University of California
Santa Cruz

Mechanistic and structural studies of spliceosome active site formation

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

MOLECULAR, CELL and DEVELOPMENTAL BIOLOGY

By

Andrew J. MacRae

June 2018

The dissertation of Andrew J. MacRae is
Approved:

Professor Melissa S. Jurica, Chair

Professor Harry F. Noller

Professor Michael Stone

Tyrus Miller
Vice Provost and Dean of Graduate Studies

Table of Contents

List of figures

Abstract:

Mechanistic and structural studies of spliceosome active site formation

Andrew MacRae

The spliceosome is a eukaryotic molecular machine composed of uridine-rich small nuclear RNAs (U snRNA) and dozens of proteins that facilitate an RNA processing event known as pre-RNA splicing. By catalyzing the removal of intronic RNA the spliceosome plays an critical role in determining the final RNA sequence from pre-RNA transcripts. Proper removal of introns in pre-mRNA maintains the coding register of mature mRNAs that are eventually translated into functional proteins by the ribosome. Accurate splicing is necessary to repeatedly generate mRNAs coding for the same protein across multiple cells. Additionally, alternative splicing, the regulated inclusion and exclusion of specific exons in the final mRNA, greatly expands the coding potential of eukaryotic genomes. Introns and spliced non-coding RNAs also serve important cellular roles. During splicing the spliceosome undergoes multiple timed rearrangements to properly define the boundaries of each intron and generate the spliceosome active site. Due to the spliceosome's structurally transient nature the precise functions and mechanisms driving many rearrangements remains undefined.

Chemical probing is a powerful tool used for biomolecule structural research and is capable of characterizing spliceosome structural features and changes during splicing. During my thesis I used a new protein probing protocol designed by the Jurica lab to globally map lysine reactivity within multiple conformations of the spliceosome. From our data and published cryo-electron microscopy structures of the spliceosome we predicted how the core spliceosome scaffold Prp8 interacts with U5 snRNA and its 5′ exon binding partner during active site formation. We then used our model to guide genetic experiments and tested the necessary amino acids within Prp8 for putative roles in 5′ splice site selection. Our genetic data suggests a subset of lysines that change in probe reactivity between the pre-activated and catalytic spliceosome work together with specific spliceosome protein components (Isy1 and Snu114) to position the 5′ splice site.

In addition to chemical probing, I designed experiments to purify a more conformationally homogenous spliceosome sample for cryo-electron microscopy structure determination. During this work I added a lithium salt wash to my spliceosome purification procedure hoping the salt would wash away loosely associated proteins that bind the spliceosome sub-stoichiometrically. If correct, then 2D particle reconstruction from negative stain electron micrographs will show a more clear average of spliceosomes with higher molecular detail than spliceosomes purified without lithium washes. During the course of this project I successfully purified catalytic spliceosomes, washed them with lithium acetate then subject them to

analysis via negative stain electron microscopy. From electron micrographs I generated two dimensional averaged particles that exhibited highly similar features to spliceosome purified without lithium. Lightly crosslinking the spliceosome complexes together prior to imaging was additionally used to determine if lithium exposure washed away spliceosome components destabilized during sample preparation. Unfortunately due to difficulties in cryo-EM sample preparation and imaging I never was able to generate a high resolution spliceosome structure from cryo-EM data.

Dedication

To all the life in this world that have talked, listened or inspired me to study biology.

Acknowledgements

I would like to acknowledge my mentor and faculty advisor Dr. Melissa Jurica for her continual guidance, thoughtfulness and conversation that generated this thesis. All Jurica lab member past and present in some manner shaped the work presented here.

My thesis committee members Dr. Harry Noller and Dr. Michael Stone provided essential critique to my thesis. I particularly appreciate Dr. Noller for dozens of thoughtful and challenging discussions regarding my thesis. Dr. Manny Ares, Dr. Alan Zahler, Dr. Jeremy Sanford also engaged me in beneficial conversation throughout my thesis.

The following collaborators directly worked on integral aspects of the work presented in this thesis. Dr. Megan Maylerle performed the yeast genetic experiments described in chapter 3 with guidance from Dr. Christine Guthrie. Dr. Robert Chalkely performed all mass spectrometry experiments. Dr. Alma Burlingame played a pivotal advisory role in mass spectrometry experiments. Dr. Eva Robinson and Dr. Patricia Coltri designed the initial lysine acetylation protocol.

The molecular graphics displayed in chapters 2 and 3 and corresponding analyses were performed with the UCSF Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco.

I would like to acknowledge my all of my family for their continual support of my scientific endeavors, specifically my brother Ian MacRae.

Finally I would like to acknowledge the National Institutes of Health for funding my thesis research.

Chapter 1

Introduction

1.1 The central dogma of molecular biology

DNA is the central repository for genetic information. In eukaryotic cells, DNA is stored within the membrane bound compartment named the nucleus. Cells use the information within DNA to build cellular components, including enzymes. Enzymes in turn are responsible for facilitating and regulating chemical reactions throughout the cell. The complete flow of information from DNA to enzyme is referred to as the central dogma of molecular biology and occurs through multiple RNA intermediates. From a chemical perspective, RNA is essentially identical to DNA except RNA contains a hydroxy group at 2′ carbon atom of its ribose ring. This hydroxy group imbues RNA with the ability to function as an enzyme as well as store information in an analogous manner to DNA. The 2′ hydroxy is also responsible for RNA's decreased stability relative to DNA.

During eukaryotic gene expression RNA polymerase uses DNA as a template to generate mRNA and non-coding RNAs. A mRNA contains the nucleotide sequence corresponding to the coding region within the gene from which it was transcribed and is used by the ribosome to make a protein product based on sequential RNA base triplets, each RNA triplet encoding for an amino acid. However, RNA must be processed after it is generated and before it moves on to its final functional role in the

cell. RNA processing events include capping, polyadenylation, covalent modification and pre-mRNA splicing.

1.2 The origins and necessity of pre-RNA splicing

The vast majority of human genes contain sequence information that interrupts the would be coding sequence needed by the ribosome to translate a functional protein. Within a pre-mRNA these sequences are referred to as introns and are chemically excised by a protein RNA (RNP) machine known as the spliceosome. Pre-mRNA splicing functions to remove non-coding information and connect coding information from pre-mRNAs. High fidelity splicing is essential to eukaryotic gene expression because if intron boundaries are improperly defined the final processed RNA will not contain the correct sequence. Depending on the classification of RNA mis-splicing can be incredibly damaging to cells. In the case of mRNAs mis-splicing can result in an altered coding potential via frame shifts that can change the primary amino acid sequence of a cellular protein. Failure to properly splice out introns in non-coding RNAs is likely to alter the secondary and or tertiary structure of the RNA an in turn alter its functional potential.

Data suggesting the existence of splicing was published as early as 1977 when it was shown that mRNAs generated from adenovirus-2 infected cells were not 100% complementary to the DNA from which it was transcribe (Berget, Moore et al. 1977). Instead the mRNA appeared to base pair with segmented stretches of DNA (as visualized by electron microscopy). DNA lacking complementarity to RNA looped out,

forming structures known as R loops. Later, it was shown that a class of U-rich small nuclear RNAs (U snRNA) stably associate with a specific set of proteins and likely play a role in pre-RNA splicing. By demonstrating the U1 snRNA and associated proteins (U1 snRNP) are necessary for the splicing of viral RNAs in adenovirus-2 infected cells (Yang, Lerner et al. 1981) U snRNPs role in splicing began to solidify. As the catalogue of spliced eukaryotic RNAs expanded researchers hypothesized as to why eukaryotic genomes contain so much DNA coding for introns.

In addition to spliceosomal introns there exits classes of self-splicing introns (group I and group II). First discovered in Tetrahymena (Kruger, Grabowski et al. 1982) self-splicing introns are autocatalytic RNA elements capable of catalyzing their own removal from primary RNA transcripts. Many RNA biologists hypothesize the spliceosomes and group II introns share a common evolutionary ancestor. This hypothesis is strengthened given the high amount of structural similarity between the U snRNAs and many group II introns (Toor, Keating et al. 2008). Additionally, the essential spliceosome scaffold protein Prp8 holds much structural similarity to maturases that stabilize group II intron active sites and reverse transcribe free RNA introns back into genomic DNA (Galej, Oubridge et al. 2013).

1.3 The chemistry of pre-RNA splicing and intron sequence features

Two chemical steps are needed to remove each intron. During the 1st step an adenosine residue within the intron nucleophilically attacks the 5′ end of the intron. This severs the 5′ end of the intron from the upstream exon and is the first catalytic

step of splicing. This step additionally forms a branched adenosine, as the residue is now covalently bound to three separate nucleotide residues (Ruskin, Krainer et al. 1984). Following, the most 3′ residue of the 5′ exon nucleophilically attacks the most 5′ end of the downstream 3′ exon. This is the second catalytic step of splicing and effectively ligates the 5′ and 3′ exons together while releasing the now branched intron.

Considering that intron removal is essential to eukaryotic life it is surprising that introns appear to have very few conserved primary sequences. This begs the question as to how does a spliceosome recognize the boundaries of an intron and the branch adenosine without error? There are four main sequences within introns that assist in intron identity. All introns start and end with a 5′ and 3′ splice site consensus sequences. Both of these sequences include an invariable dinucleotide, GU at the 5′ end and AG at the 3′ end, because of this all introns begin and end with G. the 5′ and 3′ splice sites include more than their respective invariant dinucleotides but vary across introns and species. Introns also contain branch sequences that house the branch adenosine. A polypyrimidine tract is also found just downstream of the branch.

As described above the branch adenosine is the residue that nucleophilically attacks the 5′ splice site during the first catalytic step of splicing. Together the branch and PPYT attract early spliceosome assembly factors including the U2 snRNP. Recognition of the correct branch and PPYT is critical as it has been demonstrated that

mutations within U2 snRNP proteins that alter the spliceosomes ability to use the correct branch results in mis-splicing (Tang, Rodriguez-Santiago et al. 2016).

1.4 A summary of spliceosome assembly and catalysis

RNA splicing is an incredibly dynamic process. During splicing the spliceosome undergoes the rapid gain and loss of protein and RNA components throughout assembly and active site formation (Figure 1). Initially the U1 and U2 snRNP bind to the 5′ splice site and the branch point adenosine, respectively (Rogers and Wall 1980, Black, Chabot et al. 1985). This complex is referred to as spliceosome A. At this point the tri-snRNP (containing U4, U5 and U6 snRNA) joins and forms the pre-activated spliceosome (spliceosome B) (Hashimoto and Steitz 1984, Black and Pinto 1989). This stage of assembly is of particular note because the tri-snRNP contains Prp8. Prp8 is an essential scaffold that structures much of the spliceosome active site. Decades of genetic and biochemical work has shown that Prp8 interacts with many spliceosome components in multiple regulatory ways (Galej, Nguyen et al. 2014). Following spliceosome B formation the active site is generated. Spliceosome active site formation includes many large scale rearrangements such as the unwinding of U4 and U6 snRNA duplex (facilitated by the spliceosome helicase Brr2) and generating extensive base pairing between U6 and U2 snRNA (Madhani and Guthrie 1992). By this point U1 and U4 snRNA have dissociated from the spliceosome. This complex is referred to as an activated spliceosome B.

Now that the spliceosome active site has been formed the branch adenosine is positioned to attack the 5′ splice site and form the intron lariat (Rauhut, Fabrizio et al. 2016). Following the first step of splicing the catalytic spliceosome (spliceosome C) rearranges to accommodate the 3′ splice site into the active site such that the 5′ splice site can attack the 3′ exon (Fica, Oubridge et al. 2017). Both catalytic steps are facilitated by U2 and U6 snRNA along with two magnesium ions (Fica, Tuttle et al. 2013), thus the spliceosome is a true ribozyme. Additionally, the U5 snRNA is believed to interact with the end of 5′ exon and beginning of the 3′ exon (Newman and Norman 1992) likely partially contributing to their stable position within the active site. Finally there are a number of proteins associated with the spliceosome that are implemented in specific aspects of splicing. Examples include 1st and 2nd step splicing factors, early assembly factors and a number of proteins that regulate prp8 conformation and or function. Many of these proteins are helicases that likely regulate RNA structure in an ATP dependent manner.

1.5 The current status of spliceosome structural knowledge

For most of the history of splicing research the community lacked high resolution models. The first instances of gross molecular spliceosome morphology emerged from cryo-electron microscopy (cryo-EM) in the mid 2000's. While the structures were low resolution (30 angstrom) they provided initial glimpses of relative snRNP positioning and global architecture (Jurica, Sousa et al. 2004). For a decade following those initial structures researchers used cryo-EM to generate additional

informative low resolution models. Various research groups successfully generated x-ray structures individual proteins and pieces of snRNPs (Vidovic, Nottrott et al. 2000, Liu, Luyten et al. 2001, Pomeranz Krummel, Oubridge et al. 2009). It took large technical advancements in both hardware and software of EM before near atomic models could be generated from cryo-EM data sets. EM advancements include direct electron detectors, a movie mode that reduces the noise of sample molecular motion and better particle averaging techniques during model reconstruction (Koning, Koster et al. 2018).

In 2015 the Nagai lab was the first to present and publish a high resolution model of a large component of the spliceosome, the tri-snRNP at 5.9 angstrom (Nguyen, Galej et al. 2015). The tri-snRNP is a complex of U4, U5 and U6 snRNA that has yet to assemble over an intron. Once this work was published three labs (Nagai, Shi and Lurhmann) primarily generated all of the current structures of tri-snRNPs and stalled spliceosome intermediates from yeast and human systems. Currently, their exists over 15 published spliceosome models under 8 angstroms and certainly more to come. Going from zero to over a dozen high resolution models of spliceosomes has given the splicing community so much structural information it is challenging to decide where and how to best test models regarding regulated spliceosome conformational change and individual spliceosome component function. For instance comparing the structures of pre-activated and catalytic spliceosomes highlights the structural gymnastics that must occur to generate the spliceosomes active site.

1.6 Spliceosome activation and catalysis

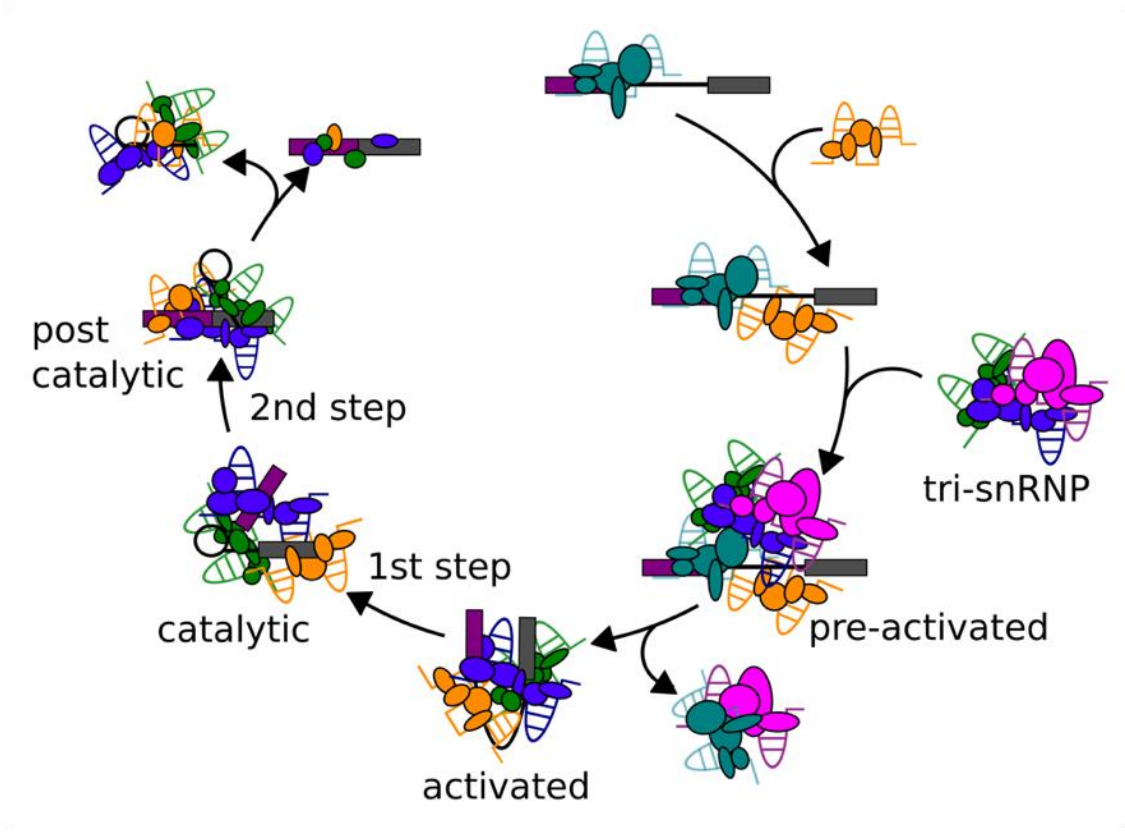
After the tri-snRNP joins the U1 and U2 snRNPs on an intron the spliceosome RNA network must be vastly remodeled to form the active site. During remodeling U4/U6 snRNA interactions are destabilized by the helicase Brr2 (Raghunathan and Guthrie 1998), a region of U6 (ACAGAGA helix) base pairs with the 5′ splice site (Sawa and Shimura 1992) and base pairing between U2 and U6 snRNA structures three key U6 residues referred to as the catalytic triad that interact with two catalytic magnesium ions (Wu and Manley 1991). As seen in multiple cryo-EM reconstructions the large scaffolding protein Prp8 along with a handful of other factors stabilizes much of this snRNA active site and move the yet to be spliced pre-RNA into the proper orientation for the first and second steps of catalysis. The only other snRNA in the spliceosome active site formation is U5. While U5 snRNA is believe to position and or stabilize the pre-RNA into the active site via direct interaction with the 5′ and 3′ exons (Newman and Norman 1992, Cortes, Sontheimer et al. 1993) much less is known about how U5 snRNA activity is regulated, relative to U2 and U6 snRNA function.

In addition to the above three snRNAs and Prp8 there are many other protein factors that regulate active site formation and catalysis. Genetic experiments implement Snu114 (a predicted GTPase) as a regulator of activation through Prp8. Spliceosome structure reconstructions model Snu114 directly contacting Prp8, further supporting its putative role regulating Prp8 conformation. The SF3B complex is a U2 associated group of proteins responsible for identifying the branch adenosine

within an intron. Just prior to activation SF3B1 houses the branch adenosine over 20 angstroms away from the active site. By activation the branch adenosine is displaced into the active site and structured for the 1st step of splicing catalysis. Given that there exist static images of spliceosome complexes with these proteins modeled we are now in a position to more precisely determine how these and other proteins function to organize, support and or rearrange the spliceosome active site.

1.7 Figures

Figure 1: Cartoon model of spliceosome assembly and catalysis. snRNPs correspond to the following color scheme; U1 teal, U2 orange, U4 magenta, U5 green and U6 blue. 1st and 2nd steps are labeled and defined by the two chemical steps of splicing. The pre-mRNA is colored with a purple 5′ exon and grey 3′ exon with a line signifying the intron. Various spliceosome intermediates are labeled.



Chapter 2

A lysine probing method to measure conformational change in RNPs

2.1 Abstract

High-resolution modeling of macromolecular complexes presents a unique and challenging barrier to structural biologists. Because of the varying stability and difficulty of homogeneous purification of multi-megadalton complexes it is likely that multiple methods in combination will be necessary to best understand molecular mechanisms driving regulated conformational change. Here we present a lysine chemical probing strategy combined with tandem mass-spectrometry to globally map reactive and non-reactive lysine residues within purified macromolecular complexes specifically ribonucleo-protein complexes. Our method can also gauge RNP structural and or conformational heterogeneity. To quantitatively validate lysine probing patterns reflect structure we probed the *E. coli* 30S ribosome subunit and compared the results to crystallographic models. We additionally provide a comparative analysis of probing results from multiple conformations of purified human spliceosomes. The presented results demonstrate this strategy is an advantageous orthogonal biochemical tool to accompany other high-resolution models of macro-molecular complexes.

2.2 Introduction

Chemical probing is a well-established method for studying protein and RNA structure. There exist many commercially available reagents that uniquely modify specific amino acids with well-defined respective chemistries (Vachet, 2009). For ribonucleo-protein complexes (RNPs), lysine accessibility is particularly informative because positively charged lysines often interact with and structure the negatively charged phosphate backbone of nucleic acids. With large RNPs it is often challenging to determine where important conformational changes exist. Lysine residues that change from interacting to not interacting, or vice versa, with negatively charged species are putative sites of regulated RNP conformation. The loss or gain of lysine-RNA interactions can potentially regulate RNA kinetics and result in stabilizing or destabilizing competing local RNA structure. With this in mind we sought to create a new amino acid probing strategy that would identify changes in lysine interactions between unique conformations of a RNP.

If the probing method does not modify all lysines residues a problem arises when the acetylation patterns of two conformations of a RNP are compared. Because proteins are trypsinized prior to tandem mass spectrometry (MS/MS) directed lysine acetylation state identification and trypsin cannot cut after acetyl-lysine, changes in lysine reactivity patterns between RNP conformations will generate different MS/MS peptides. The differences in MS/MS peptides lowers your ability to analyze the state of the same lysine across conformations because not all peptides can be measured

via MS/MS. Reasons for this difference include sequence dependent variation in peptide ionization. Because of this it is necessary to modify all lysine residues in a RNP to generate the same MS/MS peptides across multiple conformations of an RNP. To accomplish this we developed a two-stage acetylation assay that results in full acetylation of all lysine residues and discerns residues as reactive or non-reactive in the native RNP state.

First we incubate a natively assembled RNP with a lysine acetylation reagent (sulfo-NHS) to acetylate any reactive lysine residues. Following the first acetylation step the RNP is denatured via SDS-PAGE and proteins are acetylated with a deuterated acetylation reagent (D6-anhydride) while still fixed in the acrylamide matrix. D6-anhydride contains all deuterated hydrogens in the acetyl groups it eventually chemically passes to the lysine side chains. Thus the second acetylation step will deutero-acetylate any lysine that went unacetylated during the first probing step. Presumably, a lysine will not be acetylated during the first acetylation step because it is un-accessible to the acetylation reagent or otherwise interacting with a negatively charged species (positively charged lysine residues do not have any available lone pairs of electrons necessary for acetylation chemistry). After fully acetylating purified RNPs MS/MS can globally map which lysines are acetylated or deutero-acetylated (Figure 2) because the duel acetylation tactic results in a measurable three atomic unit difference between MS/MS peptides. Changes in the modification state (i.e. acetyl to or from deutero-acetyl) of a given lysine between two conformations suggests the

lysine is losing or gaining an interaction partner. To test for functionality of lysines that loose or gain interacting partners individual lysines can be targeted for mutagenesis to search for phenotypes.

2.3 Results

Lysine probing is reflective of RNP structure as determined by x-ray crystallography

To demonstrate that acetylation patterns are reflective of structure we subjected purified 30S *E. coli* ribosome subunits to our two step lysine acetylation method (Figure 2). The ribosome was chosen because it is a large RNP with available well defined high-resolution structures. A single MS/MS run measured 108 individual lysine residues within the 30S ribosome subunit. Of those 108 lysine residues, 30 were acetylated, 64 were deuterated and 13 were mixed. Mixed being a scenario where that lysine was found to contain near equal amounts of both modifications, suggesting local conformational flexibility or heterogeneity. We next physically mapped the position of all MS/MS lysines within the 30S subunit of a 70S *E. coli* ribosome (Noeske, Wasserman et al. 2015) as determined by x-ray crystallography. Visually most deuterated lysines (cyan) contact the RNA interface of the 30S subunit (Figure 3). To quantitatively gauge how the acetylation pattern correlates with the x-ray generated model we took 3 measurements from the crystal structure per lysine measured by MS/MS predicted to correlate with either lysine acetylation state. These measurements include; distance to nearest atom, B-factor and the solvent accessible surface area (SASA). If acetylation patterns reflect structure then all three values

should skew lower for deuterated lysines. This is because deuterated lysine side chains that are either buried within the complex or involved in charge-charge interactions are more rigid, defined in three dimensional space and likely to be close to other atoms. As predicted we found that all three numbers trend lower for deutero-acetyl-lysines when compared to acetyl-lysines (Figure 3).

Two-stage chemical probing can measure lysine differences and similarities between multiple spliceosome intermediates

In this chapter I compare replicate data sets after probing multiple purified spliceosome intermediates (Figure 4A) we demonstrated that acetyl-lysine and deutero-acetyl-lysine can be resolved via MS/MS traces (Figure 4B-C). Knowing that we can measure both lysines we next validate the acetylation reaction efficiency. By ensuring complete acetylation of all lysines, the two-stage modification strategy guarantees that trypsin digestion will generate comparable peptides from spliceosomes in different conformations for MS/MS analysis. Accordingly, nearly all peptides from spliceosome proteins that we sequenced by MS/MS ended in arginine (Figure 3D). Finally, we found the majority of individual lysine residues have the same acetylation state when comparing replicate samples of three different spliceosome assembly states, indicating that the modification patterns are reproducible across multiple preparations (Figure 4E). In the proceeding chapter

2.4 Discussion

The data described above demonstrates the method measures RNP structure, is reproducible and can more easily compare lysine modification states between differing RNP conformations than a method using only a single probing step. These metrics are important to clearly define prior to using the method to probe spliceosomes and or other RNPS and demonstrate its ability to measure changes in structure. Given that the splicing field and other researchers studying changes in RNP structure will see an increasing volume of high resolution structures in the near future it is important to continue to use biochemical methods such as chemical probing to test and add detail to RNP models. Detail specifically related to individual component movement will help sort out which individual gears drive movement of large and complex biological machines.

2.5 Materials and methods

Chemical probing by two-step acetylation

In the first acetylation reaction, one to five pMol of purified 30S *E. coli* ribosome subunits or human spliceosomes was incubated in 5 mM N-hydroxysulfosuccinimide (Pierce Technology) at room temperature for one hour. Reactions were quenched with 1/10th volume 1 M Tris pH 7.9, and then subjected to SDS-PAGE. After staining with Coomassie-G (5% w/v aluminum sulfate 14-18 hydrate, 10% v/v ethanol, 0.02% w/v CBB G-250, 2% v/v phosphoric acid), the entire lane containing 30S ribosome subunits or spliceosome proteins was excised into 6-9 slices. For the second acetylation reaction, gel slices were washed with shaking agitation in

1) water for 10 minutes at 37°C, 2) 10 mM ammonium bicarbonate at room temperature, 3) 50 mM ammonium bicarbonate:acetonitrile (1:1) twice for 45 minutes at 37°C, 4) water twice, and 5) 100% acetonitrile twice at 37°C for 5 minutes. After complete removal of the acetonitrile, 20 µl of D6-acetic anhydride (Acros Organics) mixed with 40 µl of 100 mM ammonium bicarbonate was added and completely absorbed by the gel. The gel slices were then submerged in 100 mM ammonium bicarbonate, and the pH adjusted to 7-8 with 1 M ammonium bicarbonate. After 60 minutes at 37°C, the gel slices were washed with water three times and submitted for MS/MS analysis.

30S ribosome heat activation

Purified 30S *E. coli* ribosomal RNA and proteins (provided by Laura Lancaster of the Noller lab) in tris-HCl pH 7.5, 100mM NH_4Cl, 10mM $MgCl_2$, 6mM BME were assembled into 30S subunits via incubation at 42° for 15 mins. After reconstitution 30S ribosomes subunits were buffer exchanged via Slide-A-Lyzer mini dialysis device (Thermo Fisher Scientific) into the following: 20mM hepes pH 7.5, 100mM KCl, 10mM $MgCl_2$ and 1mM DTT. 30S ribosomes were buffer exchanged for 5 hours at 4°.

Spliceosome purification

See chapter 3 material and methods.

Protein digestion and mass spectrometric analysis

For protein digestion, gel pieces were washed in 25 mM ammonium bicarbonate / 60% v/v acetonitrile. Disulfide bonds were reduced using 10 mM

dithiothreitol in 25 mM ammonium bicarbonate for 30 minutes at 60˚C, then free sulfhydryls were alkylated using 20 mM iodoacetamide in 25 mM ammonium bicarbonate for an hour. Gel pieces were washed again using 25 mM ammonium bicarbonate / 60% v/v acetonitrile, and then digested overnight using 200 ng TPCK-treated porcine trypsin (Promega) in 25 mM ammonium bicarbonate. Peptides were extracted using 50% v/v acetonitrile/ 1% v/v formic acid. Extracted peptides were dried down by vacuum centrifugation, and then resuspended in 0.1% v/v formic acid prior for mass spectrometric analysis.

Peptides were analyzed by LC-MS/MS using a NanoAcquity (Waters) UPLC system interfaced with either an LTQ Orbitrap Velos or QExactive Plus (both Thermo) mass spectrometer. Peptides were separated using an integrated 75 µm x 15 cm PEPMAP reverse phase column and spray tip (EasySpray source) using a gradient from 2 to 27% v/v solvent B (0.1% v/v formic acid in acetonitrile) over solvent A (0.1 % acetonitrile in water) at a flow rate of 400 nL/min for 27 min. Using the Velos, survey scans were measured at a resolution of 30,000 full width at half maximum (FWHM). The six most intense precursor ions were automatically selected for HCD fragmentation analysis by data-dependent acquisition and measured at a resolution of 7500 FWHM. For QExactive data, survey scans were measured a resolution of 70,000 FWHM, and were followed by data-dependent acquisition of the top ten most intense precursors at a resolution of 17,500 FWHM.

Raw data was converted to mgf format peak list files using in-house software based on the Raw_Extract script in Xcalibur version 2.4, and then searched using Protein Prospector version 5.18.22 against a database of all human proteins in SwissProt downloaded on May 9th 2016 supplemented with entries for E. coli maltose-binding periplasmic protein, enterobacteria phage MS2 coat protein and pig trypsin, plus decoy versions of all of these sequences (a total of 20204 entries searched). Carbamidomethylation of cysteine residues was searched as a constant modification, and variable modifications considered were acetylation or deuteroacetylation of uncleaved lysines, methionine oxidation, pyroglutamate formation from N-terminal glutamines, and protein N-terminal methionine removal and/or acetylation. Results were thresh-holded to an estimated 1% protein FDR according to target:decoy database searching (Elias and Gygi 2007). Intensities of peaks were extracted from the raw data using Protein Prospector by summing together signal over a period -10 to +20 seconds from the time the peak was selected for MSMS.

Mapping MS/MS peptides to 30S subunit and quantitative graphs

Lysine residues were mapped and colored using chimera. Distance to nearest atom was measured from the terminal nitrogen of the lysine side chain and measured by chimera. SASA and B-factor were additionally measured and pooled via chimera. Graphs were generated using R studio.

2.6 Figures

Figure 2: Two step acetylation outline. Top, cartoon model of the two-step lysine acetylation protocol. Assembled RNPs are acetylated then proteins are separated via SDS-PAGE and in gel deutero-acetylated. Bottom, a cartoon diagram showing differentially modified lysines. The lysine interacting with the phosphate will not be subject to acetylation but is subject to deutero-acetylation due to the charge state of the lysine in the native RNP. The lysine not interacting with a phosphate will be subject to acetylation (Ribosome image taken from Noller lab page (RNA)).

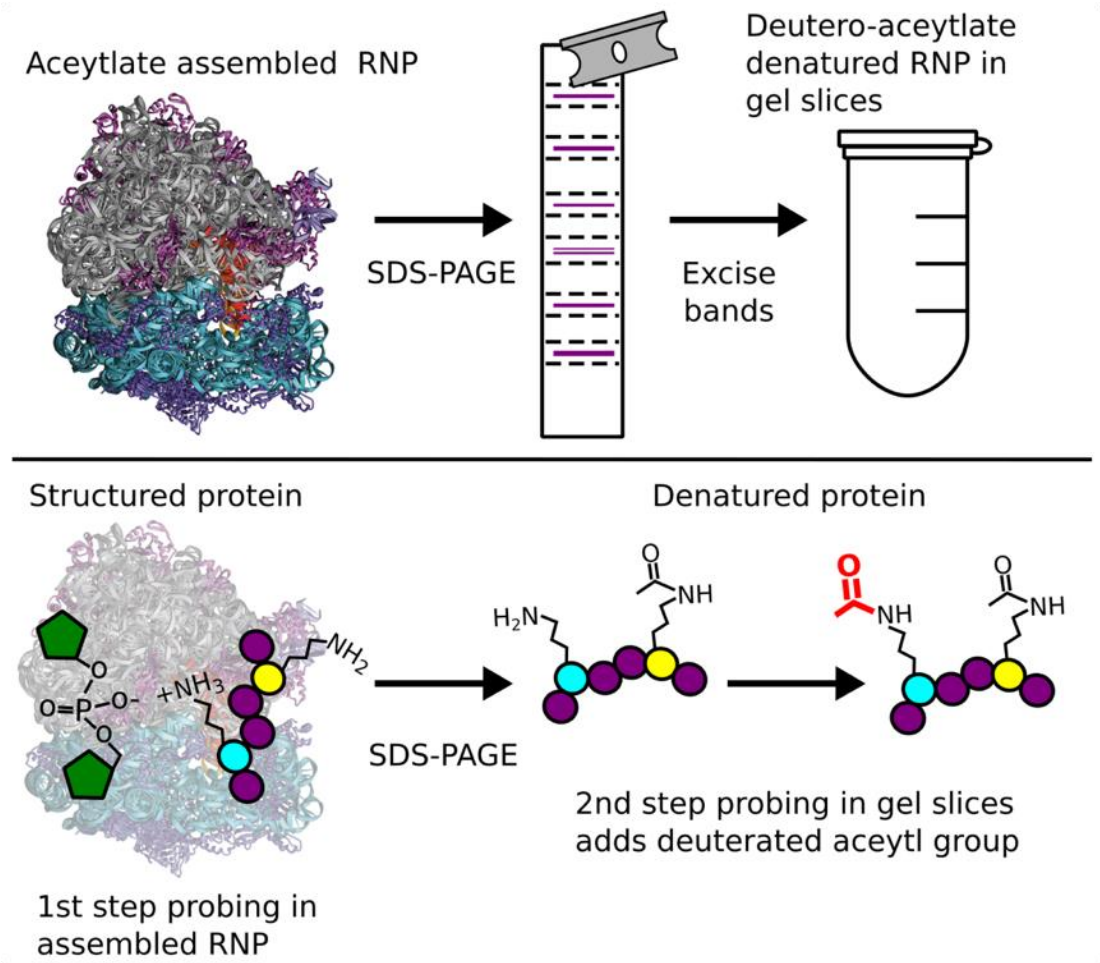


Figure 3: Left, various views of MS/MS lysines mapped to the 30S subunit of the 70S *E. coli* ribosomes (Noeske, Wasserman et al. 2015) colored by acetylation state. Cyan is deutero-acetyl-lysine, yellow is acetyl-lysine and orange is mixed. Graphs depicting the number of lysine residues vs distance to nearest atom, B-factor or SASA (solvent accessible surface area). Bars are colored based on the acetylation state of MS/MS lysines and further separated if the side chain density was poor. Values calculated via Chimera (Pettersen, Goddard et al. 2004)

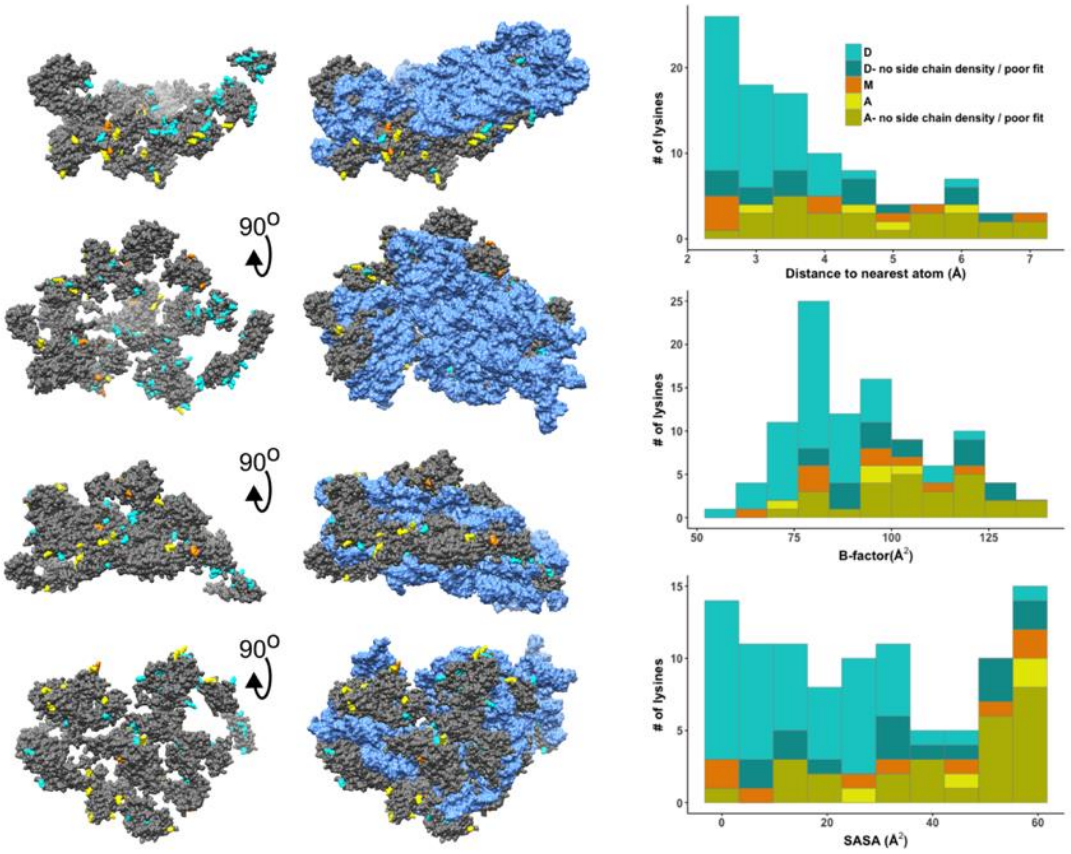
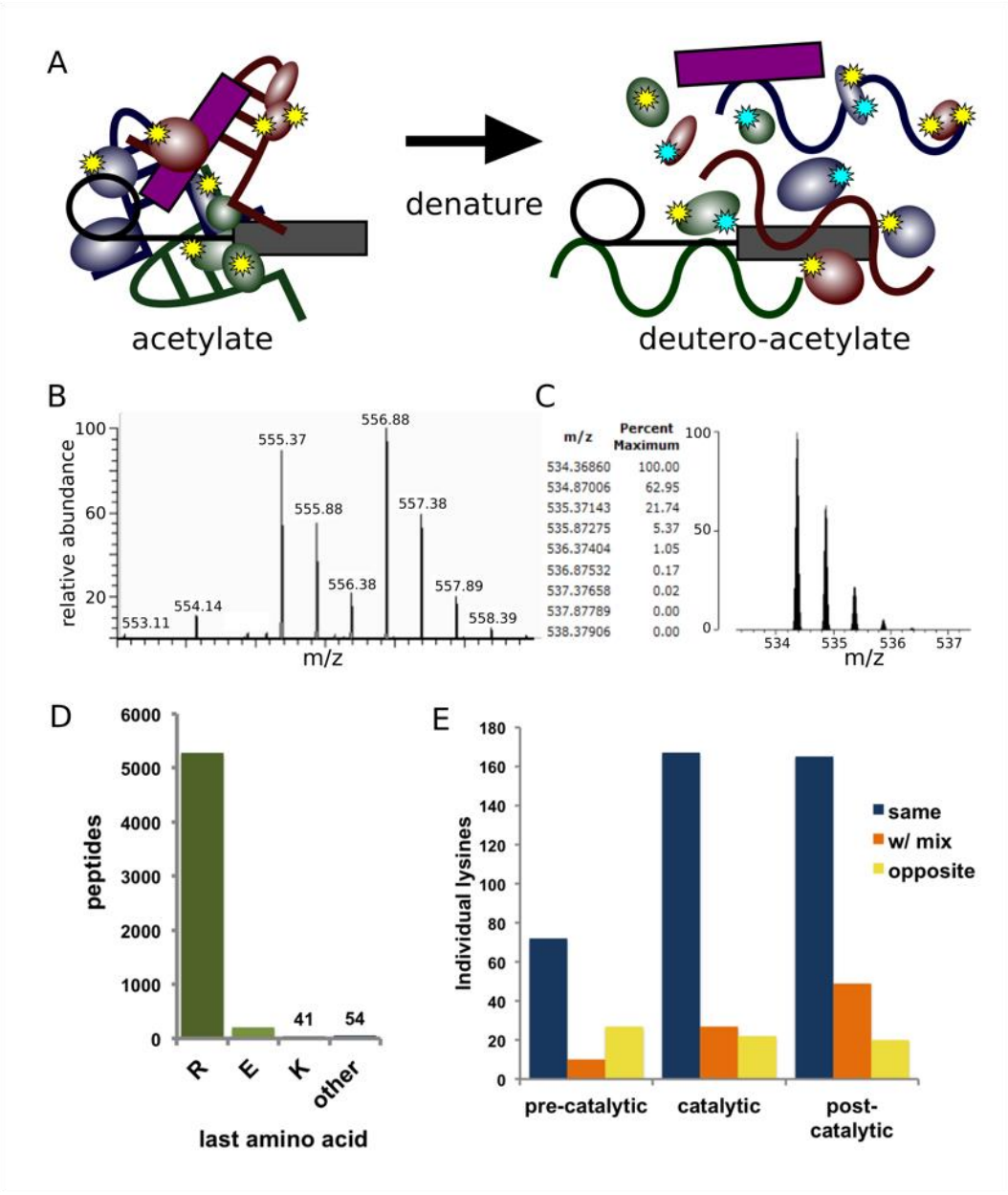


Figure 4: Two-step lysine probing of spliceosomes. A. Schematic of probing strategy: Yellow stars are acetyl lysine while cyan stars are deutero-acetyl lysine. B. Sample MS spectra containing two doubly charged peptides of m/z 555.37 and m/z 556.88 corresponding to the peptide VLIGVGKLLR from splicing factor 3B subunit 3. In the m/z 555.37 version, the lysine is modified with an acetyl group, whereas in the m/z 556.88 version, the lysine is modified with a deuteron-acetyl group, leading to a 3 Dalton mass difference. C. Theoretical modeling of the predicted isotope pattern for the lighter version using MS-Isotope (Chalkley et al. 2005) indicates a fourth isotope peak would contribute 5% peak intensity to the first isotope of the deuterated monoisotopic peak signal, and have an insignificant effect on overall quantification. D. Frequency of C-terminal amino acid identity for all unique MS/MS measured peptides from pre-catalytic and catalytic spliceosomes. E. Frequency of individual lysines identified with the same, mixed or opposite modification state between replicates samples of pre-catalytic and catalytic and post-catalytic spliceosomes.


A
denature
acetylate
deutero-acetylate
B
relative abundance
m/z
553.11
554.14
555.37
555.88
556.38
556.88
557.38
557.89
558.39
C
m/z
Percent Maximum
534.36860 100.00
534.87006 62.95
535.37143 21.74
535.87275 5.37
536.37404 1.05
536.87532 0.17
537.37658 0.02
537.87789 0.00
538.37906 0.00
D
peptides
last amino acid
R
E
K
other
41
54
E
Individual lysines
same
w/ mix
opposite
pre-catalytic
catalytic
post-catalytic

Chapter 3:

Prp8 positioning of U5 snRNA is linked to 5′ splice site recognition

3.1 Abstract

Prp8 is an essential protein that regulates spliceosome assembly and conformation during pre-mRNA splicing. Recent cryo-EM structures of the spliceosome model Prp8 as scaffolding the spliceosome's catalytic U snRNA components. Using a new amino acid probing strategy, we identified a dynamic region in human Prp8 that is positioned to stabilize the pre- mRNA in the spliceosome active site through interactions with U5 snRNA. Mutagenesis of the identified Prp8 residues in yeast indicates a role in 5′ splice site recognition. Genetic interactions with spliceosome proteins Isy1, which buttresses the intron branch point, and Snu114, a regulatory GTPase that directly contacts Prp8, further corroborate a role for the same Prp8 residues in substrate positioning and activation. Together the data suggest that adjustments in interactions between Prp8 and U5 snRNA help establish proper positioning of the pre-mRNA into the active site to enhance 5′ splice site fidelity.

3.2 Introduction

Pre-mRNA splicing is an essential step of eukaryotic gene expression, which results in the removal of intron sequences from pre-mRNA transcripts. This RNA

processing event is orchestrated by a dynamic ribonucleoprotein (RNP) complex named the spliceosome. During spliceosome assembly, specific landmark sequences within introns must be properly recognized in order to distinguish intron sequences from coding exon sequences. Following early assembly, the spliceosome structures its active site in a process referred to as activation. Activation allows 1st step chemistry of splicing to proceed, in which the spliceosome severs the 5′ splice site while simultaneously generating an intron lariat intermediate. This is followed by 2nd step chemistry, where the 3′ splice site is cleaved concomitant with exon ligation.

Recent cryo-electron microscopy (cryo-EM) structures of the spliceosome provide an unprecedented view of the organization of spliceosome components before and after active site formation (Galej, Wilkinson et al. 2016, Bertram, Agafonov et al. 2017, Bertram, Agafonov et al. 2017, Zhang, Yan et al. 2017). Differences between the stable conformations represented by the cryo-EM models indicate that a number of large and small structural rearrangements are required to generate the catalytic core of the spliceosome, which is primarily composed of three uridine-rich small nuclear RNAs (U snRNAs). Specifically, U2 snRNA positions the intron branch point, U6 snRNA holds the 5′ end of the intron and catalytic metals, and U5 snRNA interacts with the 5′ exon. These snRNAs are all supported by the large scaffold protein Prp8. Prp8 also interacts with several proteins that are known to regulate different steps of spliceosome assembly and catalysis, placing it at the nexus of structural rearrangements that allows the spliceosome to specifically position splice

site consensus sequences at intron boundaries with high fidelity (Umen and Guthrie 1996, Liu, Query et al. 2007, Mayerle and Guthrie 2016). Changes in Prp8 structure correlate with spliceosome progression, but additional biochemical analysis is required to delineate which Prp8 interactions are critical for snRNA active site formation and pre-mRNA splice site identification. Using chemical probing, we have identified a dynamic interaction between Prp8 and U5 snRNA, and functionally link the residues involved to specific splicing factors in yeast. Combined, our data serve as platform for additional mechanistic interpretation of spliceosome structural features.

3.3 Results

Probing the spliceosome

We applied the probing method to purified human pre-activated spliceosomes containing all five snRNPs (B-like complex) (Bertram, Agafonov et al. 2017) and human spliceosomes captured after the first step of catalysis (C/C* complex) (Jurica, Licklider et al. 2002) to assess structural differences before and after the RNA catalytic core is established (Figure 5A; Supplemental Figure 1). We interpreted lysines modified with an acetyl group as being in a solvent accessible position in native spliceosomes. In contrast, lysines modified with a deutero-acetyl group were presumably protected during the first round of chemical probing by molecular interactions. Although the spliceosome undergoes significant gain and loss of proteins along the transition from the pre-activation state (B complex) through activation (Bact complex) to the catalytic states (C and C* complexes), we were able to compare the modification state of 91

lysine residues from twelve proteins identified in both samples. Over 70% of these lysines maintained the same modification state, while 17% showed a distinct change. Modification of the remaining portion was variable among replicates.

Aiming to relate changes in lysine reactivity to structure, we examined the location of differentially modified lysine residues shared in the two available cryo-EM models of human spliceosomes in a pre-activation state (B complex (Bertram, Agafonov et al. 2017)) and a catalytic state remodeled for exon ligation (C* complex (Zhang, Yan et al. 2017)). Based on the presence of both Isy1, a C complex factor, and Prp22, a C* complex factor, in the MS/MS peptides, the catalytic complex that we purified likely represents a mixture of spliceosomes in both catalytic states. The lysine residues are from U5 snRNP proteins that remain with the spliceosome from pre-activation through catalysis: seven in Prp8 and one each in EFTUD2 and SNRNP40. Interestingly, they all were identified as solvent accessible in pre-activated spliceosomes, and as protected in catalytic spliceosomes. The change in reactivity suggests that side chains acquire an interaction partner at some point between the pre-activation state and splicing catalysis. The cryo-EM structures clearly bear out this prediction for Prp8 residues K68, which is located in interaction interfaces that Prp8 acquires with Bud31 during activation. Likewise, U5 40K protein residue K286 sits in an interaction interface with Prp8, which is present in the catalytic C* spliceosome, but not the pre-activation B complex.

The structural basis for the change in modification state for the five other differentially modified lysine side chains is not as clear-cut. Four conserved Prp8 lysines, K595, K597, K1300 and K1306, are notable in that they are located within 16 Å of each other in the region of Prp8 surrounding stem loop I of U5 snRNA (U5 SLI) as it interacts with the 5′ exon of pre-mRNA (Figure 5C-E; Supplemental Figure 2). Consistent with our interpretation of the probing data, the four lysine residues appear in contact with RNA or protein in the cryo-EM reconstruction of the human catalytic spliceosome (C* complex; (Zhang, Yan et al. 2017)), which includes side chains in the Prp8 model. Residues K595 and K597 are situated in a turn between two alpha helices in the N-terminal domain of Prp8, and in position to stabilize a highly negative groove created by a bend in U5 SLI towards the 5′ splice site (Figure 5D). In the C* spliceosome structure the terminal nitrogen of K595 is modeled as 2.9 and 3.1 Å, respectively, from C45 and A30 phosphate groups of U5 snRNA, which are only 4.0 Å apart. K597 is closer to the phosphate groups of U46 and A30 (3.3 and 3.9 Å, respectively (Figure 5E)). K1300 and K1306 are in the reverse transcriptase/endonuclease (RT/EN) domain of PRP8. K1306 sits between the phosphates of U5 snRNA residue C39 and a nucleotide in the 5′ exon. K1300 is surrounded by other amino acids in Prp8 with no clear access to solvent.

Based on the chemical probing results, we anticipated the same four Prp8 lysine residues to appear in a solvent accessible conformation in the cryo-EM structure of human pre- activated spliceosome (Bertram, Agafonov et al. 2017).

Unfortunately, the limited resolution of the cryo- EM density precluded modeling of side chains and their potential interactions. Compared with the catalytic structure, residues K1300 and K1306 are significantly shifted relative to U5 snRNA due to a conformational change between the RT/EN and N-terminal domains of Prp8. K595 and K597 remain in the same orientation relative to U5 snRNA and are positioned to stabilize a similar bend in U5 SLI (Supplemental Figure 3). The discordance between chemical probing results and the position of these lysine residues in the pre-activated spliceosome structure may indicate that the solvent accessible conformation is too flexible or too brief to be observed by cryo-EM. Likewise, our probing method cannot discern lysine side chains that are continuously solvent accessible from those that shift between solvent accessible and protected conformations over the time scale of the probing experiment. It is also worth noting that the two pre-mRNAs sequence used to assemble pre-activation complexes differed, which may result in different conformations of the complexes.

Differentially modified lysine residues genetically interact with multiple factors involved in 5' splice site recognition

If Prp8 residues K595 and K597 modulate U5 SLI structure, then neutralizing their positive charge will alter the conformation of U5 SLI and affect splicing. To test this hypothesis, we mutated the homologous residues (K670 and K672) to alanine in S. cerevisiae. In yeast spliceosomes cryo-EM structures, these residues interact with a similar bend in U5 SLI in a pre- activated B complex (Plaschka, Lin et al. 2017) and both

C and C* catalytic complexes (Galej, Wilkinson et al. 2016, Wan, Yan et al. 2016, Fica, Oubridge et al. 2017, Yan, Wan et al. 2017), suggesting that any functional role is evolutionarily conserved (Supplemental Figure 3). Yeast harboring the prp8-K670/672A allele are viable, and show no differences in growth relative to wild type PRP8 (Supplemental Figure 4A and B). However, prp8-K670/672A mutants do exhibit increased intron retention of two endogenous pre-mRNAs as measured by RT-qPCR, indicating the lysine to alanine mutations modestly perturb splicing (Figure 6A).

To assess whether the prp8-K670/672A allele impacts splicing fidelity and/or efficiency, we used an ACT1-CUP1 reporter system that directly links intron removal to the ability of yeast to grow in the presence of copper (Lesser and Guthrie 1993) (Figure 6B). With the reporter, the prp8-K670/672A strain grew identically to wild type PRP8 yeast, except when the reporter contained a guanine to adenosine substitution at the 5th position of the 5' splice site consensus (G5A). RT-PCR confirmed a small decrease in reporter splicing (Supplemental Figure 4C). By weakening interactions between the intron and U6 snRNA, the G5A mutation has been suggested to make the 5' splice site "slippery", thereby destabilizing the intron's position in the active site of the spliceosome (Parker and Guthrie 1985). Exacerbation of the G5A phenotype by prp8-K670/672A may be due to additional "slack" in the position of the upstream exon when U5 SLI conformation is disturbed. The effect appears to be specific to 5' splice site recognition because Prp8-K670/672A yeast have no phenotype when the branch point or 3' splice site is mutated in the ACT1-CUP1 reporter (Supplemental Figure 4).

To further explore the role of the Prp8 lysine residues in spliceosome function, we searched for genetic interactions between prp8-K670/672A and proteins involved in splicing activation and 1st step chemistry. Isy1 is a non-essential protein that is implicated in stabilizing the 1st step active site conformation of the spliceosome to help maintain fidelity during transition to 2nd step chemistry (Villa and Guthrie 2005). Consistent with this model, Isy1 contacts intron residues flanking the branch point in structures of S. cerevisiae C spliceosome caught immediately after 1st step chemistry (Galej, Wilkinson et al. 2016) (Supplemental Figure 5). Like prp8-K670/672A, isy1Δ yeast are viable with no growth phenotype, but show reduced splicing efficiency in vivo (Dix, Russell et al. 1999). Strikingly, when Isy1 is deleted, the prp8-K670/672A mutation becomes lethal (Figure 6C). This synthetic lethality suggests that the spliceosome as a whole cannot compensate for a combined loss of interactions supporting the 1st step active site conformation.

The U5 snRNP protein Snu114 is a GTPase that is proposed to regulate early snRNA rearrangements involved in spliceosome activation (Brenner and Guthrie 2005, Frazer, Lovell et al. 2009). Its influence on those rearrangements is likely indirect considering that Snu114 sits at a considerable distance from the RNA catalytic core in cryo-EM structures of the spliceosome, where it primarily contacts the N-terminal domain of Prp8 (Supplemental Figure 5). Our prp8- K670/672A allele exacerbates temperature sensitivity of snu114-40 (M842R) and snu114-50 (E910G, C928R) alleles (Brenner and Guthrie 2005) (Figure 6D), suggesting that Snu114's influence on splicing

is mediated, at least in part, by these Prp8 lysine residues. The same Snu114 alleles also show genetic interactions with Prp8 mutations linked to activation (Brenner and Guthrie 2005), as well as with deletions in U5 internal loop I (U5 ILI) at the base of U5 SLI (ΔC79-A81 and ΔA78-A81, ΔC11, ΔC112-G113) (Frazer, Lovell et al. 2009). Considering these data in conjunction with the position of the lysines relative to SLI in the cryo-EM structures, we suggest that the role of Snu114 in spliceosome activation is linked to Prp8's hold on U5 snRNA.

3.4 Discussion

Cryo-EM structures of human spliceosomes reveal the central role that Prp8 plays in supporting RNA interactions at the heart of the splicing machinery (Bertram, Agafonov et al. 2017, Bertram, Agafonov et al. 2017, Zhang, Yan et al. 2017). It augments U2 and U6 snRNA interactions with the intron and with each other as the branch point region is juxtaposed to the 5' splice site. Prp8 also cradles much of U5 snRNA, and encloses the 5' exon as it interacts with U5 SLI. U5 snRNA contacts with the exon are thought to assist in docking 5' splice site into the spliceosome catalytic core (Turner, Norman et al. 2004). Consistent with that model, mutations that alter the alignment of U5 SLI and the 5' exon inhibit the second step of splicing (O'Keefe and Newman 1998).

Differences between the structures of human pre-activated and catalytic spliceosomes imply large-scale rearrangements in U2 and U6 snRNA conformations and their interactions with the intron to create the active sites for 1st and 2nd step

chemistries (Bertram et al. 2017a; Zhang et al. 2017). Although more subtle, U5 snRNA interactions with the 5′ exon are also different before and after catalytic activation. Specifically, the last nucleotide of the exon has a different U5 snRNA base pairing partner (U41 instead of U40) in the human catalytic vs pre- activation complex, and the number of base interactions decreases from five to three. (Figure 7B). Chemical probing supports a transient loss and gain of interactions for both RNA and protein in this region during spliceosome activation. Our protein probing indicates Prp8 lysine residues in the area surrounding the loop of U5 SLI are accessible to an acetylating reagent, and then become protected by 1st step catalysis. Bases in the loop of U5 snRNA SLI also show increased protection to chemical probing following pre-activation (Anokhina, Bessonov et al. 2013). To explain these changes in reactivity, and the difference in U5 snRNA interactions with pre-mRNA visualized by cryo-EM, we hypothesize that a transient conformation of the pre-activation spliceosome exists. In our model, Prp8 partially disengages the U5 SLI, allowing it to reposition relative to the 5' exon, which aids in docking the 5' splice site into the 1st step active site.

Given this model it is not surprising that lysines mediating U5 SLI contacts genetically interact with factors previously implicated in spliceosome catalytic core formation, i.e. activation. Individually, the 5′ splice site G5A mutation impairs 5′ splice site positioning by removing a pre-mRNA/U6 contact, and Isy1 null yeast lose contacts between Isy1 and the intron branch. The Snu114 mutants exhibit temperature

sensitive growth, presumably from problems activating the spliceosome. In combination with our Prp8 lysine mutations these phenotypes are more severe, indicating that cooperation of these discrete molecular interactions supports the 1st step active site and is a piece of the larger mechanism used by the spliceosome to control accurate positioning of the pre-mRNA.

3.5 Materials and methods

In vitro assembly of pre-activated and catalytic spliceosomes

Catalytic and pre-activated pre-mRNA substrates are derivatives of the AdML transcript, and tagged with three MS2 sites in the intron or the 3′ end, respectively. T7 runoff transcription was used to generate G(5′)ppp(5′)G-capped radiolabeled pre-mRNA, which was gel purified and pre-incubated with a 50-fold excess of MS2-MBP fusion protein. In vitro splicing reactions contained 10 nM pre-mRNA, 80 mM potassium glutamate, 2 mM magnesium acetate, 2 mM ATP, 5 mM creatine phosphate, 0.05 mg/ml tRNA, and 40% HeLa cell (Biovest International) nuclear extract.

Pre-activated spliceosomes were generated in vitro by first forming the exon definition complex containing U1/U2/U4/U5/U6 snRNAs on a substrate that consists of an exon proceeded by a branch point and 3 splice site, and then adding 100 fold excess 5′ splice site oligo (AAGGUAAAGUAU) as previously described (Schneider et al. 2010), but with the following changes: 1) the exon definition complex was formed for 8 minutes prior to 5′ splice site oligo addition, and then incubated an additional 7

minutes, and 2) the 3′ exon of the pre-mRNA lacked a 5′ splice site. Catalytic spliceosomes arrested after 1st step chemistry and containing U2/U5/U6 snRNAs were accumulated in vitro using a pre-mRNA with a mutant GG 3′ splice site, and excess unspliced pre-mRNA from catalytic splicing reactions were depleted by DNA oligo directed RNase H digestion as previously described (Jurica, Licklider et al. 2002).

pre-mRNA sequences are as follows

catalytic pre-mRNA:

5′GGGAGACCGGCAGATCAGCTTGGCCGCGTCCATCTGGTCATCTAGGATCTGATATCATCG
ATGAATTCGAGCTCGGTACCCCGTTCGTCCTCACTCTCTTCCGCATCGCTGTCTGCGAGGGC
CAGCTGTTGGGGTGAGTACTCCCTCTCAAAAGCGGGCATGACTTCTGCCCTCGAGCGATAT
CCGTACACCATCAGGGTACGAGCTAGCCCATGGCGTACACCATCAGGGTACGACTAGTAGA
TCTCGTACACCATCAGGGTACGGAATTCTCTAGACTCGAGTTATTAACCCTCACTAAAGGCA
GTAGTCAAGGGTTTCCTTGAAGCTTTCGTACTAACCCTTTCCCTTTTTTTTCCTCTCTCTCTGG
CCATGGGTCGACGTTGAGGACAAACTCTTCGCGGTCTTTCCAGTACTCTTGGA3′

pre-activated pre-mRNA:

5′GGGCAAGGGTTTCCTTGAAGCTTTCGTACTAACCCTTTCCCTTTTTTTTCCTCTCTCTCTGG
CCATGGGTCGACGTTGAGGACAAACTCTTCGCGGTCTTTCCAGTACTCTTGGATCCGATATC
CGTACACCATCAGGGTACGAGCTAGCCCATGGCGTACACCATCAGGGTACGACTAGTAGAT
CTCGTACACCATCAGGGTACGGAATTCTCT3′

Spliceosome Purification

Catalytic spliceosomes were fractionated by size exclusion chromatography using Sephacryl S-400 resin (GE Healthcare) in SCB2H-N buffer (20mM HEPES pH 7.9, 150mM KCl, 5mM EDTA, 0.05% v/v NP-40). Pre-activated spliceosomes were fractionated in 10-30% v/v glycerol gradients in G-75 buffer (20mM HEPES pH 7.9, 75mM KCl and 1.5mM MgCl2) for 12 hours at 4oC in a SW-41 rotor at 25,000 RPM in a Beckman ultracentrifuge. Corresponding fractions of catalytic and pre-activated spliceosomes were pooled and bound to amylose resin via MS2:MBP. Pre-activated spliceosomes were washed with G75 buffer, and catalytic spliceosomes were washed with SCB2H (20mM HEPES pH 7.9, 150mM KCl, 5mM EDTA). Spliceosome complexes were then eluted with their respective wash buffers containing 10 mM maltose.

Chemical probing by two-step acetylation

See chapter 2 methods section

Protein digestion and mass spectrometric analysis

See chapter 2 methods section

Yeast Plasmids

Plasmids used in this study are indicated in table below. Plasmid pRS313-PRP8-WT, pRS314-PRP16 plasmids, ACT-CUP reporter plasmids, and pTB-SNU114 plasmids have been previously described (Lesser and Guthrie 1993, Query and Konarska 2004, Brenner and Guthrie 2005, Mayerle and Guthrie 2016). Plasmid pRS313-prp8-K670,672A was made from pRS313-PRP8-WT using Quikchange Mutagenesis (Agilent

Technologies). All plasmids have been deposited in the Addgene Vector Database, www.addgene.org.

PLASMIDS		
PRP8-RP	Prp16	Snu114
PRS313-*PRP8*	pRS314-*PRP16*	pTB2-pRS314-*SNU114*
PRS313-*PRP8-K670,672A*	pRS314-*prp16-1*	pTB99-pRS314-*snu114-30*
	pRS314-*prp16-101*	pTB100-pRS314-*snu114-40*
		pTB101-pRS314-*snu114-50*

Strains and Growth Assays

Strains used in this study are listed in table below. All strains were constructed using standard genetic methods. Double mutant strains were constructed via 5FOA shuffling, as described previously (Guthrie and Fink 2002). Relative growth was assessed by diluting an overnight culture to an OD600 of 0.1 and allowing it to grow to mid-log phase, an approximate OD600 of 0.5. Strains were then serially diluted 1:5 from a starting OD600 concentration of 0.1 and spotted onto rich media, minimal media lacking Ura, or minimal media containing 5FOA and then allowed to grow at 37˚C, 30˚C, 25˚C, and 17˚C before being photographed.

STRAIN	GENOTYPE
YMEG030	*MAT α prp8::LYS cup1::ura3-52 his leu lys ura ade yCP50/PRP8*
YMEG410	*MAT α prp8::LYS cup1::ura3-52 isy1::kanMx6 his leu ura ade yCP50/PRP8*
YTT120	*MAT α prp8::ADE prp16::LYS cup1::ura3-52 ade lys ura trp his yCP50/PRP8 pSE360-PRP16*
YTB108	*MAT α prp8::LYS snu114::KanMX6 his leu lys trp ura yCP50/PRP8 pRS316-SNU114*

ACT-CUP reporter Assays

Prp8 and ACT-CUP reporter plasmids (WT, G1A, U2A, G5A, C256A, BrC, BrG, U301G, A302G, G303-4C) were transformed into strain yMeg030 using standard methods (Guthrie and Fink 2002). Saturated overnight cultures grown in –Leu media were diluted to an approximate OD600 of 0.1 and allowed to grow again to an OD600 of approximately 0.5 before being diluted to an OD600 of 0.1 and spotted onto –Leu plates supplemented with 0 to 1.5 mM CuSO4. Plates were placed at 30°C and allowed to grow for 2-3 days and photographed.

RNA isolation and RT-qPCR

RNA isolation and detection were carried out a previously described (Mayerle and Guthrie 2016). Briefly, cultures were grown at 30°C overnight to saturation and then diluted to an OD600 of approximately 0.1. Diluted cultures were allowed to grow to an approximate OD600 of 0.5, pelleted, washed briefly with water, and then snap frozen at -80°C for further processing. Total cellular RNA was isolated using hot acid phenol followed by ethanol precipitation, and 20 µg total RNA was treated with 15 µL RNase free DNaseI (NEB) in a total volume of 250 µL for 1 hour at 37°C, and then re-extracted with phenol-chloroform. Superscript III Reverse Transcription System (Invitrogen) dN9 primers (Life Technologies) were used in all reverse transcription reactions. qPCR was performed using NEB Taq Polymerase and the gene-specific primers listed in table below, or previously described NSP1 and SEC17 primers. qPCR cycle conditions were 95°C 3 min, 39 cycles of 95°C 15 sec, 55°C 30 sec, 72°C 15 sec,

followed by 1 cycle at 72°C for 3 min. Three biological replicates using total RNA isolated from separate biological cultures were assayed in technical triplicate.

PRIMER	SEQUENCE
5'EXON INTRON F	5' tttttcttcccaagatcgaaa 3'
5'EXON INTRON R	5' ttaaatgggatggtgcaagc 3'
INTRON 3'EXON F	5' gcttcattctttttgttgctatattatatgtttagaggttg 3'
INTRON 3'EXON R	5' cattggcactcatgaccttc 3'
JUNCTION F	5' gaattaacaatggattctggttgc 3'
JUNCTION R	5' cattggcactcatgaccttc 3'
3'EXON F	5' ggtgtaacagcgacgacaaa 3'
3'EXON R	5' tttcccagagcagcatgatt 3'

3.6 Tables and Figures

Figure 5: Prp8 lysine residues surrounding U5 snRNA change accessibility between pre-activation and catalysis. A. Schematic of spliceosome assembly highlighting involvement of U snRNPs in the formation of the catalytic core during activation, and the two steps of splicing chemistry. B. Left, schematic of Prp8 domain structure and relative positions of four differentially modified lysine residues that spatially located near U5 snRNA in the spliceosome. RT/EN indicates the large reverse transcriptase / endonuclease domain. Right, secondary structure of U5 snRNA highlighting SLI interaction with the 5′ exon of pre-mRNA. Red indicates the region of U5 snRNA shown in panel C. C. Position of Prp8 (grey) side chains K595, K597, K1300, and K1306 (cyan) interactions with U5 SLI (blue) contacting the 5′ exon (gold) U5 SLI heteroatoms are colored red. D. View of Prp8 K595 and K597 in close proximity to the phosphate backbone of U5 SLI. Distances (Å) from lysine terminal nitrogens and non-bridging oxygens of phosphates of U5 snRNA C45, U46 and A30 are indicated by the dotted red lines. Panels C and D are derived from the cryo-EM model of the human catalytic (C*) spliceosome (Zhang, Yan et al. 2017).

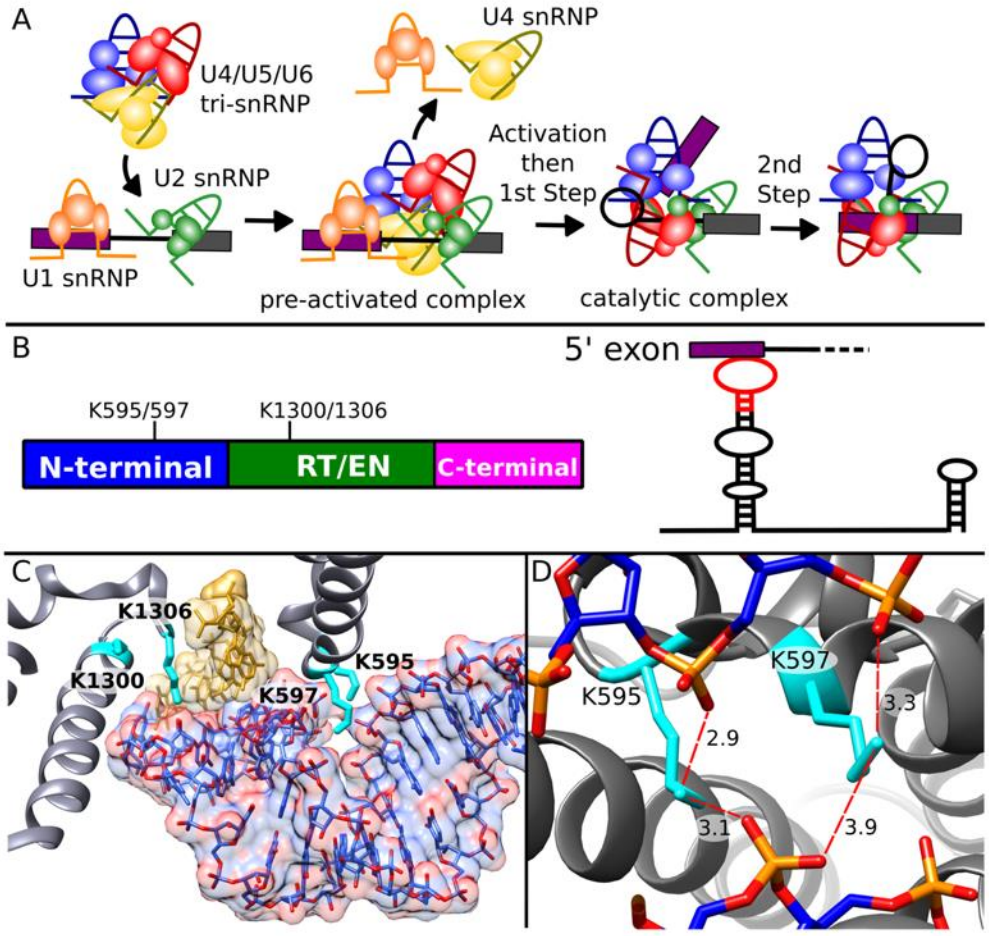


Figure 6: prp8-K670/672A genetically interacts with 5′ splice site and splicing factors involved in activation and 1st step chemistry. A. qRT-PCR analysis of relative NSP1 and SEC17 pre-mRNA transcript abundance in prp8-K670/672A (gray bars) and PRP8 (black bars). Primers that span the intron-exon junction were used to quantify pre-mRNA abundance, while primers that amplify a portion of the 3′ exon present in both the pre- and mature mRNA were used to quantify total RNA levels. Fraction unspliced is calculated as Fpre-mRNA / Ftotal. The average of three biological replicates is shown. Error bars represent standard error of the mean. B. The ACT1-CUP1 splicing reporter. The ACT1 intron is fused to the CUP1 coding sequence such that splicing efficiency can be directly assessed by relative yeast growth on copper-containing media. The 5′ and 3′ and branch site consensus sequences are shown. Numbers indicate position of specific nucleotide where the first nucleotide of the intron is defined as 1. Prp8-K670/672A and PRP8 yeast strains carrying either a WT or G5A ACT1-CUP1 reporter were serially diluted and spotted onto plates containing varying concentrations of Cu2+ and allowed to grow at 30°C for 2 days. C. Double mutant isy1Δ prp8Δ strains carrying either prp8-K670/672A or PRP8 on a HIS plasmid and PRP8 on a URA plasmid were serially diluted and spotted onto 5FOA or –ura plates and grown at 30°C. 5FOA prohibits growth of URA+ strains. prp8-K670/672A is lethal in an isy1Δ background. D. Double mutant snu114Δ prp8Δ complemented with prp8-K670/672A or PRP8 in combination with SNU114, or snu114-30, snu114-40, or snu114-50 C-terminal truncation alleles were serially diluted and spotted onto rich media and allowed to grow at 37°C and 25°C.

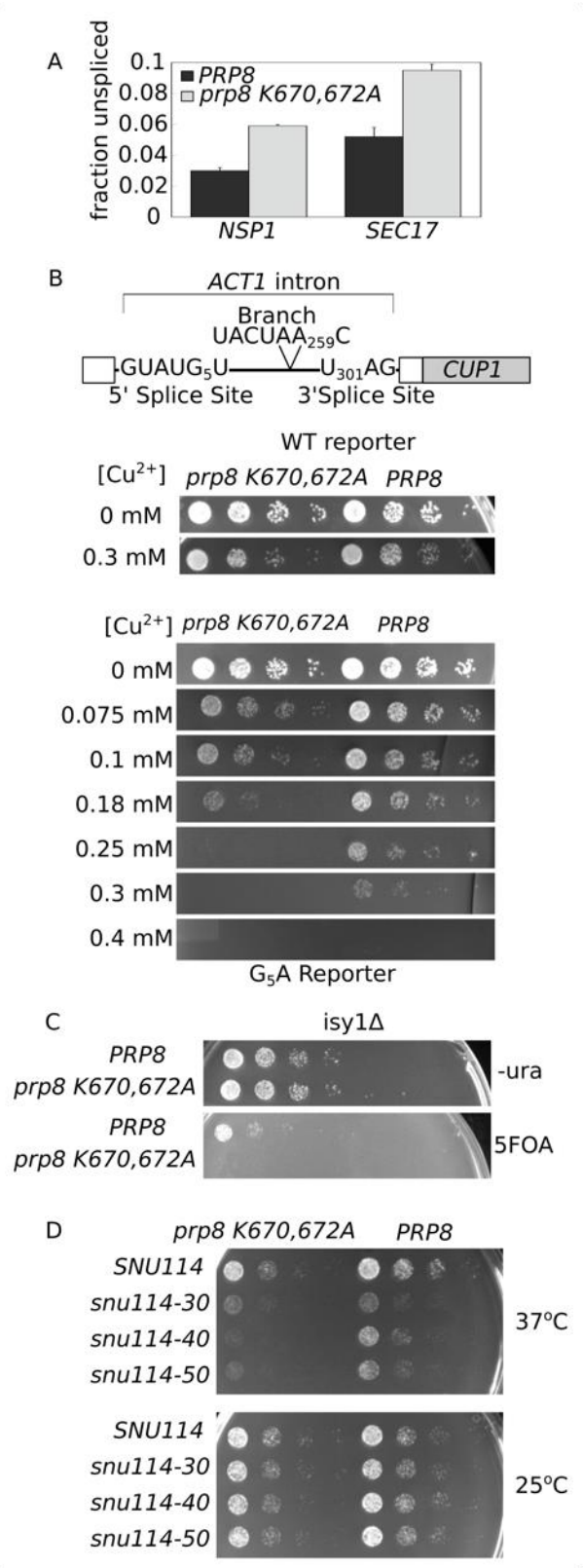
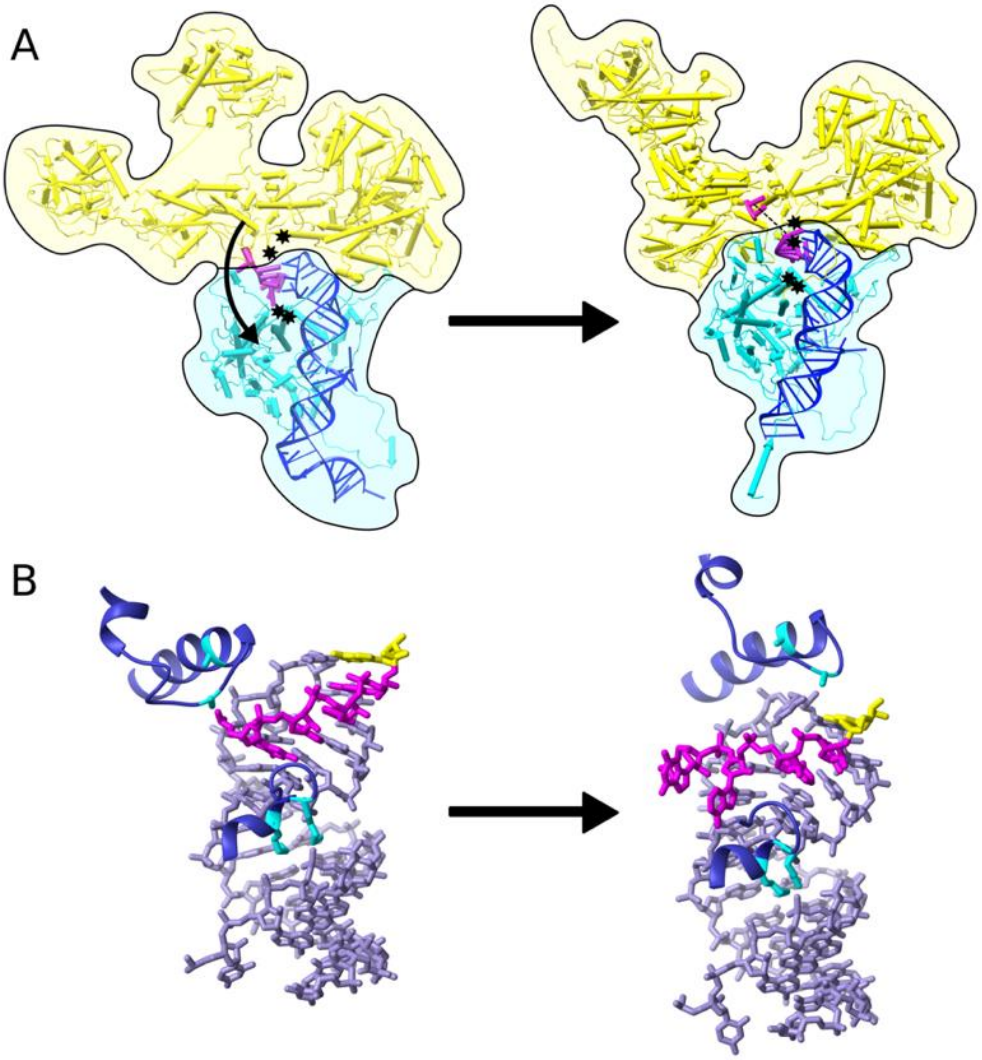
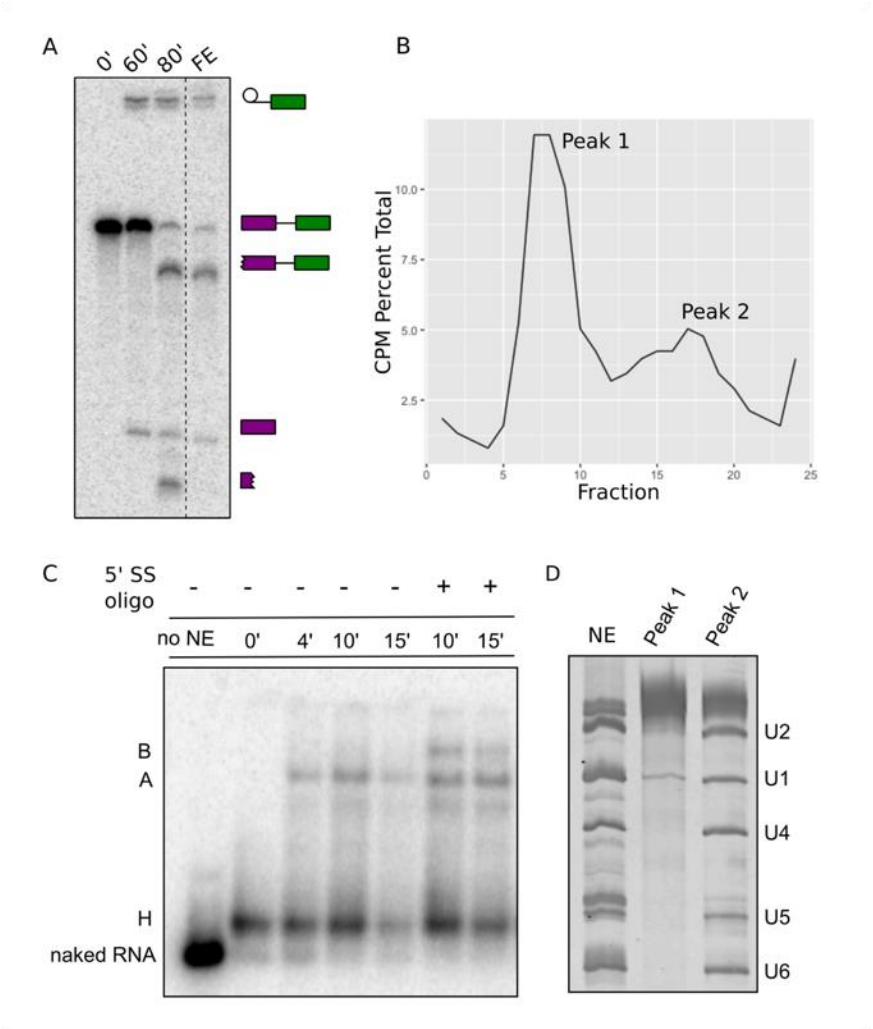

A
fraction unspliced
0.1
0.08
0.06
0.04
0.02
0
PRP8
prp8 K670,672A
NSP1
SEC17
B
ACT1 intron
Branch
UACUAA259C
GUAUG5U
U301AG
CUP1
5' Splice Site
3'Splice Site
WT reporter
[Cu2+]
prp8 K670,672A
PRP8
0 mM
0.3 mM
[Cu2+]
prp8 K670,672A
PRP8
0 mM
0.075 mM
0.1 mM
0.18 mM
0.25 mM
0.3 mM
0.4 mM
G5A Reporter
C
isy1Δ
PRP8
prp8 K670,672A
-ura
PRP8
prp8 K670,672A
5FOA
D
prp8 K670,672A
PRP8
SNU114
snu114-30
snu114-40
snu114-50
37°C
SNU114
snu114-30
snu114-40
snu114-50
25°C

Figure 7: Model of Prp8, U5 SLI, and 5′ exon interactions as the human spliceosome undergoes activation and 1st step chemistry. A. Global domain movements of Prp8 body (yellow) relative to N-terminal domain (cyan) in human pre-activation (Bertram, Agafonov et al. 2017) (left) and human C* (Zhang, Yan et al. 2017) (right) spliceosome models showing U5 snRNA (blue) and 5′ exon (magenta). The positions of differentially modified lysine residues K1300 and K1306 (upper) and K595 and K597 (lower) are indicated by black stars. B. Close-up view of the relative positions of U5 SLI (lavender), 5′ exon (magenta), and the same lysine side chains (cyan) of Prp8 (purple) in pre-activation (left) and catalytic (right) spliceosome conformations. The last nucleotide of the 5′ exon is highlighted in yellow.



3.7 Supplemental Figures

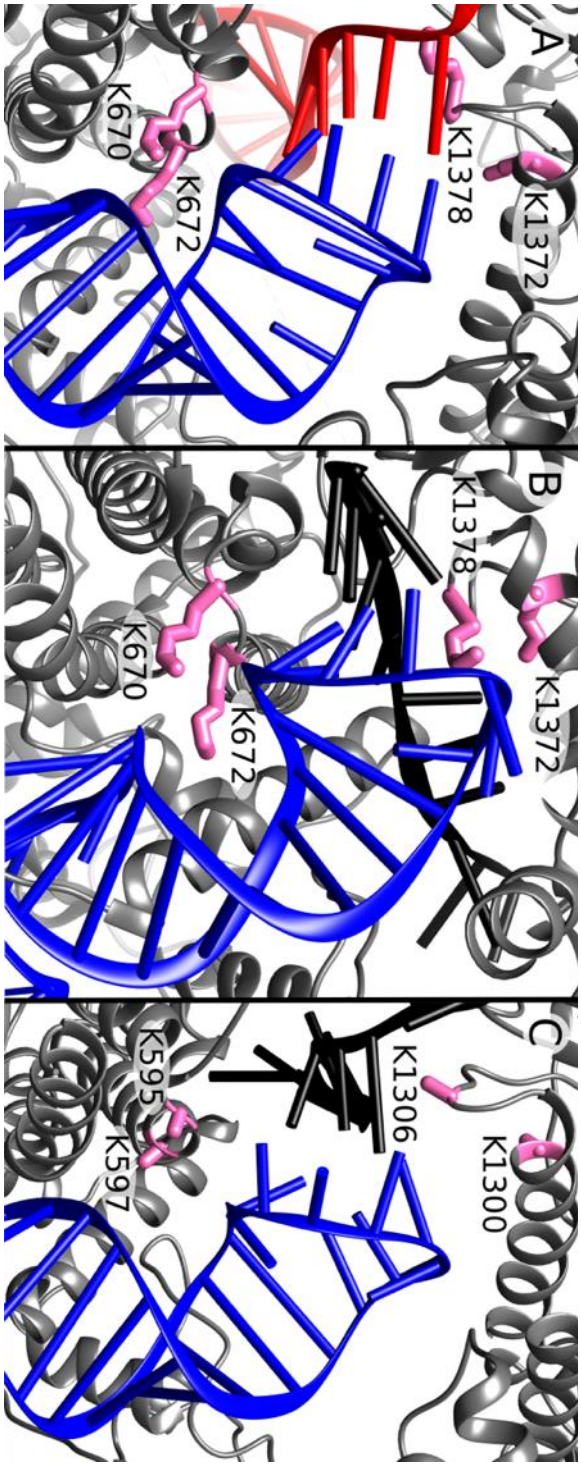
Supplemental Figure 1: Pre-activated and catalytic spliceosome purifications. A. Acrylamide Denaturing gel of radio labeled pre-mRNA during *in vitro* assembly (lanes 1-3) and the final elution (FE) after following size exclusion and amylose affinity chromatography. B. CPM percent total vs fraction of fractionated 10-30% glycerol gradient centrifugation of *in vitro* assembled pre-activated spliceosomes. Fraction 1 is the top of the gradient. Right peak corresponds to pre- activated spliceosomes. C. Native agarose gel of *in vitro* assembled pre-activated spliceosomes with and without the addition of the RNA oligo containing a 5′ splice site consensus. D. Denaturing acrylamide gel of glycerol gradient peaks as seen in B. stained for total RNA to visualize snRNA components.



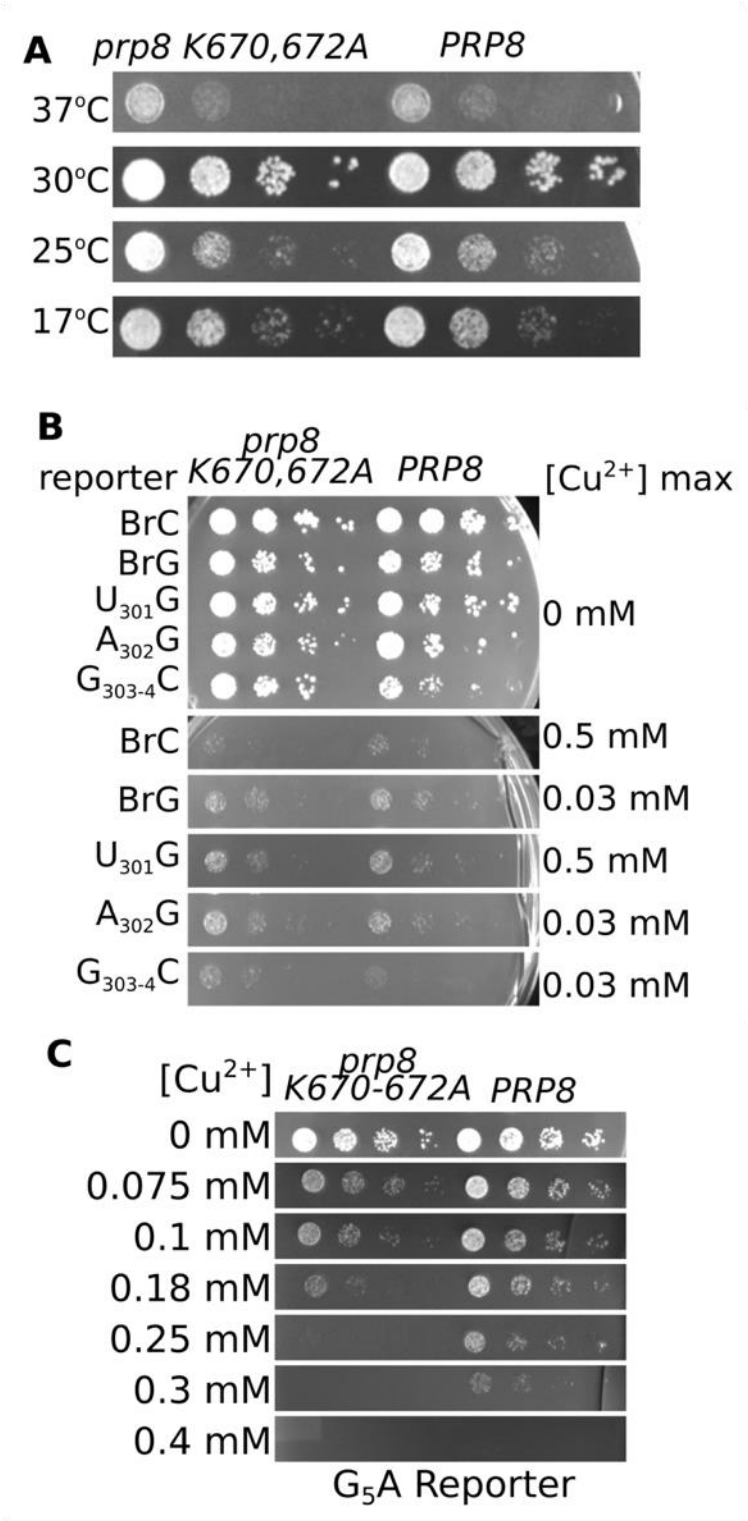
Supplemental Figure 2: Mass spectrometry results for Prp8 peptides containing acetylated lysine residues 595, 597, 1300 and 1306 for catalytic and pre-activation spliceosomes.

Spliceosome sample	PRPF8 peptide	Protein Modifications	m/z	z	ppm	E-value
Catalytic (1)	YKYKLMR	Acetyl:2H(3)@595; Acetyl:2H(3)@597	546.3182	2	9.4	0.014
Catalytic (2)	YKYKLMR	Acetyl:2H(3)@595; Acetyl:2H(3)@597	546.3153	2	4.1	0.013
Pre-activation	YKYKLMR	Acetyl@595; Acetyl@597	543.2938	2	-0.82	7.60E-04
Catalytic (1)	IKIGLNSKMPSR	Acetyl:2H(3)@1300; Acetyl:2H(3)@1306; Oxidation@1307	725.4275	2	7.8	3.1E-04
Catalytic (1)	IKIGLNSKMPSR	Acetyl:2H(3)@1300; Acetyl:2H(3)@1306	717.4291	2	6.5	1.3E-06
Catalytic (2)	IKIGLNSKMPSR	Acetyl:2H(3)@1300; Acetyl:2H(3)@1306	717.4249	2	0.67	4.4E-06
Pre-activation	IKIGLNSKMPSR	Acetyl@1300; Acetyl@1306	714.4069	2	1.8	2.3E-04
Pre-activation	IKIGLNSKMPSR	Acetyl@1300; Acetyl@1306; Oxidation@1307	722.403	2	-0.06	2.3E-08

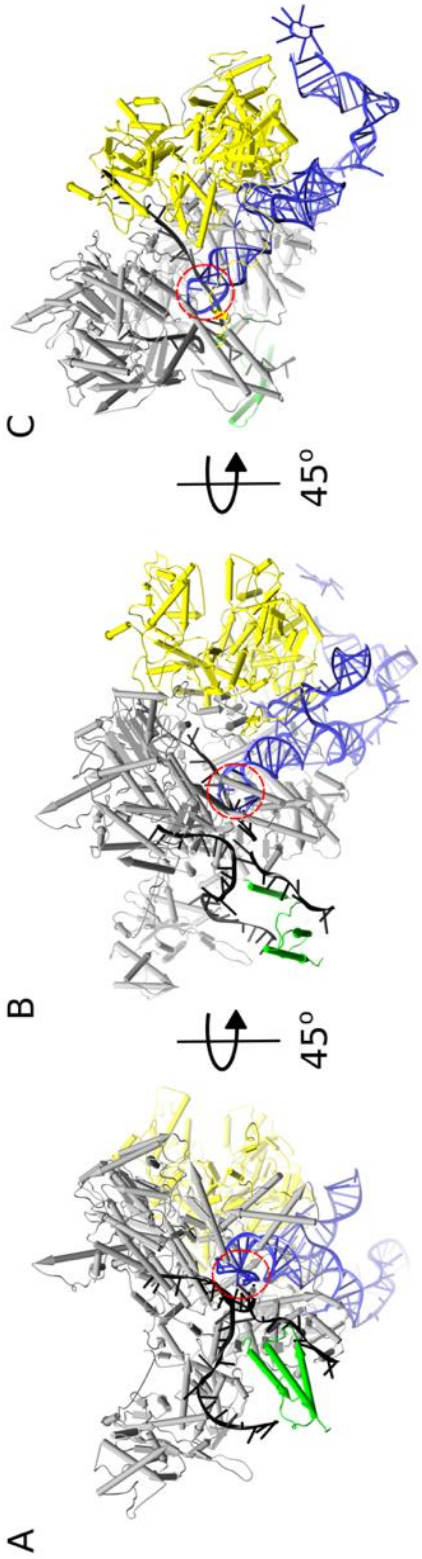
Supplemental Figure 3: Structural organization of S. cerevisiae Prp8 lysines 670, 672, 1372 and 1378 and human Prp8 lysines 595, 597, 1300 and 1306. A. S. cerevisiae pre-activated spliceosome assembled in vitro with 2 uM ATP (Plaschka, Lin et al. 2017). B. S. cerevisiae catalytic spliceosome assembled in vitro with 2 mM ATP (Galej, Wilkinson et al. 2016). C. Human pre-activated spliceosome (side chains un-modeled due to resolution limitations) assembled in vitro with 2 mM ATP (Bertram, Agafonov et al. 2017).


A
K1372
K1378
K672
K670
B
K1372
K1378
K672
K670
C
K1300
K1306
K595
K597

Supplemental Figure 4: A. Growth of serially diluted *Prp8-K670/672A* yeast relative to WT Prp8 yeast. B. ACT1-CUP1 splicing reporter assays of *Prp8-K670/672A* yeast containing various branch and 3′ splice site consensus mutations. BrC and BrG are mutations of the pre-mRNA branch point adenosine. Mutations at 301-304 are mutations to the pre-mRNA 3′ splice site consensus. C. RT-qPCR analysis of G5A reporter pre-mRNA and spliced mRNA was performed as described in the main text, with the following clarifications. Four different primer sets were used. The 5'Exon-Intron primer set (F: tttttcttcccaagatcgaaa; R: ttaaatgggatggtgcaagc) amplicon spans the 5'exon-intron boundary and reports on the level of pre-mRNA. The Intron-3'exon primer set (F: gcttcattctttttgttgctatattatatgtttagaggttg; R: cattggcactcatgaccttc) amplicon spans the intron-3'exon boundary and reports on the level of pre-mRNA and lariat intermediate. The junction primer set (F: gaattaacaatggattctggttgc; R: cattggcactcatgaccttc) amplicon spans the exon-exon mRNA junction, and reports on the abundance of fully spliced mRNA. The 3'exon primer set (F: ggtgtaacagcgacgacaaa; R: tttcccagagcagcatgatt) amplicon is contained within the 3' exon and reports on total reporter RNA abundance, as the 3'exon is present in pre-mRNA, lariat intermediate, and spliced mRNA product. Total RNA prepared from 3 separate biological replicates was analyzed in technical triplicate. The mean relative abundance of each amplicon from each primer set was normalized to that of wild type yeast expressing a wild type ACT-CUP reporter and plotted on a log scale. Error bars are SEM.


A
prp8 K670,672A
PRP8
37°C
30°C
25°C
17°C
B
prp8
reporter K670,672A PRP8
[Cu2+] max
BrC
BrG
U301G
A302G
G303-4C
0 mM
BrC
0.5 mM
BrG
0.03 mM
U301G
0.5 mM
A302G
0.03 mM
G303-4C
0.03 mM
C
[Cu2+]
prp8
K670-672A PRP8
0 mM
0.075 mM
0.1 mM
0.18 mM
0.25 mM
0.3 mM
0.4 mM
G5A Reporter

Supplemental Figure 5: Structural organization of Prp8, Isy1 and Snu114 in *S. cerevisiae* catalytic spliceosome (Galej, Wilkinson et al. 2016). A. Structure of Prp8 (grey), Isy1 (green), Snu114 (yellow) and U5 snRNA (Blue). B. 45 degree rotation of panel A. C. 45 degree rotation of panel B.


C
45°
B
45°
A

Chapter 4:

Stabilizing spliceosomes with lithium salt

4.1 Abstract

At the time of this research a limiting factor in spliceosome structure determination involved being able to biochemical purify stable homogenous spliceosomes. Homogeneity refering to composition and conformation. In order to purify a more homogeneous catalytic spliceosome I incorporated a lithium acetate wash to a standardized spliceosome purification protocol using size exclusion chromatography followed by affinity purification. I hypothesized that lithium ions would outcompete weak interactions within that spliceosome, thus dissociating weakly interacting proteins that bind spliceosomes sub-stoichiometrically. A more conformationally homogeneous spliceosome is a better candidate for cyro-electron microscopy because structure reconstruction requires averaging hundreds of thousands of individual spliceosome particles from electron microscopy micrographs. Lithium salt altered the sedimentation of spliceosomes as measured by glycerol gradients. Negative stain electron microcopy of salt washed spliceosomes suggested lithium exposure may have removed a subset of spliceosome from its periphery. Due to technical difficulties in sample preparation and instrumentation problems the lithium washed catalytic spliceosomes were never imaged using cryo-electron microscopy.

4.2 Introduction

Monovalent cations such as lithium have been previously used to specifically dissociate components from large RNPs (Noller, 1974). I intended to use lithium to outcompete weakly interacting components within purified human catalytic spliceosomes and or generate a salt stable spliceosome catalytic core. The catalytic spliceosome complex resulting from lithium acetate washing was expected to generate less noise during two dimensional averaging after negative stain EM. In addition to increasing homogeneity it would be beneficial to purify a spliceosome complex that interacts with EM grids better than current spliceosome preps. Multiple publications have shown the certain faces of spliceosome preferentially interaction with EM grids (Golas, Sander et al. 2010, Ilagan, Chalkley et al. 2013) and that crosslinked spliceosomes have extra protein RNA density as visualized in EM micrographs. It is not known what components dissociate from the spliceosome upon contact with EM grids but I hypothesize that dissociation of these components is increasing heterogeneity of the spliceosome particles on EM grids. I believe salt washes may circumvent this problem.

First to assess if lithium alter spliceosome structure I subjected purified catalytic spliceosomes to glycerol centrifugation containing and not containing 2 M lithium. After seeing a noticeable shift in catalytic spliceosome pre-mRNA intermediates I used negative stain EM to visualize any lost components suggested by the shift in sedimentation. This should be visualized by loss of protein/RNA density

between crosslinked catalytic spliceosomes and non-crosslinked catalytic spliceosomes in EM micrographs.

4.3 Results

Lithium salt exposure shifts spliceosomes in glycerol gradients

To determine if lithium salt alters spliceosome interactions and or homogeneity I first measured the sedimentation of spliceosomes through glycerol gradients containing 2 M or 0 M lithium acetate. Catalytic spliceosomes accumulated *in vitro* with HeLa nuclear extract was layered over glycerol gradients, spun then fractionated. RNA was purified from fractions and then denatured via PAGE and relative sedimentation of spliceosomes was ascertained by quantitating the position of the 3′ exon, a catalytic spliceosome intermediate. As seen in figure 1B there is a clear shift in sedimentation of the 5′ exon from fraction 11 to fraction 6 in 0 M vs 2 M lithium, respectively. Notably the lariat intermediate-3′ exon travels with the 5′ exon, co-migration of the two pre-mRNA intermediates associated with catalytic spliceosome suggests the spliceosome active site is still intact.

Lithium exposure strips components destabilized by EM grids

While a change in sedimentation suggests components are being released from the spliceosome when in contact with 2 M lithium it does not prove that the change in sedimentation is actually caused by components being released from catalytic spliceosomes. Spliceosomes and RNPs in general may sediment through lithium containing gradients at a slower rate due to changes in gradient density upon

the addition of lithium. The sedimentation change may also be caused by unfolding of the spliceosome. To more precisely determine if component release is causing the sedimentation change, I imaged purified catalytic spliceosomes washed and unwashed with 2 M lithium acetate and compared them to similarly purified spliceosomes that were lightly crosslinked with glutaraldehyde prior to imaging. Spliceosomes exhibit differing structural morphology via EM structure determination depending on if they are crosslinked or not. Crosslinked spliceosomes (Figure 2) have excess density surrounding the periphery of the particle, as visualized on negative stain EM micrographs. This extra density is not visualized in non-crosslinked spliceosomes, presumably due to the spliceosomes being partially destabilize upon contact with an EM grid. As depicted in Figure 2 it is difficult to determine if lithium washing disrupted interactions weakened by EM grid contact as the data set used for the crosslinked lithium spliceosomes was limited. Typically a few thousand particles are needed to generate 2D averaged particles and the crosslinked lithium spliceosome data set had less than 1000 particles. Visual analysis suggests the morphology between crosslinked and non-crosslinked spliceosomes differ but without comparable sized datasets a definitive answer is not known.

4.4 Discussion

It is unclear if lithium salt washes disrupt a set of interactions within the spliceosome that become unstable upon contact with EM grids. In order to determine if these purification conditions will lead to new productive spliceosome structures

negative stain single particle averages with a larger data set is recommended. Given that this work was primarily done prior to the existence of the spliceosome structural revolution the purification scheme could be further improved by incorporating aspects used by the research groups generating high resolution models (Nagai, Luhrmann and Shi). Such additions include the use of glycerol cushions prior to glycerol gradients centrifugation and a concentrating step using molecular weight cutoff columns. Previously, members of the lab tried to concentrate spliceosomes in a similar manner but they tended to stick to the filter within the column. After reading new purification methods it became clear in order to prevent spliceosomes from sticking to the column they can be concentrated in glycerol after gradient centrifugation then directly buffer exchanged out of glycerol.

Another unfortunate reality of the project is that freezing samples in vitreous ice for cryo-EM is incredibly challenging given our resources and aging electron microscope. After working on the project for two years it became apparent I could not keep up with the progress that other spliceosome structure labs were capable of. At this point I shifted my focus on the lysine probing project described in chapters 2 and 3 and decided to use current spliceosome structures to test models regarding mechanisms of spliceosome catalysis. Given where the project left off I believe there is a solid foundation of information should a future lab employee decide they want to continue.

4.5 Materials and methods

Spliceosome glycerol gradients

30 uls *in vitro* splicing reactions were layered over 600 ul 10%-30% linear glycerol step gradients (8 mM hepes pH 7.9, 40 mM KCl, 20 mM KGlu, 2 mM MgAc) and spun in a swinging bucket SW-41 Beckman rotor with plastic tube adaptors. Lithium gradients were made from a separate stock of gradient buffer containing 2 M lithium acetate. Gradients were spun at 38,000 RPM for 2.5 hours at 4^{o} centigrade then fractionated by hand into 15 fractions. Total RNA per fraction was phenol chloroform extracted then separated via 15% PAGE (30 watts for 2 hours). Radioactive pre-mRNA intermediates were visualized with a phosopho-imager (Molecular Dynamics) and quantified using ImageQuant (Molecular Dynamics)

In vitro spliceosome assembly

Catalytic spliceosomes were assembled *in vitro* as described in chapter 3

Spliceosome purification

Catalytic spliceosomes were purified as described in chapter 3 with the following changes. Prior to the final elution from amylose columns the affinity bound spliceosomes were washed with 1 ml of 2 M LiAc in SCB2H buffer (20mM HEPES pH 7.9, 150mM KCl, 5mM EDTA). Following lithium wash the column was washed with SCB2H to remove residual lithium then eluted with maltose in the same manner as described in chapter 3.

Negative stain electron microscopy grid prep, data collection and 2D particle averaging

Unfixed samples for electron microscopy were prepared by spotting 3 µL of purified complexes onto carbon-coated copper grids following plasma discharge, followed by staining with 2% (w/v) uranyl acetate solution. For fixed samples, purified complexes were incubated in 0.1% EM grade glutaraldehyde, 10 mM Hepes-NaOH pH 8, 150 mM NaCl, 0.02% NP40, 1 mM MgAc, 1 mM imidazole, 2 mM EGTA, 10 mM mercaptoethanol for 30 min at 4°C. The remaining glutaraldehyde in the crosslinking reaction was quenched with 90 mM glycine (pH 7.9) for an additional 30 min at 4°C. The samples were then immediately used to prepare grids in the same way as unfixed complexes.

Samples were analyzed using a JEOL 1230 microscope operating at 120 kV. Micrographs were exposed under low-dose conditions at 60,000× magnification on a 4 K × 4 K Gatan Ultrascan CCD camera, resulting in micrographs sampled at 4.9 Å per pixel. For each complex, approximately 2000 particles were selected manually from micrographs using the EMAN2 program Boxer (Galaz-Montoya, Flanagan et al. 2015). EMAN2 was additionally used for 2 × 2 pixel averaging, filtering, centering, and iteratively alignment and classification with default settings.

4.6 Figures

Figure 1: Gradient centrifugation of catalytic spliceosomes in 2M or 0M lithium. Top, image of separated radioactive pre-mRNA intermediates associated with the catalytic spliceosome (i.e. severed 5′ exon and lariat intermediate still attached to the downstream 3′ exon) purified from the gradient fractions. 0 on the X axis is the top of the gradient. RNA intermediated are diagramed to the right. Bottom, quantitation of the 5′ exon across 0M and 2M lithium gradients. Purple is 2M lithium and red is 0M lithium.

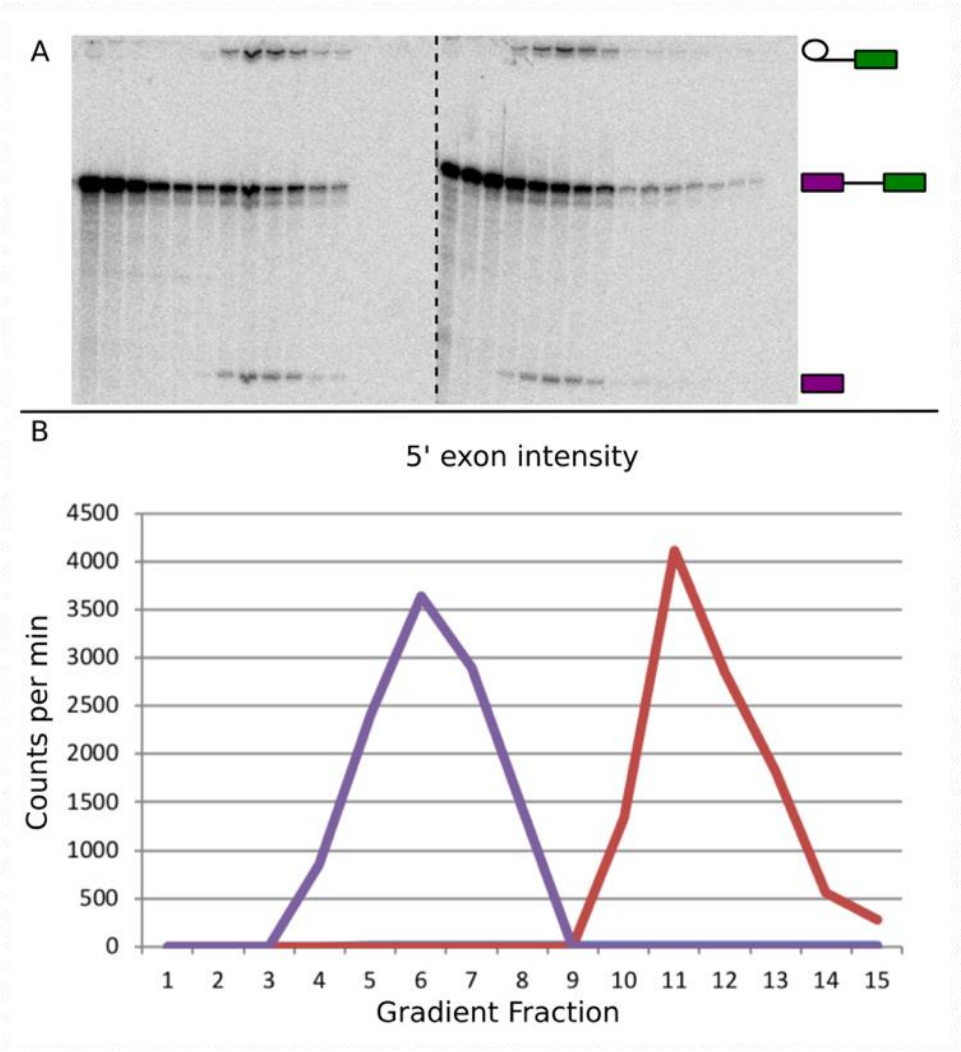
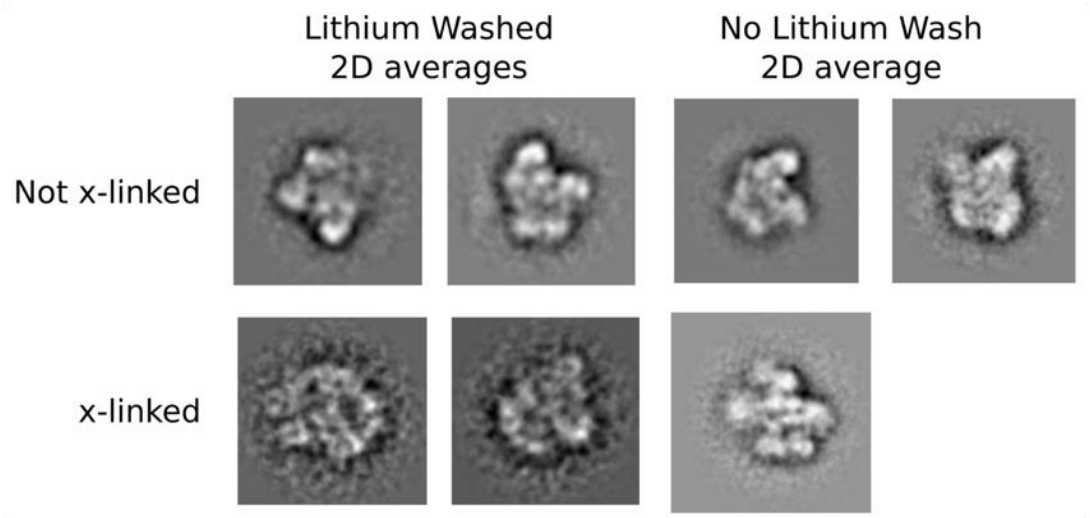


Figure 2: 2D averages of lithium washed catalytic spliceosomes to non-lithium washed catalytic spliceosome. The predominant 2 views are provided for each complex except crosslinked no lithium washed catalytic spliceosome. No lithium wash spliceosome averages not x-linked taken (Ilagan, Chalkley et al. 2013). X-linked No lithium was spliceosome average was taken from (Golas, Sander et al. 2010)



References

Anokhina, M., S. Bessonov, Z. Miao, E. Westhof, K. Hartmuth and R. Luhrmann (2013). "RNA structure analysis of human spliceosomes reveals a compact 3D arrangement of snRNAs at the catalytic core." EMBO J **32**(21): 2804-2818.

Berget, S. M., C. Moore and P. A. Sharp (1977). "Spliced segments at the 5' terminus of adenovirus 2 late mRNA." Proc Natl Acad Sci U S A **74**(8): 3171-3175.

Bertram, K., D. E. Agafonov, O. Dybkov, D. Haselbach, M. N. Leelaram, C. L. Will, H. Urlaub, B. Kastner, R. Luhrmann and H. Stark (2017). "Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation." Cell **170**(4): 701-713 e711.

Bertram, K., D. E. Agafonov, W. T. Liu, O. Dybkov, C. L. Will, K. Hartmuth, H. Urlaub, B. Kastner, H. Stark and R. Luhrmann (2017). "Cryo-EM structure of a human spliceosome activated for step 2 of splicing." Nature **542**(7641): 318-323.

Black, D. L., B. Chabot and J. A. Steitz (1985). "U2 as well as U1 small nuclear ribonucleoproteins are involved in premessenger RNA splicing." Cell **42**(3): 737-750.

Black, D. L. and A. L. Pinto (1989). "U5 small nuclear ribonucleoprotein: RNA structure analysis and ATP-dependent interaction with U4/U6." Mol Cell Biol **9**(8): 3350-3359.

Brenner, T. J. and C. Guthrie (2005). "Genetic analysis reveals a role for the C terminus of the Saccharomyces cerevisiae GTPase Snu114 during spliceosome activation." Genetics **170**(3): 1063-1080.

Cortes, J. J., E. J. Sontheimer, S. D. Seiwert and J. A. Steitz (1993). "Mutations in the conserved loop of human U5 snRNA generate use of novel cryptic 5' splice sites in vivo." EMBO J **12**(13): 5181-5189.

Dix, I., C. Russell, S. B. Yehuda, M. Kupiec and J. D. Beggs (1999). "The identification and characterization of a novel splicing protein, Isy1p, of Saccharomyces cerevisiae." RNA **5**(3): 360-368.

Fica, S. M., C. Oubridge, W. P. Galej, M. E. Wilkinson, X. C. Bai, A. J. Newman and K. Nagai (2017). "Structure of a spliceosome remodelled for exon ligation." Nature **542**(7641): 377-380.

Fica, S. M., N. Tuttle, T. Novak, N. S. Li, J. Lu, P. Koodathingal, Q. Dai, J. P. Staley and J. A. Piccirilli (2013). "RNA catalyses nuclear pre-mRNA splicing." Nature **503**(7475): 229-234.

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

Galaz-Montoya, J. G., J. Flanagan, M. F. Schmid and S. J. Ludtke (2015). "Single particle tomography in EMAN2." J Struct Biol **190**(3): 279-290.

Galej, W. P., T. H. Nguyen, A. J. Newman and K. Nagai (2014). "Structural studies of the spliceosome: zooming into the heart of the machine." Curr Opin Struct Biol **25**: 57-66.

Galej, W. P., C. Oubridge, A. J. Newman and K. Nagai (2013). "Crystal structure of Prp8 reveals active site cavity of the spliceosome." Nature **493**(7434): 638-643.

Galej, W. P., M. E. Wilkinson, S. M. Fica, C. Oubridge, A. J. Newman and K. Nagai (2016). "Cryo-EM structure of the spliceosome immediately after branching." Nature **537**(7619): 197-201.

Golas, M. M., B. Sander, S. Bessonov, M. Grote, E. Wolf, B. Kastner, H. Stark and R. Luhrmann (2010). "3D cryo-EM structure of an active step I spliceosome and localization of its catalytic core." Mol Cell **40**(6): 927-938.

Hashimoto, C. and J. A. Steitz (1984). "U4 and U6 RNAs coexist in a single small nuclear ribonucleoprotein particle." Nucleic Acids Res **12**(7): 3283-3293.

Ilagan, J. O., R. J. Chalkley, A. L. Burlingame and M. S. Jurica (2013). "Rearrangements within human spliceosomes captured after exon ligation." RNA **19**(3): 400-412.

Jurica, M. S., L. J. Licklider, S. R. Gygi, N. Grigorieff and M. J. Moore (2002). "Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis." RNA **8**(4): 426-439.

Jurica, M. S., D. Sousa, M. J. Moore and N. Grigorieff (2004). "Three-dimensional structure of C complex spliceosomes by electron microscopy." Nat Struct Mol Biol **11**(3): 265-269.

Koning, R. I., A. J. Koster and T. H. Sharp (2018). "Advances in cryo-electron tomography for biology and medicine." Ann Anat **217**: 82-96.

Kruger, K., P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling and T. R. Cech (1982). "Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena." Cell **31**(1): 147-157.

Lesser, C. F. and C. Guthrie (1993). "Mutational analysis of pre-mRNA splicing in Saccharomyces cerevisiae using a sensitive new reporter gene, CUP1." Genetics **133**(4): 851-863.

Liu, L., C. C. Query and M. M. Konarska (2007). "Opposing classes of prp8 alleles modulate the transition between the catalytic steps of pre-mRNA splicing." Nat Struct Mol Biol **14**(6): 519-526.

Liu, Z., I. Luyten, M. J. Bottomley, A. C. Messias, S. Houngninou-Molango, R. Sprangers, K. Zanier, A. Kramer and M. Sattler (2001). "Structural basis for recognition of the intron branch site RNA by splicing factor 1." Science **294**(5544): 1098-1102.

Madhani, H. D. and C. Guthrie (1992). "A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome." Cell **71**(5): 803-817.

Mayerle, M. and C. Guthrie (2016). "Prp8 retinitis pigmentosa mutants cause defects in the transition between the catalytic steps of splicing." RNA **22**(5): 793-809.

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

Nguyen, T. H., W. P. Galej, X. C. Bai, C. G. Savva, A. J. Newman, S. H. Scheres and K. Nagai (2015). "The architecture of the spliceosomal U4/U6.U5 tri-snRNP." Nature **523**(7558): 47-52.

Noeske, J., M. R. Wasserman, D. S. Terry, R. B. Altman, S. C. Blanchard and J. H. Cate (2015). "High-resolution structure of the Escherichia coli ribosome." Nat Struct Mol Biol **22**(4): 336-341.

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

Parker, R. and C. Guthrie (1985). "A point mutation in the conserved hexanucleotide at a yeast 5' splice junction uncouples recognition, cleavage, and ligation." Cell **41**(1): 107-118.

Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin (2004). "UCSF Chimera--a visualization system for exploratory research and analysis." J Comput Chem **25**(13): 1605-1612.

Plaschka, C., P. C. Lin and K. Nagai (2017). "Structure of a pre-catalytic spliceosome." Nature **546**(7660): 617-621.

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

Query, C. C. and M. M. Konarska (2004). "Suppression of multiple substrate mutations by spliceosomal prp8 alleles suggests functional correlations with ribosomal ambiguity mutants." Mol Cell **14**(3): 343-354.

Raghunathan, P. L. and C. Guthrie (1998). "RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2." Curr Biol **8**(15): 847-855.

Rauhut, R., P. Fabrizio, O. Dybkov, K. Hartmuth, V. Pena, A. Chari, V. Kumar, C. T. Lee, H. Urlaub, B. Kastner, H. Stark and R. Luhrmann (2016). "Molecular architecture of the Saccharomyces cerevisiae activated spliceosome." Science **353**(6306): 1399-1405.

RNA, C. f. M. B. o. from http://rna.ucsc.edu/rnacenter/ribosome.html.

Rogers, J. and R. Wall (1980). "A mechanism for RNA splicing." Proc Natl Acad Sci U S A **77**(4): 1877-1879.

Ruskin, B., A. R. Krainer, T. Maniatis and M. R. Green (1984). "Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro." Cell **38**(1): 317-331.

Sawa, H. and Y. Shimura (1992). "Association of U6 snRNA with the 5'-splice site region of pre-mRNA in the spliceosome." Genes Dev **6**(2): 244-254.

Tang, Q., S. Rodriguez-Santiago, J. Wang, J. Pu, A. Yuste, V. Gupta, A. Moldon, Y. Z. Xu and C. C. Query (2016). "SF3B1/Hsh155 HEAT motif mutations affect interaction with the spliceosomal ATPase Prp5, resulting in altered branch site selectivity in pre-mRNA splicing." Genes Dev **30**(24): 2710-2723.

Toor, N., K. S. Keating, S. D. Taylor and A. M. Pyle (2008). "Crystal structure of a self-spliced group II intron." Science **320**(5872): 77-82.

Turner, I. A., C. M. Norman, M. J. Churcher and A. J. Newman (2004). "Roles of the U5 snRNP in spliceosome dynamics and catalysis." Biochem Soc Trans **32**(Pt 6): 928-931.

Umen, J. G. and C. Guthrie (1996). "Mutagenesis of the yeast gene PRP8 reveals domains governing the specificity and fidelity of 3' splice site selection." Genetics **143**(2): 723-739.

Vidovic, I., S. Nottrott, K. Hartmuth, R. Luhrmann and R. Ficner (2000). "Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment." Mol Cell **6**(6): 1331-1342.

Villa, T. and C. Guthrie (2005). "The Isy1p component of the NineTeen complex interacts with the ATPase Prp16p to regulate the fidelity of pre-mRNA splicing." Genes Dev **19**(16): 1894-1904.

Wan, R., C. Yan, R. Bai, G. Huang and Y. Shi (2016). "Structure of a yeast catalytic step I spliceosome at 3.4 A resolution." Science **353**(6302): 895-904.

Wu, J. A. and J. L. Manley (1991). "Base pairing between U2 and U6 snRNAs is necessary for splicing of a mammalian pre-mRNA." Nature **352**(6338): 818-821.

Yan, C., R. Wan, R. Bai, G. Huang and Y. Shi (2017). "Structure of a yeast step II catalytically activated spliceosome." Science **355**(6321): 149-155.

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

Zhang, X., C. Yan, J. Hang, L. I. Finci, J. Lei and Y. Shi (2017). "An Atomic Structure of the Human Spliceosome." Cell **169**(5): 918-929 e914.