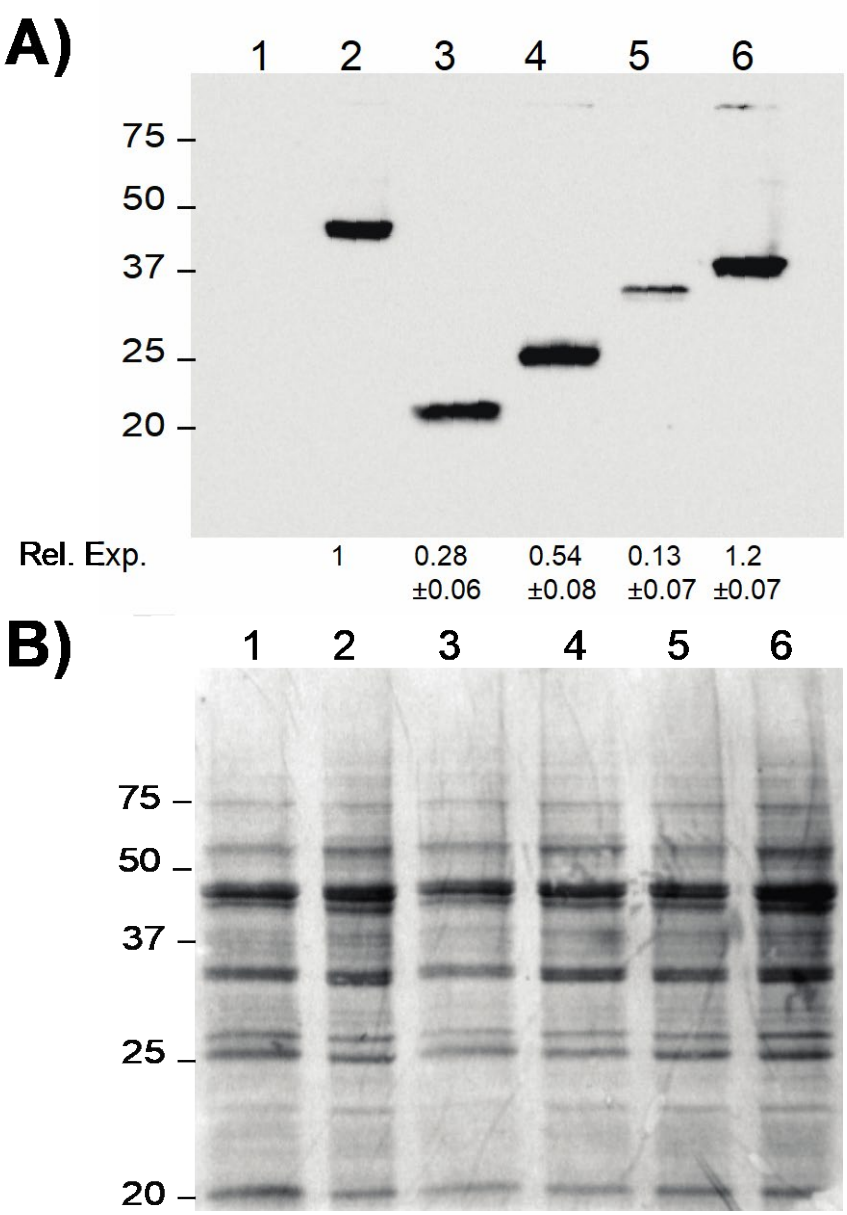


Figure 5.5 Western blot analysis of Ecm2 truncation mutant expression.

(A) Western blot results (B) Ponceau stain of the blot shown in (A). For both panels, lane 1 is lysate from a yeast strain expressing $Ecm2^{WT}$ (no Flag epitope tag); lane 2 is $Ecm2^{WT}$-3xFLAG; lane 3 is $Ecm2^{1-144}$-3xFLAG; lane 4 is $Ecm2^{1-198}$-3xFLAG; lane 5 is $Ecm2^{1-266}$-3xFLAG, and lane 6 is $Ecm2^{1-326}$-3xFLAG. Relative expression levels are shown below the Western blot and normalized against total protein contain as visualized by Ponceau staining. Expression levels represent the averages from three replicate experiments ±SD.

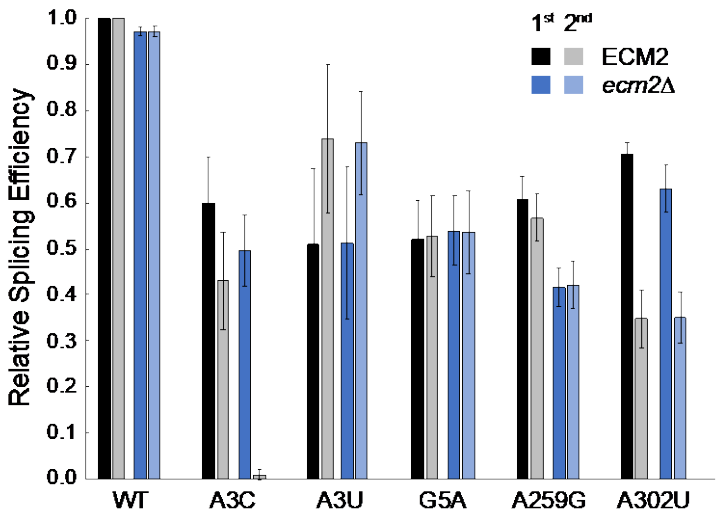


Figure 5.6 Primer extension analysis of 1st- and 2nd-step splicing efficiencies for ACT1-CUP1 reporters in the presence and absence of ECM2.

1st-step efficiencies were determined as the ratios of (lariat intermediates + mRNA) to total RNA. 2nd-step efficiencies were determined as the ratios of mRNA to total products from the 1st step. Efficiencies were normalized to the WT ACT1-CUP1 reporter in the ECM2 strain. Bar heights represent the averages from three replicate experiments, and error bars represent the standard deviations.

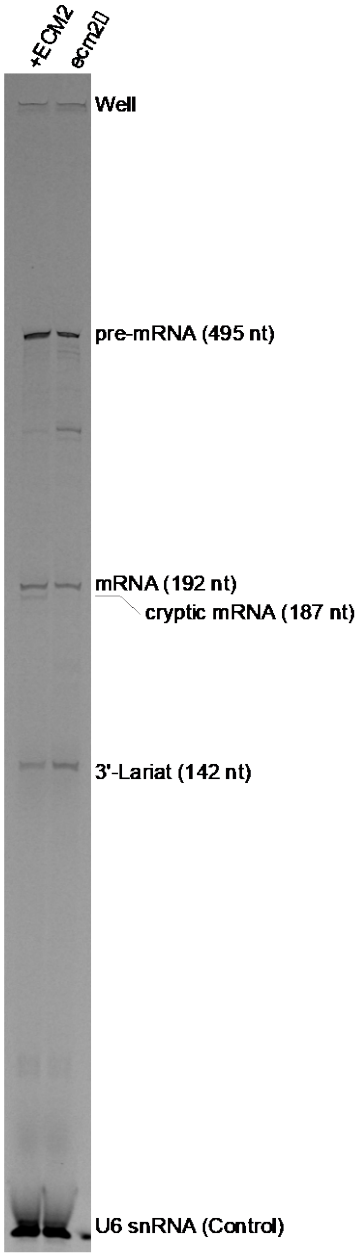


Figure 5.7

Uncropped gel image of a representative primer extension analysis of cryptic 5' SS usage. The region containing the mRNA is magnified in Fig. 5.4D. The primer extension stop located between the pre-mRNA and mRNA is reproducible but has not been attributed to a specific RNA species.

Genetic Interactions Between Ecm2 and the U6 snRNA 1st- and 2nd-Step Alleles

Like *ecm2D*, a number of alleles of the U6 snRNA and Prp8 suppress Prp16 ATPase mutations, have limited impact on splicing of ACT1-CUP1 reporters harboring consensus SS, and can promote or block splicing of reporters with particular SS substitutions (McPheeters 1996; Query and Konarska 2004; Liu et al. 2007; Mayerle et al. 2017; Eysmont et al. 2019). Many of these mutants have been categorized as 1st- or 2nd-step alleles since, in addition to causing *ts* or *cs* phenotypes, they promote one of the catalytic steps of splicing over the other (**Fig. 5.8A**). Since *ecm2D* and 1st- and 2nd-step alleles share common features, we tested for genetic interactions between these alleles and *ecm2D*. We first examined interactions with the U6 snRNA U57C and U57A mutations which promote the 1st and 2nd steps, respectively (McPheeters 1996; Liu et al. 2007).

The U6-U57A mutation had no effect on yeast growth at 16, 23, or 30°C in the absence of Ecm2 or in the presence of $Ecm2^{WT}$ or $Ecm2^{1-143}$ (which contains only the U6-binding domain, **Fig. 5.8B** and **Fig. 5.9**). The U6-U57C mutation also had no impact on growth at 16 or 23°C but was slower growing at 30°C. U6-U57C yeast containing only the empty URA3 plasmid displayed a slower-growing phenotype compared to those containing plasmids for $Ecm2^{WT}$ or $Ecm2^{1-143}$.

At higher temperatures, the impacts of the U6 mutations were more pronounced. When yeast contained the 1st-step, U6-U57C allele, we observed a strong synthetic sick interaction with *ecm2*D at 34°C that was partially rescued by expression of $Ecm2^{WT}$ or $Ecm2^{1-143}$, with the former producing stronger rescue than the latter (**Fig. 5.8B**). At 37°C only the strain expressing $Ecm2^{WT}$ was viable in the presence of U6-U57C. In contrast, we observed a weaker synthetic genetic interaction between *ecm2D* and the 2nd-step allele, U6-U57A, at 34°C (**Fig. 5.8B**). As with U6-U57C, this interaction was more pronounced at 37°C and yeast containing U6-U57C along with the truncated $Ecm2^{1-143}$ protein grew much more slowly than yeast containing $Ecm2^{1-143}$ and U6-

WT. Combined, we conclude that synthetic genetic interactions at 34° and 37°C between Ecm2 and U6-U57C elicit stronger phenotypes than those between Ecm2 and U6-U57A**.**

The interactions of *ecm2D* with these U6 mutants are most similar to those of 1st-step alleles in other splicing factors like Prp8 (Liu et al. 2007). When U6 mutations are present, loss of Ecm2 promotes the 1st step of splicing and presence of Ecm2 promotes the 2nd step. These results differ from those obtained upon deletion of the non-essential factor Isy1 (**Fig. 5.1C**): *isy1D* is synthetic lethal with U57A and suppresses the *ts* phenotype of U57C (Villa and Guthrie 2005). Thus, Isy1 appears to act as a 1st-step splicing factor when U6 is mutated, consistent with Isy1 release prior to the 2nd step (Plaschka et al. 2019), while Ecm2 can act as a 2nd-step factor and is consistent with its presence throughout both catalytic stages of splicing (**Fig. 5.1A**).

Genetic Interactions Between Ecm2 and Prp8 1st- and 2nd-Step Alleles

Genetic interactions between *ecm2D* and Prp2, Prp16, and U2 stem II and Ecm2-control of 5' SS usage support a role for Ecm2 in the 1st step of splicing. However, genetic interactions with U6-U57 mutants and results using the A3C splicing reporter support an additional role for Ecm2 in the 2nd step. We next tested if *ecm2D* would show genetic interactions with 1st- or 2nd-step alleles of Prp8 or both. We individually combined *ecm2D* with two 1st-step alleles of Prp8 ($Prp8^{R1753K}$ or $Prp8^{E1960K}$) or two 2nd-step alleles ($Prp8^{P986T}$ or $Prp8^{V1870N}$). For the 1st-step alleles, deletion of Ecm2 weakly suppressed the *cs* phenotype of $Prp8^{R1753K}$ and strongly suppressed the *cs* phenotype of $Prp8^{E1960K}$ (**Fig. 5.8C**). Neither Prp8 1st-step allele was able to completely correct the *ts* phenotype of *ecm2D*; although, slightly improved growth was observed at 37°C for yeast containing $Prp8^{E1960K}$ (**Fig. 5.8C**).

When *ecm2D* was combined with the 2nd-step alleles, we observed slightly improved growth at 37°C for yeast containing $Prp8^{P986T}$. A stronger genetic interaction was observed with the $Prp8^{V1870N}$. This 2nd-step allele exacerbated the *ts* phenotype of *ecm2D*, causing a strong

growth defect at 30ºC and no growth at 37ºC (**Fig. 5.8C**). The growth defect of Prp8^{V1870N} was partially corrected at 30ºC by combining 1st and 2nd Prp8 alleles (Prp8^{V1870N,E1960K}). However, this also resulted in stronger growth defects at other temperatures.

The *cs* suppression we observe of the Prp8^{E1960K} 1st-step allele and *ts* exacerbation of the Prp8^{V1870N} 2nd-step allele are consistent with *ecm2D* acting as a 2nd-step allele and facilitating exit of the spliceosome from the 1st-step catalytic conformation. Both the Prp8^{E1960K} and Prp8^{V1870N} substitutions are located within Prp8's RNaseH domain. Like U2 stem II, the RNaseH domain is both highly dynamic and toggles between alternate structures (Schellenberg et al. 2013; Mayerle et al. 2017). Ecm2 may impact Prp8-RNaseH dynamics or vice versa to support the 1st-step reaction. This is in juxtaposition to the results obtained with the U6 mutants, which were consistent with Ecm2 having a role in the 2nd step.

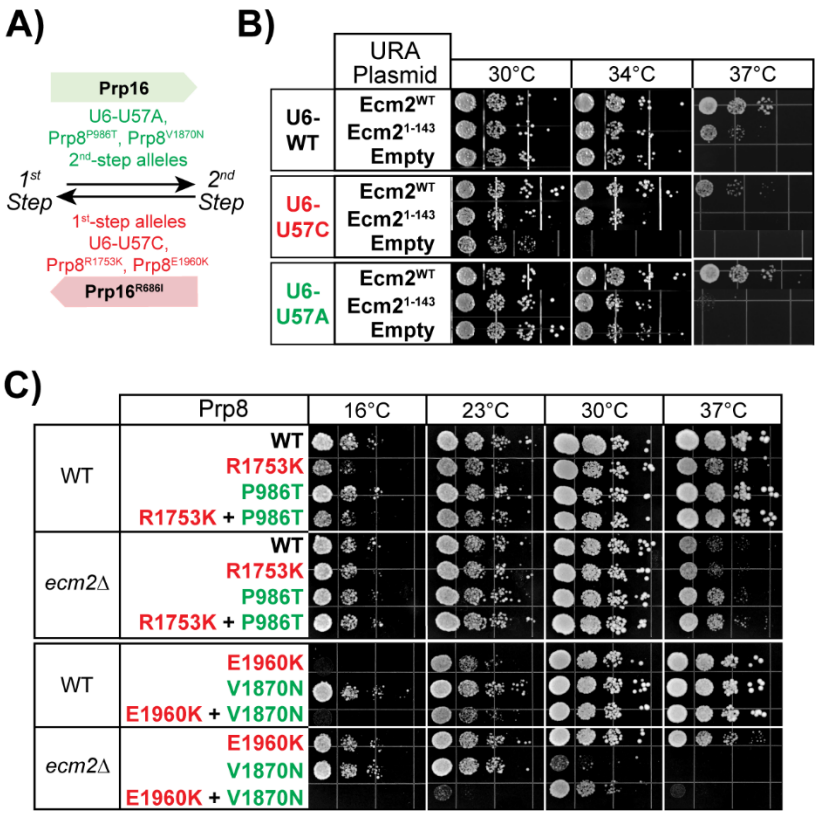


Figure 5.8 Genetic Interactions between Ecm2 and U6 or Prp8 1st- and 2nd-Step Alleles.

(A) Illustration of how alleles of Prp8, Prp16, and U6 function to promote the 1st or 2nd step of splicing. (B) A 1st- or 2nd-step allele of U6 (U57C or U57A, respectively) was combined with URA3 plasmids either lacking or coding for Ecm2 variants in *ecm2D* yeast. The strains were then tested for suppression or exacerbation of temperature-dependent growth phenotypes. Yeast were plated on -URA dropout media and imaged after 2 days of growth. (C) 1st- and 2nd-step alleles of Prp8 (red and green, respectively) were combined with *ecm2D* and tested for suppression or exacerbation of temperature-dependent growth phenotypes. Yeast were plated on YPD media and imaged after 3 (23, 30, or 37ºC) or 10 (16ºC) days of growth.

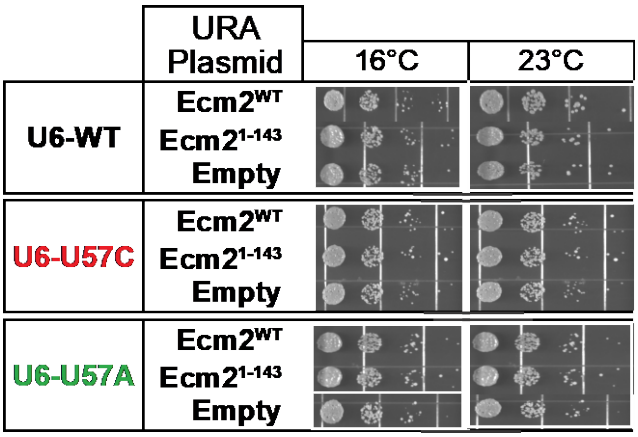


Figure 5.9

A 1st- or 2nd-step allele of U6 (U57C or U57A, respectively) was combined with URA3 plasmids either lacking or coding for Ecm2 variants in *ecm2D* yeast. The strains were then tested for suppression or exacerbation of temperature-dependent growth phenotypes. Yeasts were plated on -URA dropout media and imaged after 3 (23ºC) or 6 (16ºC) days of growth.

DISCUSSION

Using genetic and biochemical assays of splicing in cells, we have shown that yeast Ecm2 impacts multiple steps during the catalytic phase of splicing and that loss of Ecm2 perturbs how the spliceosome processes pre-mRNAs containing non-consensus SS. Deletion of *ECM2* results in genetic interactions with several structural switches in the spliceosome including U2 stem II, the RNaseH domain of Prp8, and the ATPases that control entry to and exit from the 1st step (Prp2 and Prp16, respectively). In sum, our data show that Ecm2 plays significant roles in spliceosome catalysis in addition to a function during activation (Xu and Friesen 2001).

Ecm2 Modulates the Catalytic Steps of Splicing

Our results support functions for Ecm2 during both catalytic steps in splicing. The differing genetic interactions we observe between *ecm2D* and U6 or Prp8 mutants suggest a more complicated mechanism from that of other alleles that exhibit more consistent genetic interactions (for example, a 2nd-step allele of *cef1* suppresses phenotypes of both 1st-step *prp8* and U6 alleles) (Query and Konarska 2012). Our results could be explained by distinct and genetically separable functions for Ecm2 during the 1st and 2nd catalytic steps with Prp8 mutations revealing a role in the former and U6 mutations revealing a role in the latter. Since Ecm2 has only been observed to make contact with U2 stem IIb in 1st-step cryo-EM structures, it is possible that this interaction contributes to Ecm2's distinct functions during each catalytic step.

Several of our observations with Ecm2 are similar to those previously reported for Cwc2 and Isy1 (Villa and Guthrie 2005; Rasche et al. 2012; Hogg et al. 2014). Cwc2, Ecm2, and Isy1 form a highly interconnected network of interactions with one another, the U6 snRNA, the intron, and a number of other splicing factors (**Fig. 5.1C**, **5.6A**) (Galej et al. 2016; Wan et al. 2016). All three proteins can suppress Prp16 mutations and have synthetic lethal interactions with active site U2/U6 helix Ia or Ib mutations (Xu et al. 1998; Villa and Guthrie 2005; Hogg et al. 2014).

While deletion of Cwc2, Isy1, or Ecm2 results in distinct growth phenotypes (for example, *cwc2*D is lethal at 30°C but *ecm2*D shows minimal changes in growth at that temperature), functional spliceosomes can form in yeast in the absence of one or a subset of these factors under permissive growth conditions (Xu and Friesen 2001; Villa and Guthrie 2005; Hogg et al. 2014). In addition, loss of Ecm2, loss of Isy1, or mutation of Cwc2 results in specific splicing defects in reporter pre-mRNAs with non-consensus SS (**Fig. 5.4B**) (Villa and Guthrie 2005; Hogg et al. 2014). This implies that spliceosomes missing one of these factors possess different substrate preferences and fidelity phenotypes. This is intriguing since *ecm2D* only results in a growth defect at high temperatures and splicing of pre-mRNAs containing consensus SS remains efficient (**Fig. 5.4B**). Thus, it is possible that yeast could bias the splicing of particular pre-mRNAs with non-consensus SS by regulating the Cwc2, Ecm2, and/or Isy1 content of spliceosomes without significantly compromising cellular splicing efficiency. This possibility has previously been proposed by Villa and Guthrie, who noted that deletion of Isy1 results in reduced fidelity of 3' SS selection (Villa and Guthrie 2005).

While there is some overlap in how Isy1 or Ecm2 loss or Cwc2 mutation impact splicing of non-consensus SS, the proteins also exert unique influences of the spliceosome. For example, *isy1D* and the Cwc2^{F183D} mutant improve splicing of reporter pre-mRNAs containing a A302U 3' SS, but *ecm2D* only minimally changes A302U splicing (**Fig. 5.4B**). In addition, *ecm2D* changes Cu^{2+} tolerance with the G5A reporter but this is unaffected by *isy1D* or Cwc2^{F183D} (Villa and Guthrie 2005; Hogg et al. 2014). Cellular splicing could thus be optimized for specific SS by independently controlling Cwc2, Ecm2, and Isy1 stoichiometry with spliceosomes.

These factors might also impact mRNA isoform production since Ecm2 additionally permits usage of an alternative 5' SS (**Fig. 5.4D**). Interestingly, the non-essential yeast splicing factor Bud31 is required for use of an alternative 5' SS in the SRC1 mRNA (Saha et al. 2012), and Bud31 directly contacts the U6 snRNA, Ecm2, and Cwc2 in the yeast spliceosome (Plaschka

et al. 2019). Bud31 and Ecm2 could permit promiscuous 5' SS use by similar mechanisms, although this has not yet been studied. In summary, spliceosomes may be fine-tuned by the presence or absence of non-essential splicing factors like Ecm2, and currently little is known about how the compositions of spliceosomes vary inside cells.

Consequences of a Dynamic Ecm2/U2 Stem II Interaction During Splicing

The spliceosome contains a number of proposed switches in which components toggle between one conformation or another at different stages of the reaction (Abelson 2017). The U2 snRNA contains several of these switches including a U2 stem IIa-to-IIc conformational change during activation (van der Feltz and Hoskins 2019). In addition, it has also been proposed that stem IIc switches transiently back to stem IIa between the catalytic steps of splicing before re-forming IIc during the 2nd step (Hilliker et al. 2007; Perriman and Ares 2007). This mechanism was based in part on the observation that stem II substitutions that destabilize stem IIc (or stabilize stem IIa) can suppress *cs* alleles of Prp16. Our observations that *ecm2D* also suppresses Prp16 *cs* alleles (**Fig. 5.2**) and Ecm2 contacts U2 stem IIb in C complex may provide an alternate explanation.

We propose that Prp16 indirectly disrupts the Ecm2/stem II interaction during remodeling of the spliceosome between the 1st and 2nd steps. Eliminating or weakening this interaction by stem II mutation can suppress Prp16 *cs* alleles by destabilizing the 1st-step conformation. This explanation is supported by cryo-EM structural data in which a transient contact between U2 stem IIb/c and the C-terminal RRM of Ecm2 is observed in 1st-step complexes (B* and C complexes) but not those preceding or following (B^{act} and C* complexes) (Galej et al. 2016; Rauhut et al. 2016; Wan et al. 2016; Yan et al. 2016; Fica et al. 2017; Wan et al. 2019). While additional structural information is needed for the on-pathway intermediates during 2nd-step active site assembly, stem IIa has not yet been observed in C* spliceosomes and accommodation of stem IIa in these complexes may be incompatible with binding of splicing factors (Prp17) and U2 snRNP

interactions with Syf1 (Wan et al. 2016; Fica et al. 2017; Liu et al. 2017; Yan et al. 2017; Fica et al. 2019). In light of these observations, stem IIc could remain intact throughout catalysis, and IIc-to-IIa toggling occurs later during spliceosome disassembly or U2 snRNP reassembly (Yan et al. 1998; Rodgers et al. 2016b). Regardless, further work is needed to characterize the short-lived intermediates that form during the 1st- to 2nd-step transition.

The viability of *ecm2D* and stem IIbD strains (**Fig. 5.3B**) (Ares and Igel 1990; Xu and Friesen 2001) show that the Ecm2/stem II interaction is not essential for yeast splicing. It is notable, however, that *ecm2D* exhibits synthetic lethal interactions with multiple stem II mutations. This includes, to our knowledge, the first genetic data showing synthetic lethality with the U2-2,4 mutant, which stabilizes stem IIa. Stabilization of stem IIa by U2-2,4 also results in a more significant growth defect than deletion of stem IIc (DCC, **Fig. 5.3B**), suggesting that failure to disrupt stem IIa is of greater consequence for the spliceosome than failure to form stem IIc in the absence of Ecm2. Therefore, these results complement previous evidence for the importance of stem IIa/IIc toggling (Hilliker et al. 2007; Perriman and Ares 2007) and highlight the need for stem IIa unwinding for efficient splicing. It is possible that the Ecm2/stem II interaction only becomes limiting for splicing when stem IIa/c toggling is disturbed or when the active site is destabilized by SS mutations.

Implications for Human RBM22 and Wrapped Intron Formation

The evolutionary histories of Cwc2, Ecm2, and RBM22 have not been studied, and it is uncertain how RBM22 may have evolved to functionally replace both proteins. Based on sequence alignments and crosslinking studies, it has been proposed that Cwc2 and RBM22 share a common function in binding U6 and interacting with the spliceosome's catalytic elements (Rasche et al. 2012). However, when fragments of the yeast and human C complex spliceosomes are aligned, RBM22 most closely mimics the interactions of Ecm2 with the U6 snRNA. In terms of U6 interaction, we believe that RBM22 and Ecm2, not Cwc2, are closer structural homologs.

Both RBM22 and Cwc2 bind the intronic RNA just downstream of the 5' SS. The RBM22/intron interaction contains an unusual and distinctive structure not observed with Cwc2. In human C and P complex spliceosomes, RBM22 completely encircles the intron (**Fig. 5.10B)** (Zhan et al. 2018; Fica et al. 2019). It is unlikely that this wrapped intron structure would form by threading of the intron through RBM22. Insights from Ecm2 provide a plausible mechanism for its formation. The C-terminal RRM of RBM22 could transiently interact with U2 stem IIb/c—analogous to the interaction between Ecm2 and stem II in yeast (**Fig. 5.10C**). This could open RBM22 for intron binding and enable subsequent wrapping of the intron after disruption of the RRM/stem II interaction.

Analyses of cryo-EM structures reveal that movement of RBM22 towards U2 stem II is not occluded by presence of other splicing factors and stem IIb is within an accessible distance for the RRM, assuming structural flexibility of the linker between the ZNF and RRM domains. There is some biochemical evidence for a RBM22/U2 snRNA interaction: anti-RBM22 antibodies can immunoprecipitate (IP) small amounts of the U2 snRNA from C complex spliceosomes after proteinase K treatment and without co-IP of the U5 or U6 snRNAs—consistent with a direct interaction (Rasche et al. 2012). We note, however, that in recent structures of human pre-B^{act} spliceosomes RBM22 and the intron were modeled with the same intertwined topology as found in later stage complexes despite the presence of U2 stem IIa/b and a large distance between RBM22 and U2 stem II (~150 Å) (Townsend et al. 2020). While it is possible that the limited resolution of the electron density map within the RBM22 region prevented modeling of alternate conformations, the structures nonetheless suggest that further experiments are needed to determine if intron wrapping by RBM22 occurs spontaneously (*i.e,* without a U2 snRNA stem II chaperone) and if an interaction between the RRM of RBM22 and stem IIb/c can occur during transitions between pre-B^{act} and B^{act} spliceosomes.

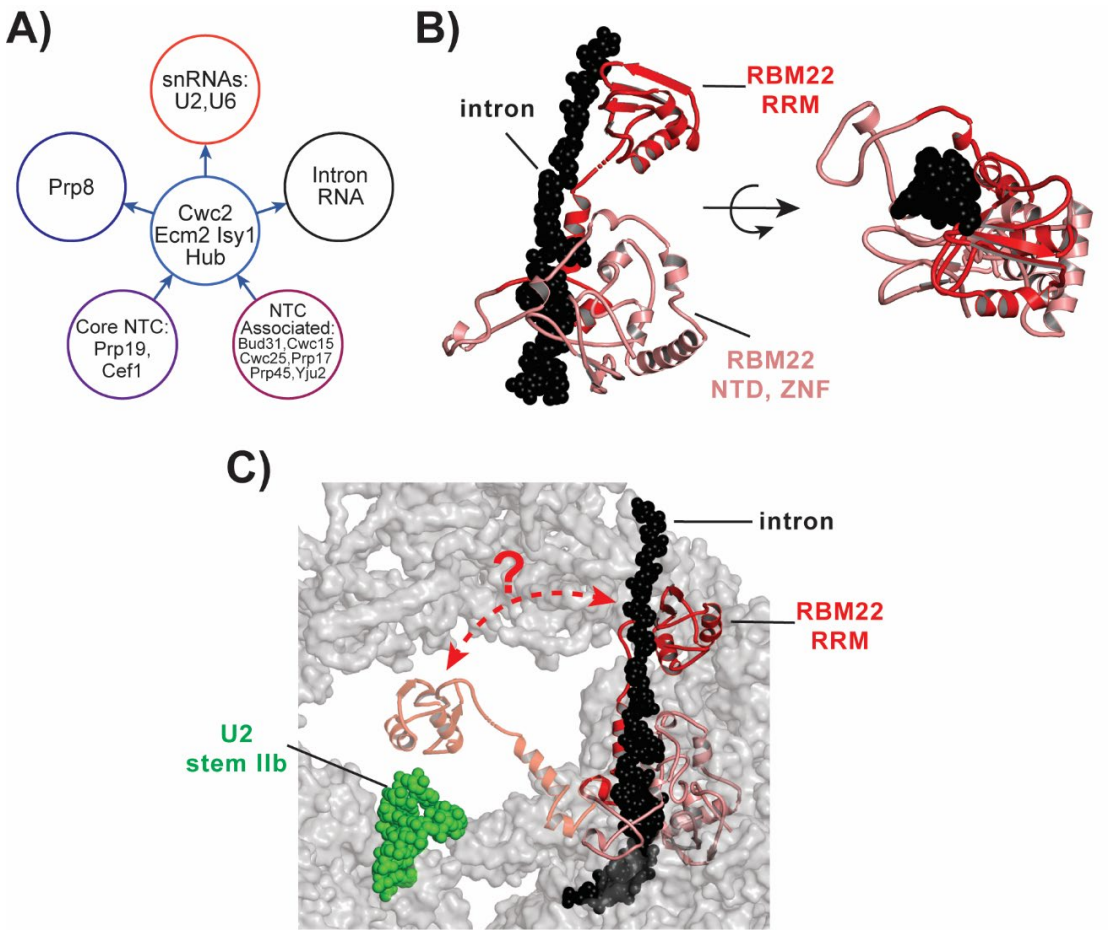


Figure 5.10 The Cwc2/Ecm2/Isy1 Interaction Network and Structure of Human RBM22.

(A) A large number of splicing factors interact with Cwc2, Ecm2, and/or Isy1 suggesting that these proteins form a network hub for modulating spliceosome activity. In this model, regulatory signals could flow into the hub from the NTC and NTC-related proteins and outwards to the spliceosome active site consisting of the intron, Prp8, and U2/U6 snRNAs. (B) Two views of the cryo-EM structure of RBM22 from a human C complex spliceosome. Domains of RBM22 are noted and intronic RNA downstream of the 5' SS is shown in black spacefill. Note that RBM22 completely encircles the RNA. Structure from 6EXN.pdb. (C) Hypothetical model for formation of the structure shown in panel (B). The RRM domain of RBM22 could make transient contact with human U2 stem IIb to allow for docking of the intron and subsequent wrapping. Structures in panels (B) and (C) are from 5YZG.pdb. The hypothetical model in panel (C) was created using PyMol.

REFERENCES

Abelson J. 2017. A close-up look at the spliceosome, at last. Proc Natl Acad Sci U S A 114: 4288-4293.

Ares M, Igel AH. 1990. Lethal and temperature-sensitive mutations and their suppressors identify an essential structural element in U2 small nuclear RNA. Genes Dev 4: 2132-2145.

Carrocci TJ, Paulson JC, Hoskins AA. 2018. Functional analysis of Hsh155/SF3b1 interactions with the U2 snRNA/branch site duplex. *RNA* **24**: 1028-1040.

Carrocci TJ, Zoerner DM, Paulson JC, Hoskins AA. 2017. SF3b1 mutations associated with myelodysplastic syndromes alter the fidelity of branchsite selection in yeast. *Nucleic Acids Res* **45**: 4837-4852.

Eysmont K, Matylla-Kulińska K, Jaskulska A, Magnus M, Konarska MM. 2019. Rearrangements within the U6 snRNA Core during the Transition between the Two Catalytic Steps of Splicing. *Mol Cell* **75**: 538-548.e533.

Fica SM, Oubridge C, Galej WP, Wilkinson ME, Bai XC, Newman AJ, Nagai K. 2017. Structure of a spliceosome remodelled for exon ligation. *Nature* **542**: 377-380.

Fica SM, Oubridge C, Wilkinson ME, Newman AJ, Nagai K. 2019. A human postcatalytic spliceosome structure reveals essential roles of metazoan factors for exon ligation. *Science* **363**: 710-714.

Galej WP, Wilkinson ME, Fica SM, Oubridge C, Newman AJ, Nagai K. 2016. Cryo-EM structure of the spliceosome immediately after branching. *Nature* **537**: 197-201.

Goldstein AL, McCusker JH. 1999. Three new dominant drug resistance cassettes for gene disruption in Saccharomyces cerevisiae. *Yeast* **15**: 1541-1553.

Hilliker AK, Mefford MA, Staley JP. 2007. U2 toggles iteratively between the stem IIa and stem IIc conformations to promote pre-mRNA splicing. *Genes Dev* **21**: 821-834.

Hogg R, de Almeida RA, Ruckshanthi JP, O'Keefe RT. 2014. Remodeling of U2-U6 snRNA helix I during pre-mRNA splicing by Prp16 and the NineTeen Complex protein Cwc2. *Nucleic Acids Res* **42**: 8008-8023.

Hogg R, McGrail JC, O'Keefe RT. 2010. The function of the NineTeen Complex (NTC) in regulating spliceosome conformations and fidelity during pre-mRNA splicing. *Biochem Soc Trans* **38**: 1110-1115.

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

Kastner B, Will CL, Stark H, Lührmann R. Structural Insights into Nuclear pre-mRNA Splicing in Higher Eukaryotes.

-. 2019. Structural Insights into Nuclear pre-mRNA Splicing in Higher Eukaryotes. *Cold Spring Harb Perspect Biol* **11**.

Kaur H, Groubert B, Paulson JC, McMillan S, Hoskins AA. 2020. Impact of cancer-associated mutations in Hsh155/SF3b1 HEAT repeats 9-12 on pre-mRNA splicing in Saccharomyces cerevisiae. *PLoS One* **15**: e0229315.

Lardelli RM, Thompson JX, Yates JR, Stevens SW. 2010. Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. *RNA* **16**: 516-528.

Lesser CF, Guthrie C. 1993a. Mutational analysis of pre-mRNA splicing in Saccharomyces cerevisiae using a sensitive new reporter gene, CUP1. *Genetics* **133**: 851-863.

-. 1993b. Mutations in U6 snRNA that alter splice site specificity: implications for the active site. *Science* **262**: 1982-1988.

Liu L, Query CC, Konarska MM. 2007. Opposing classes of prp8 alleles modulate the transition between the catalytic steps of pre-mRNA splicing. *Nat Struct Mol Biol* **14**: 519-526.

Liu S, Li X, Zhang L, Jiang J, Hill RC, Cui Y, Hansen KC, Zhou ZH, Zhao R. 2017. Structure of the yeast spliceosomal postcatalytic P complex. *Science* **358**: 1278-1283.

Mayerle M, Guthrie C. 2017. Genetics and biochemistry remain essential in the structural era of the spliceosome. *Methods* **125**: 3-9.

Mayerle M, Raghavan M, Ledoux S, Price A, Stepankiw N, Hadjivassiliou H, Moehle EA, Mendoza SD, Pleiss JA, Guthrie C et al. Structural toggle in the RNaseH domain of Prp8 helps balance splicing fidelity and catalytic efficiency.

-. 2017. Structural toggle in the RNaseH domain of Prp8 helps balance splicing fidelity and catalytic efficiency. *Proc Natl Acad Sci U S A* **114**: 4739-4744.

McGrail JC, Krause A, O'Keefe RT. 2009. The RNA binding protein Cwc2 interacts directly with the U6 snRNA to link the nineteen complex to the spliceosome during pre-mRNA splicing. *Nucleic Acids Res* **37**: 4205-4217.

McPheeters DS. 1996. Interactions of the yeast U6 RNA with the pre-mRNA branch site. *RNA* **2**: 1110-1123.

Parker R, Guthrie C. 1985. A point mutation in the conserved hexanucleotide at a yeast 5' splice junction uncouples recognition, cleavage, and ligation. *Cell* **41**: 107-118.

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

Perriman RJ, Ares M. 2007. Rearrangement of competing U2 RNA helices within the spliceosome promotes multiple steps in splicing. *Genes Dev* **21**: 811-820.

Plaschka C, Newman AJ, Nagai K. 2019. Structural Basis of Nuclear pre-mRNA Splicing: Lessons from Yeast. *Cold Spring Harb Perspect Biol* **11**.

Query CC, Konarska MM. 2004. Suppression of multiple substrate mutations by spliceosomal prp8 alleles suggests functional correlations with ribosomal ambiguity mutants. *Mol Cell* **14**: 343-354.

-. 2012. CEF1/CDC5 alleles modulate transitions between catalytic conformations of the spliceosome. *RNA* **18**: 1001-1013.

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

Rauhut R, Fabrizio P, Dybkov O, Hartmuth K, Pena V, Chari A, Kumar V, Lee CT, Urlaub H, Kastner B et al. 2016. Molecular architecture of the Saccharomyces cerevisiae activated spliceosome. *Science* **353**: 1399-1405.

Rodgers ML, Didychuk AL, Butcher SE, Brow DA, Hoskins AA. 2016a. A multi-step model for facilitated unwinding of the yeast U4/U6 RNA duplex. *Nucleic Acids Res* **44**: 10912-10928.

Rodgers ML, Tretbar US, Dehaven A, Alwan AA, Luo G, Mast HM, Hoskins AA. 2016b. Conformational dynamics of stem II of the U2 snRNA. *RNA* **22**: 225-236.

Romero-Calvo I, Ocón B, Martínez-Moya P, Suárez MD, Zarzuelo A, Martínez-Augustin O, de Medina FS. 2010. Reversible Ponceau staining as a loading control alternative to actin in Western blots. *Anal Biochem* **401**: 318-320.

Saha D, Banerjee S, Bashir S, Vijayraghavan U. 2012. Context dependent splicing functions of Bud31/Ycr063w define its role in budding and cell cycle progression. *Biochem Biophys Res Commun* **424**: 579-585.

Schellenberg MJ, Wu T, Ritchie DB, Fica S, Staley JP, Atta KA, LaPointe P, MacMillan AM. 2013. A conformational switch in PRP8 mediates metal ion coordination that promotes pre-mRNA exon ligation. *Nat Struct Mol Biol* **20**: 728-734.

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

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

Semlow DR, Blanco MR, Walter NG, Staley JP. 2016. Spliceosomal DEAH-Box ATPases Remodel Pre-mRNA to Activate Alternative Splice Sites. *Cell* **164**: 985-998.

Sikorski RS, Hieter P. 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. *Genetics* **122**: 19-27.

Townsend C, Leelaram MN, Agafonov DE, Dybkov O, Will CL, Bertram K, Urlaub H, Kastner B, Stark H, Lührmann R. 2020. Mechanism of protein-guided folding of the active site U2/U6 RNA during spliceosome activation. *Science* **370**.

Umen JG, Guthrie C. 1995. A novel role for a U5 snRNP protein in 3' splice site selection. *Genes Dev* **9**: 855-868.

van der Feltz C, Hoskins AA. 2019. Structural and functional modularity of the U2 snRNP in pre-mRNA splicing. *Crit Rev Biochem Mol Biol* **54**: 443-465.

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

Wagner JD, Jankowsky E, Company M, Pyle AM, Abelson JN. 1998. 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.

Wahl MC, Will CL, Lührmann R. 2009. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**: 701-718.

Wan R, Bai R, Yan C, Lei J, Shi Y. 2019. Structures of the Catalytically Activated Yeast Spliceosome Reveal the Mechanism of Branching. *Cell* **177**: 339-351.e313.

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

Wang Y, Guthrie C. 1998. PRP16, a DEAH-box RNA helicase, is recruited to the spliceosome primarily via its nonconserved N-terminal domain. *RNA* **4**: 1216-1229.

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

Wlodaver AM, Staley JP. 2014. The DExD/H-box ATPase Prp2p destabilizes and proofreads the catalytic RNA core of the spliceosome. *RNA* **20**: 282-294.

Xu D, Field DJ, Tang SJ, Moris A, Bobechko BP, Friesen JD. 1998. Synthetic lethality of yeast slt mutations with U2 small nuclear RNA mutations suggests functional interactions between U2 and U5 snRNPs that are important for both steps of pre-mRNA splicing. *Mol Cell Biol* **18**: 2055-2066.

Xu D, Friesen JD. 2001. Splicing factor slt11p and its involvement in formation of U2/U6 helix II in activation of the yeast spliceosome. *Mol Cell Biol* **21**: 1011-1023.

Xu YZ, Query CC. 2007. Competition between the ATPase Prp5 and branch region-U2 snRNA pairing modulates the fidelity of spliceosome assembly. *Mol Cell* **28**: 838-849.

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

-. 2017. Structure of a yeast step II catalytically activated spliceosome. *Science* **355**: 149-155.

Yan C, Wan R, Shi Y. 2019. Molecular Mechanisms of pre-mRNA Splicing through Structural Biology of the Spliceosome. *Cold Spring Harb Perspect Biol* **11**.

Yan D, Perriman R, Igel H, Howe KJ, Neville M, Ares M. 1998. CUS2, a yeast homolog of human Tat-SF1, rescues function of misfolded U2 through an unusual RNA recognition motif. *Mol Cell Biol* **18**: 5000-5009.

Zhan X, Yan C, Zhang X, Lei J, Shi Y. 2018. Structure of a human catalytic step I spliceosome. *Science* **359**: 537-545.

Chapter Six

Conclusions and Future Directions

SUMMARY

Combining our single molecule results with existing knowledge about spliceosome activation provides a clearer picture of how activation may occur. In this chapter, I will discuss a temporal model for yeast spliceosome activation by framing them in the context of early, middle, and late transitions as described in the Introduction. This model suggests several interesting experiments that could be done to further elucidate the activation process.

Early activation– Prp28-mediated Transition of the pre-B to B Complex

After the tri-snRNP is recruited to the pre-spliceosome (A complex), the initial pre-B intermediate will form before undergoing a transition to B complex. During this step, Prp28 activity results in 5'SS transfer from base pairing with the U1snRNA (U1 snRNP) to base pairing with sites near the ACAGAGA sequence of the U6 snRNA. This likely occurs prior to B complex protein (BCP) association.

Direct monitoring on Prp28 dynamics relative to other factors would further our understanding of early activation. Some potential experiments include monitoring the dynamics of Prp28 relative to the BCP for understanding how spatially mutually exclusive factors are coordinated during this transition. We have added a SNAP tag to the terminus of Prp28. However, the fast on-off dynamics of Prp28 makes it difficult to determine the state of the spliceosome (data not shown). More sophisticated single molecule experiments that can monitor multiple spliceosome subunits simultaneously or the use of Prp28 (or other splicing factor) mutants may help to elucidate this process.

Middle activation– Brr2 mediated Transition of the B to B^{ACT} Complex

The transition from B to B^{ACT} represents the most compositionally and conformationally dynamic step, which involves the exchange of ~50 factors. As I identified multiple intermediates with characteristic compositions during the transition, I am going to discuss this transition as

occurring with a series of sub-steps. The new intermediates were named based on their characteristic compositions ($B^{\Delta U4\Delta NTC}$, $B_{NTC+BCP+Lsm}$).

Transition from B to $B^{\Delta U4\Delta NTC}$

Recruitment of the BCP could signify the initial step in this transition. It remains unknown how their recruitment is coupled to U1 snRNP release and Prp28 action. Thus, I cannot rule out the possibility of early B intermediate containing U1 snRNP and Prp28 as well. Shortly after the stable integration of the BCP, Brr2 helicase activity may be stimulated to unwind U4/U6 stem I. Then Dib1, di-snRNP factors (except for U6 snRNA and Lsm2-8) and tri-snRNP specific proteins (Snu66 and Prp6) are released by an as-yet undefined mechanism. Meanwhile, both Lsm2-8 and the BCP, that may function to stabilize U2/U6 helix II and the 5'SS/U6 duplex respectively, are retained in the spliceosome until the recruitment of the NTC.

As my work has only tested two U4 snRNP proteins (Prp3/Prp4 dimer) and Snu66, it remains uncertain whether other factors (Dib1, Prp6, Prp31, Snu13, U4 snRNA, U4 Sm) are released together as well. Even though Prp31, Snu13, Prp3/Prp4, U4 Sm and U4 snRNA are all components of U4/U6 di-snRNP, they bind to U4 snRNA with different affinities. Some potential experiments for ascertaining their coupled dynamics include monitoring the dynamics of U4 snRNA, Prp31, Prp6 and Dib1 relative to Prp3/Prp4. Given their locations in the B intermediate, any uncoupled dynamics (i.e., ordered rather than simultaneous release of these components) may suggest a role in facilitating the U2 snRNP movement toward $Prp8^{NTD}$ in the next step.

Transition from $B^{\Delta U4\Delta NTC}$ to $B_{NTC+BCP+Lsm}$

Upon the release of U4 and potentially other factors, a previously uncharacterized spliceosome state ($B^{\Delta U4\Delta NTC}$) appears. To the best of my knowledge, it features the presence of the BCP and Lsm2-8 and the absence of Prp3/Prp4, Snu66, and Syf1/Cef1 (NTC proteins). In relationship to structures in previous and later spliceosome states, additional structural

rearrangements including movement of U2 snRNP toward $Prp8^{NTD}$ and the initial formation of U2/U6 catalytic center (including the formation of U6 ISL, Ia, Ib helixes) may also occur during this state.

It could be beneficial to obtain a cryo-EM structure for the intermediate stalled at this transition step and compare it to earlier and later intermediates for further understanding the dynamic changes occurring during this transition.

Transition from $B_{NTC+BCP+Lsm}$ to B^{ACT}

At the beginning of this transition, the $B^{\Delta U4\Delta NTC}$ intermediate can stably bind to NTC proteins, generating another previously uncharacterized $B_{NTC+BCP+Lsm}$ complex. Further conformational changes and recruitment of other factors (B^{ACT}-specific factors, NTR proteins, Prp17, Cwc21 and Cwc22) that have overlapping binding sites with the BCP and Lsm2-8, may help to promote irreversible release of the BCP and Lsm2-8 to form the initial B^{ACT} complexes (defined by the absence of BCP and Lsm2-8).

Studying how different classes of proteins (NTR, RES, B^{ACT}-specific, Prp2, Spp2, Cwc21, Cwc22) are recruited at this stage can further our understanding in how an allosteric network forms between interactors near the 5'SS, BPS (Branch Point Site) and catalytic center. The ordered recruitment of these factors could potentially indicate the order opening of the corresponding binding sites. One useful landmark event for the formation of B^{ACT} is the recruitment of B^{ACT}-specific factors (Cwc24 or Cwc27, the latter is not essential). Key experiments that could lead to a better understanding of this stage are described as follows. Monitoring the recruitment of Cwc24 relative to release of BCP could inform about how the switch of protection at 5'SS is coordinated between step-specific factors. Monitoring the recruitment of Cwc24 relative to NTC could inform about how the protection signal around the 5'SS is temporally coupled to the formation of a new spliceosome conformation. As NTC and NTR proteins make up the key

scaffolding structures around the catalytic triplex, the poorly understood recruitment order of NTR proteins (e.g. Bud31, Cwc2, Ecm2) relative to the NTC can also be monitored for understanding the sequential order of stabilization interactions along U6 snRNA.

Late activation stage – Prp2-mediated Transition from B^{ACT} to B^{*}

No single molecule experiments from this thesis can yet illuminate the transition from B^{ACT} to B*. Translocation of Prp2 starting at intron downstream of BPA in a 3' to 5' direction leads to freeing the branch point adenosine from the SF3b subcomplex (specifically Hsh155/SF3B1) of U2 snRNP and freeing the scissile phosphate at the 5'SS from Cwc24 and Prp11 (a protein in SF3a subcomplex of U2 snRNP). Following this, several newly recruited factors (Isy1, Prp17, Yju2 and Cwc25) are hypothesized to bring the branch point and 5'SS closer to the catalytic center for the 1^{st} step of splicing. Some potential interesting experiments include monitoring the release of U2 snRNP proteins (Hsh155) relative to the recruitment of step1-specific factors (Isy1 or Cwc25) for understanding how the exposure of splicing signals is coordinated with the stabilization effects of step-1 specific factors (Crawford et al., 2013).

OUTLOOK

Understanding the Molecular Mechanism of Forming the U2/U6/Lsm2-8 Complex

Results from Chapter 2 have shown that U2 and U6 sequences comprising helix II can form a complex with Lsm2-8. Important regions promoting the formation of this complex are the poly-U tail at the 3'-end of U6 sequence and the base-pairing region of U2 and U6. The mechanism of poly-U recognition by Lsm2-8 has been studied previously. Previous psoralen cross-linking experiments done by a former lab member, Margaret Rodgers, showed that U2/U6 likely form base-pairing interactions. In the future, several approaches can be taken to better understand the process of forming U2/U6/Lsm2-8 complex. Obtaining a cryo-EM or crystal structure for the in vitro assembled U2/U6/Lsm2-8 can help directly identify the presence of U2/U6 duplex and identify the contact surface between the duplex and Lsm2-8. Alternatively, the duplex could be crosslinked to the protein to identify interaction sites. To better understand the process of forming the complex, the structure can be used as a guide for designing smFRET experiments. FRET donor and acceptors can be separately placed on U6 and U2 oligos for monitoring the duplex formation process. Experiments could potentially be designed to answer if an intermediate of U2 non-specific binding on U6/Lsm2-8 forms preceding to base-pairing interaction establishes and if the base-pairing interaction starts from the outer side or inner side of Lsm2-8. Successful detection likely relies on if the timescale of these processes matches the resolution of the experiment. Given the great similarities between Lsm2-8 and the well-studied bacterial Hfq, it would be interesting to test a specific model of Lsm2-8 binding to the stem loop structure in the U2 oligo (corresponding to the 5' stem loop in full length U2 snRNA) prior to facilitating its opening and reannealing with U6 oligo.

Characterization and Isolation of the tri-snRNP Complex Present at 2 mM ATP

Results from Chapter 3 led me to propose the presence of two states of tri-snRNP (tri-snRNP-HI and tri-snRNP-LO) depending on the concentration of ATP, characterized by the presence and absence of B complex proteins. As their presence in the tri-snRNP can potentially affect the action of Prp28 and locking the 5' stem of U6 snRNA in the exon channel, experiments need to be done to carefully distinguish the tri-snRNP present at low vs. high [ATP]. Initial tests can be done to monitor the dynamics of Prp28 and Sad1 during *in vitro* splicing. As Prp28 and Sad1 were initially missing from purified yeast tri-snRNP and my model proposes that they should be simultaneously present in yeast tri-snRNP as they are in the human complex, monitoring the dynamics of Prp28 relative to Sad1 under activation and inhibition conditions could provide a starting point. I expect their simultaneous presence on pre-mRNA molecules under normal activation conditions (2mM ATP) but not under activation inhibition conditions (0.05mM ATP).

One possible strategy for purifying tri-snRNP-HI could be depleting B complex proteins in *in vitro* assembled splicing reactions to prevent its integration into the tri-snRNP, and then carrying out sedimentation analysis to isolate tri-snRNP fractions. If successful, a cryo-EM structure of tri-snRNP-HI can be obtained and compared with tri-snRNP-LO to understand their interconversion.

Understanding the Mechanism of 5' exon Localization to the Exon Channel.

In my proposed model for the transition from tri-snRNP-HI to tri-snRNP-LO, the U6 snRNA is used as a substrate of Prp28 in the absence of pre-mRNA and placed in the exon channel. This model can be supported by cryo-EM data for tri-snRNP-HI and identification of the RNA occupying the exon channel. If the 5' stem of U6 snRNA is still in the channel, then it indicates an unknown molecular mechanism for switching from the 5' stem of U6 snRNA to the 5' exon in the channel at the stage of forming the B^{ACT} intermediate. If exon mimic sequence of yeast Prp28 is in the channel, then it is consistent with the human tri-snRNP architecture.

Understanding the Roles of the BCP in Spliceosome Activation

Results from Chapter 4 showed that Spp381$^{151\text{-}180}$ is the most essential region for yeast growth. If deleting this region does not just affect the mutants' cellular expression and localization, then it is most likely to affect yeast growth by impacting the *in vivo* splicing processes. If the mutant impacts either or both cellular expression and localization, then we would need to determine if correcting expression and localization still leads to a growth defect in order to assess if the mutants are also impacting the splicing process. Given the only known function of Spp381 in yeast is during spliceosome activation, the lethality phenotype conferred by deleting this region could result from impairing either the formation of B intermediate or the transition to B^{ACT} intermediate. The potential interaction partners of Spp381$^{151\text{-}180}$ include Prp38, Prp8^{Endo}, and Prp8$^{2090\text{-}2150}$ (a linker region between Prp8^{RH} and Prp8$^{Jab1/MPN}$). The Prp38/Spp381 interaction has been shown to be important for *in vivo* splicing. I also showed that either C-terminus or N-terminus region along with Spp381$^{151\text{-}180}$ is required for the proper cellular function of Spp381. Either terminus structurally extends to other parts of spliceosome, each potentially serving as an independent anchor point for stabilizing Spp381$^{151\text{-}180}$/Prp38 at the corresponding sites.

Even though the recruitment and release timings of BCP suggest that they play important roles in the Brr2-mediated activation step, these roles remain poorly understood. More data for the BCP can be found in studies of the homologous factors in the human spliceosome. Mapping the locations of these proteins in both human and yeast spliceosome reveals their potential roles centering around at least four themes. These could potentially provide a starting point for designing experiments and understanding their roles in activation in the future.

The first theme is positioning of the Brr2 helicase. Human BCP SMU1 and RED (no homologues in yeast) are placed between the U2 snRNP and Brr2. Human FBP21 appears to be placed near Brr2 (N-RecA-2 domain), potentially preventing access to U4/U6 helix I. While yeast does not have FBP21 homologue, the N terminus of Snu23 appears to localize to a similar location

as the FBP21 interaction site on Brr2 (N-RecA-2) and may further extend to the ATP-binding pocket of Brr2. Though only limited regions of human MFAP1 have been resolved by cryo-EM, its cryptic orthologue Spp381 in yeast has its α2 helix run along a groove formed between Brr2 and its regulatory domain Prp8$^{Jab1/MPN}$. Experiments can be designed to understand how each potential Brr2 interaction region in the BCP affects splicing and activation.

The second theme is interactions with the 5'SS/U6^{ACAGAGA} region. In the current yeast B complex structure, the 5'SS has not yet established an essential interaction with the A^{47}CAGAGA^{53} sequence of U6 and instead has limited interactions with the tip of a U6 stem structure (U6$^{38-40}$), at the base of which is the A^{47}CAGAGA^{53} site. Snu23 (which contains a nucleic-acid binding ZnF domain) and Prp38 appear to make interactions with part of the U6 stem structure (U6$^{37-46}$). In contrast, studies of the human B complex spliceosome show that the 5'SS/U6^{ACAGAGA} RNA duplex has formed and is stabilized by FBP21, hPrp38 and hSnu23. FBP21 appears to interact with the human 5'SS/A^{41}CAGAGA^{47} duplex and hPrp38 together with hSnu23 appear to interact with an extended duplex formed upstream of the U6 A^{41}CAGAGA^{47} sequence. In a cryo-EM structure of a later human spliceosome complex (pre-B^{ACT1}), the 5'SS/A^{41}CAGAGA^{47} duplex is almost enclosed by hPrp38, hSnu23 and hWBP11 (another human B-specific factor). Experiments can be designed to understand whether the proposed stabilization of 5'SS/U6 duplex by BCPs is important for splicing and activation.

The third theme is interactions of the C-terminus of yeast Spp381 (or the C-terminus of human MFAP1 together with UBL5) to interact with the RNA present in the exon exit channel. Though my results have shown that deleting the C-terminus of Spp381 has little impact for normal yeast growth, the stabilization effect may become important under other extreme conditions.

Finally, the fourth theme is bridging the gap between Prp8^{NTD} and Prp8^{Endo} using Prp38 and Spp381/MFAP1 (and potentially Snu23). As the Prp8^{NTD} and Prp8^{Endo} forms a gate to the exon channel, some components of the BCP could serve as a “gatekeeper” restricting access to

exon channel. Experiments can be done to test if the gate is stabilized by BCP and whether this interaction is important for splicing and activation.

In a sum, after recruitment of the BCP, they are mostly organized around the gate of the exon channel formed by $Prp8^{NTD}$ and $Prp8^{Endo}$. The function to potentially stabilize the 5'SS and the entract of the channel and either the 5' exon or U6 5' stem at the exit of the channel. In addition, ySnu23 or hFBP21/hSnu23 can potentially serve as a link for connecting these interactions to the N-terminal helicase cassette of Brr2 responsible for ATP-dependent U4/U6 unwinding during activation. Conceivably, upon proper positioning of the BCP, signaling (such as post-translational modifications) may be passed through the network via ySnu23 or hFBP21/hSnu23 to activate the helicase activity of Brr2. Preliminary tests could be done to test the importance of the N-terminus of yeast Snu23. Previously, phosphorylation/dephosphorylation, ubiquitylation/deubiquitylation and acetylation have been implicated in regulating different stages of splicing (Chanarat and Mishra, 2018; Mermoud et al., 1994; Wahl et al., 2009). It would be interesting to identify potentially interesting PTM sites in the BCP and test their effects on splicing and activation.

Towards Building a More Complete Kinetic Map

Given the known cryo-EM structures of key spliceosome intermediates, building a kinetic map for splicing is essential for gaining a deeper understanding on splicing mechanism. Two immediate advantages of having a spatial-temporal map are listed.

Currently available cryo-EM structures of spliceosomes represent snapshots of stable intermediates stalled in the process of splicing but with unknown transition rates. Knowing the transition rates between intermediates can inform about the relative energy landscape between different transitions. A faster transition indicates an energetically favorable process under assay conditions, such as in the transition from recruiting B complex proteins to releasing U4 snRNP proteins. A slower transition can suggest either a multi-step process or an energetically

unfavorable process. For example, the BCP stays much longer in an activation intermediate without U4 snRNP proteins than with U4 snRNP proteins, consistent with a two-step transition from $\mathrm{B}_{\mathrm{BCP}}^{\Delta\mathrm{U4}\Delta\mathrm{NTC}}$ to $\mathrm{B}_{\mathrm{BCP+NTC}}$ and then to $\mathrm{B}_{\mathrm{NTC}}$. Such kinetic information is hidden from simply examining snapshots of spliceosomes.

Many cryo-EM structures of the spliceosome have been obtained by attempting to remove the conditions required for transitions between steps, such as using ATPase mutants, varying ATP or magnesium concentrations, inhibiting phosphatase activity, and other strategies. For a complicated and dynamic process such as the transition from B to B^{ACT}, it remains challenging to stall intermediates after a specific sub-step. Currently, key intermediates are still undefined for further understanding of this process. Monitoring the dynamics of splicing factors relative to landmark events in splicing can help reconstruct the order of events that occur in real time, thus simplifying the complicated splicing process into set of transitions mapped between intermediates with characteristic compositions. An added benefit is identifying previously unknown intermediates and providing potential strategies for stalling these complexes for cryo-EM or other studies. My results suggest the existence of at least two uncharacterized intermediates ($\mathrm{B}_{\mathrm{Lsm+BCP}}^{\Delta\mathrm{U4}\Delta\mathrm{NTC}}$, $\mathrm{B}_{\mathrm{Lsm+BCP+NTC}}$) in activation.

In addition, questions arise regarding how to relate cryo-EM snapshots of spliceosome intermediates to a kinetic mechanism. A cryo-EM structure of spliceosome intermediate represents an average solution to an ensemble of single spliceosome particles with similar structural states, potentially likely corresponding to the most stable state prior to the transition to the next intermediate. A single interval in the kinetic map describes an average time spent by an intermediate with characteristic compositions, likely corresponding to a time series of the compositionally similar intermediate but with evolving structural states (rugged energy landscapes). A cryo-EM structure of a stalled spliceosome intermediate likely corresponds to one of these structural states within an interval of the kinetic map. To illustrate different possible

structural states within an interval, I showed examples of initial and end structural states in the interval for $B_{Lsm+BCP}^{\Delta U4\Delta NTC}$ (See Chapter 3, **Figure 3.21**). Its initial structural state ($B_{Lsm+BCP}^{\Delta U4\Delta NTC}$, step 1) may resemble the cryo-EM structure of the previous B intermediate. Its end structural state ($B_{Lsm+BCP}^{\Delta U4\Delta NTC}$, step 2) may more closely resemble the cryo-EM structure of latter B^{ACT} intermediate.

Key experiments for better understanding of activation are listed in the above **SUMMARY** section. However, similar methods can be extended to other stages of the splicing cycle.

Isolating Spliceosomal Intermediates by Mutagenizing Step-specific Factors

Step-specific factors can be potential mutagenesis targets since their functions are required for specific transitions during splicing. I have shown an example of mutagenizing one of BCP component, Spp381. The strongest phenotype I observed is from mutagenizing the PEST region of a minimally functional Spp381. In the future, this mutant strain can potentially be used for stalling spliceosome activation *in vitro*. These *in vitro* assembled spliceosomes can be isolated and characterized to understand at which stage the spliceosome is stalled. If BCP recruitment is affected, the intermediate may be stalled prior to BCP integration with compositional similarity to the pre-B spliceosome. If the BCP is present in the isolated intermediate, this could indicate a loss of function during activation. Cryo-EM structures can potentially be obtained to further understand how the stalled intermediate fits into the activation process.

Potential applications of the Kinetic Map

A kinetic map derived from analyzing the orders of observed events in single molecule experiments can provide an independent model of splicing process. The information can be used to cross-verify the information from other models, such as the ones built from cryo-EM structures. The comparison from different models can be informative, for they provide information from different perspectives. Cryo-EM structures provide a spatial map of stable interactions within spliceosome intermediates. However, the kinetic map provides relative temporal information of

how long different proteins spend in splicing-related intermediates, regardless of the stability of interactions within those intermediates. One example for illustrating the usefulness of comparing different models can be from the study of the BCP. While the BCP was previously regarded as a components of the yeast tri-snRNP as evidenced from biochemical and structural studies, my kinetic map shows that they are most often recruited after the tri-snRNP is integrated into the spliceosome under splicing conditions. The kinetic map can provide insights into splicing from the perspective of real-time splicing dynamics.

The kinetic map can also be used as a reference for understanding the effects of different conditions, mutants, or inhibitors on specific splicing stages as compared to the normal conditions. My results showed that the BCP is mostly recruited to the spliceosome together with the tri-snRNP under activation inhibition conditions (0.05mM ATP) but are recruited to the spliceosome after tri-snRNP integration under normal activation conditions (2mM ATP). A negative regulation step can lead to abnormal binding events (such as changes in characteristic lifetimes of intermediates, abnormal recruitment, or release events) of factors acting downstream of the step but have little to none effect for factors acting upstream. I expect this reference map would be a useful source for identifying the exact negatively regulated steps in the splicing process, thus helping us elucidate complicated splicing regulatory mechanisms in the future.

Possible Improvements for Improving the Kinetic Map.

Given the great advantages of constructing a kinetic map for splicing, more improvements can be made in the future to make the process more efficient, accurate, and high throughput. To avoid the inconsistent kinetic parameters from tagging different splicing factors, alternative labeling strategies can be considered, such as reconstituting extracts by adding purified and site-specifically labeled proteins into extracts with proteins of interest depleted. This system could potentially be more controllable in terms of the amounts of labeled proteins present than the approaches used here. It may also allow introduction of multiple proteins simultaneously with site-

specific labeling and higher labeling efficiency. To improve the accuracy of the map, higher temporal resolution could be needed to distinguish simultaneous events from ordered events. This would need to be accompanied by an increase in the data analysis speed such as finding better algorithms to identify desired binding events automatically and accurately such as those recently implement in Tapqir (Ordabayev et al., 2022).

REFERENCES

Chanarat, S., and Mishra, S.K. (2018). Emerging Roles of Ubiquitin-like Proteins in Pre-mRNA Splicing. Trends Biochem Sci *43*, 896-907. 10.1016/j.tibs.2018.09.001.

Crawford, D.J., Hoskins, A.A., Friedman, L.J., Gelles, J., and Moore, M.J. (2013). Single-molecule colocalization FRET evidence that spliceosome activation precedes stable approach of 5' splice site and branch site. Proc Natl Acad Sci U S A *110*, 6783-6788. 10.1073/pnas.1219305110.

Mermoud, J.E., Cohen, P.T., and Lamond, A.I. (1994). Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism. EMBO J *13*, 5679-5688.

Ordabayev, Y.A., Friedman, L.J., Gelles, J., and Theobald, D.L. (2022). Bayesian machine learning analysis of single-molecule fluorescence colocalization images. Elife *11*. 10.7554/eLife.73860.

Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701-718. 10.1016/j.cell.2009.02.009.