

Structural and Functional Investigation of Four Essential pre-mRNA Splicing Proteins

by

Erin Laine Garside

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Department of Biochemistry
University of Alberta

© Erin Laine Garside, 2019

Abstract

Exons, protein coding regions in eukaryotic genes, are interrupted by non-coding introns that must be removed from transcribed pre-mRNA prior to translation into protein. Introns are removed, and exons ligated, in a process called pre-mRNA splicing. Conserved sites within the intron, the 5' and 3' splice sites (SS) and the branch point sequence, define the boundaries of the intron. Intron removal and exon ligation at correct splice sites is essential for the translation of functional protein from the resulting mRNA.

Chemically, pre-mRNA splicing occurs via two transesterification reactions. In the first step, branching, the 2' hydroxyl of an adenosine within the branch point sequence attacks the 5' SS to generate a 2'-5' phosphodiester linkage and a free 5' exon with a 3' hydroxyl. In the second step, exon ligation, the 3' hydroxyl of the 5' exon attacks the 3' SS to generate ligated exons and a branched intron.

Pre-mRNA splicing is catalyzed by the spliceosome, a multi-megaDalton complex composed of five distinct small nuclear ribonucleoprotein particles (snRNPs) and other protein complexes, which assembles on every intron to be removed. Each snRNP features a U-rich small RNA stabilized within a ring of Sm proteins and a set of snRNP-specific proteins. U1 snRNP recognizes the 5' SS via base-pairing with its small nuclear RNA (snRNA). U2 snRNA forms an imperfect duplex with the branch point sequence. The un-base-paired branch A is excluded from the duplex and selected as the nucleophile for the first step of splicing. SF3b,

a subcomplex of U2 snRNP that dissociates from the spliceosome prior to the first step, stabilizes this duplex and keeps the branch A sequestered prior to the branching reaction. U4 snRNA in the U4/U6.U5 tri-snRNP base-pairs with the catalytic residues of U6 snRNA to prevent premature formation of the spliceosome active site. U5 snRNP, the largest snRNP, contains many regulatory proteins required for the progression of the spliceosome through the splicing cycle.

This thesis investigates the role of four essential snRNP proteins: U1C, Sap49, Snu13, and Prp8.

U1C is a component of the U1 snRNP. It features an N-terminal zinc finger and stabilizes the duplex between the 5' SS and U1 snRNA. Positively charged residues of U1C are thought to interact with the 5' SS during its initial recognition. We designed constructs to cross-link U1C to the 5' SS to study the nature of this interaction.

Sap49, part of the SF3b complex of U2 snRNP, is shown to cross-link to the intron upstream of the branch A. Mutations in Sap49 that cause haplo-insufficiency are associated with the acrofacial dysostosis Nager syndrome. A recently described I84R mutation has been linked to a mild form of Nager syndrome. We failed to express recombinant I84R Sap49, and the corresponding mutation is not viable in *S. cerevisiae* cells. We propose that this mutant is not stable, leading to degradation and haplo-insufficiency resulting in Nager syndrome.

Snu13 is a small, highly conserved protein found in both the U4 snRNP and the non-spliceosomal box C/D snoRNP. It is maintained in the highly reduced

spliceosome of the red alga *C. merolae*. Our work shows this Snu13 is very similar to other Snu13s, both in its structure and its binding to U4 snRNA. These results suggest that results from studies of the *C. merolae* are relevant to the human spliceosome.

Prp8 is the largest, most highly conserved spliceosomal protein. It features an RNase H-like (RH) domain that regulates progression through the splicing cycle. Sequence analysis of the *C. merolae* Prp8 predicts the absence of a highly conserved 17 amino acid insertion in the RH domain, which we confirmed by X-ray crystallography. Other proteins observed interacting with this insertion in structures of the spliceosome throughout the splicing cycle are also predicted to be absent. Organisms with few introns, including *C. merolae*, and fewer spliceosome components also lack the insertion.

RH metal binding in the human RH domain has been visualized crystallographically. Analysis of this metal binding suggests it is not required for splicing; however, a positive charge, either a metal or a nearby arginine, may play a role.

Taken together, these findings increase our understanding of pre-mRNA splicing and the spliceosome.

Preface

Parts of this thesis have been performed in collaboration with other research groups and published previously.

Work in Chapter 3 was performed in collaboration with Daniel Pomeranz Krummel currently at the Winship Cancer Institute, Emory School of Medicine in Atlanta, Georgia, and published as

Reversibly constraining spliceosome-substrate complexes by engineering disulfide crosslinks
by P McCarthy, E Garside, Y Meschede-Krasa, A MacMillan and D Pomeranz Krummel. *Methods*, 125, 25-35. 2017.

I designed and created the U1C mutants, the modified hairpins and the modified pre-mRNA splicing substrate. I also depleted the extracts of U1 snRNP and performed the splicing experiments. P. McCarthy reconstituted the U1 snRNP with both wild type and mutant U1C and sent me the samples.

Work in Chapter 4 was performed in collaboration with Oana Caluseriu's lab in the Departments of Medical Genetics and Pediatrics at the University of Alberta. Dr. Caluseriu diagnosed the individual with Nager syndrome, and discovered the I84R mutation of Sap49. Karim Atta, of the MacMillan lab, cloned the set of human Sap49 mutants. Ayat Omar, also of the MacMillan lab, cloned the *S. cerevisiae* mutants and generated the mutant strains. I performed the *S. cerevisiae* stress tests. I also cloned the Sap145 fragments, expressed the

recombinant human Sap49 and 145 proteins, and performed the gel shift experiments.

Work in Chapter 5 was performed in collaboration with Stephen Rader in the Department of Chemistry, University of Northern British Columbia, and published as

Conserved structure of Snu13 from the highly reduced spliceosome of
Cyanidioschyzon merolae
by CS Black, EL Garside, AM MacMillan and SD Rader.
Protein Sci, 25 (4), 911-916. 2016.

C. Black cloned, expressed and purified Snu13. I then crystallized the Snu13 and solved its structure. C. Black performed the fluorescence polarization experiments and assembled the K_D comparison chart.

Work in Chapter 6 was performed in collaboration with TA Whelan and NM Fast of the Biodiversity Research Center and Department of Botany, University of British Columbia. I cloned, expressed, purified, crystallized and solved the structure of the *C. merolae* Prp8 RH domain. TA Whelan selected taxa for sequence comparison. I performed sequence alignment and analysis. TA Whelan and I assembled the table comparing genome size and intron number to conservation of the β -hairpin insert. TA Whelan performed the BLAST searches for Aar2 and Snu66.

The hT1783A/S hR1865A double mutant structures in Chapter 7 were previously published in Garima Mehta's thesis (2015). G. Mehta cloned,

expressed, purified and crystallized the Prp8 RH domain mutants. I solved the structure. G. Mehta created the mutant *S. cerevisiae* strains. I performed the growth tests and the primer extensions. Reanalysis of the data lead to conclusions in this thesis distinct from those in G. Mehta's thesis.

Unless otherwise noted, I created all images using Pymol and PowerPoint. I also assembled all the tables. The following figures are re-printed with permission.

Figure 3-2A was originally printed in Kondo et al., 2015 (Figure 1-figure supplement 2) and is licensed under CC BY 4.0.

Figure 3-14B was originally printed in McCarthy et al., 2017 (Figure 7.B). Permission is not required for use.

Figure 4-7 was originally printed in Lansinger & Rayan, 2015 (Figures 1.B and 3.A). Reprinted from The Journal of Hand Surgery, 40 (4), by Y Lansinger and G Rayan, Nager Syndrome, pp 851-854, copyright 2015, with permission from Elsevier.

Figure 5-1C was originally printed in Vidovic et al., 2000 (Figure 1). Reprinted from Molecular Cell, 6 (6), by I Vidovic, S Nottrott, K Hartmuth, R Luhrmann, and

R Ficner, Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment, pp 1331-1342, copyright 2000, with permission from Elsevier.

Figure 5-10A was originally printed in Stark et al., 2015 (Figures 3.B and Supplemental Figure 2.A). Permission is not required for use.

Figure 5-10B was originally published in Black et al., 2016 (Figure 3). Reprinted with permission from John Wiley and Sons.

Figure 6-1C was originally published in Stark et al., 2015 (Figure 1). Permission is not required for use.

Table of Contents

Abstract	ii
Preface	v
Table of Contents	ix
List of Figures	xiii
List of Tables	xvi
Chapter 1: An Introduction to pre-mRNA Splicing	1
1-1. Pre-mRNA Splicing	2
1-2. Alternative Splicing	3
1-3. snRNPs	6
1-4. Additional Splicing Proteins	8
1-5. The Splicing Cycle	12
1-6. Splicing Fidelity	16
1-7. Splicing and Human Diseases	18
1-7a. Mutations in Splice Sites/RNA	19
1-7b. Mutations in Spliceosomal Proteins	21
1-7c. Splicing and Cancer	22
1-8. Splicing Studies	24
1-9. Summary	26
Chapter 2: A Structural Investigation of the Spliceosome	31
2-1. Spliceosomal Structures	32
2-2. Prp8	32
2.3. Tri-snRNP	34
2-4. Spliceosome Structures	37
2-5. Yeast Structures	44
2-5a. A Complex	45
2-5b. B Complex	46
2-5c. B ^{act} Complex	51
2-5d. C Complex	54
2-5e. C* Complex	57
2-5f. P Complex	58
2-5g. Intron Lariat Spliceosome Complex	61
2-6. Human Structures	63
2-6a. B Complex	63
2-6b. C* Complex	75
2-7. Group II Introns	77
2-8. Two-Metal Mechanism	81
2-9. Group II Introns and the Spliceosome	83
2-10. Summary	86
Chapter 3: 5' Splice Site Recognition by U1C and the U1 snRNP	87
3-1. U1 snRNP	88

3-2. U1 snRNP Structure	89
3-3. U1C	94
3-4. 5' Splice Site Recognition	98
3-5. U1 snRNP in the A Complex	100
Results	
3-6. U1C does not Bind RNA	102
3-7. Modified RNA Hairpin	104
3-8. Modified pre-mRNA	107
3-9. Splicing in the Absence of U1 snRNP	110
3-10. Spliceosome Reconstitution	111
3-11. Discussion	111
3-12. Materials and Methods	114
3-12a. Protein Expression	114
3-12b. RNA Transcription and purification	115
3-12c. Radiolabelling	116
3-12d. Electrophoretic Mobility Shift Assays (EMSAs)	116
3-12e. RNA Synthesis	117
3-12f. Cross-linking and Reduction Time Courses	118
3-12g. RNA Ligation	119
3-12h. Splicing	120
3-12i. U1 snRNP Depletion	120
3-12j. Binding Reconstituted U1 snRNP to pre-mRNA	121
Chapter 4: The Interaction of Sap49 with Sap145 and RNA	122
4-1. SF3b	123
4-2. Sap49 and Sap145	125
4-3. Structures of Sap49, Hsh49, and Cus1	129
4-4. Nager Syndrome	133
4-5. Sap49 Depletion in Frogs	135
Results	
4-6. Yeast Mutants	136
4-7. Protein Expression	138
4-8. Binding of Sap49 to RNA	139
4-9. Discussion	141
4-10 Materials and Methods	145
4-10a. <i>S. cerevisiae</i> Mutants	145
4-10b. <i>S. cerevisiae</i> Growth Tests	145
4-10c. Expression of MBP-tagged Sap49	146
4-10d. Co-expression of His-tagged Sap49 and the Sap145 fragment	147
4-10e. Hot RNA Transcription	147
4-10f. Electrophoretic Mobility Shift Assays (EMSAs)	148

Chapter 5: The U4 snRNP Protein Snu13 from <i>C. merolae</i>	149
5-1. Snu13	150
5-2. Snu13 and U4 snRNA	152
5-3. Snu13 and Prp31	154
5-4. Snu13 in the Tri-snRNP and B Complex	155
Results	
5-5. <i>C. merolae</i> Snu13 Structure	160
5-6. <i>C. merolae</i> Snu13 and RNA	161
5-7. Discussion	164
5-8. Materials and Methods	166
5-8a. Cloning, expression, and protein purification	166
5-8b. Crystallization and Structure Determination	166
5-8c. Fluorescence Polarization	166
Chapter 6: The RH Domain of <i>C. merolae</i> Prp8	167
6-1. Splicing in <i>C. merolae</i>	168
6-2. The RNase H Domain of Prp8	174
6-3. The RH Domain of Prp8 in the Spliceosome	176
Results	
6-4. The RH Domain of <i>C. merolae</i> Prp8	183
6-5. Conservation of the β -hairpin in Other Species	187
6-6. Discussion	192
6-7. Materials and Methods	195
6-7a. Identification, Cloning and Expression of the Prp8 RH Domain from <i>C. merolae</i>	195
6-7b. Crystallization	195
6-7c. Data Collection and Processing	196
6-7d. Model Building and Refinement	196
Chapter 7: Metal Binding by the Human Prp8 RH Domain	197
7-1. The Active Sites of the Spliceosome	198
7-2. Prp8 and the Spliceosomal Equilibrium	201
Results	
7-3. Structures of the hT1783A/S hR1865A Double Mutants	204
7-4. Phenotypes of yT1855A/S	206
7-5. Structures of hQ1894E/S hR1865A Double Mutants	210
7-6. Discussion	212
7-7. Materials and Methods	221
7-7a. Protein Expression and Purification	221
7-7b. Crystallization	221
7-7c. Data Collection and Processing	222
7-7d. Model Building and Refinement	222
7-7e. Creation of Mutant <i>S. cerevisiae</i> Strains	222
7-7f. Copper Resistant Strains	223

7-7g. Growth Assays	223
7-7h. Primer Extension	224
Chapter 8: Conclusions and Future Directions	226
8-1. Conclusions and Future Directions	227
References	234

List of Figures

Chapter 1	
1-1. Pre-mRNA splicing	4
1-2. Alternative splicing	5
1-3. snRNP composition	9
1-4. Spliceosome assembly	12
Chapter 2	
2-1. A structural overview of Prp8	33
2-2. Structures of the <i>S. cerevisiae</i> and human tri-snRNP	35
2-3. Details of the <i>S. cerevisiae</i> tri-snRNP	36
2-4. <i>S. cerevisiae</i> A complex	45
2-5. <i>S. cerevisiae</i> B complex	47
2-6. Details of the <i>S. cerevisiae</i> B complex	48
2-7. <i>S. cerevisiae</i> B ^{act} complex	50
2-8. RNA core of the <i>S. cerevisiae</i> B ^{act} spliceosome	52
2-9. Protein components of the <i>S. cerevisiae</i> B ^{act} spliceosome	53
2-10. <i>S. cerevisiae</i> C complex	54
2-11. Details of the <i>S. cerevisiae</i> C complex	56
2-12. <i>S. cerevisiae</i> C* complex	57
2-13. Details of the <i>S. cerevisiae</i> C* spliceosome	58
2-14. <i>S. cerevisiae</i> P complex	59
2-15. Details of the <i>S. cerevisiae</i> P complex	60
2-16. <i>S. pombe</i> ILS complex	61
2-17. Details of the <i>S. pombe</i> ILS complex	63
2-18. Human B complex	73
2-19. Details of the human B complex	74
2-20. Human C* complex	75
2-21. Details of the human C* complex	76
2-22. Group II intron	78
2-23. The spliceosome and group II intron active sites have conserved metal binding sites	81
Chapter 3	
3-1. Early stages of spliceosome assembly	89
3-2. Structure of the human U1 snRNP	91
3-3. Cryo-EM structure of the <i>S. cerevisiae</i> U1 snRNP	93
3-4. U1C sequence alignment	95
3-5. Structure of U1C	97
3-6. U1C and the U1 snRNA/5' SS duplex	99
3-7. U1C in the A complex	101
3-8. EMSA of U1C and the 5' SS	103

3-9. Design of disulfide cross-linking substrates	105
3-10. Cross-linking between a modified hairpin and U1C mutants	106
3-11. Modified PIP	108
3-12. The modified intron does not splice	109
3-13. The modified intron splices in a U1 snRNP independent construct	110
3-14. The reconstituted U1 snRNP does not rescue splicing in a U1 snRNP depleted extract	112
Chapter 4	
4-1. Structure of SF3b	124
4-2. Structure of the human Sap 49	126
4-3. Sequence alignment of Sap49 from six species	127
4-4. Interaction of Hsh49 with Cus1	129
4-5. Cus1 and Hsh49 in the A complex	131
4-6. Hsh49 and Cus1 in the B ^{act} structure	132
4-7. Phenotype of Nager syndrome	134
4-8. Growth tests of Hsh49 mutants	138
4-9. Expression of Sap49	140
4-10. Sap49 pulls down Sap145	141
4-11. Sap49 EMSAs	142
Chapter 5	
5-1. Conservation of Snu13 and K-turn motifs	151
5-2. Snu13 binds the K-turn of U4 snRNA	153
5-3. Interactions of Snu13 with Prp31	155
5-4. Snu13 at the heart of the U4/U6 di-snRNP	156
5-5. Rearrangement of the N-terminus of Prp31 upon incorporation into the tri-snRNP	157
5-6. Prp6, Prp3, and Prp4 in the tri-snRNP	158
5-7. The conserved hydrophobic binding pocket of Snu13	159
5-8. Structure of <i>C. merolae</i> Snu13	161
5-9. Sequence alignment of Snu13 from six organisms	163
5-10. <i>C. merolae</i> Snu13 binds U4 snRNA	164
5-11. Partial potential <i>C. merolae</i> tri-snRNP	165
Chapter 6	
6-1. Splicing in the red alga <i>C. merolae</i>	171
6-2. <i>C. merolae</i> Intron Removal	173
6-3. RH domains of RNase H and Prp8	175
6-4. Crystal structure of <i>S. cerevisiae</i> Prp8 with Aar2	177
6-5. Interactions of the RH domain of Prp8 in the tri-snRNP and the spliceosomal B complex	178

6-6. Prp8 RH interactions in the B ^{act} and C spliceosomal complexes	179
6-7. Interactions of the Prp8 RH domain in the C* complex	182
6-8. Prp8 RH domain interactions in the P and ILS complexes	183
6-9. Location of the RH domain relative to the Prp8 core during the splicing cycle	184
6-10. Sequence alignment and secondary structure prediction for the Prp8 RH domains of human (HS), <i>S. cerevisiae</i> (SC) and <i>C. merolae</i> (CM)	185
6-11. Crystal structure of the <i>C. merolae</i> Prp8 RH domain	186
6-12. Metal binding by Prp8	187
6-13. Sequence alignment of the Prp8 b-hairpin region from 35 taxa	191
 Chapter 7	
7-1. Equilibrium between the two steps of splicing	200
7-2. Prp8 alleles in the b-hairpin stabilize one of two monomers	203
7-3. The hT1783S/A hR1865A double mutants of the Prp8 RH domain	205
7-4. Phenotype of the yT1855A and yT1855S mutants is wild type	208
7-5. The hQ1894E hR1865S double mutant of the RH domain of Prp8	211
7-6. The hQ1894S hR1865A double mutant of the Prp8 RH domain	212
7-7. β -hairpin flexibility	215
7-8. Conservation of the metal binding site and hR1865	216
7-9. Model of the catalytic and transitional alleles of the Prp8 RH domain	218

List of Tables

Chapter 1

- 1-1. Conservation of seven spliceosomal proteins among four species 25
- 1-2. Comparison of protein components in human, *S. cerevisiae*, *S. pombe*, and *C. merolae* spliceosomes 27-29

Chapter 2

- 2-1. Protein and RNA components in the structures of the yeast spliceosome 38-43
- 2-2. Protein and RNA components in the structures of human spliceosomes 63-73

Chapter 4

- 4-1. Sap49 mutation in Nager syndrome patients 134

Chapter 5

- 5-1. Data collection and refinement statistics for *C. merolae* Snu13 162

Chapter 6

- 6-1. *C. merolae* introns 172
- 6-2. Data collection and refinement statistics for the RH domain of *C. merolae* Prp8 189
- 6-3. Intron number, hairpin conservation, and presence of Aar2 and Snu66 in 16 species 192

Chapter 7

- 7-1. Data collection and refinement statistics for four human Prp8 RH domain mutants 213

Chapter 1
An Introduction to pre-mRNA Splicing

1-1. Pre-mRNA Splicing

Eukaryotic genomes feature a split gene structure where protein-coding exons are interrupted by non-coding sequences called introns. Following transcription, introns in the precursor mRNA (pre-mRNA) are removed and exons are ligated together to produce mature mRNA in a process called pre-mRNA splicing (reviewed in Will & Luhrmann, 2011; Dunn & Rader, 2014). The resulting mRNA is then exported from the nucleus and is translated into protein. Correct splicing is required for the production of functional proteins. Many human diseases, including cancers, are caused by improper splicing (reviewed in Chabot & Shkreta, 2016; Anna & Monika, 2018).

Introns have three conserved sequences (Figure 1-1A): a GURAGU at the 5' splice site (5' SS) at the 5' end of the intron (and 3' end of the preceding exon); a branch point sequence (BPS) with the consensus sequence UACUAAC in *S. cerevisiae* (YUNAY in humans); and a 3' splice site (3' SS) composed of an AG dinucleotide at the 3' end of the intron (and 5' end of the following exon). Introns may also include a variable length poly-pyrimidine tract (PPT) composed of uridines (U) and cytidines (C) between their BPS and 3' SS (Senapathy et al., 1990).

Chemically, pre-mRNA splicing occurs via two sequential transesterification reactions (Figure 1-1). In the first step of splicing, the 2' hydroxyl of the branch A, UACUAAC, (bolded and underlined) within the BPS attacks the phosphodiester linkage preceding the first GU at the 5' SS to generate a free 5' exon and a lariat intron still attached to the 3' exon. The circularized, lariat

structure of the intron is due to three phosphodiester bonds on the branch A: standard 5' and 3' linkages to the upstream and downstream regions of the intron respectively, and the new 2'-5' bond formed by the branch A and the G at the 5' SS. Following the first step of splicing, the newly released 3' hydroxyl at the end of the 5' exon attacks the AG dinucleotide at the 3' SS to release the lariat intron and ligate the two exons (Jacquier, 1990). Phosphorothioate substitutions (replacing one of the non-bridging oxygens on the backbone phosphate with sulfur) at the 5' and 3' SS to generate chiral substrates showed an inversion of stereochemistry for each step of splicing, consistent with an in-line S_N2 reaction (Moore & Sharp, 1993).

Spliceosome assembly occurs at each intron to be removed. Most genes in higher eukaryotes have multiple introns, and the splicing process is repeated for each intron. Once the introns have been removed and the exons ligated the mRNA is exported to the cytosol for translation.

1-2. Alternative Splicing

Constitutive splicing refers to the most common splicing pattern for a given pre-mRNA. However, in alternative splicing different transcripts of the same gene are spliced using distinct splice sites to produce an array of spliced mRNA (reviewed in Nilsen & Graveley, 2010). The resulting change in mRNA sequence can lead to the production of different protein isoforms. Alternative splicing increases genome diversity without a corresponding increase in genome size. Levels of alternative splicing correlate with organism complexity, with an

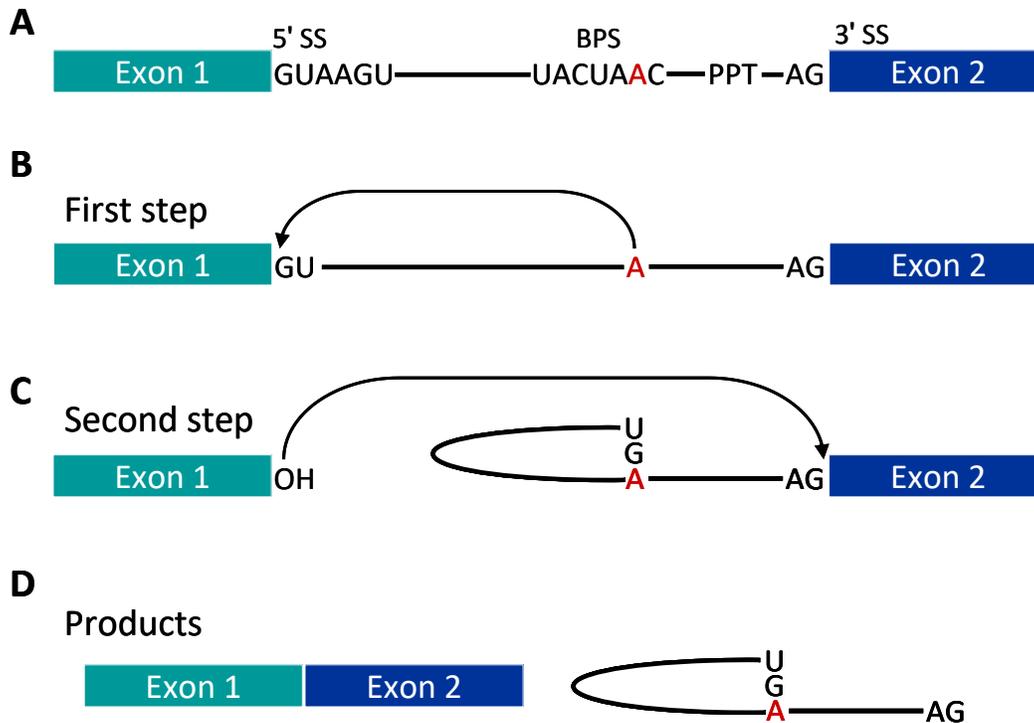


Figure 1-1. Pre-mRNA splicing. A. Structure of pre-mRNA showing conserved sequences within the intron, including 5' splice site (5' SS), 3' splice site (3' SS), and branch point sequence (BPS) with the branch adenosine (branch A) in red. B. The first step of splicing occurs when the 2' hydroxyl of the branch A attacks the 5' SS. C. The products of the first step of splicing are the free 5' exon with a 3' hydroxyl and a lariat intron attached to the 3' exon. The 3' hydroxyl of the 5' exon attacks the 3' SS in the second step of splicing. D. The products of the second step of splicing are the ligated exons and a lariat intermediate.

estimated 95% of human genes being alternatively spliced (Pan et al., 2008; Wang et al., 2008). There are a number of ways a transcript may be alternatively spliced (Figure 1-2). Exons may be included or excluded (exon skipping), including mutually exclusive exons. Introns may be removed, or become part of an exon within the transcript (intron retention). Using a different 5' or 3' SS changes the length of the resulting introns (alternate splice site selection; Black, 2003).

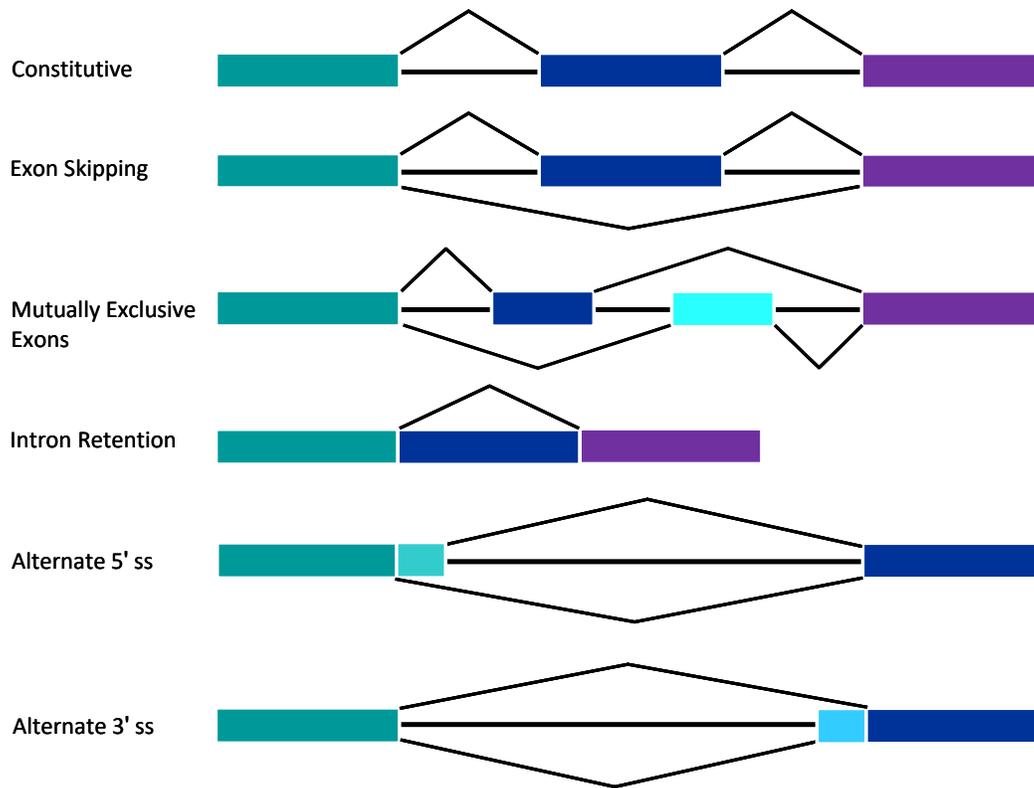


Figure 1-2. Alternative splicing. A number of alternative splicing types exist. Exons are shown in coloured boxes; introns in black. Sequences to be spliced out are indicated by chevrons.

Alternative splicing is highly regulated by a network of trans-acting splice site activators and silencers that influence splice site selection by either blocking or recruiting splicing factors at specific splice sites. The equilibrium between activators and silencers directs splice site selection (Chen & Manley, 2009; Smith & Valcarcel, 2000). Mutations that alter the balance of spliced isoforms, or activate novel splice sites (cryptic splice site activation) often lead to disease (see below; Ohno et al., 2015; Ward & Cooper, 2010).

1-3. snRNPs

Pre-mRNA splicing is catalyzed by a large multi-megaDalton complex, called the spliceosome, composed of both protein and RNA. Five small nuclear ribonucleoprotein particles (snRNPs) U1, U2, U4, U5 and U6, assemble on each intron to be removed (Brow, 2002; Hoskins et al., 2011). Each snRNP is composed of the eponymous U-rich small nuclear RNA (snRNA) with a specific secondary structure and encircled by a heptameric ring of Sm (or Lsm in U6 snRNP) proteins (Figure 1-3; Hinterberger et al., 1983; Bringmann & Luhrmann, 1986). Other snRNP-specific proteins associate with the snRNA and Sm core. These proteins may be structural, or have catalytic or regulatory activity.

U1 snRNP and U2 snRNP act individually. U4, U5 and U6 snRNPs are assembled into the U4/U6.U5 tri-snRNP, with extensive base-pairing between the U4 and U6 snRNAs (Nguyen et al., 2015). In addition to the snRNPs, other splicing proteins and protein complexes that are not initially associated with a U-rich snRNA act to regulate the spliceosome.

In humans, U1 snRNP consists of U1 snRNA, the Sm core, and three snRNP-specific proteins: U1A, U1C and U1-70k (Pomeranz Krummel et al., 2009). U1 snRNA is well conserved and features an invariant, short, single-stranded region at the 5' end. U1A and U1-70k are structural proteins that stabilize the U1 snRNP. *S. cerevisiae* contains six additional U1 associated proteins. In humans, homologues of these proteins are trans-acting factors that regulate splicing (Li et al., 2017).

The U2 snRNP is composed of the U2 snRNA, the Sm core and twelve U2-specific proteins. U2 snRNA toggles between two mutually exclusive stem loops, stem IIa and stem IIc, during the splicing cycle (Rodgers et al., 2016). Most of the proteins in U2 snRNP are arranged into two sub-complexes, SF3a and SF3b, that promote U2 snRNP binding to the pre-mRNA and then dissociate prior to the first step of splicing. SF3a is composed of three proteins: Prp21, Prp11 and Prp9 (Lin & Xu, 2012). In humans, SF3b is composed of seven proteins: Sap155, Sap130, Sap145, Sap49, Sap14a/p14, Sap14b/PHF5a and Sap10. They interact with the branch site to keep it sequestered before the first step of splicing (Golas et al., 2003). Sap155 contains 22 HEAT repeats, and tightly binds p14. In addition, U2 snRNP has two structural proteins, U2-A' and U2-B'', that remain associated with the spliceosome throughout the splicing cycle (Casparly & Seraphin, 1998).

U4 and U6 snRNAs are extensively base-paired with each other to form helices I and II, separated by the U4 snRNA internal stem loop (Hardin et al., 2015). The ACAGAGA sequence in U6 snRNA that recognizes the 5' SS is located at the base of a stem loop (Nguyen et al., 2016). While U4 snRNA contains the Sm core common to the other snRNPs, U6 snRNA is encircled by the Lsm core, which is composed of homologues to the Sm proteins (Zhou et al., 2014). There are five proteins associated with the U4/U6 di-snRNP: Snu13 and Prp3 are structural proteins; Prp31 is important for formation of the tri-snRNP; and Prp4 is required for incorporation of the tri-snRNP into the spliceosome (Nguyen et al., 2016).

The U5 snRNP is the largest snRNP. Like the other snRNPs, U5 snRNA is

associated with the Sm core. Seven U5 snRNP-specific proteins are then assembled onto the snRNP (Nancollis et al., 2013). Prp8 is a large, scaffolding protein at the core of the spliceosome. It interacts with Brr2, an RNA helicase, and Snu114, a GTPase that regulates Brr2. The U5 snRNP also includes Prp6, which stabilizes the tri-snRNP, and the helicase Prp28 (Sander et al., 2006). Dib1 prevents premature spliceosome activation (Schreib et al., 2018).

Three tri-snRNP specific proteins are added when the U5 snRNP associates with the U4/U6 di-snRNP. Sad1 prevents the dissociation of U5 snRNP from the tri-snRNP. Snu66 is a long protein with unknown function. 27k is a 27 kD protein that may help recruit the tri-snRNP to the spliceosome (Nguyen et al., 2016).

1-4. Additional Splicing Proteins

A number of other proteins not complexed with snRNA are also involved in splicing. Some bind to the spliceosome in complexes, such as those associated with the nineteen complex (NTC), and others bind individually. *S. cerevisiae* spliceosomes are associated with about 100 proteins, and human spliceosomes with over 200 proteins. The main non-snRNP associated spliceosomal proteins are discussed below.

U2AF is a heterodimer of U2AF65 and U2AF35. It, along with splicing factor 1 (SF1) recognize the 3' end of the intron. U2AF65 binds the PPT while U2AF35 interacts with the 3' SS. SF1 binds the BPS (Huang et al., 2002).

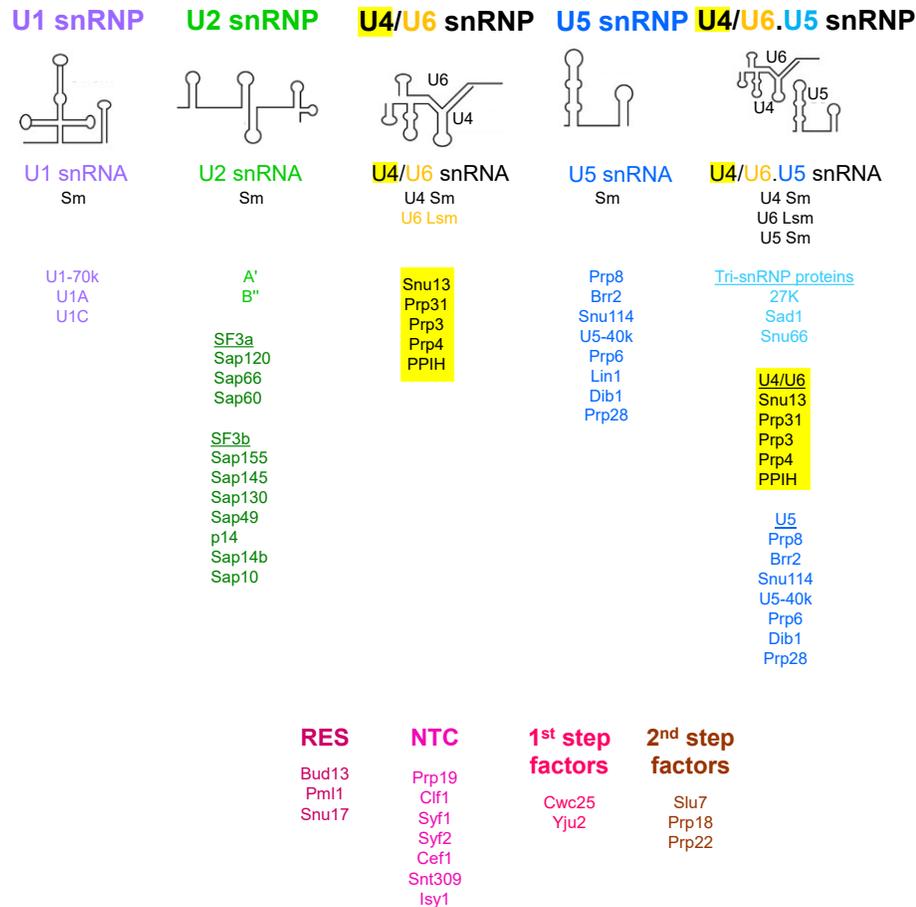


Figure 1-3. snRNP composition. Composition of the snRNPs and other associated complexes. RES – retention and splicing; NTC – nineteen complex. Adapted from Will & Luhrmann, 2011

The SR protein family is composed of proteins featuring repeating serine-arginine (SR) dipeptides. They activate splicing at specific sequences by binding exon splicing enhancers (ESEs) to recruit the spliceosome to nearby splice sites. To date, a core set of 12 SR proteins has been identified in humans (Zhou & Fu, 2013).

Heterogeneous ribonucleoprotein particles (hnRNPs) act in opposition to SR proteins: instead of activating splice sites they repress them. hnRNPs bind to

exon silencing sequences (ESSs) to block spliceosome assembly at these splice sites. Humans express 13 different hnRNP families (Busch & Hertel, 2012).

The retention and splicing complex (RES) consists of three proteins and is required for the first step of splicing. Bud13 and Pml1 bind Snu17 in a cooperative manner (Wysoczanski & Zweckstetter, 2016). RES stabilizes the spliceosome prior to activation and is required for efficient progression of the spliceosome through the early stages of the splicing cycle (Bao et al., 2017a). It also prevents premature binding of Prp2 to the spliceosome, acting to delay transition to an active spliceosome.

The nineteen complex (NTC) is named for the tetramer of Prp19 at its core, and is composed of eight additional proteins: Cef1, Clf1, Syf1, Syf2, Isy1, Snt309, and Ntc20. It joins the spliceosome after the tri-snRNP, and is involved in the regulation of splicing (Hogg et al., 2010). While it is required for the dissociation of the Lsm proteins from U6 snRNA, it stabilizes the association of U5 and U6 snRNPs with the spliceosome following the loss of U4 snRNP (Chan et al., 2003).

There are additional proteins that associate with the NTC, including Cwc25 (Tseng et al., 2017) and Yju2 (Liu et al., 2007a); following the release of SF3a and SF3b, they enable the first step of splicing to occur, and are then removed after the first transesterification reaction. Other proteins associated with the NTC in the nineteen related complex (NTR) are Cwc2, Bud31, Cwc15, Ecm2, Prp45, and Prp46. Cwc2 cross-links to U6 snRNA in both steps of splicing, stabilizes an interaction between U2 and U6 snRNAs, and antagonizes Prp16 (McGrail et al.,

2009). The proteins Cwc21, Cwc22, Cwc23, Cwc24, Cwc25, Cwc27, Prp17, Prp22, Slu7, and Spp2 also bind to the spliceosome (Schmidt et al., 2014).

The core of the exon junction complex (EJC), present in higher eukaryotes, is composed of four proteins: eIF4AIII, Magoh, Y14, and MLN51. It is deposited about 24 nucleotides upstream of the exon-exon junction during the splicing cycle, and improves export of the spliced mRNA from the nucleus to the cytosol (Woodward et al., 2017).

The spliceosome disassembly complex, sometimes also called the Ntr, consists of three proteins (Ntr1, Ntr2, and Prp43) and is required for the release of snRNPs from the spliceosome following exon ligation (Su et al., 2018).

Progression through the splicing cycle, and proofreading of splice sites are controlled by RNA helicases (Koodathingal & Staley, 2013). Prp5, a helicase required for pre-spliceosome formation, remodels U2 snRNA to facilitate recognition of the branch point sequence, and proofreads the resulting duplex (Liang & Cheng, 2015). The U5 snRNP protein Prp28 displaces U1 snRNP from the 5' splice site and allows U6 snRNA to base-pair in its place (Staley & Guthrie, 1999). Prp2 and its cofactor Spp2 activate the spliceosome by promoting the removal of SF3a and SF3b (Warkocki et al., 2009). Prp16 is an RNA helicase that reorganizes the spliceosome for exon ligation following intron branching (Burgess & Guthrie, 1993). Prp22 promotes the second step of splicing, and plays a role in exon release (Schwer, 2008). Prp43 is required for spliceosome disassembly and release of the spliced mRNA (Tsai et al., 2005).

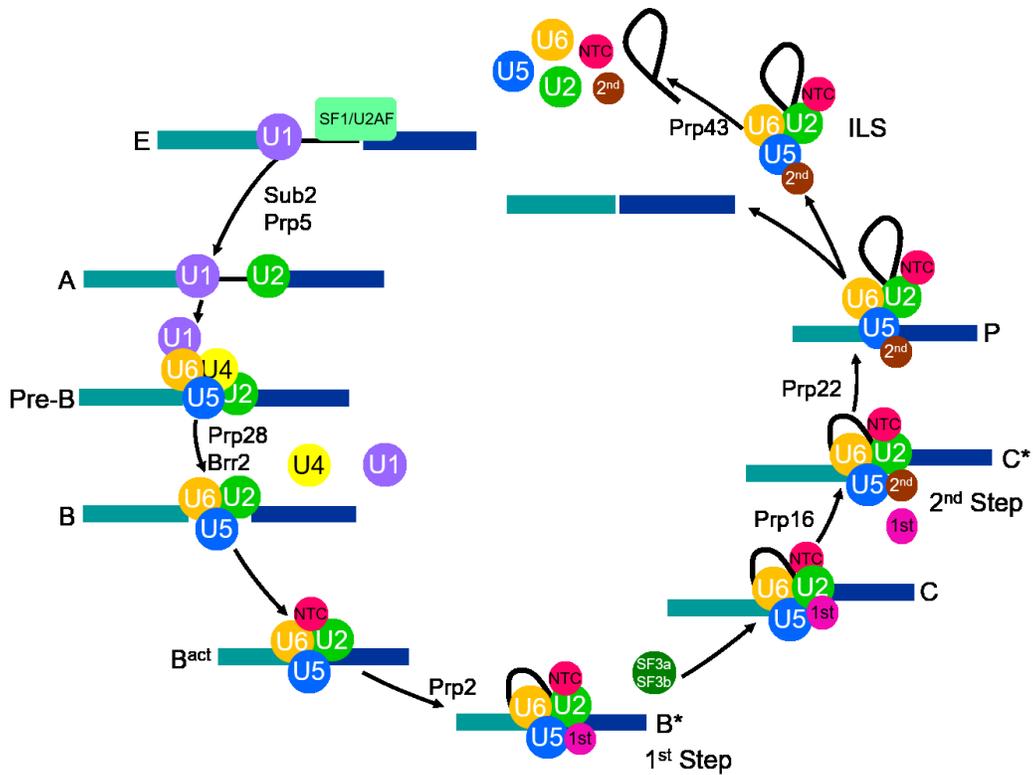


Figure 1-4. Spliceosome assembly. Assembly and disassembly of a spliceosome on an intron. Colours are consistent with Figure 1-3; based on Will and Luhrmann, 2011.

1-5. The Splicing Cycle

Prior to splicing, SR proteins and hnRNPs bind the pre-mRNA transcript to form H complex and direct spliceosome assembly (Bennett et al., 1992). U2AF and SF1 bind the 3' SS and BPS respectively (Huang et al., 2002).

Canonically, the splicing cycle (Figure 1-4) begins when the 5' end of U1 snRNA in U1 snRNP base-pairs with the intron's 5' SS (Siliciana & Guthrie, 1988) to form early (E) complex (commitment complex (CC) in yeast; Michaud & Reed, 1993). U1C is proposed to stabilize this interaction (Rossi et al., 1996). These ATP-independent associations commit the pre-mRNA to splicing, with the intron

defined by the U1 snRNP and U2AF identification of splice sites.

U2AF and SF1 recruit U2 snRNP to the intron. U2 snRNP associates with the intron in the first ATP-dependent step, giving rise to the A complex (Crawford et al., 2013). Sub2 acts to displace U2AF and SF1 from the BPS and 3' SS (Kistler & Guthrie, 2001). SF3a holds U2 snRNP in the closed conformation that makes it a substrate for Prp5 (Wiest et al., 1996). Prp5 hydrolyses ATP to remodel U2 snRNA into an open conformation, freeing the branch binding region. U2 snRNA adopts stem IIa, and the branch binding region base-pairs with the BPS, forming an imperfect duplex (Rodgers et al., 2016). The branch A is excluded from the helix, selecting it as the nucleophile for the first step of splicing (Query et al., 1996; Newby & Greenbaum, 2001). SF3b contacts this bulged A, and isolates it until the spliceosome is fully assembled (Schellenberg et al., 2011; Plaschka et al., 2017).

The U4/U6.U5 tri-snRNP binds the A complex to form the B complex. Docking of the tri-snRNP to the spliceosome activates the Prp28 and Brr2 helicases (Raghunathan & Guthrie, 1998). Prp28 allows exchange of U1 snRNA with U6 snRNA at the 5' SS, possibly by destabilizing the interaction between U1C and the 5' SS (Staley & Guthrie, 1999; Mohlmann et al., 2014). U1 snRNP is released from the spliceosome. Nucleotides in the loop of the U5 snRNA stem 1 base-pair with the 3' end of the 5' exon (Cortes et al., 1993). Snu114 exchanges its bound GDP for GTP, which activates Brr2 (Small et al., 2006). Brr2, further regulated by the C-terminus of Prp8, unwinds the U4/U6 stem I (Kuhn et al., 2002a; Mozaffari-Jovin et al., 2013). It is not known how stem II is unwound. After this structural

rearrangement, U4 snRNA is released from the spliceosome, along the U4/U6 di-snRNP proteins, including the Lsm core of U6 (Cheng & Abelson, 1987; Chan & Cheng, 2005; Hoskins et al., 2011). The loss of the U4 snRNA allows U6 snRNA to base-pair with U2 snRNA and form the U2/U6 helices Ia and Ib (Madhani & Guthrie, 1992; Fortner et al., 1994; Ryan & Abelson, 2002). The U6 snRNA ACAGAGA stem loop denatures, and the U6 internal stem loop (ILS), which is vital for splicing catalysis, forms (Madhani & Guthrie, 1992; Fortner et al., 1994). Because the U2/U6 snRNA duplex is mutually exclusive with the U4/U6 snRNA duplex, the latter has been proposed as a regulatory structure to prevent the premature formation of the U2/U6 snRNA duplex in the spliceosomal active site (Brow & Guthrie, 1989).

The NTC, NTR, and RES complexes are recruited to the spliceosome after U1 and U4 snRNPs are released, giving rise to the B^{act} complex (Bessonov et al., 2010). Further structural rearrangement gives rise to the B* complex. Prp2 releases the RES complex in the transition to B*, and displaces SF3a and SF3b to expose the nucleophilic 2' hydroxyl on the branch A (Warkocki et al., 2009; Bao et al., 2017b). U5 loses Prp6, Dib1, Snu23 and Spp381 (Schmidt et al., 2014). Cwc24 and Cwc27 of the NTC are also released, which allows for the binding of first step factors Yju2 (Liu et al., 2007a) and Cwc25 (Tseng et al., 2017) to activate the first step of splicing. The 5' SS and BPS are brought into close proximity. U2 snRNA stem IIa adopts the mutually exclusive stem IIc, which promotes the first step of splicing (Hilliker et al., 2007). Branching of the intron occurs.

Subsequent rearrangements of B* complex, driven by Prp16, produce the C complex (Umen & Guthrie, 1995a; Smith et al., 2008). Prp16 unwinds the U6 snRNA ISL and displaces Yju2 and Cwc25. The second step factors Slu7, Prp18, and Prp22, whose association with the spliceosome was blocked by Yju2 and Cwc25, now bind (Tseng et al., 2011). Cwc22 recruits the EJC (Steckelberg et al., 2015). The newly formed 2'-5' branch product of the first step of splicing is moved out of the active site (Schwer & Guthrie, 1992). U2 snRNA stem IIc toggles back to stem IIa for rearrangement of the substrates, then back to stem IIc for the second step of splicing (Hilliker et al., 2007). The 5' SS and 3' SS are brought together by U5 and U6 snRNAs, and the U6 ISL reforms (Konarska et al., 2006). The C* complex is formed, and the second step of splicing takes place to ligate the exons (Fica et al., 2017; Yan et al., 2017). Rearrangements of the C* complex give rise to the P complex.

Following splicing, Prp22 releases the mature mRNA from the spliceosome by disrupting the interaction between U5 snRNA and the mRNA (Schwer, 2008). The mRNA is exported from the nucleus into the cytosol for translation.

Release of mature mRNA produces the intron lariat spliceosome (ILS) which is then disassembled (Chen et al., 2014). Prp43 is recruited to the spliceosome and, regulated by Ntr1 and Ntr2 (Tsai et al., 2007), promotes spliceosomal disassembly (Tsai et al., 2005). Brr2 and Snu114 facilitate the unwinding of the U2/U6 snRNA duplex (Small et al., 2006). U2, U5 and U6 snRNPs and the other protein components are released to be reassembled and

recycled in a new round of splicing. The lariat intron is debranched by Dbr1 and degraded.

Other pathways to a mature spliceosome are possible. In humans, with multiple, long introns of varying size, and exons of consistent length (~250 bases), the exon may be defined first. In an exon-defined complex, U1 snRNP interacts with the U2 snRNP 5' to it on the upstream intron (Berget, 1995). This complex is then resolved into the intron-defined spliceosome described above and splicing occurs. Additionally, the minor spliceosome, composed of U11, U12, U4atac and U6atac snRNPs (analogous to U1, U2, U4 and U6 snRNPs respectively), along with U5 snRNP, removes about 1% of human introns with distinct 5' and 3' splice sites (Turunen et al., 2013).

1-6. Splicing Fidelity

Many proteins in the spliceosome act to ensure splice sites are selected appropriately. In addition to the proteins and RNAs involved directly in splice site selection, RNA helicases have been suggested to act in a kinetic proofreading model to ensure proper splicing (Semlow & Staley, 2012). Prp28, Prp22, and Prp16 act as timers to reject substrates that take too long to splice, as suboptimal substrates are cleaved more slowly than optimal ones due to a decreased stability of spliceosome association with the pre-mRNA (Koodathingal & Staley, 2013).

Prp16 proofreading is based on a competition between the Prp16-dependent

release of the first step factor Cwc25 and the first step of splicing (Tseng et al., 2011). If a substrate splices quickly, Prp16 acts after the first step, as it should, to release Cwc25 and the spliceosome proceeds to the second step. However, if the first step is slow due to, for example incorrect 5' SS selection, Prp16 releases Cwc25 before lariat formation, and the spliceosome is disassembled by Prp43 (Koodathingal et al., 2010). Prp16 is also able to reject aberrant intermediates as it catalyzes the rearrangement of the spliceosome between the first and second steps (Villa & Guthrie, 2005). Prp28 may proofread the U6 snRNA/5' SS interaction and reject 5' SS that do not form a stable duplex with U6 snRNA (Yang et al., 2013). Prp2 may sense the protein composition in B complex. If Prp2 binds prematurely to complexes that lack RES or other proteins, it disassembles these complexes before they can proceed to activation (Warkocki et al., 2009). The disassembly factor Prp22 may be activated prior to exon ligation (Mayas et al., 2006): pulling on the 3' exon could destabilize the interaction of a sub-optimal 3' SS with the spliceosome, leading to spliceosome disassembly before exon ligation. The spliceosome disassembly factor Prp43 is also able to disassemble spliceosomes with aberrant products or otherwise stalled spliceosomes (Pandit et al., 2006).

If pre-mRNAs are spliced incorrectly, they may be degraded by nonsense-mediated decay (NMD) once in the cytosol (Lykke-Andersen & Jensen, 2015). NMD is activated by premature termination codons (PTC), which are stop codons more than 55 nucleotides upstream of an exon junction complex. The EJC is deposited on mRNA about 24 nucleotides upstream of the 5' SS (Le Hir et al.,

2000). As stop codons are usually found in the final exon, a termination codon upstream of an EJC could indicate improper splicing. EJCs are removed from the mRNA by the ribosome during translation. Any EJCs remaining on the mRNA following termination of translation recruit proteins to degrade the mRNA.

1-7. Splicing and Human Diseases

Improper splicing produces mRNAs that do not encode for the intended protein. Altered splicing patterns can lead to mRNAs with deletions or insertions, frame shift mutations, or premature stop codons. Decay of these mRNAs, for example by NMD, leads to a decreased level of protein product (Lykke-Andersen & Jensen, 2015). If translated, the resultant mutant proteins often have a deleterious effect on the cell.

Mutations within pre-mRNAs can change their splicing patterns. In addition to mutated splice sites themselves, mutations in regulatory sites within the pre-mRNA can activate cryptic splice sites, or silence canonical ones, leading to incorrect alternatively spliced mRNAs (Matlin et al., 2005). Additionally, mutations in spliceosomal proteins can lead to the recognition of incorrect splice sites and production of incorrect mRNAs.

A large number of human diseases are caused by, or associated with, aberrations in splicing (Rahman et al., 2015; Chabot & Shkreta, 2016; Anna & Monika, 2018): mutated splice sites are implicated in at least 15% of all genetic diseases (Matlin et al., 2005). An understanding of splicing can shed light on how

splicing has been disrupted and give insights into the nature of the resulting proteins. A handful of diseases associated with mis-spliced mRNA are discussed below.

1-7a. Mutations in splice sites/RNA

Cystic fibrosis (CF) is a recessive genetic disorder caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR), a $\text{Cl}^-/\text{HCO}_3^-$ channel. The resulting mutant protein leads to secretion of viscous mucus that causes problems with the digestive and pulmonary systems. Symptoms of CF include poor weight gain and growth due to an impairment in nutrient absorption, as well as difficulty breathing and frequent chest infections due to buildup of mucus in the lungs. About 70% of CF cases are caused by the deletion of phenylalanine 508. Mutations altering splicing or splice site strength are also associated with CF. For example, the PPT within intron 8 varies from 5-9 uridines. The 5U variant splices with low efficiency, causing in frame exon skipping (Chu et al., 1993). When the DNA and mRNA of a patient with CF was sequenced, it was discovered that the G893G mutation (2811 G>T) created a new splice site in exon 15 that results in the in-frame deletion of 76 amino acids. The resulting protein retains some function, resulting in a mild phenotype (Faa et al., 2010).

Duchenne muscular dystrophy (DMD) and the less severe Becker muscular dystrophy (BMD) are both X-linked myopathies due to mutations in the dystrophin gene (Bellayou et al., 2009). They cause muscle weakness and muscle wasting.

Premature stop codons and frame shift mutations in dystrophin result in DMD, while mutations that maintain the reading frame are seen in BMD. In one patient, a G>A mutation at the 3' end of the 156 bp exon 25 (GU>AU, abolishing the 5' splice site of the following intron) results in the skipping of this exon (Habara et al., 2009). However, the resulting transcript remains in frame, leading to the milder BMD. A patient with DMD also had a G>A mutation at the 3' end of exon 45 (5' splice site of the following intron). In this case, a downstream cryptic splice site is activated, leading to the loss of 32 bp and a frame shift of -1 (Habara et al., 2009).

Spinal muscular atrophy (SMA) is a recessive degenerative disease that causes loss of motor neurons and leads to muscle wasting. It is caused by mutations in SMN (survival motor neuron), a protein involved in the assembly of the Sm core on snRNAs (Price et al., 2018). A defect in splicing of the pre-mRNA transcript, due to mutations that cause skipping of exon 7, produces a truncated protein that is degraded (Wu et al., 2018). Loss of the SMN protein interferes with the formation of the snRNPs required for splicing. Thus, a splicing defect in SMN causes further splicing defects due to a decrease in snRNP levels. Spinraza, an antisense oligonucleotide that encourages inclusion of exon 7, has recently been approved by the FDA for treatment of SMA (Wan & Dreyfuss, 2017).

This is not a comprehensive list of diseases associated with mutations that alter splicing of the pre-mRNA transcript. Succinyl-CoA:3-ketoacid CoA transferase deficiency can arise due to the skipping of exons 12 and 13 in the of the 3-oxoacid CoA transferase 1 (OXCT1) transcript (Hori et al., 2013). Hemophilia

B can be caused by exon skipping due to a truncated PPT within the factor IX (9) gene (van de Water et al., 2004). Neurofibromatosis type 1 can be caused by a number of mutations in the neurofibromin (NF1) gene, 25% of which induce alternative splicing. For example, a C>G mutation within the BPS of intron 15 causes intron retention (Xu et al., 2014).

1-7b. Mutations in spliceosomal proteins

It is not only mutations in pre-mRNAs that drive aberrant splicing: splicing factors themselves may also harbour mutations that lead to disease (Faustino & Cooper, 2003). Given the ubiquity of splicing throughout all cell types, it is surprising that mutations in constitutive splicing factors tend to give rise to a localized phenotype (Lehalle et al., 2015).

Retinitis pigmentosa is a form of familial vision loss caused by a decrease in photoreceptor cells. A subset of patients have mutations in key spliceosomal proteins, including Prp8, Brr2, Prp3, Prp4, and Prp31, which are thought to disregulate Brr2's unwinding of the U4/U6 duplex (McKie et al., 2001; Vithana et al., 2001; Chakarova et al., 2002). These mutations influence alternative splicing in a tissue-specific manner, especially within the retina (Farkas et al., 2012).

Guion-Almeida mandibulofacial dysostosis is associated with mutations in Snu114 (Lines et al., 2012). Haplo-insufficiency arising from these mutations, including frameshift, missense, nonsense, and splice site mutations, causes microcephaly, dysplastic ear, and developmental delays.

Nager syndrome is an acrofacial dysostosis associated with malar hypoplasia, micrognathia and upper limb anomalies. It is caused by haplo-insufficiency of Sap49, resulting from mutations to the initiator Met, nonsense mutations, or frame-shift mutations (Bernier et al., 2012; Petit et al., 2014). A recently discovered isoleucine to arginine mutation will be discussed further in chapter 4.

1-7c. Splicing and Cancer

Alternative splicing in cancer cells preferentially produces protein isoforms that improve cell proliferation, survival, or migration. These changes to the splicing pattern may be caused by mutations in specific cancer genes themselves, or mutations in spliceosomal proteins (Urbanski et al., 2018).

Sap155 is one of the most commonly mutated splicing proteins in cancer. Mutations, especially those that cluster in the HEAT repeats of Sap155, alter 3' splice site selection (Darman et al., 2015). Recurrent mutations are commonly seen in myelodysplastic syndromes (in about 80% of cases; Malcovati et al., 2011), as well as occasionally seen in chronic lymphocytic leukemia (15%; Wang et al., 2011), pancreatic (3%; Biankin et al., 2012) and breast cancers (2%; Maguire et al., 2015). Small molecules exist that act on Sap155, including spliceostatins, sudemycins and pladienolides (Corrionero et al., 2011; Fan et al., 2011; Convertini et al., 2014; Kashyap et al., 2015). Clinical trials are now underway to test the efficacy of such drugs in the treatment of cancer.

An example of a disrupted balance between spliced isoforms leading to cancer is Mnk2, a kinase in the MAPK pathway. It has two spliced isoforms: Mnk2-a, the full length protein that includes the MAPK domain; and, Mnk2-b, a truncated protein that lacks the MAPK domain (Scheper et al., 2003). Both isoforms can phosphorylate the translation initiation factor eIF4E, which leads to cell growth. However, the MAPK domain, absent in Mnk2-b, is required to phosphorylate p38, which leads to cell death in response to stress (Maimon et al., 2014). An increase in Mnk2-b relative to Mnk2-a leads to cell growth due to the inability to activate the pro-apoptotic p38. While Mnk2-a levels are higher in normal breast, lung and colon tissue, Mnk2-b levels are higher in the corresponding tumour cells (Scheper et al., 2003).

The BCL2L1 transcript is alternatively spliced to BCL-xL (anti-apoptotic) or BCL-xS (pro-apoptotic) isoforms based on the 5' SS of exon 2 (Boise et al., 1993). An increase in the ratio of BCL-xL to BCL-xS is found in lymphoma, and neuroblastoma cell lines and primary tumours (Dole et al., 1995; Xerri et al., 1996). Increased expression of BCL-xS in cancer cell lines sensitizes these cells to radiation and chemotherapy, while increased levels of BCL-xL decreases apoptosis following chemotherapy (Li et al., 2016).

To tweak the balance between alternatively spliced mRNAs, small antisense oligonucleotides (ASOs) are under investigation to block access to the splice sites that lead to the oncogenic isoforms (Havens & Hastings, 2016). To increase their stability in cells, ASOs are modified with phosphorothioates or 2' methoxyethyls.

1-8. Splicing Studies

Splicing has been studied in a variety of organisms. This thesis will explore splicing in four systems: human, *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae* and *Cyanidioschyzon merolae*. Each system has its own benefits and limitations, but together can offer a more complete picture of splicing than any one organism on its own.

Splicing errors and mutations in splicing factors give rise to human disease. Understanding the workings of the human spliceosome is crucial in the design of treatments for these diseases. Study of the human spliceosome will provide the most accurate insights into human splicing. Human spliceosomes can be reconstituted for splicing experiments *in vitro*. However, the human spliceosome is incredibly complex, making it difficult to tease out the role of individual splicing factors and their regulation. In addition, cell culture and genetic manipulation are more difficult with human cell lines than with *S. cerevisiae*. Recombinant expression of human proteins in *E. coli* is possible, but the resulting constructs lack post-translational modifications.

S. pombe is a yeast system with a spliceosome similar to that of humans, but possessing fewer regulatory components. Core spliceosomal proteins are highly conserved, so that insights from the *S. pombe* spliceosome are applicable to humans. However, there is no reconstituted *in vitro* splicing system in *S. pombe*: work must be done either *in vivo* or with recombinantly expressed proteins. Isolated

spliceosomes in *S. pombe* extracts are predominantly in a post-catalytic lariat-intron complex (Chen et al., 2014).

S. cerevisiae features a reconstituted *in vitro* splicing system that is easy to manipulate (Ares, 2013). Genetic mutations or deletions can be made easily and tested for viability (Lesser & Guthrie, 1993; Igel et al., 1998) using a variety of assays. Both *in vivo* and *in vitro* experiments can then be performed with these mutant strains to explore the effects on splicing. The *S. cerevisiae* spliceosome is also simpler than that in humans with high conservation of splice sites. However, the *S. cerevisiae* system lacks several key splicing proteins, including the SF3b component p14 (Dziembowski et al., 2004). Spliceosomal proteins are highly conserved between humans and *S. cerevisiae*, as is the pattern of spliceosome assembly and disassembly. Insights into the *S. cerevisiae* spliceosome are generally applicable to humans.

	HS-SP	HS-SC	HS-CM	SP-SC	SP-CM	SC-CM
Prp8	72.8 (84.5)	60.0 (74.9)	30.8 (45.4)	60.6 (79.6)	29.7 (44.7)	28.3 (44.2)
Brr2	48.1 (67.9)	38.9 (59.1)	20.8 (33.7)	41.6 (60.7)	19.8 (31.7)	19.7 (31.6)
Snu114	54.5 (70.6)	33.5 (51.9)	21.5 (35.8)	34.9 (52.6)	22.4 (36.3)	21.0 (34.4)
Snu13	73.4 (85.9)	68.0 (83.6)	51.0 (68.0)	80.2 (89.7)	51.7 (68.3)	50.3 (69.7)
Sap49	37.2 (48.9)	17.5 (26.7)	9.2 (13.3)	15.1 (24.3)	9.4 (16.2)	12.3 (18.8)
Sap155	52.4 (67.0)	36.8 (49.4)	17.6 (30.5)	40.9 (56.8)	18.5 (32.7)	21.3 (37.4)
Prp3	29.0 (43.1)	17.7 (29.6)	15.4 (28.2)	21.7 (35.6)	18.9 (30.6)	18.1 (30.2)

Table 1-1. Conservation of seven spliceosomal proteins among four species. Percent sequence identity (with similarity in brackets) for spliceosomal protein homologues. HS is human, SP is *S. pombe*, SC is *S. cerevisiae*, CM is *C. merolae*. Alignments made with ClustalOmega.

C. merolae is an acidophilic red alga that lives at high temperatures and low pH (50°C, pH 1.5-3) whose genome was recently described (Matsuzaki et al., 2004). This organism has only 27 annotated introns in 26 genes, out of 4,803 genes in total. It has a much smaller complement of splicing proteins than any of the above systems, and likely purged its genome of less essential splicing proteins, leaving a core of the most fundamental ones (Stark et al., 2015). RT-PCR has shown that the pre-mRNA is spliced; however, little else is known about splicing in *C. merolae*. Protocols for studying splicing in this organism have not been developed. Thus far, information about *C. merolae*'s spliceosome results from bioinformatic studies and the expression of recombinant *C. merolae* proteins. The high temperature of its habitat may make its proteins amenable to crystallization.

1-9. Summary

It is important to study splicing. The spliceosome is a highly complex cellular machine, and improper splicing is linked to many human diseases. While our knowledge of pre-mRNA splicing and the spliceosome has grown considerably since splicing was first described, there is still much to learn.

In this thesis I will investigate a number of spliceosome associated proteins and explore their role in pre-mRNA splicing. Chapter two is an overview of the recently published cryo-EM structures of spliceosomal complexes. Chapter three investigates how U1C interacts with the 5' splice site, both alone and within the context of the U1 snRNP using RNA with a thiol tether attached to the phosphate

Complex	Human	<i>S. pombe</i>	<i>S. cerevisiae</i>	<i>C. merolae</i>	
U1 snRNP	U1-A	usp10/mud2	Mud1	-	
	U1-70K	Usp101p	Snpl	-	
	U1-C	Usp103p	Yhc1	-	
	Tia1*		Nam8	-	
		Usp105p	Prp39	-	
	Prp40*		Prp40	-	
			Prp42	-	
			Snu56	-	
	RBM25*	-	Snu71	-	
	Luc7*	Luc7/Usp106p	Luc7	-	
			Urn1	-	
			Npl3	-	
	U2 snRNP	U2-A'	Lea1	Lea1	-
		U2-B''	Msl1	Msl1	-
SF3a1/SF3a120/ Sap114		Sap144	Prp21	CMJ300C	
SF3a2/SF3a66/ Sap62		Sap62	Prp11	CMH102C CMN095C	
SF3a3/SF3a60/ Sap61		Sap61	Prp9	CMQ406C	
SF3b1/SF3b155/ Sap155		Prp10	Hsh155	CMB002C	
SF3b2/SF3b145/ Sap145		Sap145	Cus1	CMT357C	
SF3b3/SF3b130/ Sap130		Prp12/Sap130	Rse1/Prp12	CML103C	
SF3b4/SF3b49/ Sap49		Sap49	Hsh49	CME063C	
SF3b6/SF3b14b/ PHF5a		Ini1	Rds3	CMS014C	
SF3b14a/p14		Sap14	-	-	
SF3b10		Sap10	Ysf3	-	
U4/U6 snRNP		Prp3/90K	Cwf2/Prp3	Prp3	CMT170C
		Prp4	unnamed	Prp4	-
	Prp31	Prp31	Prp31	-	
	Snu13/15.5K	Snu13	Snu13	CMP335C/ Snu13	
	PPIH		-	-	
	U5 snRNP	Prp8/220K	Spp42/Cwf6	Prp8	CMH168C
Brr2/200K		Brr2	Brr2	CML192C	
Snu114/116K		Cwf10	Snu114	CMK208C	
Dib1/15K/Txn4a		Dim1	Dib1	CMN033C CMS018C	
40K		Cwf17/Spf38	-	-	
Prp28		Prp28	Prp28	-	
Prp6		Prp1	Prp6	-	
tri-snRNP		Snu66/SART1	Snu66	Snu66	-
	Sad1		Sad1	-	
	27K				
Sm core	SmB/B'	SmB/B'	SmB/B'	CMK022C	

	SmD1	SmD1	SmD1	CMF084C
	SmD2	SmD2	SmD2	CMN302C
	SmD3	SmD3	SmD3	CMM065C
	SmE	SmE	SmE	CMM109C CMH215C
	SmF	SmF	SmF	CMQ171C
	SmG	SmG	SmG	CMO342C
Lsm core	Lsm1		Lsm1	CMT394C
	Lsm2		Lsm2	CMB130C
	Lsm3		Lsm3	CMT262C
	Lsm4		Lsm4	CMG061C CMT545C
	Lsm5		Lsm5	CMP159C
	Lsm6		Lsm6	CMP138C
	Lsm7		Lsm7	CMP206C
	Lsm8		Lsm8	-
RES	Cwc26	Cwf26	Bud13	-
	SNIP1		Pml1	-
	Snu17/RBMX2		Snu17/Ist3	-
NTC	Prp19	Cwf8	Prp19	-
	CDC5L	Cdc5	Cef1	CMR098C
	CRN1	Cwf4	Clf1	-
	ISY1	Cwf12/Isy1	Isy1	-
	SYF1	Cwf3	Syf1	-
	GCIP-IP/Syf2	Syf2	Syf2	-
	SPF27/DAM1	Cwf7	Snt309	-
			Ntc20	-
NTR	RBM22	Cwf2	Cwc2	-
	CWC15		Cwc15	-
	RBM22		Ecm2	-
	Prp45/SKIP	Prp45	Prp45/Cwf13	-
	PRL1/pml/ Prp46	Prp5	Prp46	CMR305C
	G10/Bud31	Cwf14	Bud31/Dub31	CMG014C
First step factors	CWC25/ CCDC49		Cwc25	-
	CCDC94/ CCDC130	Cwf16	Yju2	CMN267C

Second step factors	hPRP22	Prp22	Prp22	CMG044C
	PRP17/CDC40		Prp17	-
	Prp18	Prp18	Prp18	-
	Slu7		Slu7	-
B Complex	hSNU23	Snu23	Snu23	-
	hPRP38	Prp38	Prp38	CMJ144C
			Spp381	-
B ^{act} Complex	CWC27	Cwf27	Cwc27	-
	DHX16		Prp2	-
	CWC22		Cwc22	-
	RNF113	Cwf24	Cwc24	-
			Spp2	-
Disassembly	TFIP11	Ntr1 like	Ntr1	-
			Ntr2	-
	DXH15		Prp43	CMM048C
EJC	Magoh		-	-
	eIF4A3/DDX48		Fal1	CMK028C
	Y14/RBM8		-	-
Helicases	Prp16	Prp16	Prp16	CMO384C
	Prp22/DHX8	Prp22	Prp22	CMG044C
Other	U2AF ⁶⁵	U2AF ⁵⁹	Mud2	CMS438C
	U2AF ³⁵	U2AF ²³	U2AF1	-
	SF1	SF1	Msl5	CM1292C
	UAP56		Sub2	CME073C
	SRRM2		Cwc21	-
			Cwc23	-
		Cwf1		-
		Cwf19		-
		Cyp2		-
		Cwf15		-
		Cwf11		-
	DBR1	Dbr1	Dbr1	CMK205C
SR	Rsp31			CMO009C
	SRSF2			CML202C
hnRNP	hnRNP H53			CMF163C

Table 1-2. Comparison of protein components in human, *S. cerevisiae*, *S. pombe* and *C. merolae* spliceosomes (previous pages). Homologous proteins from distinct organisms often have different names. This table compares the name and presence of splicing proteins from four species. Legend: –: the homologue was not found in that organism via extensive BLAST search. A blank space means the homologue may be present but has not been identified. The *S. cerevisiae* U1 snRNP is larger than the human U1 snRNP; * indicates proteins present in yeast U1 snRNP, but are not associated with the human U1 snRNP, rather acting as splicing factors. Table adapted from Stark et al., 2015.

backbone. Chapter four discusses a newly identified mutation of Sap49 that causes Nager syndrome, and biochemically characterizes its effect. The interaction of Sap49 with Sap145 within SF3b, and their combined interaction with RNA will also be explored in chapter four. Chapter five describes the structure of the *C. merolae* Snu13 and its binding to U4 snRNA. The structure of the RH domain of the *C. merolae* Prp8, which lacks a highly conserved β -hairpin insert, is presented in chapter six, along with a discussion of the role of the β -hairpin within the spliceosome. Finally, the ability of the RH domain of human and *S. cerevisiae* Prp8 to bind metal ions, as well as the effect of that metal binding on yeast growth are considered in chapter seven. Taken together, the work of this thesis will improve our understanding of protein-protein and protein-RNA interactions within the spliceosome.

Chapter 2

A Structural Investigation of the Spliceosome

2-1. Spliceosomal Structures

As structure and function are inherently linked, structures of spliceosomal proteins and complexes will help further our understanding of the splicing process. Until recently, the spliceosome as a whole was not amenable to structural study due to its large size and flexibility in both composition and conformation. Structural investigation was limited to NMR study of spliceosomal protein domains, X-ray crystallography of individual spliceosome components alone or bound to one or two other splicing factors, and low resolution cryo-EM. Improvements in single molecule cryo-EM techniques have recently allowed the visualization of high resolution spliceosome complexes captured throughout the splicing cycle.

2-2. Prp8

Prp8 is a highly conserved (60% identity between *S. cerevisiae* and humans) 280 kDa protein found in the tri-snRNP (Hodges et al., 1995). It cross-links to the 5' SS, 3' SS, BPS, U5 snRNA and U6 snRNA, placing it at the heart of the spliceosome. Mutations in Prp8 can compensate for mutations within the splice sites and branch site that cause splicing defects (Grainger and Beggs, 2005). It is composed of three regions: the N-terminal (NT) domain; the core, consisting of a reverse transcriptase (RT) domain with a thumb region, a linker domain, and, the endonuclease (EN) domain; as well as the C-terminal region with RNase H (RH) and Jab/MPN (Jab) domains (Galej et al., 2013; Figure 2-1). As expected from the high sequence conservation, Prp8 from human, *S. cerevisiae*, and *S. pombe* have

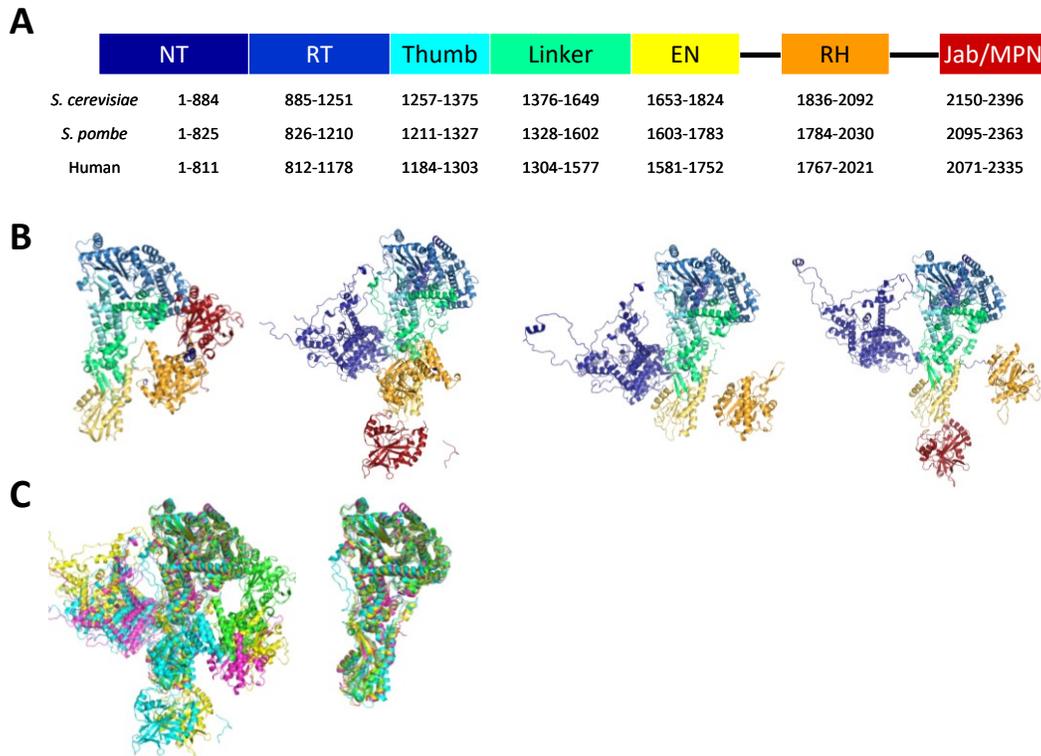


Figure 2-1. A structural overview of Prp8. A. Prp8 is composed of seven domains: N-terminal (NT), reverse transcriptase (RT), thumb/X (thumb), linker, endonuclease (EN), RNase H (RH) and Jab/MPN domains. Amino acid boundaries for each domain shown below. B. Structure of Prp8s left to right: *S. cerevisiae* X-ray structure (PDB 4I43), *S. cerevisiae* B^{act} cryo-EM structure (PDB 5GM6), *S. pombe* ILS cryo-EM structure (PDB 3JB9), and human B cryo-EM structure (PDB 5O9Z). Domains coloured as in A. C. Prp8 contains a highly conserved structural core. Left, overlay of above four full length Prp8s; right, overlay of Prp8 core. 4I43 green, 5GM6 cyan, 3JB9 magenta, 5O9Z yellow.

the same domain organization and fold.

The NT domain was not crystallized with the core and C-terminal regions of Prp8, and its structure was unknown until revealed by the cryo-EM spliceosome structures (Yan et al., 2015). This domain is composed of two lobes that are primarily α -helical, and is connected to the RT domain by a flexible linker. The RT domain contains palm, fingers, and thumb. It is similar to the maturase encoded by and associated with the self-splicing Group II intron (discussed below; Zhao &

Pyle, 2016). The EN domain contains the characteristic endonuclease fold, with five β -strands and three α -helices. The residues that bind catalytic metals in functional endonucleases, while conserved in Prp8, can be mutated to uncharged amino acids with no effect on yeast viability, suggesting that the EN domain of Prp8 does not cleave nucleic acid (Galej et al., 2013). The linker domain is mainly α -helical and connects the RT and EN domains. A flexible linker connects the RH domain to the EN domain. The RH domain contains a five-stranded β -sheet buttressed by two α -helices. A 17 amino acid insertion between β 1 and β 2 of the RNase H fold forms a β -hairpin (Pena et al., 2008; Ritchie et al., 2008; Yang et al., 2008). The RH domain also features five α -helices C-terminal to the RNase H fold. The Jab/MPN domain is connected to the RH domain via a long, flexible linker. Jab/MPN is named for its similarity to the MPN superfamily, which deubiquitinates proteins (Tran et al., 2003). However, the Jab/MPN domain of Prp8 appears to lack deubiquitination activity. Instead, it is proposed to be a mediator of protein-protein interactions. Within U5 snRNP, the Jab/MPN domain interacts tightly with Brr2 (Mozaffari-Jovin et al., 2013), and the long linker allows flexibility for Brr2 positioning.

2-3. Tri-snRNP

The cryo-EM structure of the *S. cerevisiae* tri-snRNP was solved to 3.7 Å (Nguyen et al., 2016; Figure 2-2). U4 and U6 snRNAs form two stems (I and II) separated by the 5' U4 stem loop (Figure 2-3A). The catalytic residues of U6

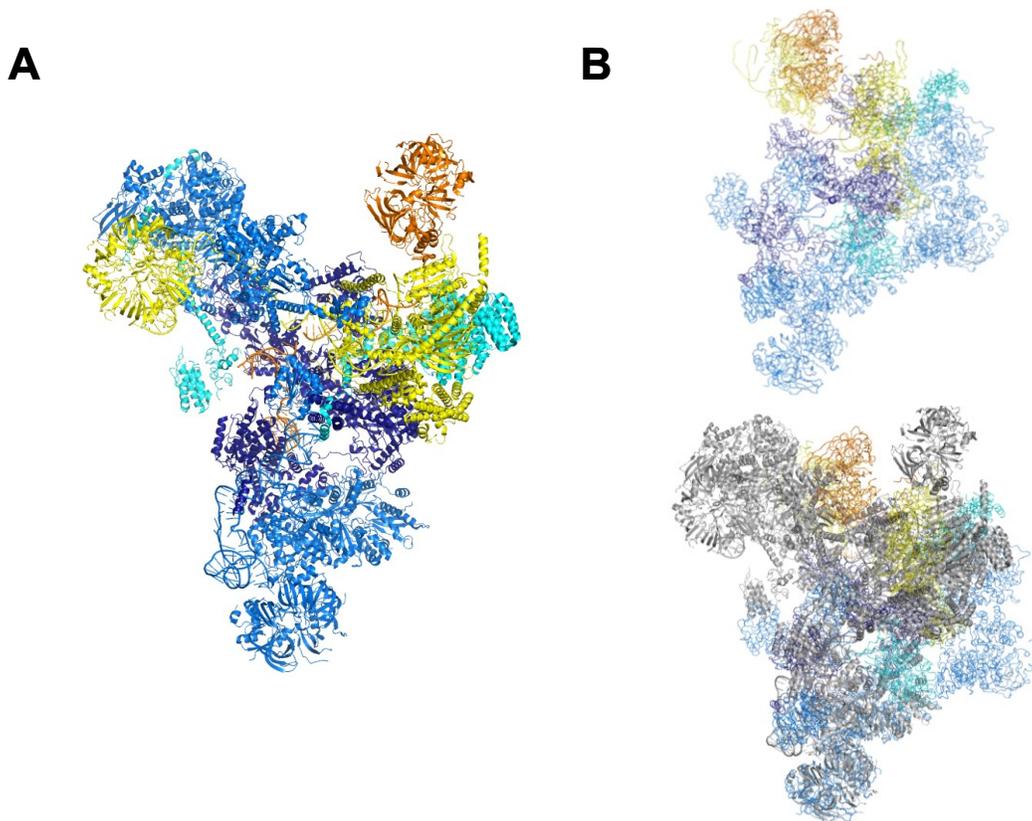


Figure 2-2. Structures of the *S. cerevisiae* and human tri-snRNP. A. Cryo-EM structure of the *S. cerevisiae* tri-snRNP (PDB 5GAN) coloured according to table 2-1. B. Cryo-EM structure of the human tri-snRNP (PDB 3JCR; top). Bottom: alignment of the human tri-snRNP (coloured) with the *S. cerevisiae* tri-snRNP (grey).

snRNA base-pair with U4 snRNA. Upstream of that duplex, U6 snRNA forms a 5' stem loop structure distinct from the internal stem loop (ISL) that will form at the active site of the spliceosome. The 5' of the ACAGA sequence that binds the 5' SS forms the tail end of this 5' hairpin. U6 snRNA also base-pairs with the top of stem loop 1 of U5 snRNA, which will base-pair with the 3' end of the 5' exon. Prp3 recognizes the U4/U6 snRNA duplex and interacts with the single stranded 3' tail of U6 snRNA (Figure 2-3B).

Prp8 scaffolds the snRNA structure. The U5-U6 snRNA interaction is

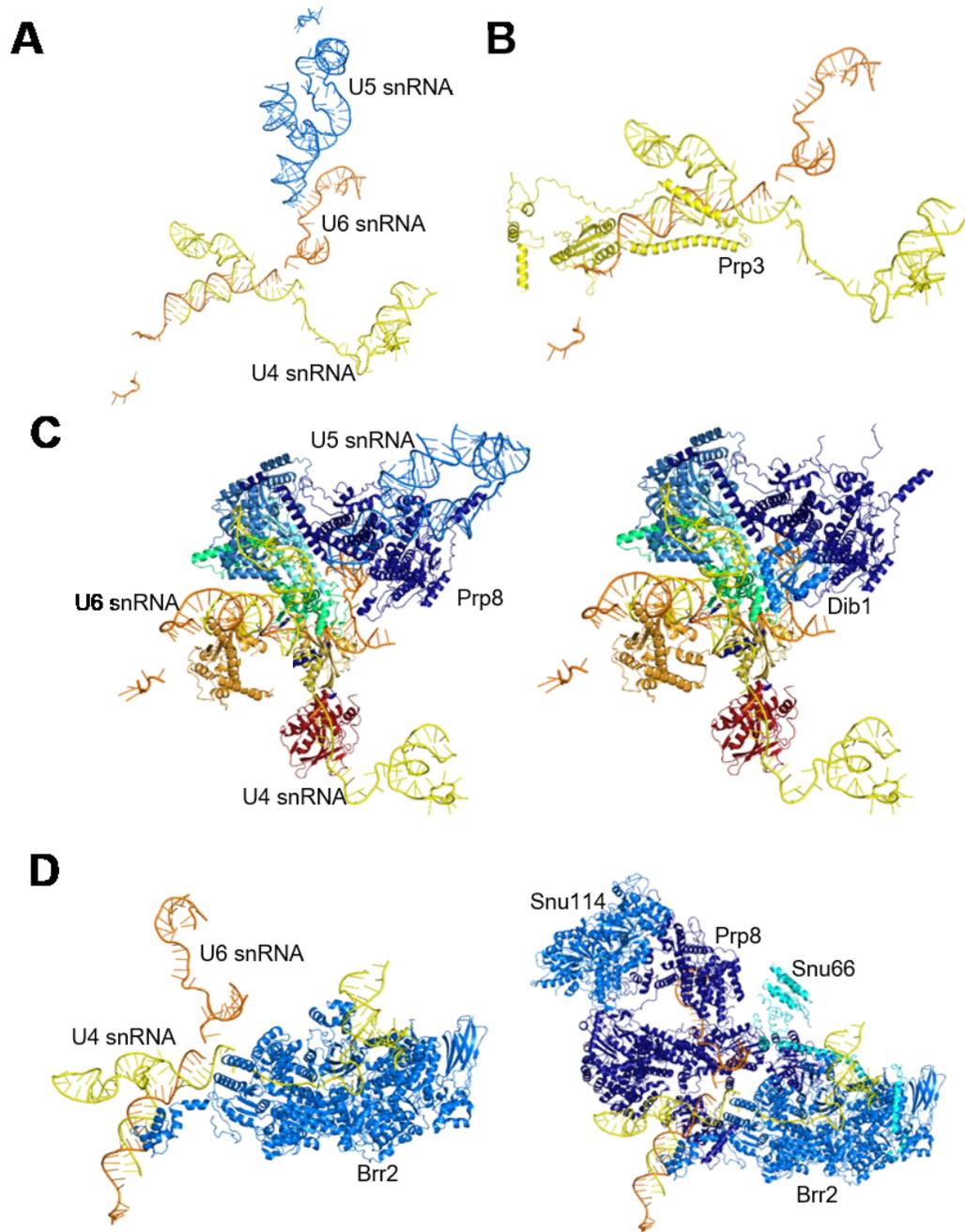


Figure 2-3. Details of the *S. cerevisiae* tri-snRNP A. RNA core of the *S. cerevisiae* tri-snRNP. U6 snRNA is extensively base-paired with U4 snRNA. Loop 1 of U5 snRNA interacts with U6 snRNA. B. Prp3 stabilizes the U4/U6 snRNA duplex. C. The RNA core of the tri-snRNP is bound by Prp8 (left) with Dib1 bridging the NT and linker domains of Prp8 (right). D. Brr2 binds U4 snRNA to unwind the U4/U6 duplex (left). Right, Snu114 modulates the activity of Brr2 via Prp8 or Snu66. Colouring is consistent with Table 2-1.

sandwiched between the NT domain and the core, with the single-stranded 5' tail of U6 snRNA passing through the cavity between these regions. The U6 snRNA hairpin is nestled between the linker and EN domain on one side with the NT domain on the other. Dib1 bridges the NT and linker domains (Figure 2-3C). The single-stranded region of U4 snRNA 3' to U4/U6 stem I passes through the active site of Brr2 (Figure 2-3D). Brr2 is bound to the Jab/MPN domain of Prp8 and is encircled by Snu66. Prp8 separates Brr2 and its modulator Snu114.

The structure of the human tri-snRNP was also solved by cryo-EM at 7 Å (Agafonov et al., 2016; Figure 2-2), and is very similar to the *S. cerevisiae* structure. The organization of the RNA core is spatially conserved. A major difference between the human and *S. cerevisiae* tri-snRNPs is the location of Brr2. In the yeast structure it is in line with the RH domain of Prp8 and bound to U4 snRNA. In the human structure the Jab/MPN domain has rotated toward the linker region of Prp8 moving Brr2 away from the U4/U6 snRNA duplex.

2-4. Spliceosome Structures

The spliceosome is a complex, dynamic multi-megadalton molecular machine. Recently, improvements in cryo-EM techniques have been used to generate high resolution models of the spliceosome in various stages of the splicing cycle. The complexes as named can likely be further subdivided, as proteins are added or released in steps that are not resolved by non-structural analyses, such as gels, stalled spliceosomes, or affinity purifications. For example, the B to B^{act}

PDB	6G90	5GAN	5ZWM	5ZWO	5NRL	5LQW	5GM6	5GMK	5LJ5	5LJ3	5MQ0	5WSG	6EXN	6BK8	5YLZ	5Y88	3JB9
Group	Nagai	Nagai	Shi	Shi	Nagai	Luhrmann	Shi	Shi	Nagai	Nagai	Nagai	Shi	Nagai	Zhao	Shi	Shi	Shi (pombe)
Complex	A	Tri- snRNP	Pre-B	B	B	Bact	Bact	C	C	C	C*	C*	P	P	P	ILS	ILS
Resolution	4	3.7	3.4	3.9	7.2	5.8	3.5	3.4	10	3.8	4.2	4	3.7	3.3	3.6	3.5	3.6
Intron	I		G	G	I	9	MN	BMN	EI	EI	EI3	BNM b	EI	ei	E	E	OQ
U1																	
U1A/Mud1	A																
U1-70k/Snp1	B																
U1C/Yhc1	C																
Prp39	D																
Prp42	E																
Nam8	F																
Snu56	G																
Luc7	H																
Snu71	J																

Sm	bdefghi																
U1 snRNA	1																
U2																	
Msl1	Y		p	p	Y			a	Y	Y	Y	X	Y	s	p	p	k
Lea1	W		o	o	W			b	W	W	w	Y	W	r	o	o	j
Sm	stuvwxy		hijklmn	hijklmn	svtuwxy			esuvwxyz	klmnpqr	klmnpqr	knpqrlm	FGHKUVW	klmnpqr	klmnopq	hijklmn	hijklmn	Zbflmno
Hsh155	O		1	1	O	Q	G										
Rse1	P		3	3	P	X	F										
Cus1	Q		2	2	Q		H										
Hsh49	R		4	4	R		e										
Rds3	S		5	5	S	Y	J										
Ysf3	Z		6	6	Z	Z	K										
Prp9	T		u	u	T												
Prp11	U		v	v	U		I										
Prp21	V		w	w	V												
U2 snRNA	2		H	H	2	2	L	L	Z	Z	2	L	2	2	F	F	P
U4/U6																	
Snu13		K	M	M	K												

Prp31		F	L	L	F												
Prp3		G	J	J	G												
Prp4		H	K	K	H												
Sm		klmnpqr	PQRSTUVWXYZ	PQRSTUVWXYZ	klmnpqr												
LSm		2345678	qrstvxyz	qrstxyz	a3joz78												
U4 snRNA		V	I	I	4												
U6 snRNA		W	F	F	6	6	E	E	V	V	6	E	6	6	D	D	N
U5																	
Prp8		A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Brr2		B	D	D	B	C	B		B								
Snu114		C	C	C	C	B	C	C	C	C	C	C	C	B	C	C	B
Dib1		D	E	E	D												
Sm		bdefghj	abcdefg	abcdefg	bdhiefg	bdefghj	kihjlmg	ghijklm	bdefghj	bdefghj	bdefghj	kihjlmg	bdefghj	abcdfgh	abcdefg	abcdefg	DEFGHIJ
U5 snRNA		U	B	B	5	5	D	D	U	U	5	D	5	5	B	B	C
Tri-snRNP																	
Prp6		J	N	N	J												
Snu66		E	O	O	E												
RES																	

Bud13			Y	Y		L	W										
Pml1			Z	Z		N	U										
Ist3/Snu17			X	X		J	V	H									
NTC																	
Prp19							opqr	opqr	tuvw		tuvw	opqr	tuvw	uvwx	qrst	qrst	STUV
Clf1						R	d	d	S	S	S	d	S	T	I	I	R
Syf1						P	v		T	T	T	v	T	U	H	H	r
Syf2							f	l			y	l	y	R	K	K	
Cef1						W	c	c	O	O	O	c	O	S	J	J	W
Snt309							t	t	s		s	t	s	Y	G	G	i
Isy1									G	G							
Cwc23																T	
NTR																	
Cwc2						F	R	R	M	M	M	R	M	G	N	N	Y
Cwc15							S	S	P	P	P	S	P	H	P	P	a
Ecm2						D	Q	Q	N	N	N	Q	N	M	M		
Prp45						M	P	P	K	K	K	O	K	E	Q	Q	M
Prp46						K	O	O	J	J	J	P	J	D	O	O	K

Bud31						E	T	T	L	L	L	T	L	I	L	L	e
B																	
Snu23				W	L												
Prp38				0	M												
Spp381				9	N												
Bact																	
Cwc27							b										
Prp2						O	Y										
Cwc22						H	Z	Z	H	H	H	Z	H	L	S		
Cwc24							a										
Spp2																	
1st Step																	
Cwc25								G	F								
Yju2								F	D				D			R	
2nd Step																	
Prp22										v		V	P	W			
Prp17							n	n			o	n	o	M	T	S	g
Prp18											f	a		U			

Slu7											c		c	O	V		
Other																	
Cwc21						X	J	R	R	R	J	R	K	R			
Prp16								Q			e						
Slu17												c					
Stt11													F				
Ntr1																	U
Ntr2																	V
Prp43																	W
Pombe																	
Cwf1																	K
Cwf17																	L
Cwf19																	c
PPIL																	d
Cwf15																	h
Cwf11																	X

Table 2-1. Protein and RNA components in the structures of the yeast spliceosome (preceding pages). Summary of the composition of the high resolution cryo-EM structures of *S. cerevisiae* and *S. pombe* spliceosomal complexes whose coordinates have been released. Colouring by sub-complex consistent throughout.

transition features the loss of U1 and U4 snRNPs in two separate steps, but the term B complex includes spliceosomes with and without U1 snRNP (Brow, 2002).

2-5. Yeast Structures

The first spliceosome structure solved was the intron-lariat spliceosome of *S. pombe* (Yan et al., 2015). It was known that the majority of spliceosomes purified from *S. pombe* extracts are post-catalytic, containing U2, U6, and U5 snRNPs along with a lariat intron (Chen et al., 2014). While this over-abundance of non-splicing spliceosomes wreaked havoc on attempts to develop an *in vitro* splicing assay in *S. pombe* extract, it allowed for the purification of sufficient homogenous spliceosomes for high resolution cryo-EM. This structure is significant because it was proof of concept: spliceosomes are amenable to structure determination by cryo-EM.

Structures of *S. cerevisiae* spliceosomes quickly followed, allowing the visualization of the splicing cycle from the assembly of the A complex to the post-catalytic intron-lariat spliceosome. Components of each spliceosome structure is summarized in Table 2-1.

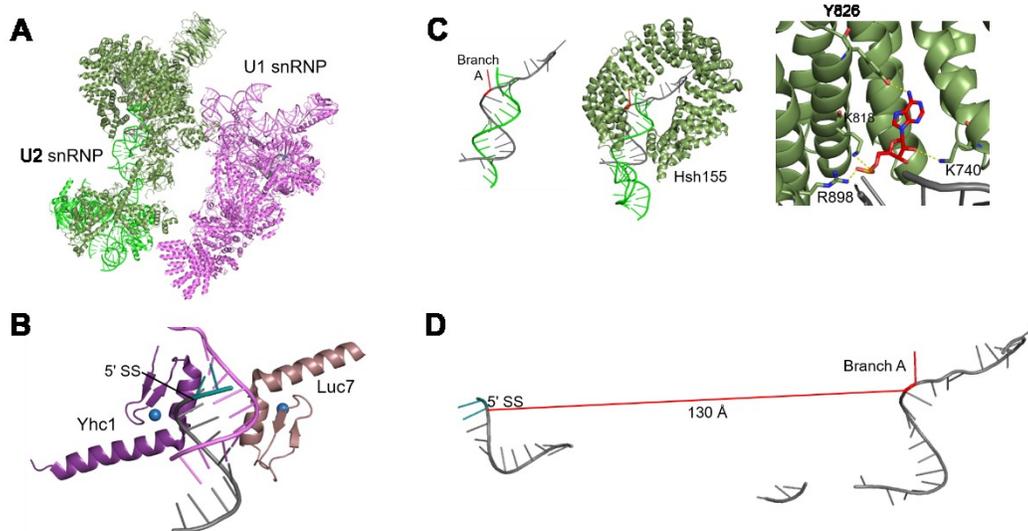


Figure 2-4. *S. cerevisiae* A complex. A. Overview of the A complex. U1 snRNP in violet, U2 snRNP in green (PDB 6G90). B. Yhc1 and Luc7 are both zinc fingers that interact with the 5' SS/U1 snRNP duplex. B. The branch A is bulged out of the duplex between the U2 snRNA and branch region (left). The Branch A is bound by Hsh155 (middle) and interacts with K740, K818, Tyr826, and R898. D. The 5' SS and branch A are separated by about 130 Å. Colouring is consistent with Table 2-1.

2-5a. A Complex

The structure of the *S. cerevisiae* A complex shows both U1 and U2 snRNPs bound to the pre-mRNA (Figure 2-4A; Plaschka et al., 2018). The resolution of the U1 snRNP is 4.0 Å, while the resolution of the U2 snRNP ranges from 4.9-10.4 Å.

The 5' end of the U1 snRNA base-pairs with the 5' SS. This duplex is stabilized by Yhc1 interacting with the pre-mRNA, and Luc7 interacting with the U1 snRNA (Figure 2-4B). Both these proteins share a similar fold: a zinc finger buttressed by a long α -helix that contacts the RNA backbone. Yhc1 also interacts with the junction between stem loops 3-3 and 3-4 of U1 snRNA. As in the *S. cerevisiae* U1 snRNP structure (Li et al., 2017), the yeast specific components of the U1 snRNP, with the exception of Luc7, bind in a cluster at the bottom of the

snRNP.

U2 snRNA forms an imperfect duplex with the branch site, with the branch A excluded from the helix. SF3b binds to this region. The HEAT repeats of Hsh155 encircle the duplex, and hydrogen-bond to the branch A: the N6 amine contacts Tyr826, while the 2' hydroxyl hydrogen-bonds to Lys740, and the backbone phosphate interacts with Lys818 and Arg898 (Figure 2-4C). A region of the intron upstream of the branch interacts with the first RRM of Hsh49, consistent with cross-linking studies (Schneider et al., 2015).

The SF3a complex also binds the helix between the intron and U2 snRNA, and is positioned near the U2 Sm core and U2 snRNA. Prp9 and Prp11 are positioned alongside the duplex while Prp21 links Prp9 and Prp11.

U1 and U2 snRNP form two hemispheres that come together to create the A complex. The branch A and 5' SS are separated by ~ 130 Å, and each sequestered by protein factors (Figure 2-4D). The interface between the two snRNPs is minimal, with contacts forming between Lea1 in U2 snRNP and Prp39 in the U1 snRNP. From this structure, it appears that recognition of the 5' SS and branch site are two independent events within the A complex. The minimal interface between U1 and U2 snRNPs may facilitate U1 snRNP dissociation from the spliceosome following the addition of the tri-snRNP.

2-5b. B Complex

The structure of the B complex was captured after U1 snRNP had

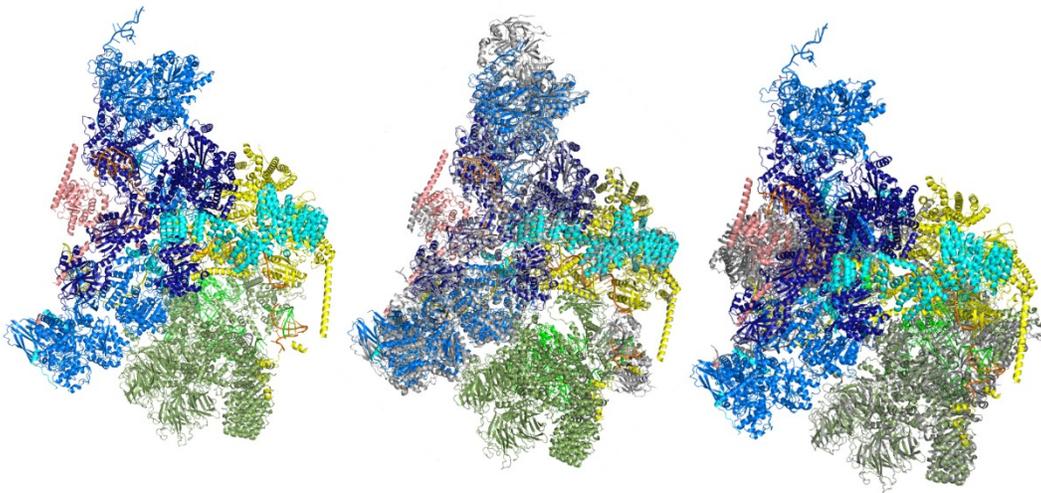


Figure 2-5. *S. cerevisiae* B complex. Cryo-EM structure of the *S. cerevisiae* B complex (PDB 5NRL; left) and overlay with the tri-snRNP structure (PDB 5GAN, grey; right). Colouring is consistent with Table 2-1.

dissociated from the spliceosome, but before Brr2 unwound the U4/U6 snRNA duplex (Plaschka et al., 2017). It is composed of U2 snRNP and the tri-snRNP, along with three B complex-specific proteins: Snu23, Prp38, and Spp381. The resolution ranges from 3.6 to 17.2 Å, with an average resolution of 7.2 Å (Figure 2-5).

Alignment of Hsh155 from the A and B complexes reveals little change in the location and arrangement of the U2 snRNP. The positions U4 snRNA and Brr2 overlap with that of the U1 snRNP.

The structure of the U4/U6 and U5 snRNAs is very similar to that in the tri-snRNP (Nguyen et al., 2016; Figure 2-6A). U4 and U6 snRNA are still base-paired to each other with Brr2 ready to unwind the duplex (Figure 2-6B). Brr2 contacts the Jab/MPN domain of Prp8. Snu114, which activates Brr2, is on the opposite side

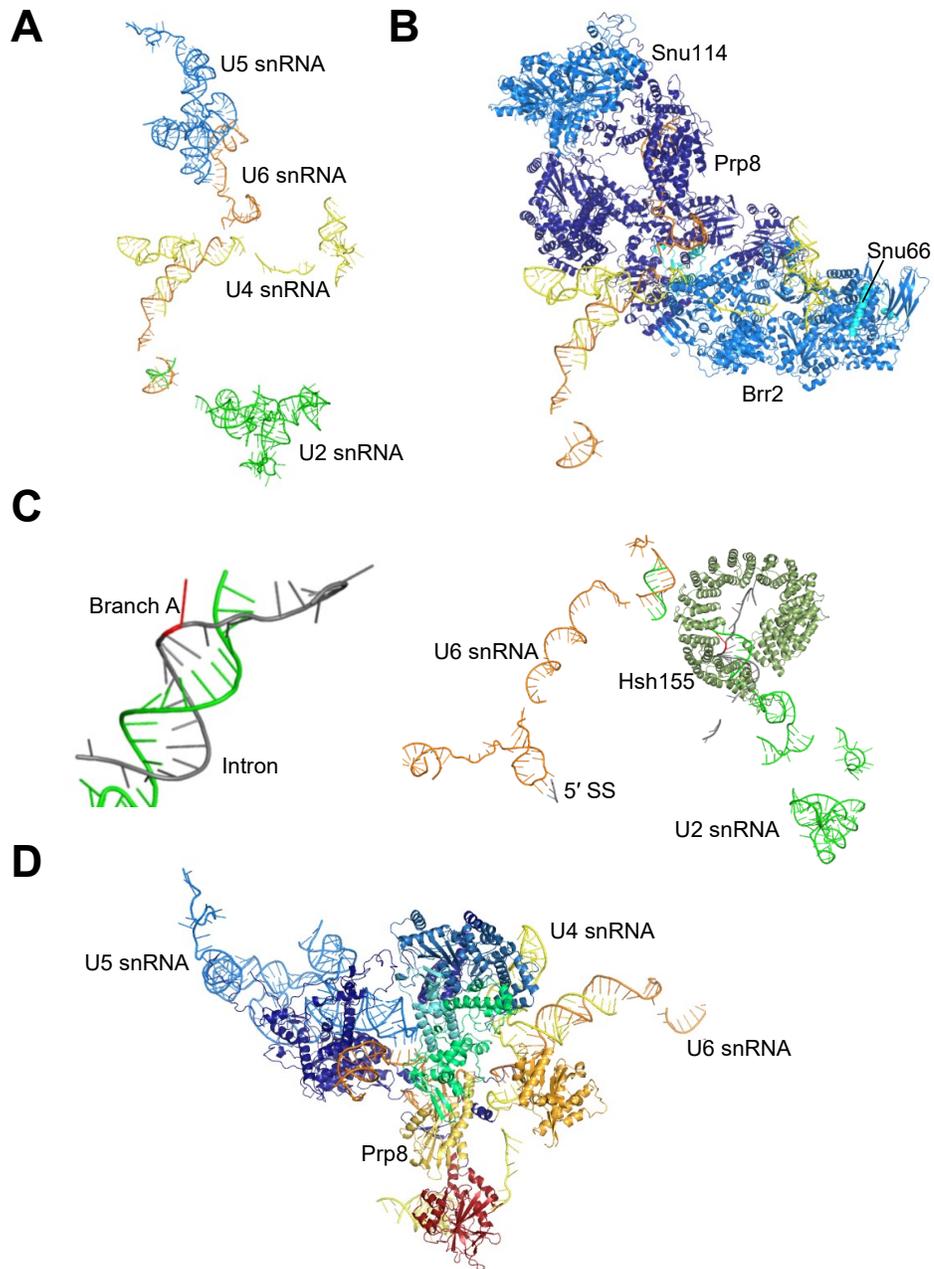


Figure 2-6. Details of the *S. cerevisiae* B complex. A. snRNA organization in the B complex. U4 and U6 snRNAs remain base-paired. U6 snRNA begins to base-pair with U2 snRNA. B. Brr2 bound to U4 snRNA, ready to unwind the U4/U6 duplex. C. The BPS forms an imperfect duplex with U2 snRNA to bulge out the branch A and select it as the nucleophile for the first step of splicing (left). Right, the branch A and 5' SS are separated by over 130 Å. Hsh155 binds the branch A to further sequester it. D. Prp8 scaffolds the RNA core. Colouring is consistent with Table 2-1.

of Prp8. The activating signal from Snu114 may be transmitted to Brr2 via Prp8 or Snu66.

The tri-snRNP binds the A complex such that Brr2 is positioned alongside U2 snRNP. The 3' end of U6 snRNA, just upstream of the Lsm binding site, base-pairs with the 5' end of U2 snRNA to form U2/U6 helix II. Binding of the tri-snRNP to U2 snRNP is stabilized by interactions between Prp3 and Hsh155 on one end, and Brr2 with Cus1 on the other end.

While the exons are not visible in this structure, three nucleotides just downstream of the 5' SS are seen base-pairing with U6 snRNA. As in the A complex, the branch point sequence (BPS) UACUAAAC is duplexed with the GUAGUA of U2 snRNA, and the branch A is bulged out (Figure 2-6C). The two reactive sites of the pre-mRNA continue to be physically separated while the spliceosome assembles to prevent premature splicing using potentially incorrect splice sites. In addition to being separated from the 5' SS, the branch A is associated with SF3b; it is nestled in the HEAT repeats of Hsh155. U2 snRNA stem IIa is just upstream of the BPS/U2 snRNA duplex.

The single-stranded region of U6 snRNA is sandwiched between the N-terminal and core domains of Prp8. The exon binding loop of U5 snRNA lies alongside the N-terminal domain and contacts U6 snRNA. The U4 stem and U4/U6 duplex lie on the opposite side of Prp8, along the RT domain. The U5 stem loop 1 is positioned between the two halves of the N-terminal domain (Figure 2-6D).

U6 snRNA is bound to the 5' SS. The 3' end of the final 5' SS binding site on U6

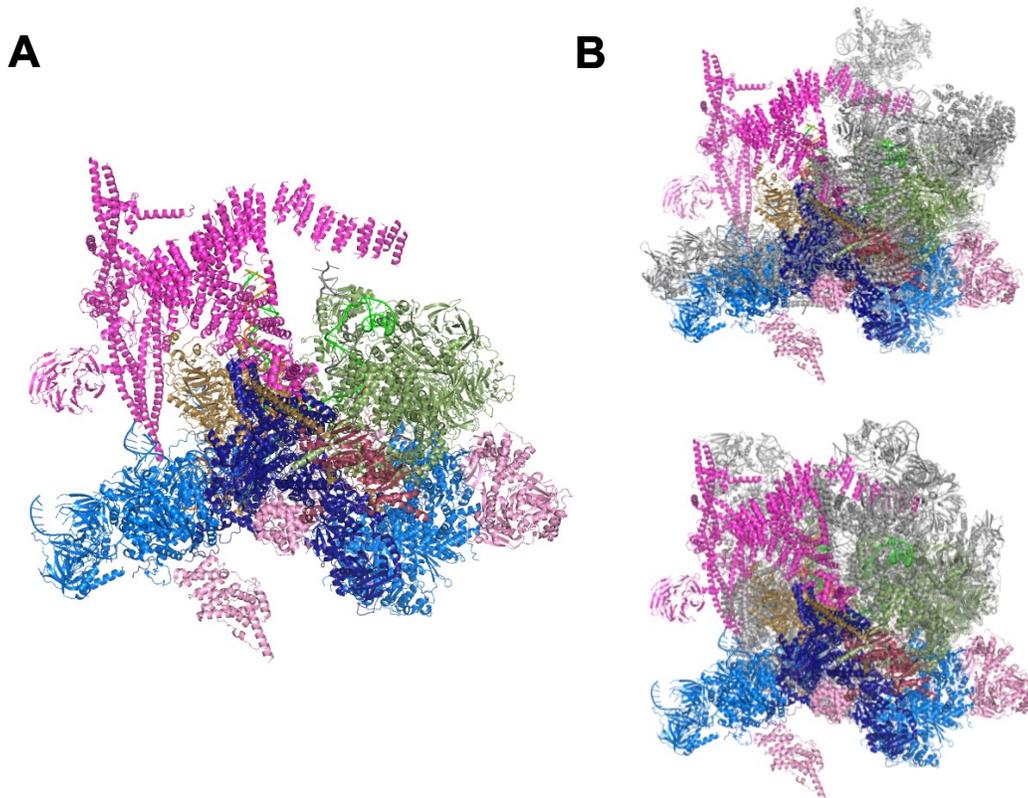


Figure 2-7. *S. cerevisiae* B^{act} complex. A. Cryo-EM structure of the *S. cerevisiae* B^{act} complex (PDB 5GM6). B. Comparison of *S. cerevisiae* B^{act} complex with the B complex (grey). Top: aligned via Prp8. Bottom: aligned via Hsh155. Colours are consistent with Table 2-1.

snRNA (ACAGAGA) is in a stem loop, and shortly downstream it is bound to U4 snRNA, keeping it from binding the 5' SS. Instead, the 5' SS is base-paired with the loop of this stem. The stem loop is stabilized by the N-terminal domain of Prp8, while the linker domain inserts into the base of the stem. A β hairpin in the linker between the RH and Jab/MPN domains lies alongside the stem loop.

2-5c. B^{act} Complex

Major structural rearrangements occur during the transition from the B to B^{act} complex (Figure 2-7). The U4 snRNP is lost, as are Dib1, and the three B-specific proteins Snu23, Prp38, and Spp381. The RES and NTC join the spliceosome, along with a number of other proteins, including Cwc27, Cwc22, Cwc24, Spp2 and Prp2. The average resolution of the B^{act} complex is 3.5 Å, with the core at 2.8-3.2 Å resolution (Yan et al., 2016).

Brr2 unwinds the U4/U6 snRNA duplex to release U4 snRNA, its associated proteins, and the U6 Lsm core, which allows the U6 ISL to form. U2 and U6 snRNA base-pair to form U2/U6 helices Ia and Ib and II. The U6 snRNA ACAGAGA stem loop straightens out to base-pair with the 5' SS. A three helix junction is formed by the U2 snRNA/BPS duplex, the duplex between U6 snRNA and the 5' SS, and the U6 ISL. The catalytic center of the spliceosome is composed of these snRNAs (Figure 2-8A). However, an N-terminal region of Prp11 lies across the active site, and blocks binding of the branch A. The exon-binding loop 1 of U5 snRNA is freed to interact with the 5' exon. This brings U5 and U6 snRNA into close proximity. The first three nucleotides of intron are single-stranded, leaving the scissile phosphate available for nucleophilic attack. The base-pairing with U5 snRNA positions the 5' SS for the first step of splicing. However, the branch A is still about 50 Å away. Hsh155 remains positioned between it and the 5' SS (Figure 2-8B).

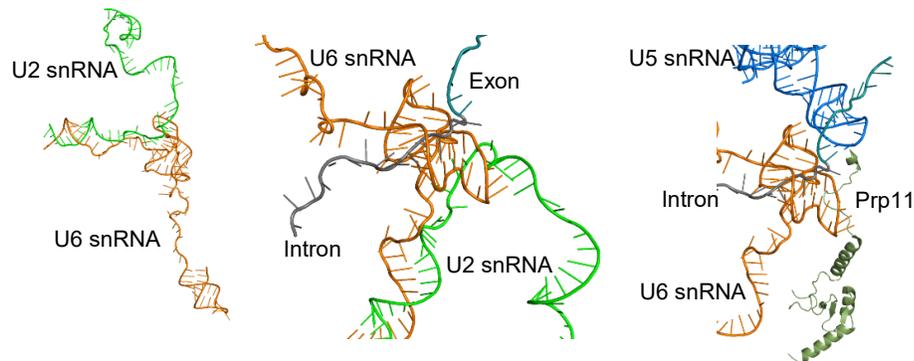
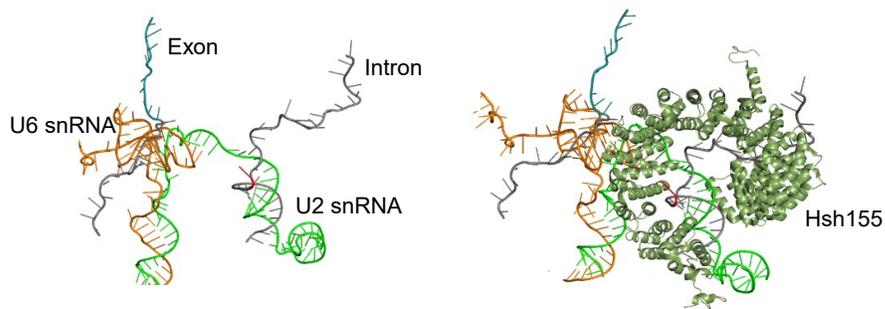
A**B**

Figure 2-8. RNA core of the *S. cerevisiae* B^{act} spliceosome. A. RNA at the core of the spliceosome. Left, U2 and U6 snRNA base-pair to form helices Ia and Ib and II; U6 snRNA forms the ISL. Center, a three-helix junction forms from the U2/U6 helix, the U6 ISL and the 5' SS base-paired to U6 snRNA. Right, Prp11 blocks binding of the branch A in the active site. B. The 5' SS and branch A remain separated in space (left) and by Hsh155 (right). Colouring is consistent with Table 2-1.

Prp8 continues to scaffold the RNA core of the spliceosome (Figure 2-9A).

The NT domain of Prp8 rotates down toward the RNA core relative to the B complex. The formation of the U6 ISL pulls the 5' end of U6 snRNA that was held between the NT and core domains of Prp8 to lie alongside the NT domain. The 5' exon replaces U6 snRNA in Prp8. The position of U5 snRNA does not change, with stem loop 1 remaining in the cleft in the N-terminal domain.

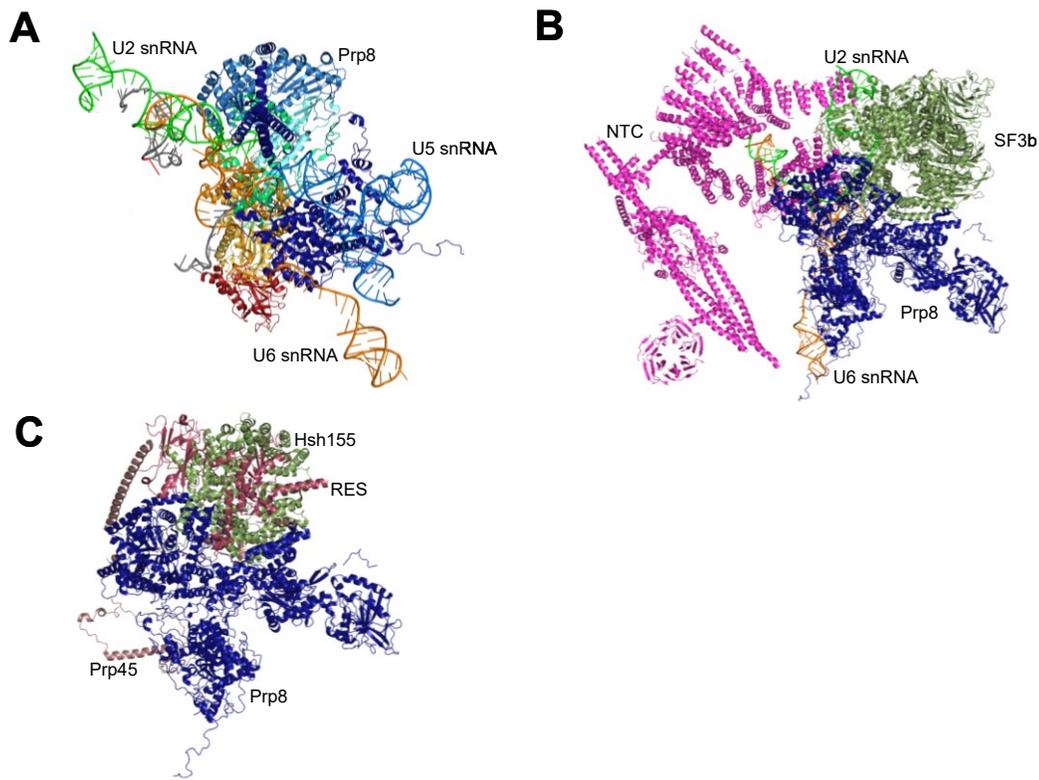


Figure 2-9. Protein components of the *S. cerevisiae* B^{act} spliceosome. A. Prp8 binds the RNA core. B. NTC (magenta) binds the spliceosome opposite SF3b. C. RES binds the spliceosome between Hsh155 and Prp8, and is stabilized by Prp45. Colours consistent with Table 2-1.

The nineteen complex (NTC) binds the spliceosome alongside the U2/U6 snRNA helix II, on the opposite side of the spliceosome from SF3b. The bulge between U2/U6 helix II and helices Ia and Ib is stabilized by Clf1 and Syf2 of the NTC (Figure 2-9B).

The RES complex binds with Bud13 interacting with the linker region of Prp8. Bud13 and Snu17 bind between Prp8 and Hsh155. Pml1 interacts with a long helix of Prp45. The RES, along with Cwc24, may help stabilize the association of SF3b with the spliceosome to prevent premature dissociation (Bao et al., 2017a;

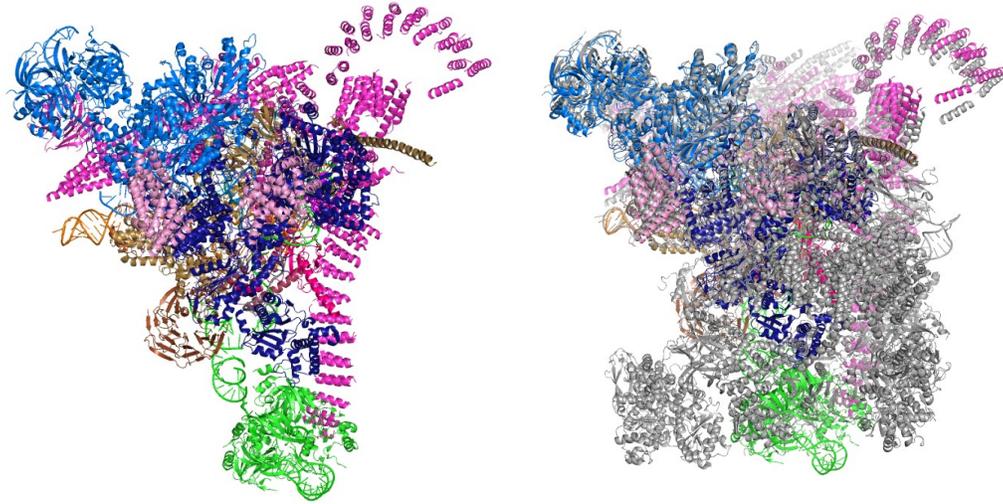


Figure 2-10. *S. cerevisiae* C complex. Cryo-EM structure of the *S. cerevisiae* C complex (PDB 5GMK; left) and overlaid with the B^{act} complex (PDB 5GM6, grey; right). Colours consistent with Table 2-1.

Figure 2-9C).

The spliceosome must undergo further rearrangements into B* complex before the first step of splicing occurs. Currently, there is no structure for the B* complex. To form this complex, SF3a and SF3b are released by Prp2, exposing the branch A; and, the spliceosome rearranges to bring the BPS to the 5' SS in the active site. The first step splicing factors Yju2 and Cw25 bind.

2-5d. C Complex

The first step of splicing occurs in the B* complex, which gives rise to the C complex (Figure 2-10). As the C complex is formed upon branching of the intron, it hints at the arrangement of the B* complex. The cryo-EM structure of C complex was captured directly after the first step of splicing, as evidenced by the presence

of the lariat intron and first step factors Yju2 and Cwc25. The C complex was solved with an average resolution of 3.4 Å, and 2.9 Å at the center of the spliceosome (Wan et al., 2016).

SF3a and SF3b dissociate prior to the first step of splicing, which allows for the rearrangement of the spliceosome and the binding of Yju2 and Cwc25. Dissociation of Prp11 with the SF3a complex opens up space alongside the U6 ISL for the U2 snRNA/BPS duplex. Removal of Prp11 also allows Cef1 to stabilize the active site created by U2 and U6 snRNAs (Figure 2-11A). Removal of Hsh155 releases the branch A, and allows the RH domain of Prp8 to swing up alongside the BPS/U2 snRNA duplex. The U2 snRNA/BPS duplex rotates around U2 snRNA to bring the branch A into proximity with the 5' SS, which remains in place in the active site. The adenine base of the branch A is hydrogen bonded to the uridine two bases 5' to it, which extends its 2' hydroxyl toward the 5' SS. The branched 5' SS product of the first step is visible in the C complex structure (Figure 2-11B).

U5 and U6 snRNAs do not change position between the B^{act} and C complexes: U5 snRNA remains bound to the 5' exon; U6 snRNA maintains the U6 ISL, its interactions with the 5' end of the intron, and its base-pairing with U2 snRNA. U2 snRNA adopts the catalytically competent I_{lc} stem. Modeled are two metal ions coordinated by the backbones of U80 and the catalytic triad (A59 G60 C61) of U6 snRNA (Figure 2-11C). U80 is bulged out of the U6 ISL and base stacks on G52 and A53. The catalytic triad is base-paired with GCU (21-23) of U2 snRNA.

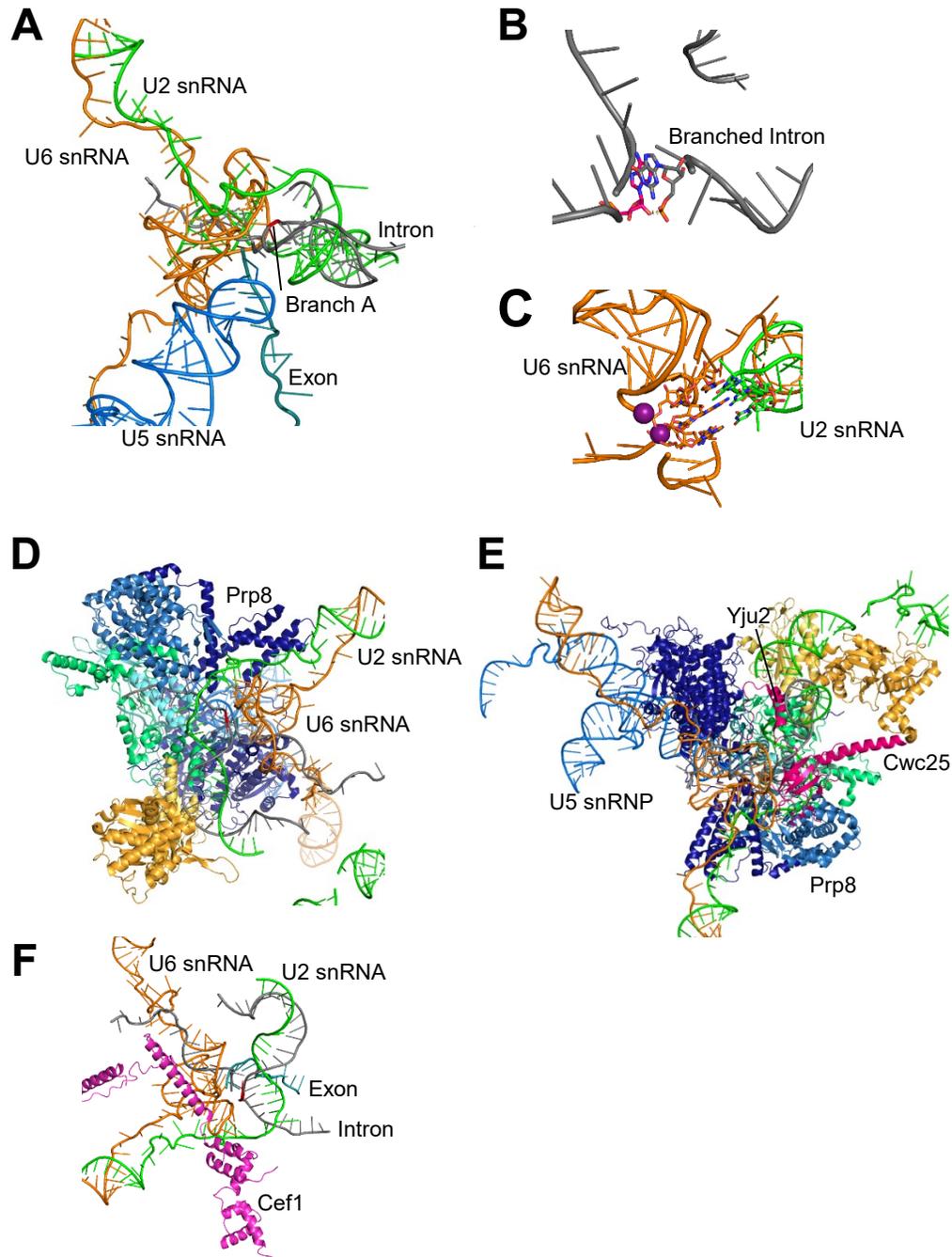


Figure 2-11. Details of the *S. cerevisiae* C complex. A. RNA organization shows rearrangement to bring the branch A into the active site. B. The first step of splicing has occurred. C. The active site of the spliceosome features two metal ions. D. Prp8 binds the RNA core. E. First step factors, Yju2 and Cwc25, are bound to the spliceosome, suggesting this complex was captured directly after branching. F. The RNA core of the spliceosome is stabilized by Cef1. Colours are consistent with Table 2-1.

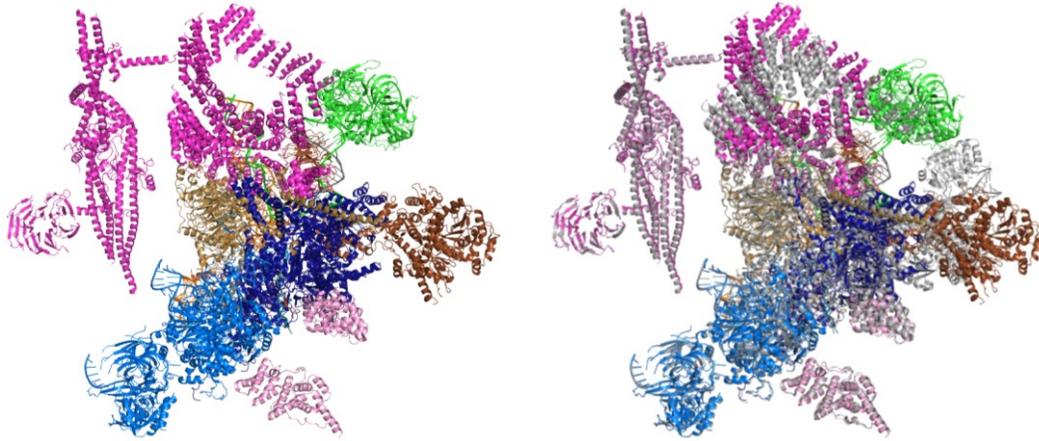


Figure 2-12. *S. cerevisiae* C* complex. Cryo-EM structure of the *S. cerevisiae* C* complex (PDB 5WSG; left) and comparison with the C complex (PDB 5GMK, grey; right).

The catalytic core is stabilized by numerous proteins. U2 and U5 snRNAs remain nestled against Prp8, with the U6 ISL in the NT domain (Figure 2-11D). Yju2 binds between the tail of U6 snRNAs following helices Ia and Ib, and the U2/BPS duplex. Its N-terminus pokes into the RNA core. Cwc25 stabilizes the U2/BPS duplex (Figure 2-11E). Isy1 lies between the BPS/U2 snRNA duplex and the helix formed by U6 snRNA and the 5' SS. Cef1 stabilizes the U6 ISL (Figure 2-11F).

2-5e. C* Complex

The C* complex is very similar to the C complex, and was solved at 4 Å resolution (Yan et al., 2017; Figure 2-12). It shows the movement of the branched intron out of the active site, and the positioning of the second step substrates for catalysis (Figure 2-13A). The first step splicing factors Yju2 and Cwc25 dissociate, along with Snu17. Second step factors Prp17, Prp18, Prp22 and Slu7 bind.

The U2/U6 snRNA catalytic core maintains its conformation in the C* complex. A rotation around A30 of U2 snRNA moves the branched A and 5' end of the intron out of the active site by 15-20 Å. The BPS/U2 snRNA duplex, and the rest of U2 snRNA, flips up to sit against the RH of Prp8. Prp17 binds alongside the distal part of the intron/U2 helix (Figure 2-13B). The β hairpin in the RH inserts into the minor groove of the BPS/U2 snRNA duplex, with the tip extending toward the U6 ISL (Figure 2-13C). The base-pairing between U6 snRNA and the 5' end of the intron is maintained, and moves with the branched A out of the core of the spliceosome. The 5' exon remains base-paired to loop 1 of U5 snRNA. Its free 3' hydroxyl is positioned to attack the 3' SS, which has moved into the active site. The 3' SS and 3' exon do not appear to be base-paired to snRNA, but are held in place by Prp8 (Figure 2-13D). The 3' end of the intron is bound between the RH and linker, and the 3' exon is bound between the RT and thumb domains and linker. The NT domain, core of Prp8 and Cef1 do not change between the C and C* complexes. The RH domain of Prp8 shifts up to interact with the BPS/U2 snRNA duplex.

2-5f. P Complex

The P complex arises from exon ligation in the C* complex. The cryo-EM structure has a resolution of 3.7 Å (Wilkinson et al., 2017 Figure 2-14). The RNA core and protein organization of the post-catalytic spliceosome is very similar to that of the C* spliceosome. The exons have been ligated, but are still held by U5

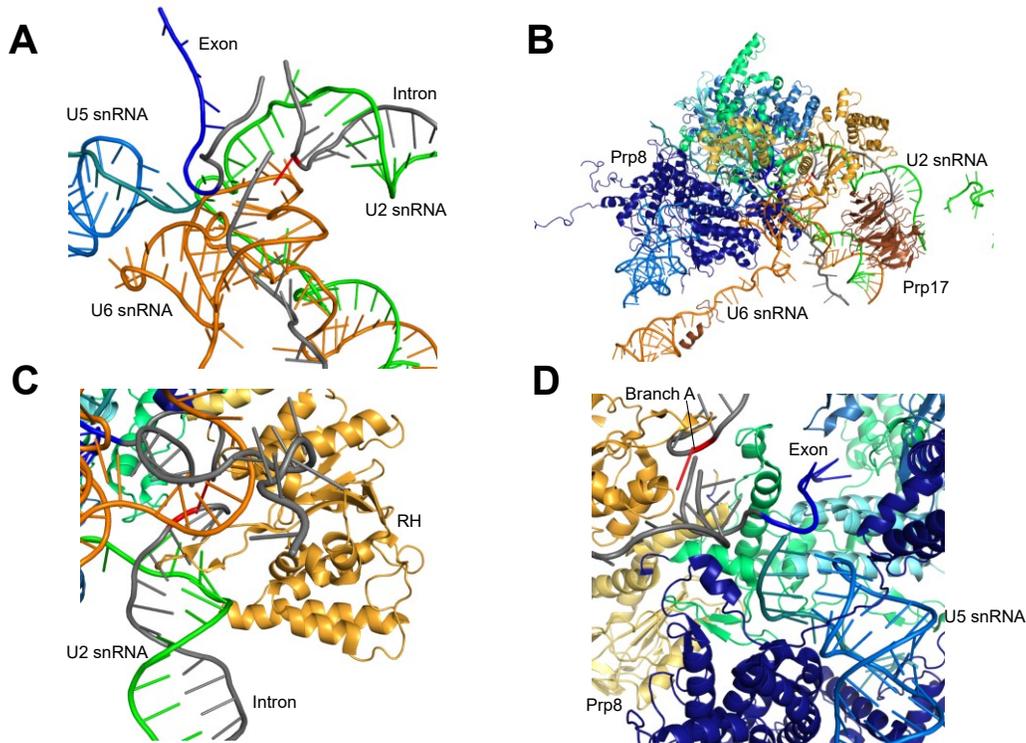


Figure 2-13. Details of the *S. cerevisiae* C* spliceosome. A. RNA core of the spliceosome. The 5' exon is base-paired to U5 snRNA while the 5' SS is base-paired with U6 snRNA. The branch A has rotated out of the active site to make way for the 3' exon. B. Prp17 stabilizes the RNA core. C. The β hairpin of Prp8's RH inserts into the minor groove of the BPS/U2 snRNA duplex. D. The 3' exon is bound by the linker of Prp8.

snRNA (5' exon) and Prp8 (3' exon). The U2 snRNA duplexed to the BPS is still separated from U6 snRNA by the hairpin in the RH domain of Prp8.

The last base of the intron (G) interacts with the first base of the intron (G) which stacks on U6 snRNA A51. The branch A interacts with the penultimate nucleotide (A) of the intron, which stacks on the following G. These interactions project the 3' hydroxyl of the 3' SS into the active site, suggesting that this is the position of the intron during exon ligation (Figure 2-15A). These interactions are not observed in the C* complex.

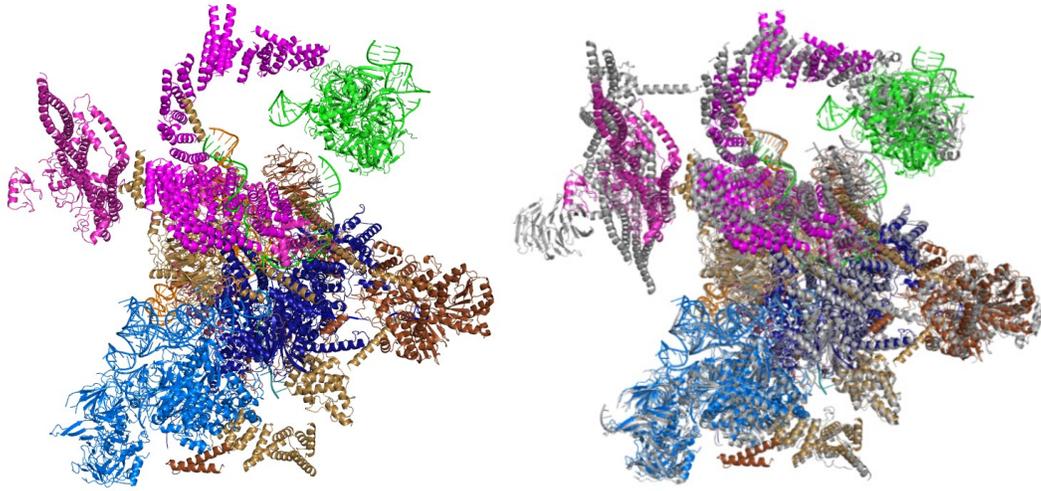


Figure 2-14. *S. cerevisiae* P complex. A. Cryo-EM structure of the *S. cerevisiae* P complex (PDB 6EXN; left) compared to the C* complex (PDB 5WSG, grey; right).

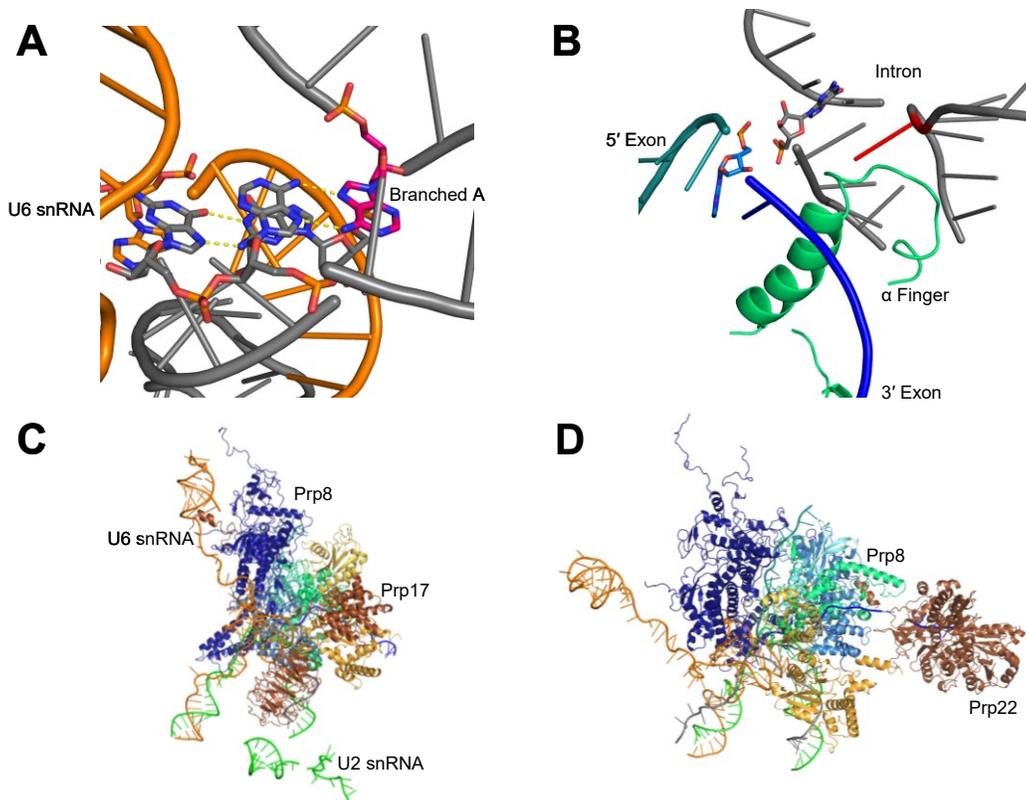


Figure 2-15. Details of the *S. cerevisiae* P complex. A. The branch A is held in place by non-canonical base-pairing with the A at the 3' SS. That A stacks on the final G of the intron, which itself interacts with the first G of the 5' SS. B. The 3' SS wraps around an α -helix in the linker of Prp8 to expose the scissile phosphate for exon ligation. C. Prp8 and Prp17 bind the RNA. D. Prp22 is positioned on the 3' exon for release of spliced mRNA.

The 3' SS and 3' exon are separated by the α -finger (1565-1610) from the linker region of Prp8. In their uncleaved state, this interaction would open the scissile phosphate to expose it for nucleophilic attack (Figure 2-15B). The β -finger of the RH domain interacts with the branch duplex on the other side of the RNA to stabilize recognition of the 3' SS.

The second step factors Prp18 and Slu7 do not interact directly with the

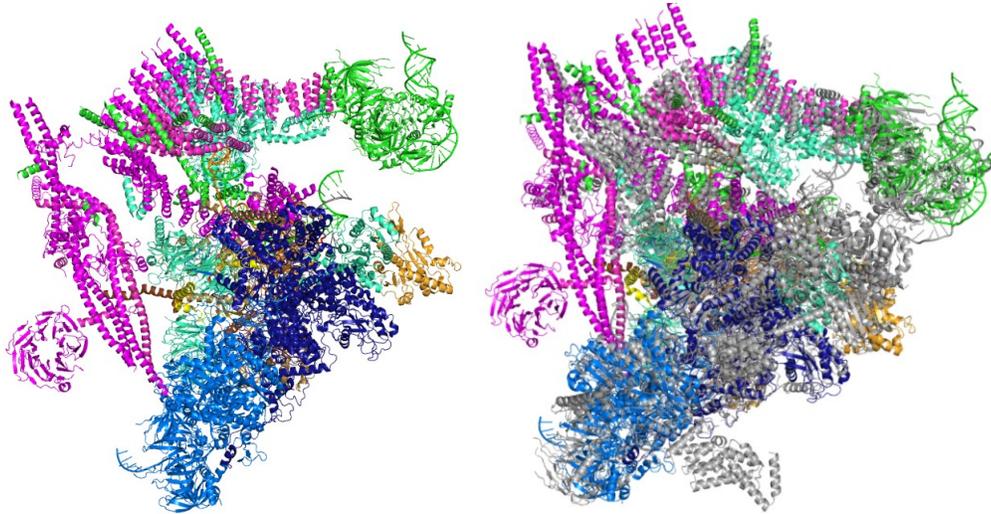


Figure 2-16. *S. pombe* ILS complex. A. Cryo-EM structure of the *S. pombe* ILS (PDB 3JB9; left) compared with the *S. cerevisiae* P complex (PDB 6EXN, grey; right).

RNA core. Instead, they act to stabilize the spliceosome as a whole for exon ligation. Prp18 maintains the conformation of the RH domain of Prp8, while Slu7 wraps around the core of Prp8 (Figure 2-15C).

Prp22 is positioned at the 3' end of the ligated mRNA, consistent with it releasing the ligated mRNA by pulling on the 3' exon (Figure 2-15D). Its association with the spliceosome is stabilized by interactions with Prp8, and an unidentified chain. The 3' exon passes over the thumb domain of Prp8 before binding to Prp22.

2-5g. Intron Lariat Spliceosome Complex

The ILS complex was the first spliceosome structure to be solved. It is from *S. pombe* rather than *S. cerevisiae*, and the resolution ranges from 2.9 Å at the core

to 8 Å at the periphery, with an average resolution of 3.6 Å (Yan et al., 2015; Figure 2-16).

The RNA core of the spliceosome is highly similar to that of the *S. cerevisiae* P complex (Figure 2-17A). Consistent with an ILS, the mRNA is not present in this structure. U5 snRNA loop 1 is not base-paired to anything. U2 and U6 snRNAs base-pairing at the catalytic core is maintained. The branched A is rotated slightly so it no longer points into the active site. Discrepancies in the bulge between helices I and II of the U2/U6 snRNA duplex could be due to species differences. The intron lariat is still bound to the spliceosome with the 5' SS base-paired with U6 snRNA and the BPS duplexed with U2 snRNA. Cef1 remains at the core of the spliceosome (Figure 2-17B). The RH domain of Prp8 is rotated down from its position in the P complex so that it is no longer engaged with the RNA (Figure 2-17C). Second step factors Slu7 and Prp18 are not visible.

The *S. pombe* spliceosome has a number of proteins not present in the *S. cerevisiae* spliceosome, including Cwf1, Cwf17, Cwf19, Cyp2, Cwf15, and Cwf11. These proteins are located throughout the spliceosome (Figure 2-17D).

Cwf15 interacts with Cwf1, and its N-terminus reaches in between the U5 and U6 snRNAs. Cwf19 bridges the linker, RH, and EN domains of Prp8. The branched A lies atop a helix of Cwf19. Although not visible, it appears that the region 3' to the branch A would pass over Cwf19 (Figure 2-17E).

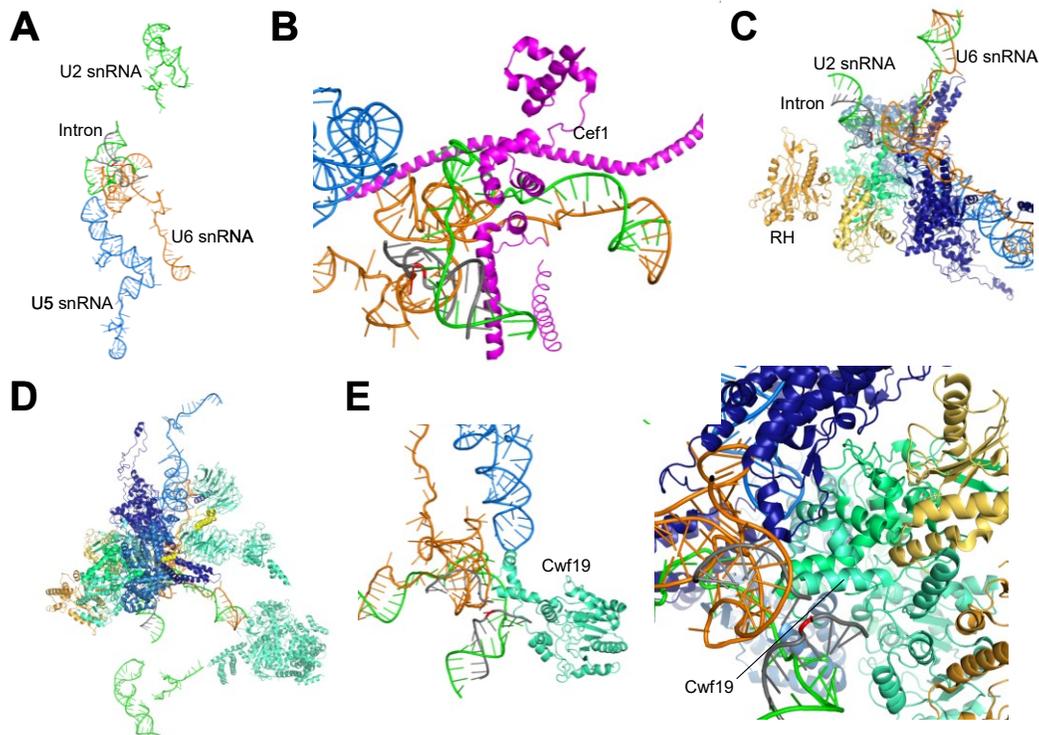


Figure 2-17. Details of the *S. pombe* ILS complex. A. RNA core of the spliceosome. B. Cef1 stabilizes the RNA core. C. Prp8 binds the RNA. The RH domain is rotated away from the BPS/U2 snRNA duplex. D. *S. pombe* specific proteins (light cyan) bound to the spliceosome. E. *S. pombe* specific Cwf19 binds the branch helix (left) and connects the branch helix to Prp8 (right).

2-6. Human Structures

The human spliceosome is more complex than that in yeast. As a result, the few human structures published are at lower resolution. To date, structures whose coordinates have been released are of the human B, B^{act}, C, and C* complexes. The components of these structures is summarized in Table 2-2.

2-6a. B Complex

The human B complex was solved to 4.5 Å resolution (Bertram et al., 2017a;

Figure 2-18). Comparison with the *S. cerevisiae* B complex shows very similar structural organization. Alignment of the human B complex Prp8 with that of *S. cerevisiae* overlays the position of the tri-snRNPs quite closely. The U2 snRNP does not align as closely when overlaid this way, suggesting conformational differences between the *S. cerevisiae* and human complexes. When aligned via Hsh155/SF3b155 the U2 snRNP SF3b aligns closely, but the tri-snRNP does not. SF3a and three of the SF3b components (SF3b45, SF3b145, and p14) are not visible in the human B complex. Helix II between U2 and U6 snRNAs has formed, but is in a different orientation relative to U2 snRNP than in the *S. cerevisiae* structure. The *S. cerevisiae* tri-snRNP and U2 snRNP are very similar to the human tri-snRNP and U2 snRNP, but they bind to different conformations or stages in B complex assembly.

U2 snRNA is base-paired with the BPS (YUNAC in humans). The branch A is nestled in SF3b155, and the distance between the branch A and 5' SS is over 150 Å (Figure 2-19A). As in the *S. cerevisiae* spliceosome, separating the reactive groups for the first step prevents premature splicing (Figure 2-19B).

Prp8 does not interact with SF3b155 in this structure. Instead, contacts between the tri-snRNP and U2 snRNP are mediated by Prp3, and the B-specific proteins Smu1 and RED, which connect Brr2 with SF3b130.

Table 2-2. Protein and RNA components in the structures of human spliceosomes (following pages). Summary of the high resolution cryo-EM structures of human spliceosomal complexes whose coordinates have been released. Colouring by sub-complex consistent throughout.

	3JCR	6AH0	6AHD	5O9Z	5Z58	5Z56	5Z57	5YZG	5XJC	5MQF
	Luhrmann	Shi	Shi	Luhrmann	Shi	Shi	Shi	Shi	Shi	Luhrmann
Complex	Tri-snRNP	pre-B	B	B	B ^{act} early	B ^{act} mature	B ^{act} late	C	C*	C*
Resolution (Å)	7	5.7	3.8	4.5	4.9	5.1	6.5	4.1	3.6	5.9
Intron		G	G	Y	G	G	G	G	G	YZ
U2										
Sap155		1	1	v	1	1	1			
Sap130		3	3	w	3	3	3			
Sap145		2	2		2	2	2			
Sap49		4	4		4	4	4			
p14		5	5		5	5	5			

Sap10		7	7	x	7	7	7			
PHF5a		6	6	y	6	6	6			
Sap114		u	u		w	u	w			
Sap62		v	v		u	v	u			
Sap61		w	w		v	w	v			
U2-A'		o	o	z	o	o	o	o	o	W
U2-B''		p	p	l	p	p	p	p	p	X
Sm		hijklmn	hijklmn	STUVWXZ	hijklmn	hijklmn	hijklmn	hijklmn	hijklmn	hijklmn
U2 snRNA		H	H	2	H	H	H	H	H	2
U4/U6										
Snu13	I	M	M	O						
Prp3	K	J	J	E						
Prp4	L	K	K	F						

Prp31	J	L	L	H						
PPIH			W	M						
Sm	opqrstu	PQRSTUV	PQRSTV	hijklmn						
LSm	2345678	qrstxyz	qrstxyz	opqrstu						
U4 snRNA	M	I	I	4						
U6 snRNA	N	F	F	6	F	F	F	F	F	6
U5										
Prp8	A	A	A	A	A	A	A	A	A	A
Brr2	C	D	D	C	D	D	D	D	D	
Snul14	B	C	C	B	C	C	C	C	C	B
Prp28	F	X								
U5-40K	D	E	E	D	E	E	E	E	E	F
Txn4a	E	O	O	J						
Sm	OPQRSTU	abcdefg								

U5 snRNA	H	B	B	5	B	B	B	B	B	5
Tri-snRNP										
Prp6	G	N	N	G						
Sad1	V	W								
Snu66			9	P						
RES										
SNIP1					X	X	X			
Snu17					Y	Y	Y			
Bud13					Z	Z	Z			
NTC										
Prp19						qrst	qrst	qrst	qrst	GHIJ
CDC5L					L	L	L	L	L	L
CRN					J	J	J	J	J	O
SYF1				I		I	I	I		M

SYF2								M	M	N
SPF27						K	K	K	K	K
ISY1								y		
NTR										
RBM22						O	O	O	O	P
Cwc15					P	P	P	P	P	R
Prp46					T	T	T	T	T	D
Prp45									R	C
Bud31						N	N	N	N	Q
B Complex										
Prp38A			Z	I						
MFAP1			0	K						
Smu1			Y	L						
Snu23			8	N						

WBP4			X	Q						
RED				R						
Bact										
Cwc22					V	V	V	V	V	T
Cwc27					z	z				
RNF113					M	M				
Prp2					x	x	x			
1 st Step										
Cwc25								X		
Yju2								Y		
2 nd Step										
Prp17						W	W	W	W	E
Slu7									Z	
EJC										

DDX48								u	u	p
MAGOHB								v	v	
RBM8								w	w	
MLN51								x	x	
Other										
Ubl5			A0							
Prp16								Z		
SRRM2						U	U	U	U	S
PRKRIP1									X	
PRP22									Y	q
PPIE						y	y	l		o
AQR						Q	Q	Q	Q	U
PPIL1						S	S	S	S	V
SKIP					R	R	R	R		

PPWD1								2		
PPIG								3		

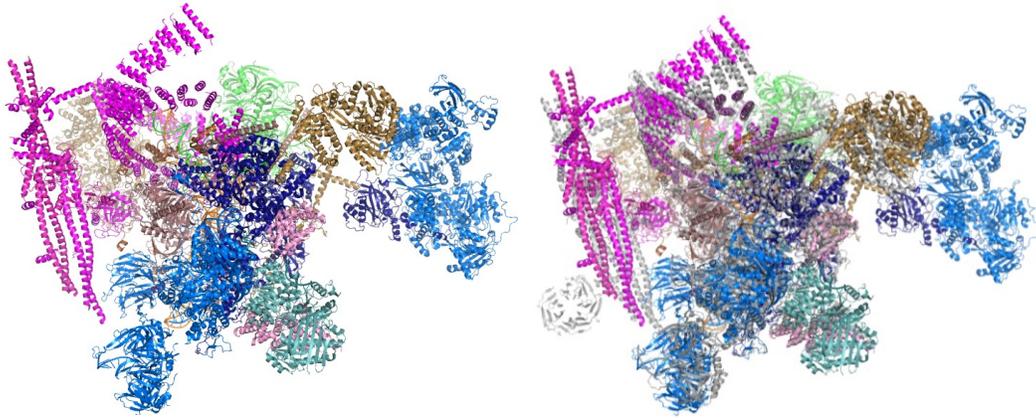


Figure 2-18. Human B complex. Cryo-EM structure of the human B complex (PDB 5O9Z; left) compared with the *S. cerevisiae* B complex aligned via Prp8 (PDB 5NRL, grey; right, top) and Hsh155 (right, bottom).

The 5' SS is base-paired with the ACAGA region of U6 snRNA, and the 3' end of the 5' exon is base-paired to U5 snRNA loop 1 (Figure 2-19C). This region of the intron is stretched into the cleft of Prp8 that in the *S. cerevisiae* structure binds the tail of U6 snRNA. This extends the backbone of the GU at the 5' SS in a way similar to the backbone in the *S. cerevisiae* B^{act} structure. Perhaps the human B structure was captured later in the splicing cycle than the *S. cerevisiae* B structure, as the base-pairing between U5 and U6 snRNAs seen in the *S. cerevisiae* structure has been replaced with the base-pairing between the 5' exon and U5 snRNA seen in B^{act} and subsequent complexes. The 5' SS is sequestered by the NT domain of Prp8 on one side and the thumb and linker domains on the other (Figure 2-19D). Similar to its homologue, Dib1 in *S. cerevisiae*, Dim1 sits between the pre-mRNA binding regions of U5 and U6 snRNAs. It sits atop the 5' SS, bridging the two lobes of the Prp8 NT domain and the linker region (Figure 2-19E).

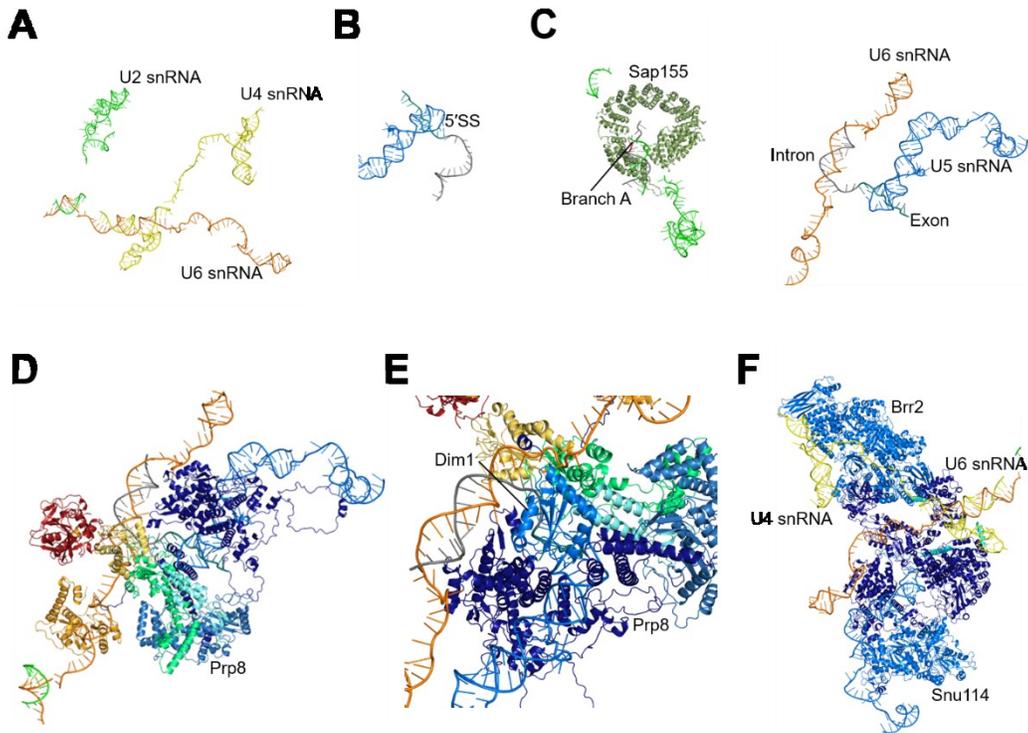


Figure 2-19. Details of the human B complex. A. RNA core of the complex. B. The 5' exon base-pairs with loop 1 of U5 snRNA while the 5' SS base-pairs with U6 snRNA. C. The 5' SS is separated from the branch A, which is bound to Hsh155. D. Prp8 bound to the RNA. E. Dim1 sits on top of the 5' SS and between the NT and linker domains of Prp8. F. Brr2 binds the U4 snRNA, ready to unwind the U4/U6 snRNA duplex. Snu114 sits on the opposite side of Prp8.

U4/U6 snRNA adopt a conformation similar to that in the yeast B complex. Helices I and II are formed, and separated by the U4 stem loop. As in the *S. cerevisiae* B complex, Brr2 is bound to U4 snRNA and ready to unwind U4/U6 helix I. Again, Brr2 and Snu114 are on opposite sides of Prp8 (Figure 2-19F). The signal to Brr2 from Snu114 to unwind the U4/U6 snRNA duplex must be transmitted via Prp8, Snu66, or another protein.

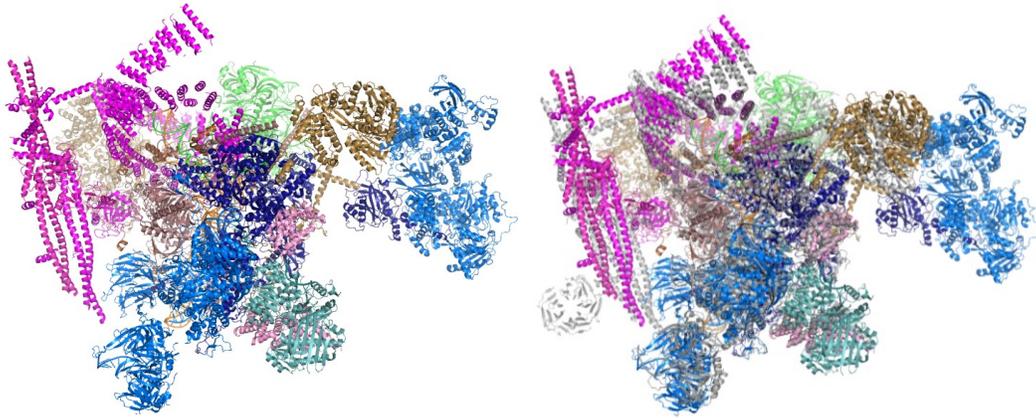


Figure 2-20. Human C* complex. Cryo-EM structure of the human C* complex (PDB 5XJC; left) compared with the *S. cerevisiae* C* complex (PDB 5WSG, grey; right).

2-6b. C* Complex

C* complex, primed for exon ligation, was solved to 3.6 Å (Bertram et al., 2017b; Figure 2-20). The RNA core of the human and *S. cerevisiae* spliceosomes is very similar. The catalytic core of the spliceosome is present: the U6 snRNA internal stem loop, and the base-pairing between U2 and U6 snRNAs. The 5' exon remains base-paired with loop 1 of U5 snRNA. The branched A intron lariat structure duplexed to U2 snRNA and U6 snRNA is moved out of the active site to about 20 Å away, opening up space in the catalytic core for the 3' exon, which is absent in this structure, to bind (Figure 2-21A).

CDC5L is at the core of the spliceosome (Figure 2-21B). It interacts with the U6 ISL and the U2/U6 helix I while also separating the BPS/U2 snRNA duplex from the 5' SS/U6 snRNA helix. The β-finger of the Prp8 RH domain lies alongside the branch A and the 5' SS/U6 snRNA duplex (Figure 2-21C).

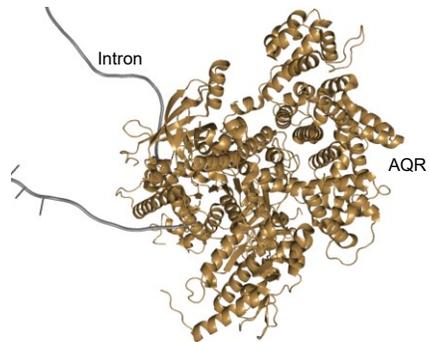
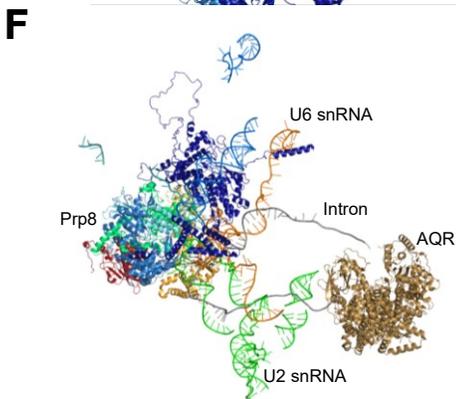
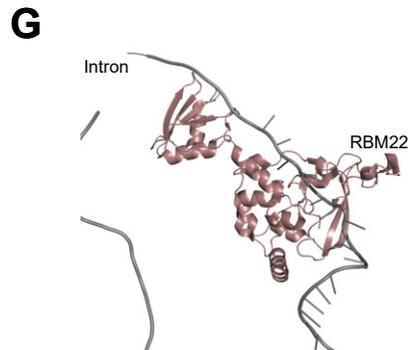
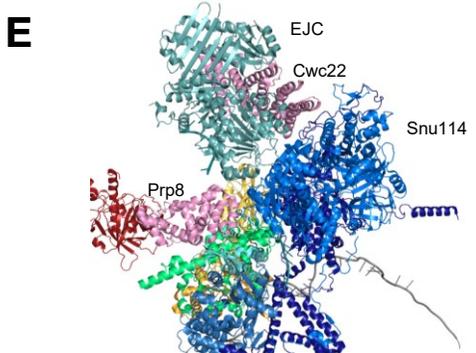
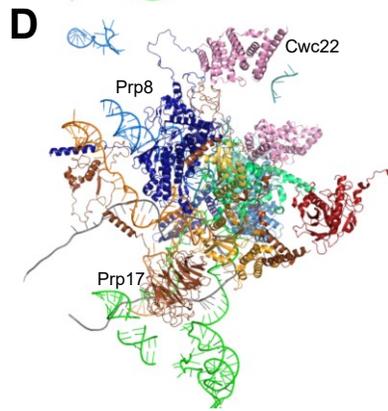
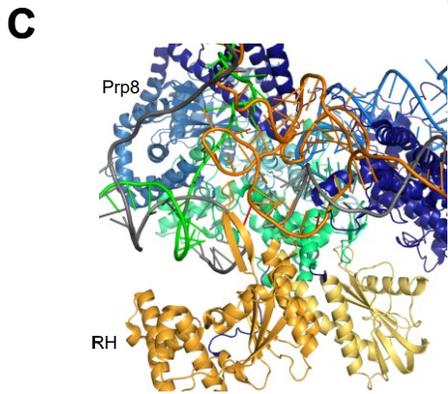
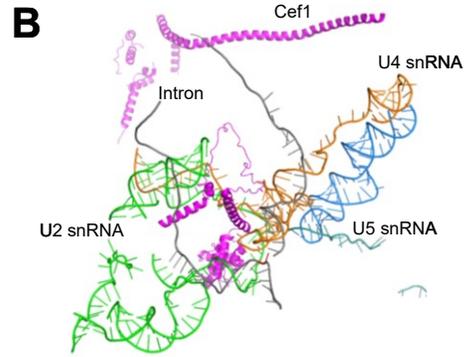
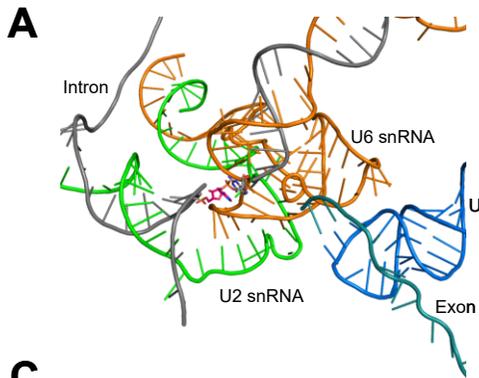


Figure 2-21. Details of the human C* complex (previous page). A. The branched intron has been moved out of the active site, and the 5' exon is in position for exon ligation. B. The active site is stabilized by Cef1. C. The RH β -finger inserts into the BPS/U2 snRNA duplex. D. Prp17 is bound to the spliceosome. E. The EJC binds the 5' exon and interacts with Cwc22, Snu114, and Prp8. F. Aquarius binds the lariat intron. G. RBM22 straddles the intron and must be removed prior to spliceosome disassembly.

The second step factors Slu7 and Prp17 stabilize the spliceosome. As in the *S. cerevisiae* C* complex, Slu7 interacts with Prp8 and other proteins in the spliceosome. C-terminal α -helices bridge the EN and RH domains, while an N-terminal region interacts with the Prp8 linker and NT domains and Cwc22. Prp17 binds an extended duplex between the BPS and U2 snRNA, while interacting with the 5' end of U6 snRNA (Figure 2-21D).

EIF4A-III of the EJC binds the 5' exon 20-25 nucleotides upstream of the 3' end of the exon, with MAGOH, Y14 and MLN51 assembled around one side and Snu114 bound to the other (Figure 2-21E). The N-terminal lobe of Cwc22 interacts with EIF4A-III, with the C-terminal lobe bound to Prp8.

The human-specific protein AQR binds to the outside of the spliceosome. It bridges components of the NTC: the bundle of Prp19 and Syf1. Although the middle of the intron lariat is not visible in the structure, AQR is in a position to bind it, or direct it away from the spliceosome (Figure 2-21F). RBM22 straddles the lariat intron, and must be removed prior to spliceosome disassembly (Figure 2-21G).

2-7. Group II Introns

Many introns do not require a spliceosome for removal. Group I and II

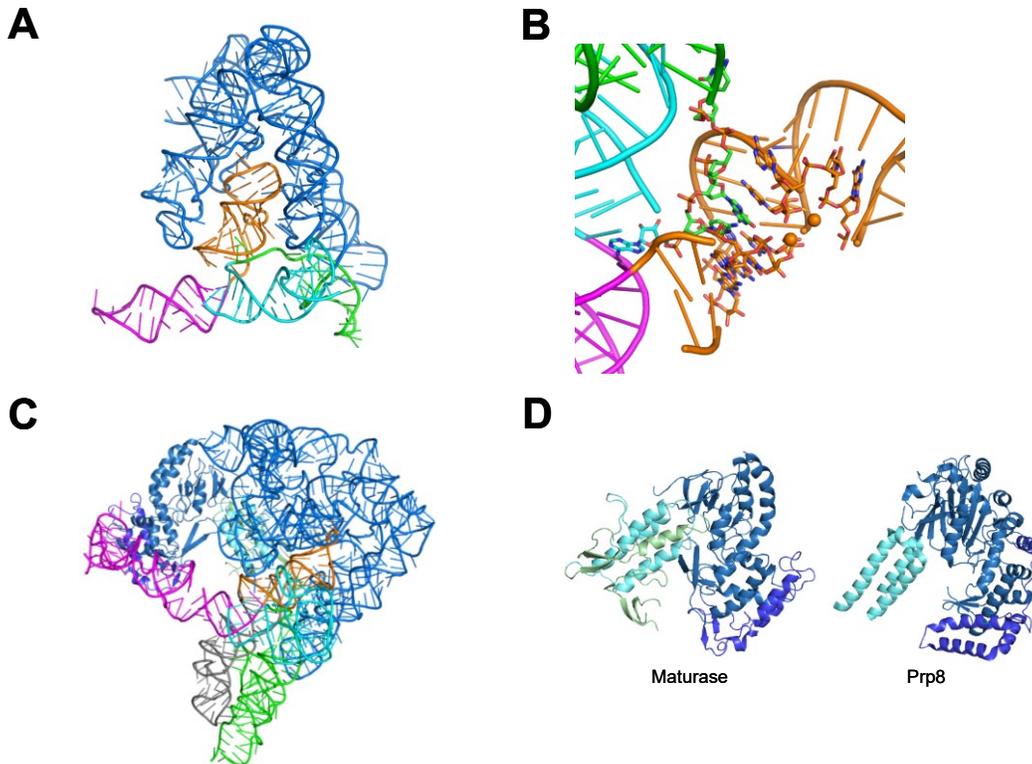


Figure 2-22. Group II intron. A. Crystal structure of a spliced group II intron (PDB 3BWP). D1 in blue, D2 in green, D3 in cyan, D4 in magenta, D5 in orange. D6 is not visible in the structure. B. Active site of the intron. Two metal ions are bound 3.9 Å apart within the D5 stem loop. C. Structure of a group II intron bound to its maturase (5G2X). D. Similarity between the group II maturase (left), and Prp8 RT (4I43; right), adapted from Qu et al 2016. Fingers/palm domain coloured blue, thumb/X in cyan.

introns are self-splicing RNAs that remove themselves from an RNA transcript to ligate their flanking exons (Cech, 1990; Pyle, 2016). The catalytic RNA is contained within the complex secondary structure of the intron (Figure 2-22A). Splicing for some of these introns is dependent on a single protein co-factor, a reverse transcriptase/maturase.

For all introns, splicing proceeds via sequential transesterification reactions; however, the nucleophile for the first step of splicing differs between intron types.

Group I introns use a free guanosine nucleotide in the first step (Cech, 1990). In the absence of their protein co-factors, Group II introns can activate a water molecule for the first step, producing a linear intron. In the presence of their maturase a branch adenosine from the 3' end of the intron acts as the nucleophile, resulting in a lariat intron similar to spliceosomal introns (Pyle, 2016). In both types of self-splicing introns the second step of splicing is analogous to that of spliceosomal splicing, with the free 3' hydroxyl of the 5' exon attacking the 3' SS for exon ligation.

Group II introns have six domains, D1-6, with catalytic activity residing within the highly conserved domain 5 (D5; Toor et al., 2008; Marcia & Pyle, 2012; Figure 2-22B). Domain 1 has exon binding sites that base-pair with the exons and position them for catalysis. It also binds the branch A, and acts as the scaffold for the intricate 3D structure of the intron. Domain 2 (D2) controls the position the branch site, and Domain 3 (D3) stimulates splicing via interactions with D5. The linker between domains 2 and 3 (J2/3) is a component of the active site. Domain 4 binds the maturase, and contains its ORF. Domain 5 contains a bulge located near the catalytic triad, and is essential for splicing. The branch A is located in domain 6.

The active site of the Group II intron features two divalent metal ions bound within the catalytic triad of D5 (Toor et al., 2008; Marcia & Pyle, 2012). The backbone phosphates in the bulge of D5 are twisted to form binding sites for these two metal ions. This strained conformation is stabilized by interactions with the

triple helix formed between a stem in D5 five nucleotides away from the bulge nucleotides and J2/3 in addition to two monovalent metal ions.

The 5' exon is positioned by base-pairing within D1. The exon binding loop is held in place by the extensive secondary structure surrounding it in D1 (Toor et al., 2008). Following the first step of splicing, the 5' SS must be removed from the active site and replaced by the 3' SS. Structural rearrangements of the intron involving J2/3 and the D5 bulge open the active site and release the bound metal ions. Interactions between D2 and D6 pull the branched A-5' SS linkage out of the active site. The 3' SS is positioned by interactions between D6 and domains 3 and 5. The active site of the spliceosome is reformed with four bound metal ions. Exon ligation occurs (Marcia & Pyle, 2012).

The Group II intron maturase contains an RT domain followed by a maturase domain (Qu et al., 2016; Zhao & Pyle, 2016; Figure 2-22C). Some maturases also encode an EN domain. Despite low sequence identity between them, structures of the maturase are more similar to the RT domain of Prp8 than the RT domains of retroviral or telomerase RTs (Figure 2-22D). The maturase binds D1 and D4 of the Group II intron.

Group II introns are mobile genetic elements that can reverse splice into the genome (Yang et al., 1996). The maturase that assisted with forward splicing reaction catalyzes the reverse reaction to insert the intron between what will become new exons (Lambowitz & Zimmerly, 2011). The reverse transcriptase of the maturase transcribes the RNA into DNA. RNase H degrades the RNA, and the

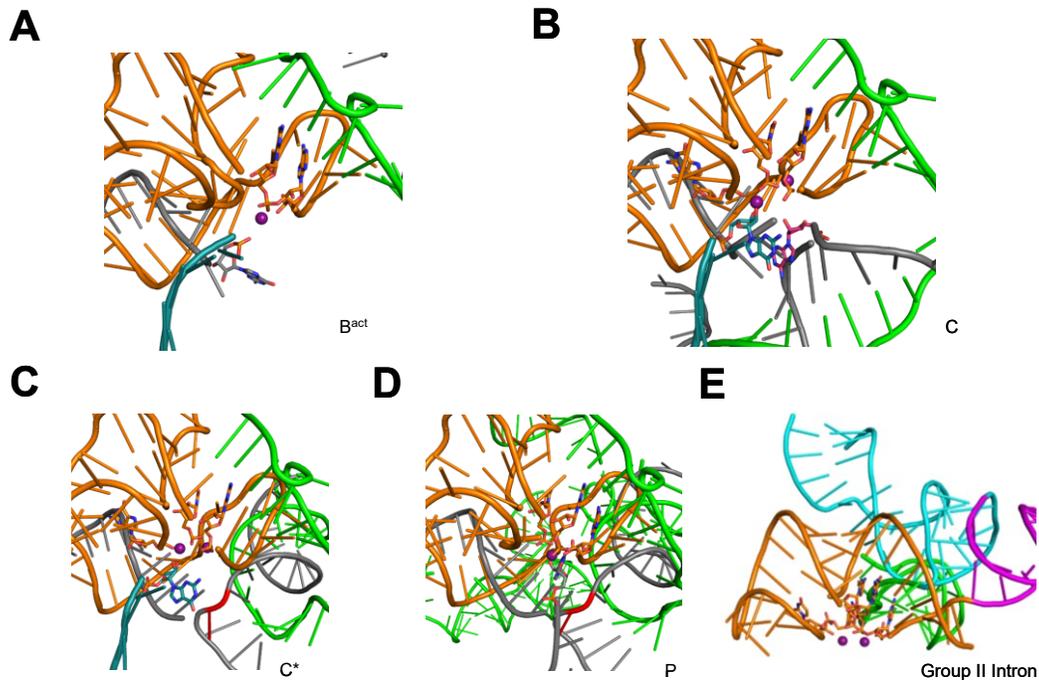


Figure 2-23. The spliceosome and the group II intron active sites have conserved metal binding sites. A. Active site of the B^{act} spliceosome with one catalytic metal visible. B. Active site of the C complex shows two metals. C. The C* active site has two visible metal ions. D. One catalytic metal is visualized in the P spliceosome. E. The active site of the group II intron bears a striking resemblance to that of the spliceosome.

DNA is filled in by normal DNA repair polymerase mechanisms.

2-8. Two-Metal Mechanism

Steitz and Steitz (1993) proposed a two metal mechanism for RNA cleavage based on the mechanism of the 3',5'-exonuclease of DNA polymerase I in *E. coli*. This enzyme produces a 5' phosphate and 3' hydroxyl without amino acid side chains participating directly in the chemistry of the reaction. Two metal ions 3.8 Å apart form inner-sphere contacts with the scissile phosphate and water nucleophile: metal 1 (M1) activates the nucleophile, while metal 2 (M2) stabilizes the transition

state and the 3' oxyanion leaving group. A number of other enzymes catalyzing phosphoryl transfer, including alkaline phosphatase, the RNase H domain of HIV reverse transcriptase, and the P1 nuclease, feature two metal ions 3.9 Å apart. Group II introns and the spliceosome likely rely on a similar two metal ion mechanism.

Phosphorothioate substitutions paired with thiophilic Mn^{2+} rescue show that U6 snRNA binds metal ions important for the catalysis of both steps of splicing (Fica et al., 2013). Specifically, non-bridging oxygens on the *S. cerevisiae* U6 snRNA G78 and U80 coordinate catalytic metals during branching, and U59 and U80 coordinate metals for exon ligation.

Of the four metals observed coordinated by U6 snRNA in the B^{act} structure (Yan et al., 2016; Figure 2-23A) three are thought to be structural. The fourth metal, thought to be the catalytic M2, is coordinated by the phosphates of A59 and G60. The phosphate of U80 is not seen to be involved in M2 coordination, and the oxygen of the 5' SS scissile phosphate is 3.5 Å away from M2. M1 is not observed in this structure, suggesting the core of the spliceosome is not fully catalytically active.

A slight shift in the backbone of U6 snRNA between the B^{act} and C complexes (Yan et al., 2016) allows M1 to bind. M1 is coordinated by G60 and U80. It is 6 Å away from the 2' hydroxyl of the branch A, and 6.5 Å from M2, suggesting a slight rearrangement following the first step of splicing. M2 is coordinated by the free 3' hydroxyl of the 5' exon (the leaving group), and phosphate oxygens of G78 and U80 (Figure 2-23B).

M1 and M2 switch roles between branching and exon ligation (Wilkinson et al., 2017; Yan et al., 2017), so that in the second step the metal that activated the nucleophile (M1) becomes the metal that stabilizes the leaving group (M2) and vice versa. In the C* complex M1 is coordinated by G78 and U80. The 3' hydroxyl of the 5' exon is 3.8 Å away. M2, 4.6 Å from M1, is coordinated by G60 and U80. The 3' exon is not visible in this structure (Figure 2-23C).

One metal is modeled in the P complex. It is likely M2, as it is coordinated by U80 A59 G60 and the 3' end of the intron lariat (Wilkinson et al., 2017; Figure 2-23D).

In the crystal structure of the Group II intron from *Oceanobacillus iheyensis* (Toor et al., 2008), M1 has two inner sphere ligands, phosphate oxygens of U375 and C377, and M2 is coordinated by phosphate oxygens from C358, G359 and C377. M1 and M2 are 3.9 Å apart, the distance between metals catalyzing a classic two-metal reaction mechanism (Figure 2-23E).

2-9. Group II Introns and the Spliceosome

Spliceosomal and Group II introns share a splicing mechanism that features branching using a bulged adenosine followed by exon ligation (Peters & Toor, 2015). In group II introns, this adenosine is selected via base-pairing within D6; the branch A in spliceosomal introns is excluded from the helix formed between the intron and U2 snRNA. Stereochemical requirements are the same for spliceosomal and group II introns as is the observed inversion of stereochemistry

(Moore & Sharp, 1993; Podar et al., 1998).

Splicing for both types of introns is catalyzed by two metal ions coordinated in a very similar manner. U6 snRNA adopts a fold, the U6 ISL, with a two nucleotide bulge separated by five nucleotides from a catalytic triad analogous to the arrangement of D5. The catalytic triad, C358 G359 and C360, of the *Oceanobacillus iheyensis* Group II intron is base-paired with G383 U384 and G385. G288 and C289, from J2/3, and C377 form base triples with the catalytic triad. UA375 and A376 are bulged out from the D5 helix (Toor et al., 2008). In comparison, the GA at the end of the U6 snRNA ACAGAGA sequence that base-pairs with the 5' exon also form base triples with the U6 snRNA catalytic triad AGC that is duplexed to U2 snRNA. A79 and U80 are bulged from the helix and bind metal ions. The 5' exon base-pairs with loop 1 of U5 snRNA in the spliceosome, and with exon binding sequences within D1 of Group II introns.

The similarities between spliceosomal and group II introns, including the similarity between the maturase and Prp8 (Qu et al., 2016; Zhao & Pyle, 2017), suggests an evolutionary relationship between these two types of introns (Vosseberg & Snel, 2017). Group II introns are able to reverse splice back into DNA and may be the evolutionary precursors to spliceosomal introns. Because the introns themselves are catalytic, the nature of the exons do not matter. The intron may then have outsourced its catalytic core to the snRNPs, leaving introns within genes that can be removed by the trans-acting RNA within the spliceosome.

2-10. Summary

Great strides have been made towards a structural understanding of the spliceosome. In the last two years cryo-EM structures have added greatly to our knowledge including further evidence that spliceosomal and group II introns share a common ancestor. However, there is still much to learn about spliceosomal regulation and structural rearrangements. The spliceosome is highly dynamic and the few static pictures available provide only snapshots of the splicing process.

Chapter 3

5' Splice Site Recognition by U1C and the U1 snRNP

Parts of this chapter have been adapted from McCarthy et al., 2017.

3-1. U1 snRNP

U1 snRNP is the first snRNP to associate with the pre-mRNA. It, along with the SF1/U2AF trimer, binds the pre-mRNA in an ATP-independent manner to form the early (E) complex in humans (commitment complex (CC) in yeast) from the non-spliceosomal H complex composed of regulatory SR and hnRNPs bound to the pre-mRNA (Figure 3-1; Michaud & Reed, 1991). *In vitro* splicing is inhibited upon depletion of U1 snRNP, and can be rescued by the addition of either U1 snRNP or purified SR proteins (Barabino et al., 1990; Crispino et al., 1994; Tarn & Steitz, 1994; Will et al., 1996).

In humans, U1 snRNP is composed of the U1 snRNA, the heptameric Sm core, and three U1-specific proteins: U1A, U1-70k and U1C. U1A and U1-70k interact directly with U1 snRNA via RNA binding domains (RBDs), while U1C is a small helical protein that binds U1-70k (Pomeranz Krummel et al., 2009) and the Sm core.

U1 snRNP requires the 5' end of its U1 snRNA to bind the 5' SS in the pre-mRNA: the 5' end of U1 snRNA base-pairs with the 5' SS. Cleavage of the 5' end of U1 snRNA or removal of the 5' SS from an intron reduces binding between U1 snRNP and pre-mRNA (Mount et al., 1983). The helicase Prp28 is required for release of U1 snRNP from the spliceosome and to allow the 5' SS to base-pair with U6 snRNA in the activated spliceosome (Staley & Guthrie, 1999; Chen et al., 2001).

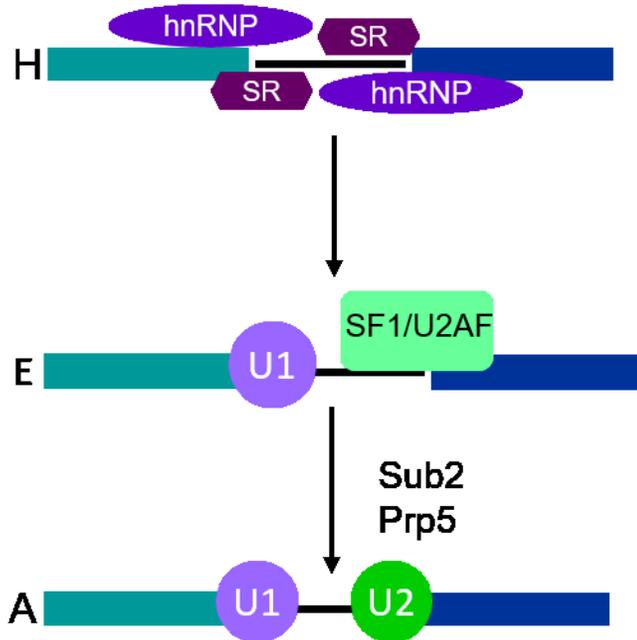


Figure 3-1. Early stages of spliceosome assembly. Prior to spliceosome assembly, SR proteins and hnRNPs bind the pre-mRNA to form H complex. U1 snRNP and the SF1/U2AF trimer are recruited to and replace the SR proteins and hnRNPs at the 5' SS and BPS respectively to form the early (E) complex. Activation of Sub2 and Prp5 allows U2 snRNP to replace the SF1/U2AF trimer at the BPS, which gives rise to the A complex.

3-2. U1 snRNP Structure

The 164 bases of human U1 snRNA fold into four stem loops and helix H (Krol et al., 1990). Helix H is formed between the nucleotides three bases upstream from the single-stranded 5' sequence that base-pairs with the 5' SS and a region between stem loop 3 and the Sm binding site (Figure 3-2A). The Sm core binds a U-rich single-stranded region between stem loops 3 and 4. U1-70k binds stem loop 1 while U1A binds stem loop 2. U1C binds to U1-70k and the Sm core in the assembled U1 snRNP rather than to the U1 snRNA.

The structure of a reconstituted U1 snRNP was initially solved at 5.5 Å resolution (Figure 3-2B; Pomeranz Krummel et al., 2009). RNA duplexes can be

visualized at this low resolution, as can α -helices and β -sheets. The structure contains U1 snRNA, the Sm core, the N-terminus of U1-70k, and the N-terminus of U1C. U1A and its stem loop were deleted, as U1A binding is not required for the formation of E complex. Instead, stem loop 2 was replaced with a kissing loop to facilitate crystallization contacts.

Proteins were located within the map using data from multiple crystals, each containing proteins individually labeled with heavy metals (Oubridge et al., 2009). Thus, the published structure has been verified by diffraction data from numerous crystals, all featuring distinct arrangements of heavy metals.

Diffraction from several different crystals in which one of the seven Sm proteins was labeled with seleno-methionine (SeMet) was used to place each Sm protein in the map. The Sm ring forms in the following order: SmE, SmG, SmD3, SmB, SmD1, SmD2, SmF.

The RBD of U1-70k interacts with U1 snRNA stem loop 1. The N-terminal 97 residues of U1-70k, which do not contain the RBD, are able to interact with U1 snRNP in the absence of the RBD. The N-terminus of U1-70k, required for U1C binding to the U1 snRNP, is predicted to be unstructured. Crystals of seven individual SeMet mutants of U1-70k were grown, and the anomalous peaks used to trace the path of its N-terminus. These crystals showed that U1-70k wraps around the Sm core and, along with SmD3, creates a binding site for U1C.

Unlike U1A and U1-70k, U1C does not contain an RBD and interacts with U1 snRNP via protein-protein interactions, rather than direct contact with U1

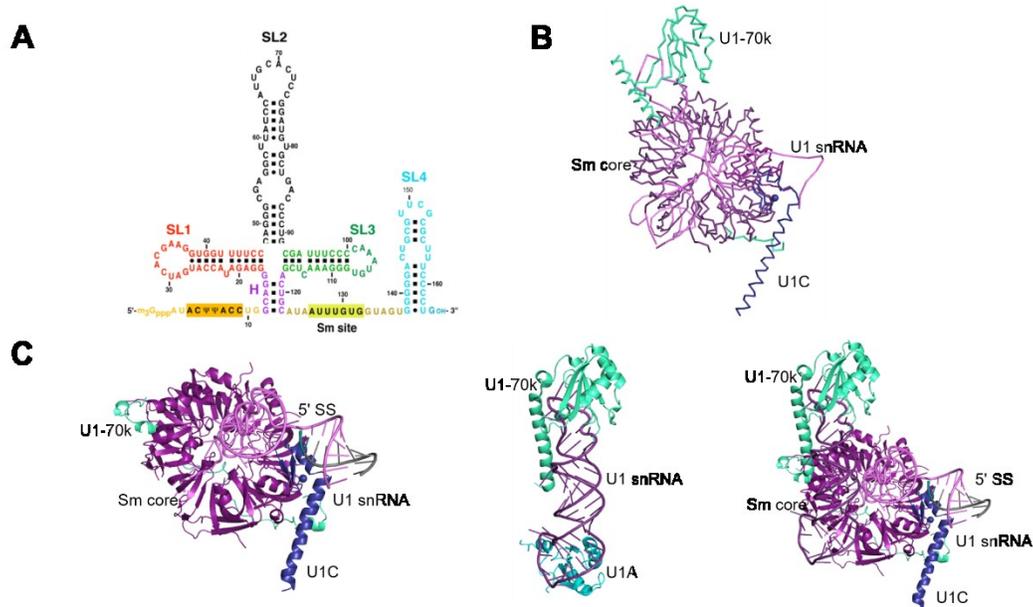


Figure 3-2. Structure of the human U1 snRNP. A. Secondary structure of the U1 snRNA consists of four stem loops (SL1-4), a single-stranded Sm binding site, and helix H between the Sm binding sites and the single-stranded 5' end of the snRNA. Originally published in Kondo et al. (2015), licensed under CC BY 4.0. B. The 5.5 Å crystal structure (PDB 3CW1) of U1 snRNP shows the Sm core (dark purple) bound to the snRNA (violet), with U1-70k (mint) binding stem loop 1 and U1C (dark blue) interacting with U1-70k and the Sm core. C. U1 snRNP was crystallized at higher resolution by splitting the particle into two pieces. Left, a portion of the U1 snRNA bound within the Sm core (PDB 4PJO). The N-terminus of U1-70k is fused to SmD1. U1C is bound, and interacting with a duplex formed between the 5' end of U1 snRNA and an oligo representing a 5' SS. Middle, the RBD of U1-70k and its long helix is bound to stem loop 1. U1A, bound to stem loop 2 hybridized to the bottom of stem loop 1, stabilizes the complex (PDB 4PKD). Right, composite structure of U1 snRNP generated by alignment to the 5.5 Å structure via U1C (PDB 4PJO) or U1-70k (PDB 4PKD).

snRNA. The N-terminal 45 residues, which include a C₂H₂ zinc finger, are sufficient for interaction with U1 snRNP (Nelissen et al., 1991). U1C binding to U1 snRNP is also dependent on the N-terminus of U1-70k and on SmD3.

U1 snRNP was then divided into two constructs to create a second set of U1 snRNP crystal structures at higher resolution (Figure 3-2C; Kondo et al., 2015). The first construct consisted of a minimal U1 snRNP, with the Sm core, U1C, a

truncated U1-70k with the N-terminal 59 residues fused to the C-terminus of SmD1, and an oligo modeling the 5' SS. Helix H of the snRNA was replaced with a kissing loop. These crystals diffracted to 3.3 Å. Each Sm protein in the Sm core interacts with one base of the U1 snRNA. The path of the N-terminus of U1-70k as determined by SeMet substitution in the lower resolution structure was confirmed: it interacts with SmD2, SmD3 and SmB. U1C's Helix 1, the zinc finger and helix 2 sit on SmD3, while helix 2 interacts with residues 1-12 of U1-70k. The second construct features the U1-70k RBD and stem loop 1. To visualize the long helix of U1-70k it was fused to U1A RBD bound to stem loop 2 fused to the bottom of stem loop 1. This structure was solved to 2.5 Å. Both structures align closely to the low resolution structure, and the resulting overlay gives a higher resolution structure of the U1 snRNP.

S. cerevisiae U1 snRNA is 568 nucleotides long, compared with 164 in humans (Kretzner et al., 1987; Kretzner et al., 1990). This U1 snRNP contains seven proteins not stably associated with the human U1 snRNP (Luc7, Nam8, Prp39, Prp40, Prp42, Snu56, and Snu71; Gottschalk, et al., 1998), four of which are associated with alternative splicing in humans: Luc7L (Luc7 in yeast), Tia1 (Nam8), Prp40 (Prp40), and RBM25 (Snu71; Forch et al., 2002; Puig et al., 2007). The cryo-EM structure of the *S. cerevisiae* U1 snRNP shows a similar organization to human U1 snRNP, with the U1-70k, U1A and U1C homologues binding around the U1 snRNA, and the additional, *S. cerevisiae*-specific proteins bound to the bottom of the snRNP (Figure 3-3; Li et al., 2017). The 5' end of the U1 snRNA,

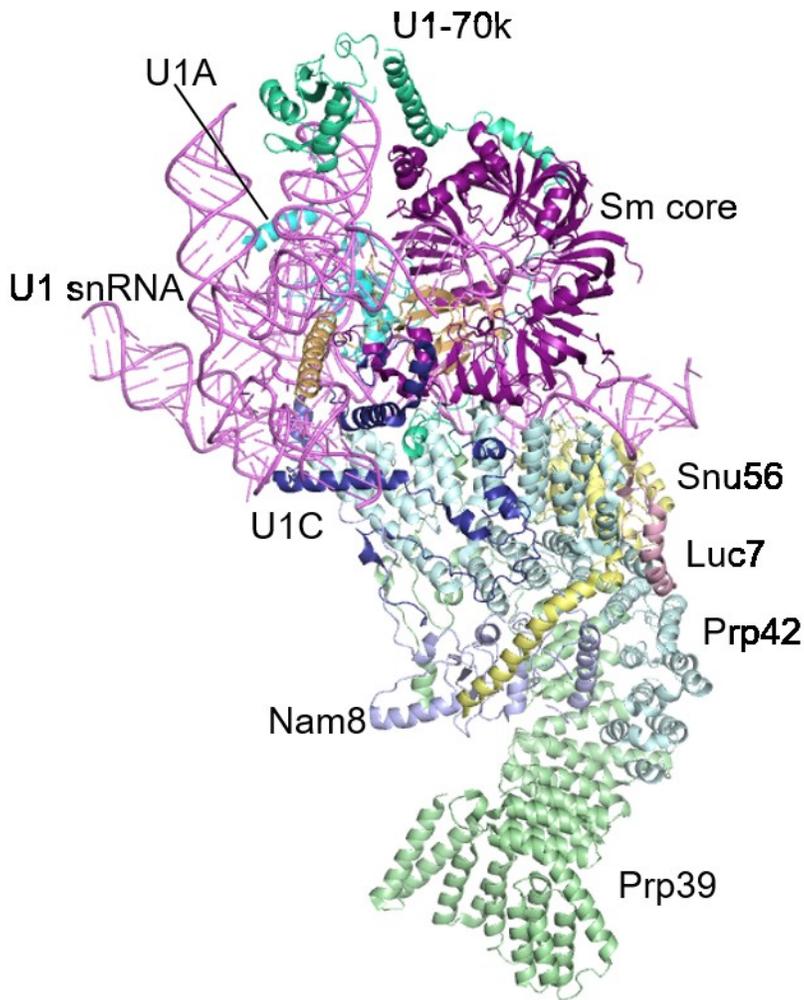


Figure 3-3. Cryo-EM structure of the *S. cerevisiae* U1 snRNP. The *S. cerevisiae* U1 snRNP contains seven proteins not seen in human U1 snRNP (Luc7 (light pink), Nam8 (light blue), Prp39 (light green), Prp40 (not seen), Prp42 (light cyan), Snu56 (light yellow), and Snu71 (not seen); (PDB 5UZ5). The additional proteins bind to the bottom of the snRNP.

the 5' SS binding site, is not visualized in this structure. Yhc1 (U1C in humans) is in a similar position to that of U1C, and likely interacts with the pre-mRNA similarly to U1C in the human U1 snRNP. The non-conserved C-terminal region of Yhc1 interacts with two additional yeast U1 snRNP proteins: Nam8 and Prp42 (possibly Prp39 in humans).

3-3. U1C

U1C in humans has been shown to be required for U1 snRNP binding to the 5' SS (Heinrichs et al., 1990; Will et al., 1996). Loss of U1C reduces binding of the U1 snRNP to pre-mRNA by about 50%, and binding can be rescued by incubation of U1 snRNPs lacking U1C with recombinant U1C prior to incubation with pre-mRNA (Heinrichs et al., 1990). In contrast, depletion of U1A decreased binding to pre-mRNA only marginally. E complex formation and splicing of U1 snRNP depleted extracts are rescued by U1 snRNPs lacking U1A, but only partially rescued by U1 snRNPs lacking both U1A and U1C (Will et al., 1996). Particles lacking U1A, U1C, and U1-70k were not able to form E complex. U1C acts in the context of U1 snRNP, as U1C alone is not able to shift H complex to E complex.

While the N-terminus of the 159 amino acid U1C is highly conserved, the C-terminus shows very low levels of sequence conservation (Figure 3-4; Muto et al., 2004). U1C truncated by the 99 C-terminal residues was able to complement U1 snRNPs lacking U1C, while U1C featuring mutations of the zinc binding residues was not (Nelissen et al., 1991). The double mutant K28G R29S was also not able to rescue formation of E complex in U1 snRNPs that lack wild type U1C (Will et al., 1996). Deletion of the final 51 amino acids of Yhc1 has no adverse effect on growth in *S. cerevisiae* (Schwer & Shuman, 2014).

Increasing the concentration of KCl decreases U1 snRNP binding to pre-mRNA (Heinrichs et al., 1990). If the interaction between the snRNP and pre-mRNA was exclusively due to base-pairing between the two, an increase in salt

1999).

The NMR structure of U1C residues 1-61 shows two β -sheets and three α -helices (Figure 3-5A; Muto et al., 2004). The zinc finger is extended by α helix 2. α -helix 3 is loosely associated with α -helix 2, and changes conformation upon binding to U1 snRNP (Figure 3-5B; Pomeranz Krummel et al., 2009; Kondo et al., 2015). Cys6, Cys9, His24 and His30 coordinate a zinc ion in a zinc finger. The five residues between His24 and His30 introduce a break in and kink the helix between His24 and Arg28. Mutagenesis in *S. cerevisiae* Yhc1 shows that while the C6A mutation is lethal, alanine mutants of the other zinc binding residues are able to rescue Yhc1 deletion *in vivo* (Schwer & Shuman, 2014). C9A and H30A were temperature sensitive, and H24A grew at all temperatures. Adjacent to the metal binding pocket is a hydrophobic core composed of Phe4, Leu13, Val20 and Arg21. A third cysteine in U1C, Cys25, is not involved in the zinc finger, nor does it form disulfide bonds (Nelissen et al., 1991).

Alanine mutants of nine surface residues of *S. cerevisiae* are able to complement a yeast strain lacking U1C (Schwer & Shuman, 2014). However, these mutants became synthetic lethal with the deletion of other, inessential splicing factors or nucleotides. Yhc1 K28A (but not N29A) is synthetic lethal with the U2AF65 homologue Mud2 deletion (Schwer & Shuman, 2014). Mud2 binds the polypyrimidine tract in the formation of the commitment complex. In the Mud2 deletion, binding of U1 snRNP to the pre-mRNA may form a commitment complex adequate for the recruitment of U2 snRNP and the formation of the A complex. In

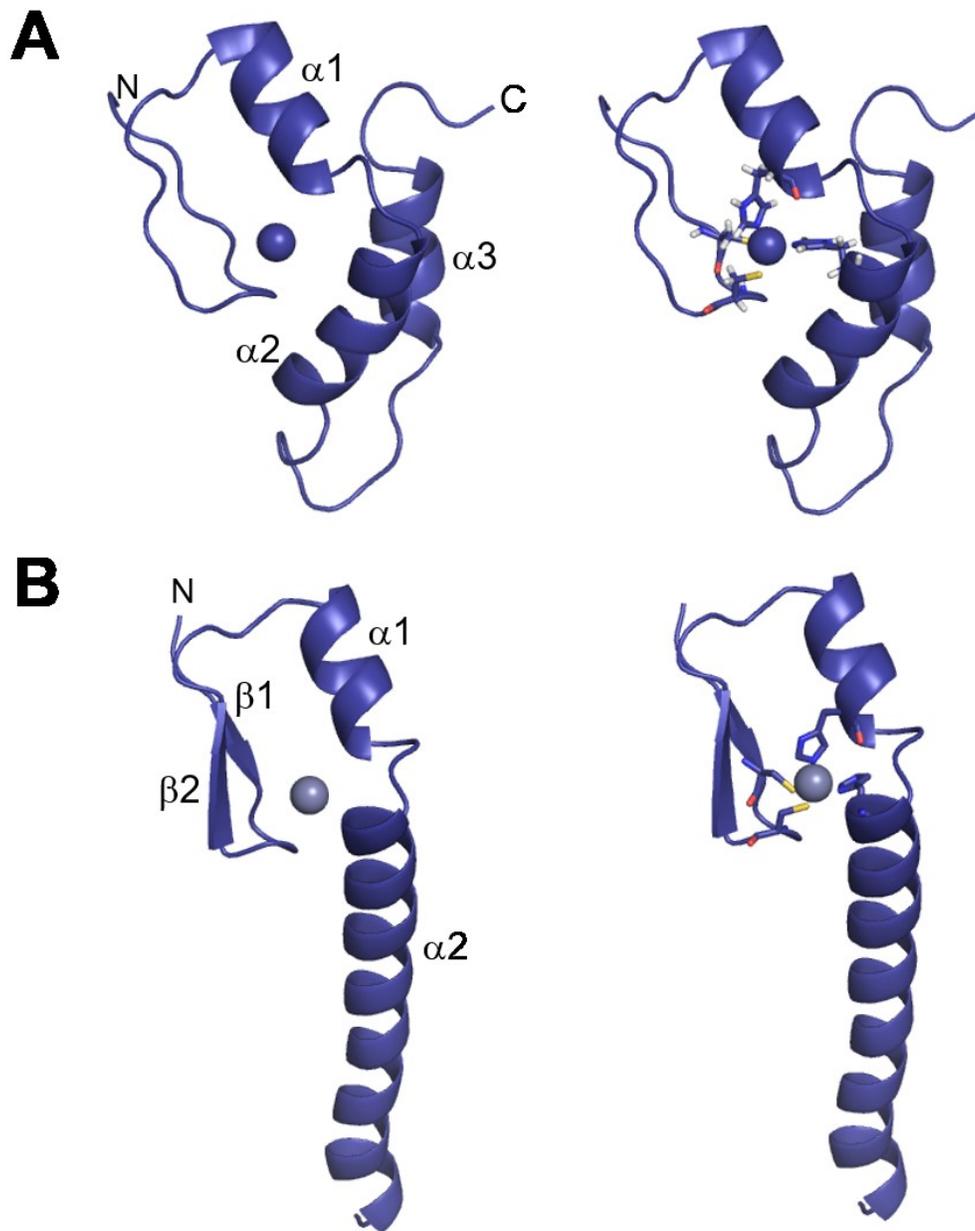


Figure 3-5. Structure of U1C. A. NMR structure of U1C 1-61 (PDB 2VRD) shows two α -helices involved in a zinc finger and a third helix folded up against the back (left). B. Structure of U1C from the crystallized minimal U1 snRNP (PDB 4JPO; left). The formation of the zinc finger is very similar between the NMR and crystal structures. However, helix $\alpha3$ in the NMR structure becomes part of an extended helix $\alpha2$ within the U1 snRNP. Zinc binding residues of U1C shown for the NMR structure (A, right) and crystal structure (B, right).

the K28A strain, the presence of Mud2 in the CC may be sufficient to overcome

any weakened binding of the U1 snRNP to the 5' SS. However, in the Mud2 deletion strain, K28A may disrupt binding of U1 snRNP to the pre-mRNA sufficiently to prevent U2 snRNP recruitment and formation of the A complex. When the linker between the 5' SS/U1 snRNA duplex and helix H of U1 snRNA is shortened by one nucleotide, it becomes synthetic lethal with K28A and severely synthetically sick with N29A (Schwer & Shuman, 2014). Deletion of the essential helicase Prp28, which disrupts the 5' SS/U1 snRNA duplex, can be bypassed by RNA mutations that destabilize base-pairing between the 5' SS and U1 snRNA. Yhc1 K28A is able to bypass the Prp28 deletion at higher temperatures, suggesting that K28 plays a role in stabilizing the 5' SS U1 snRNA duplex (Schwer & Shuman, 2014).

3-4. 5' Splice Site Recognition

The sequence at the 5' end of the U1 snRNA is AUAC $\psi\psi$ ACG (ψ is pseudo-uridine; Reddy et al., 1974; Rogers & Wall, 1980). In humans, the 5' SS consensus sequence is GURAGU. The almost invariant 5' GU dinucleotide at the 5' SS base-pairs with U1 snRNA C8 and A7 (underlined). The pseudo-uridines then base-pair with the purine (R) and A. The 3' GU base-pairs with C4 and A3 (Figure 3-6A; Mount et al., 1983; Freund et al., 2003). Mutations of the 5' SS that decrease base-pairing to U1 snRNA also decrease binding to U1 snRNP (Siliciano & Guthrie, 1988).

In the low resolution crystal structure of U1 snRNP the 5' ends of two U1

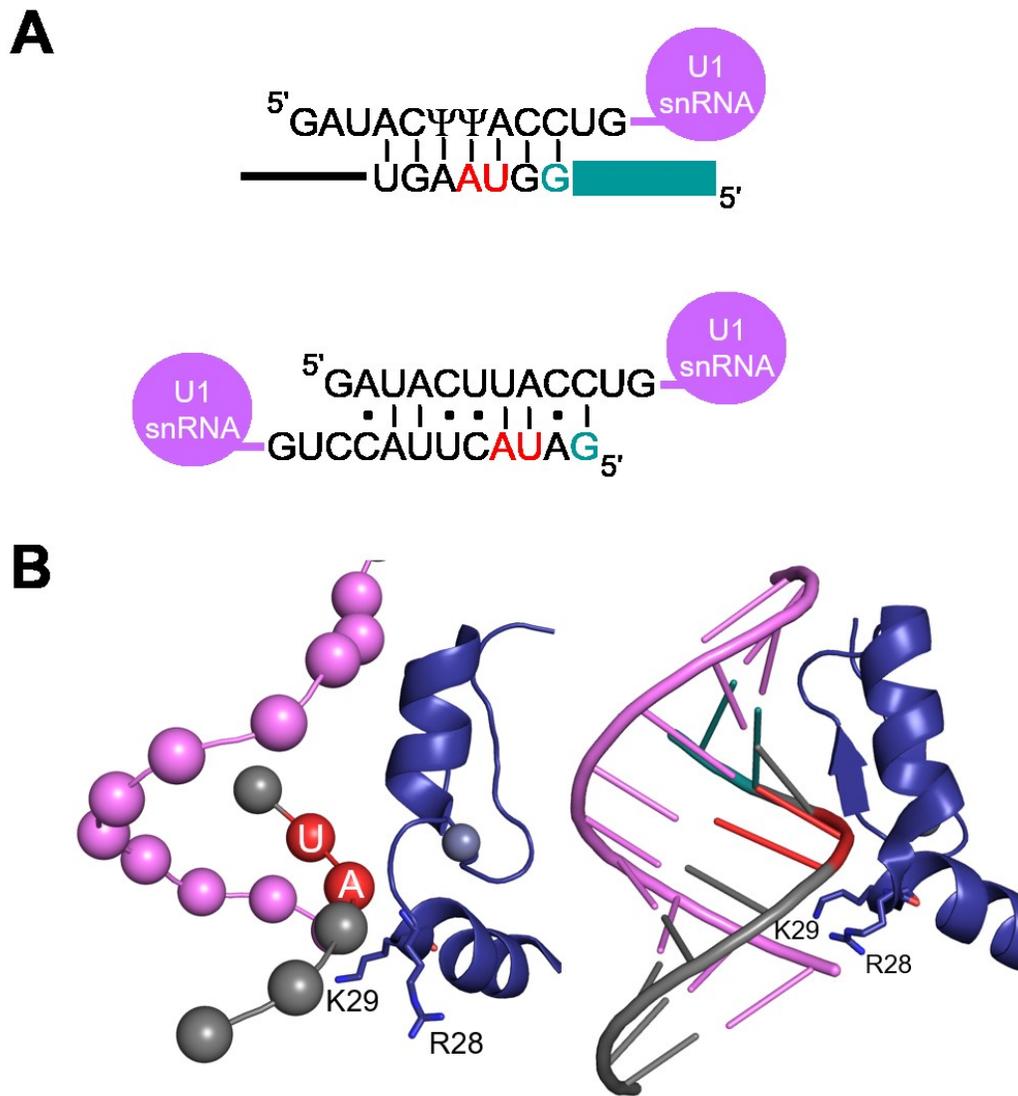


Figure 3-6. U1C and the U1 snRNA/5' SS duplex. A. Top, model of the base-pairing between U1 snRNA and the 5' SS. Cleavage occurs between the final G of the exon (teal) and the first G of the 5' SS (black). Bottom, serendipitous interaction between neighbouring molecules of U1 snRNA in the 5.5 Å crystal structure of U1 snRNP mimics the interaction between the U1 snRNA and the 5' SS. Adapted from Pomeranz Krummel et al. (2009). B. Interaction between U1C and the U1 snRNP/5' SS interaction as seen in the low resolution (left; PDB 3CW1 with U1C from PDB 2VRD aligned) and high resolution (right; PDB 4JPO) U1 snRNP crystal structures.

snRNAs in the same asymmetric unit base-pair to serendipitously mimic of the interaction of U1 snRNA with the 5' SS (Figure 3-6A; Pomeranz Krummel et al.,

2009). This interaction is recognized and stabilized by the zinc finger and N-terminal helices of U1C. In accordance with evidence that the double mutation of Arg28 and Lys29 disrupts the formation of E complex, Arg28 and Lys29 are positioned to interact with the 5' SS/ U1 snRNA duplex (Figure 3-6B). However, they are not visible due to the low resolution of this structure.

The interaction between U1 snRNA and the 5' SS was also seen in the second, higher resolution structure of the human U1 snRNP (Figure 3-6B; Kondo et al., 2015). Relative to the first structure, the duplex between the 5' SS and U1 snRNA is tilted, possibly due to the lack of constraints imposed by the second molecule of U1 snRNP in the first structure. U1C helix α 1 and strand β 1 interact with the minor groove of the 5' SS/U1 snRNA duplex. U1C makes hydrogen bonds with backbone phosphates and 2' hydroxyls of the RNA, but does not participate in base recognition. The side chains of R28 and K29 are in proximity to the duplex, but are not positioned to hydrogen-bond with the RNA in the structure. The lack of sequence-specific contacts between U1C and the pre-mRNA suggests that U1C does not select the 5' SS. Rather, U1 snRNA identifies the 5' SS, and U1C stabilizes the resulting duplex.

3-5. U1 snRNP in the A complex

The A complex is formed when U2 snRNP replaces SF1 and U2AF on pre-mRNA bound to U1 snRNP. A recent cryo-EM structure of the *S. cerevisiae* A complex shows stoichiometric amounts of U1 and U2 snRNPs (Figure 3-7A;

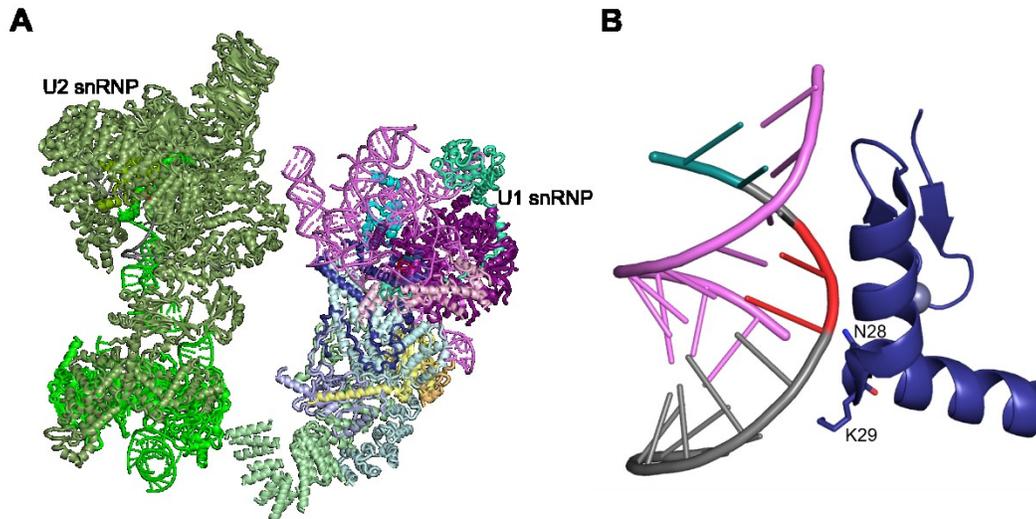


Figure 3-7. U1C in the A complex. A. Cryo-EM structure of the *S. cerevisiae* A complex (PDB 6G90) shows U2 snRNP (green) interacting with U1 snRNP (coloured as per Figure 3-3) in two interfaces. B. Interaction between U1C and the U1 snRNA/5' SS duplex is very similar to that seen in the human crystal structure.

Plaschka et al, 2018). The branch A, bulged out of a duplex between the branch point sequence (BPS) and U2 snRNA, is bound by Hsh155 and Rds3. As in the B complex, the 5' SS and BPS are kept sequestered.

Base-pairing between the 5' SS and U1 snRNA is observed in this structure (Figure 3-7B). The Yhc1 zinc finger contacts the backbone of the first four nucleotides of the 5' SS, while a previously unvisualized loop contacts the final GU. A zinc finger on Luc7 also contacts the 5' end of the U1 snRNA, similar to how Yhc1 interacts with the 5' SS. Luc7L (Luc7-like) in humans and Luc7 in *S. cerevisiae* are highly conserved, suggesting a similar recognition of the U1 snRNA for both proteins.

Results

3-6. U1C does not bind RNA

It has been shown that human U1C alone does not bind the 5' SS (Muto et al., 2004). Previously, the RNA used to test for U1C binding has been single-stranded. The U1 snRNP structure shows U1C inserting into the minor groove of the 5' SS/U1 snRNA duplex, so perhaps U1C only binds double-stranded RNA. We transcribed and labeled an RNA hairpin designed to mimic the interaction between the 5' SS and U1 snRNA.

Our U1C construct, amino acids 1-61, is based on the construct expressed for NMR studies (Muto et al., 2004). This construct binds zinc, and is incorporated into the U1 snRNP. Complementation studies show that residues 1-60 are sufficient to convert the H complex to E complex in the presence of U1C-depleted U1 snRNP (Will et al., 1996). The C-terminal region is not highly conserved, nor is it observed in proximity to the pre-mRNA and thus unlikely to be required for the U1C-RNA interaction.

We then used this double-stranded RNA and U1C construct to perform electrophoretic mobility shift assays (EMSAs). Even at concentrations of 65 μ M U1C did not bind the hairpin RNA (Figure 3-8A).

Because U1C operates within the context of U1 snRNP, we then designed constructs to mimic the protein environment of U1C in the hopes of stabilizing the interaction between U1C and the 5' SS duplex. We cloned and expressed the N-terminal 30 residues of U1-70k predicted to interact with U1C and SmD3. We also

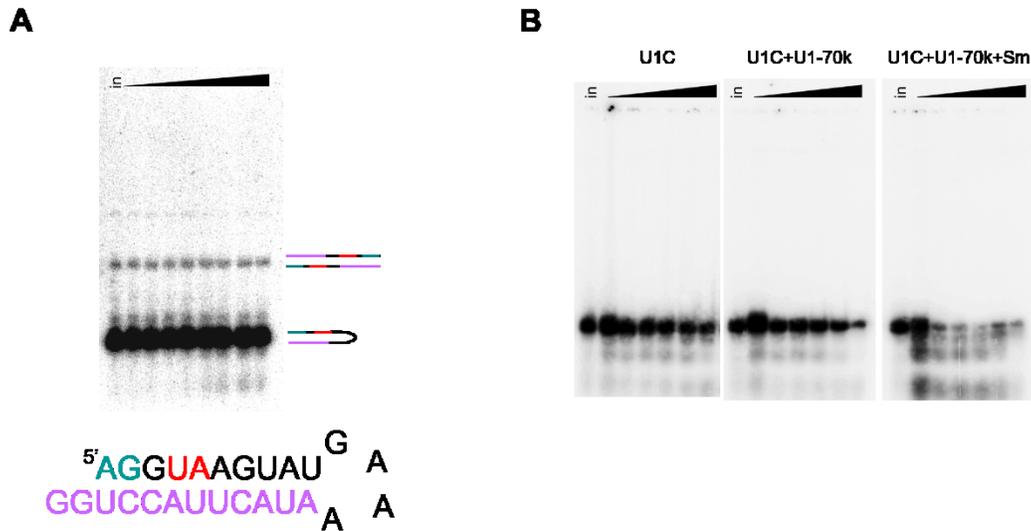


Figure 3-8. EMSA of U1C and the 5' SS. A. Top, U1C alone does not bind a duplex modeled on the helix formed between the U1 snRNA and 5' SS (bottom; colours consistent with Figure 3-6.). B. Addition of U1-70k with or without SmD3 and SmB does not allow U1C to shift the model duplex.

cloned SmD3 1-75 to interact with U1C and U1-70k, and SmB 1-91 to interact with SmD3.

There were problems with these constructs. The only complex that formed was between SmD3 and SmB. U1C was not able to pull down U1-70k or the SmD3/SmB complex. Likewise, U1-70k was not able to pull down U1C or the Sm proteins. The Sm proteins did not pull down either U1C or U1-70k. A 1:1:1:1 mixing experiment showed the proteins did not shift the 5' SS hairpin RNA (Figure 3-8B). We could not reconstitute the U1C interaction with the 5' SS using the protein fragments we expressed; something else was needed. The interface between U1C and the 5' SS duplex is small, and it is likely that the minimal U1 snRNP seen in the second crystal structure (with the Sm core, U1C, N-terminus of U1-70k, and

a truncated U1 snRNA) is approaching the smallest 5' SS binding unit. The U1 snRNA is needed to base-pair with the 5' SS, the Sm core and U1-70k are required to stabilize the U1 snRNA and the binding of U1C, and U1C stabilizes the RNA duplex.

3-7. Modified RNA hairpin

It is possible that the lack of observable stable binding between U1C and the 5' SS/U1 snRNA duplex is due to the small face of this interface. To constrain this interaction we engineered a disulfide bond between a modified RNA hairpin and cysteine mutants of U1C. Schellenberg et al. (2011) used a similar strategy to cross-link p14 to the branch A within an RNA hairpin mimicking the BPS/U2 snRNA duplex (Figure 3-9A). In a panel of p14 cysteine mutants a decrease in disulfide stability in the presence of reducing agent indicates the disulfide has formed between the protein and RNA in a less favoured orientation. A more stable disulfide suggests that thiols are able to quickly reform the reduced disulfide due to interactions between the protein and RNA holding the thiols in close proximity (Gilbert, 1995; Stanojevic & Verdine, 1995).

Based on the first U1 snRNP crystal structure we designed a panel of modified RNA hairpins and a set of U1C cysteine mutants. The RNAs featured a cystamine tether incorporated onto a backbone phosphate. The nitrogen of the cystamine replaces one of the oxygens on a backbone phosphate, with the thiol group on the other end of the carbon linker. Due to the synthesis chemistry, the

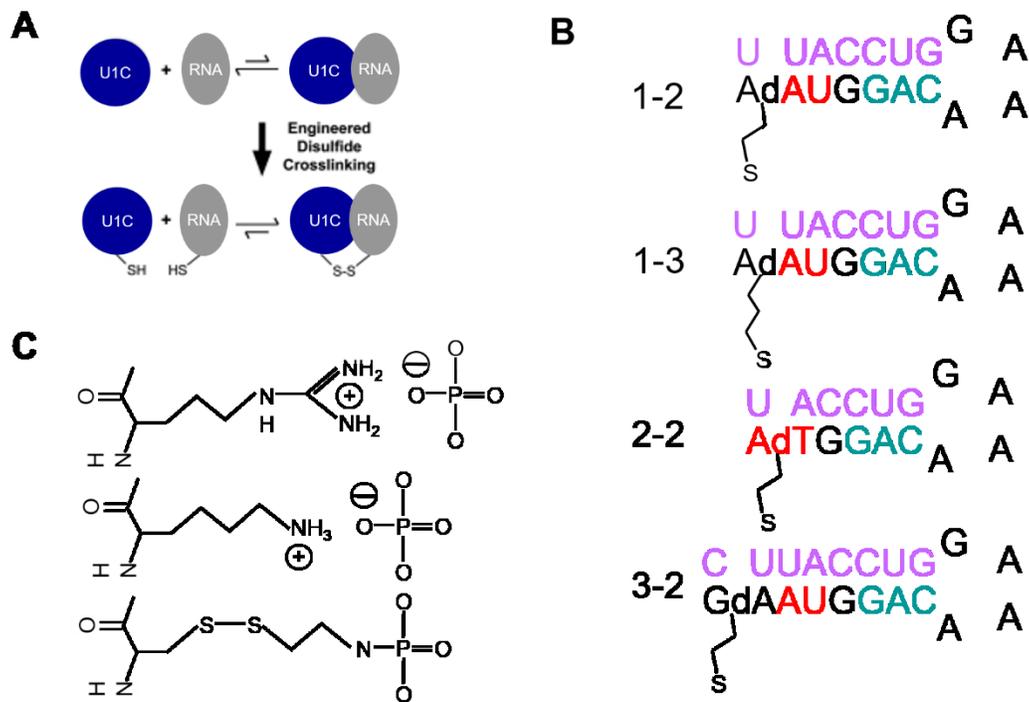


Figure 3-9. Design of disulfide cross-linking substrates. A. Schematic of an engineered disulfide between a modified RNA backbone and cysteine mutants of U1C. Adapted from Schellenberg et al. (2011). B. Modified RNA duplexes with a thiol tethered to the backbone between the last two nucleotides, using either a two (-2) or three (-3) carbon linker. Colours are consistent with Figures 3-6 and 3-8. C. The disulfide tethered to the backbone phosphate (bottom) replaces an ionic interaction between arginine (top) or lysine (middle) sidechains with a non-bridging oxygen of the phosphodiester backbone.

modification must be located on the phosphate between the final two bases, and the second to last base must be 2' deoxy. We synthesized three constructs with the thiol on a two carbon linker between the GUA*A of the 5' SS (construct 1-2: G/GUdA*A; / indicates the 5' SS, * is the tether); the GU*A of the 5' SS, (construct 2-2: G/GdT*A) and the GUAA*G (construct 3-2: G/GUAdA*G). We also synthesized construct 1 with a three carbon linker between the phosphate backbone and thiol (construct 1-3; Figure 3-9B).

The first 61 amino acids of U1C contain three cysteines: Cys6 and Cys9 that

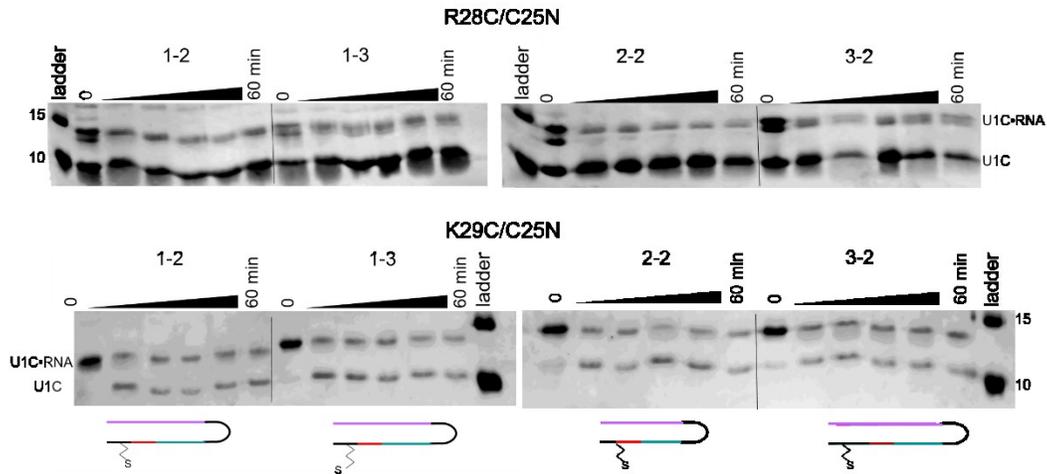


Figure 3-10. Cross-linking between a modified hairpin and U1C mutants. Reduction time course of a cross-link between U1C and a modified RNA as visualized by non-reducing SDS-PAGE. Top, R28C/C25N forms a less stable disulfide, and is reduced more quickly than that between K29C/C25N (bottom). RNA from Figure 3-9 shown in schematic at the bottom.

are involved in the zinc finger, and the surface exposed Cys25. To generate disulfides that mimic protein-RNA interactions we mutated C25N (the C25S mutant shows little effect on E complex formation; Nelissen et al., 1991) in conjunction with individual R28C and K29C mutations. The range lost by mutating the longer sidechains to the shorter cysteine is compensated for by the length of the thiol tether (Figure 3-9C). The three mutant U1Cs, C25N, R28C/C25N and K29C/C25N were stably expressed as MBP fusion proteins and purified on an amylose column. Subsequent release of the MBP by TEV protease did not destabilize the U1C.

The double mutants R28C/C25N and K29C/C25N, but not the singly mutated C25N, cross-linked to the modified RNA hairpins. Stability of the disulfide was tested by reduction time courses in which the protein-RNA complexes

were incubated in BME. More K29C/C25N forms disulfides with the RNA than R28C/C25N, and the K29C/C25N disulfides resist reduction more than the R28C/C25N (Figure 3-10). R28C/C25N seems to dimerize with itself in addition to forming a disulfide with the RNA. In *S. cerevisiae*, the K28A mutant shows more synthetic lethality with a variety of mutant proteins and RNAs than the N29A (Schwer & Shuman, 2014).

3-8. Modified pre-mRNA

In order to study the interaction of U1C with the 5' SS in the context of U1 snRNP we designed a modified pre-mRNA to trap the E complex. A reversible disulfide formed between a derivatized pre-mRNA and the cysteine of U1C within the U1 snRNP could stall dissociation of U1 snRNP, and allow for structural or functional studies.

The U1 snRNP can be reconstituted *in vitro* from recombinantly expressed proteins and *in vitro* transcribed U1 snRNA (Will et al., 1996; Pomeranz Krummel et al., 2009). We decided to reconstitute U1 snRNP with the cysteine mutants of U1C and stall spliceosome assembly by means of disulfide cross-links between the U1C and a modified pre-mRNA substrate.

We used the data from the cross-links formed between the U1C mutants and modified RNA hairpins to design a splicing substrate based on the PIP intron, a commonly used *in vitro* splicing substrate. The PIP intron contains short exons separated by an intron with canonical splice sites. It has been shown to splice *in*

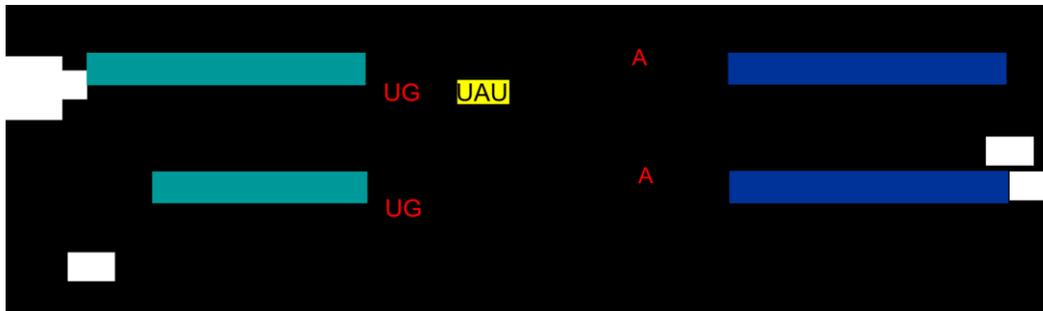


Figure 3-11. Modified PIP. Schematic of *in vitro* splicing substrate PIP based on construct 3-2. UAU (highlighted) was also deleted due to synthesis chemistry.

in vitro in HeLa derived splicing extracts (Moore & Sharp, 1992, Query et al., 1994). Because the K29C/C25N cross-link to construct 3-2 was the most stable we designed a single-stranded RNA molecule with a thiol tether on the backbone phosphate between the fourth and fifth nucleotides of the 5' SS (G/GUGdA*GU). The synthesized RNA (or an unmodified control) would then be ligated into a full length splicing construct with a commercially synthesized 5' piece and an *in vitro* transcribed 3' piece. The sequence of the 5' region of the intron is GUGAGUAUGG (5' SS underlined). Due to the synthesis protocol, requiring the modification to be between the last two nucleotides, and the *in vitro* transcription, which requires a G as the first nucleotide, we omitted UAU. The 5' end of the intron would then become GUGAGGG with a G->U substitution at the final position of the 5' SS (Figure 3-11). We expected this construct to splice even with the G->U mutation because Freund et al. (2005) showed that a construct GUAACA (with the final GU mutated to CA) is spliced.

The modified RNA ligated into the PIP intron blocked splicing (Figure 3-12). Transcribed wild type PIP, and PIP ligated with an unmodified middle piece

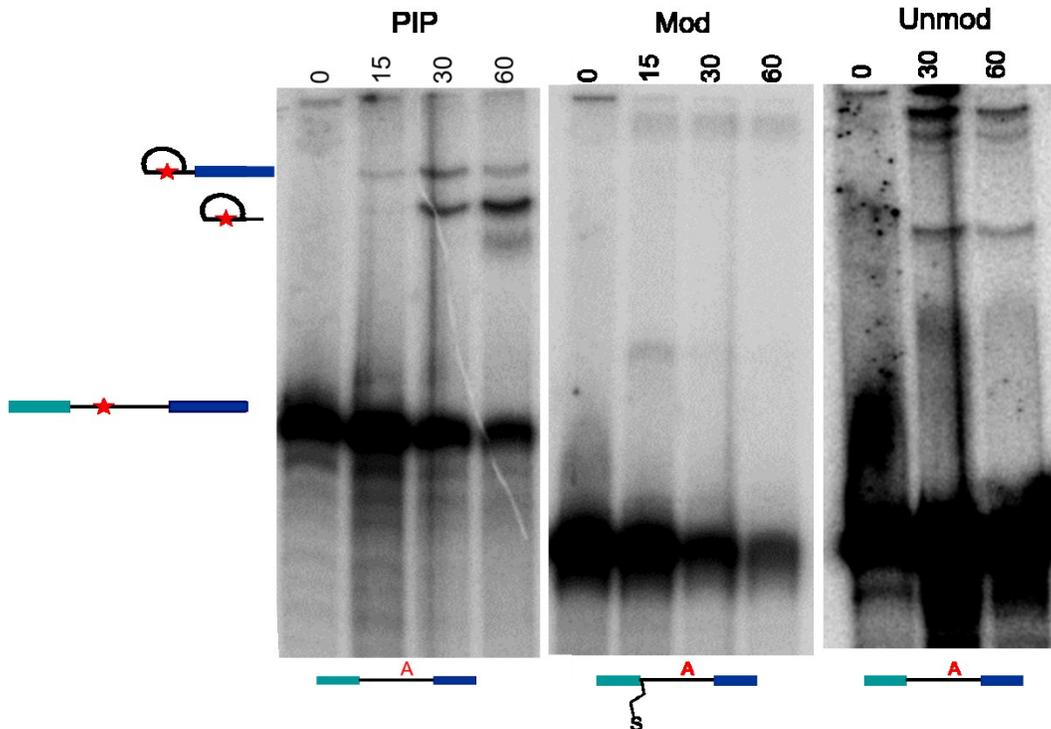


Figure 3-12. The modified intron does not splice.

Splicing time courses for three PIP substrates. Wild type transcribed PIP splices (left), as does the PIP reconstituted via ligation with an unmodified piece (right). PIP ligated with the modified 5' SS (middle) does not splice. Splicing reactions were run on a denaturing gel, with the radiolabel at the 5' end of the middle piece (indicated by a red star)

were spliced, but the modified PIP was not. The HeLa nuclear extracts contained wild type U1C, and it is possible that steric clashes between R28 or K29 blocked association of U1 snRNP with the modified RNA.

In order to confirm that the deleted UAU was not blocking U1 snRNA/5' SS binding we reworked the synthesis protocol to accommodate them. The 5' end of the intron became GUGA*GUAU. We ligated the new modification (or an unmodified control) into the PIP intron. *In vitro* splicing was not rescued by the additional UAU. The unmodified control and PIP were spliced.

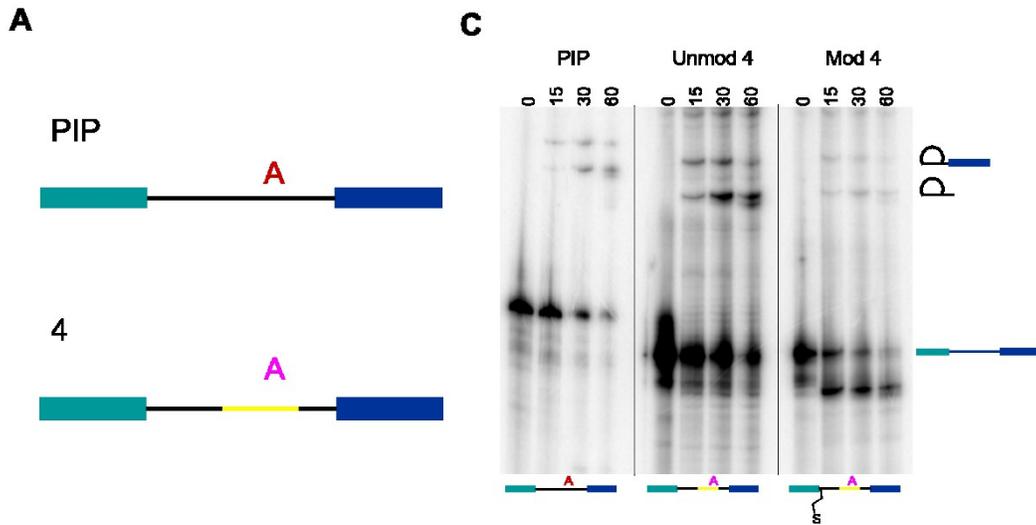


Figure 3-13. The modified intron splices in a U1 snRNP independent construct. A. The PIP/ β globin chimera (4) developed by Crispino et al. (1996) features changes around the BPS that allow splicing in the absence of U1 snRNP. B. Placing the modification within the context of the chimera allows U1 snRNP binding to be bypassed, and the pre-mRNA to be spliced.

3-9. Splicing in the absence of U1 snRNP

Some pre-mRNA sequences do not need U1 snRNP for splicing, and are able to be spliced in extracts depleted of U1 snRNP. Crispino et al. (1996) showed that a chimera of the PIP intron and another common substrate for *in vitro* splicing, the β -globin pre-mRNA, is spliced in U1 snRNP depleted extracts in a sequence-specific manner. Interestingly, it is changes to the BPS, rather than the 5' SS, that allow for this splicing.

To test if an inability to bind U1 snRNP was causing the inhibition of splicing for the modified pre-mRNA, we created the chimeric pre-mRNA with the modified oligo ligated in at the 5' SS. This new construct was spliced (Figure 3-13). The block in splicing caused by the tether was overcome with an intron sequence that does not require the binding of U1 snRNP for splicing. The tether

likely interferes with U1 snRNP binding, providing biochemical evidence that the 3' end of the 5' SS also interacts with U1 snRNP.

3-10. Spliceosome reconstitution

Next, we tried to reconstitute U1 snRNP with the R28C/C25N and K29C/C25N mutant U1Cs and test for their complementation of U1 snRNP depleted splicing extracts. We depleted U1 snRNP from HeLa nuclear extract (Crispino et al., 1994; Crispino et al., 1996). Western blots using antibodies against U1C showed a decrease of U1 snRNP (Figure 3-14A). We then attempted to rescue the depleted extract with the reconstituted U1 snRNP as a precursor to rescuing splicing of the modified intron with U1 snRNPs containing the U1C cysteine mutants (Figure 3-14B). However, splicing is inhibited at high KCl concentrations (over 60 mM), but the U1 snRNP was not stable at KCl concentrations less than 200 mM. To overcome this problem we first incubated U1 snRNP with the pre-mRNA to allow binding, then added the mixture to the U1 snRNP depleted extract for splicing. We were not able to rescue splicing in this way using the unmodified intron and wild type U1C (Figure 3-14C). As a result, we did not try complementation assays with the mutant U1 snRNPs and modified RNA.

3-11. Discussion

U1 snRNP binds the 5' SS via base-pairing between the pre-mRNA and U1 snRNA, with this interaction stabilized by U1C. Structural studies of the U1 snRNP

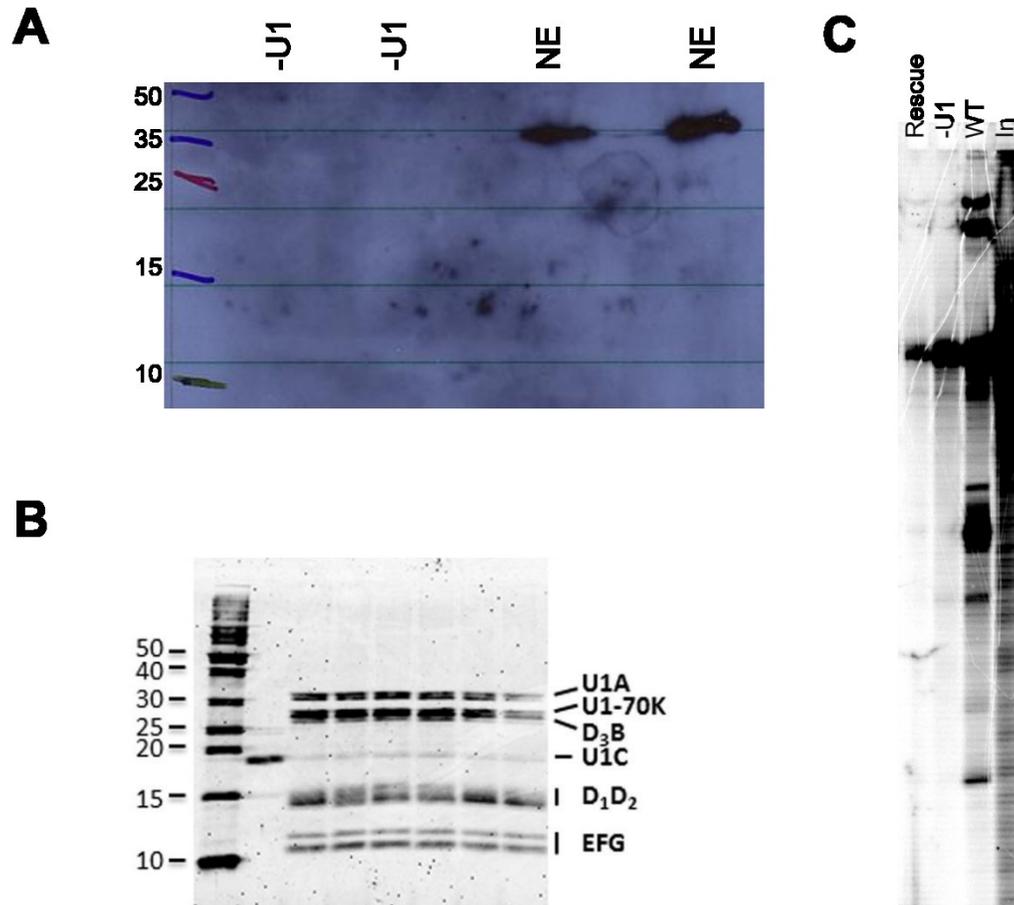


Figure 3-14. The reconstituted U1 snRNP does not rescue splicing in a U1 snRNP depleted extract. A. Western blot with anti-U1 antibodies shows depletion of U1 snRNP in HeLa nuclear extract. B. The U1 snRNP was reconstituted from *in vitro* transcribed U1 snRNA and recombinantly expressed U1 proteins (From McCarthy et al. (2017). Permission is not required to reprint). C. The reconstituted U1 snRNP did not rescue splicing in the U1 snRNP depleted extract.

suggest U1C recognizes the resulting duplex, but does not contact the pre-mRNA in a sequence specific manner. Arg28 and Lys29 are positively charged residues in position to interact with the U1 snRNA/5' SS duplex. However, structural evidence of this interaction is absent from the available crystal structures, due to the low resolution of the first U1 snRNP structure, and a lack of an interaction in the high resolution structure.

The recently published *S. cerevisiae* spliceosomal A complex, featuring the U1 and U2 snRNPs bound to the pre-mRNA shows the backbone of Lys28 hydrogen bonding to the phosphate backbone of the middle uridine of the 5' SS (GUAUGU), the position suggested by our cross-linking experiments. The positively charged side chain does not appear to be contacting the RNA. The side chain of Asn29 does contact the adenosine of the 5' SS.

Using the stability data from the interaction between the RNA hairpins and mutant U1C, we designed splicing constructs with a modified backbone. Ligation of modified oligos into pre-mRNA substrates showed that this modification blocked splicing in a U1-dependent manner, likely due to steric hindrance between the tether and long side chains of Arg28 and Lys29. Sequences able to splice in the absence of U1 snRNP were not inhibited by this modification, suggesting it blocks the U1 snRNP interaction, but do not affect the subsequent steps of spliceosome assembly and splicing.

Taken together, these experiments suggest an interaction between the human U1C Arg28 or Lys29 with the 5' SS/U1 snRNA duplex that has not been visualized in the available structures. Different binding or splicing conditions may be able to rescue U1 snRNP depleted extract with the reconstituted U1 snRNP, and a cross-link between U1C and the pre-mRNA could be used to stall spliceosome assembly.

3-12. Materials and Methods

3-12a. Protein expression

Human U1C residues 1-61 were cloned into the pMal plasmid with a cleavable TEV site. Mutants were generated by overlapping PCR and cloned into pMAL. U1C was transformed into *E. coli* Rosetta cells for protein expression. Cells were grown in LB supplemented with 5 g/L dextrose at 37°C until an OD₆₀₀ of 0.6, and induced with 1 mM IPTG and 1 mM ZnCl₂ overnight at 18°C. Cells were harvested by centrifugation, frozen at -20°C for 30 minutes, and resuspended in U1C lysis buffer (20 mM HEPES pH 7.5, 500 mM NaCl, 1 M urea, 5 mM BME) on ice. Cells were lysed by sonication, and cleared by centrifugation. Lysate was run over an amylose column, washed with U1C lysis buffer and eluted in U1C elution buffer (lysis buffer with 20 mM maltose). Elutions were concentrated, TEV cleaved overnight at 4°C, and run over a Superdex 75 column in standard gel filtration buffer (SGFB; 20 mM Tris pH 8, 100 mM NaCl, 5 mM BME). Protein-containing fractions were concentrated and stored at 4°C.

Human U1-70k residues 1-30 were cloned into the pGEX plasmid with a TEV cleavage site and transformed into *E. coli* Rosetta cells. Cells were grown at 37°C until an OD₆₀₀ of 0.6, induced 1 mM IPTG and grown overnight at 18°C. Cells were harvested by centrifugation, frozen at -20°C for 30 minutes and resuspended in GST lysis buffer (50 mM Tris pH 8, 100 mM NaCl, 5 mM BME). Cells were lysed by sonication and cleared by centrifugation. Cleared lysate was run over a GST column and washed with GST lysis buffer. Protein was eluted with GST

elution buffer (GST lysis buffer with 20 mM glutathione). Concentrated protein was run on the Superdex 75 column in SGFB. Fractions containing protein were concentrated and stored at 4°C.

Human SmB (residues 1-91) and SmD3 (residues 1-75) were cloned into the His-tagged and untagged sites of pACYC respectively. Plasmids were transformed into *E. coli* BL-21 gold cells. Cells were grown at 37°C until an OD₆₀₀ of 0.6, and induced with 1 mM IPTG overnight at 18°C. Cells were harvested by centrifugation and frozen at -20°C for 30 minutes. Cells were resuspended in Ni lysis buffer (50 mM Tris pH 8, 100 mM NaCl, 5 mM BME, 20 mM imidazole) and lysed by sonication. Lysate was cleared by centrifugation and run over a Ni-NTA column, washed with Ni lysis buffer and eluted with Ni elution buffer (Ni lysis buffer with an additional 200 mM imidazole). Protein was concentrated and run on the Superdex 75 column in SGFB. Fractions containing protein were concentrated and stored at 4°C.

3-12b. RNA transcription and purification

RNA was transcribed from double-stranded PCR templates. Each 400 µL reaction contained 1x transcription buffer, 1 mM each riboNTP, 5 µL PCR template, and 5 µL T7 RNAP at 37°C for 4 hours. The final Mg²⁺ concentration is 6 mM.

Following transcription, RNA was run on a 15-20% (depending on product size) 19:1 acrylamid:bisacrylamide denaturing gel with 8 M urea in 1x TBE. RNA

was visualized by UV shadow and sliced from the gel. Gel slices were incubated, shaking, overnight in 400 μ L 0.3 M sodium acetate and 100 μ L phenol. The RNAs were filtered through spin columns to remove gel, and the supernatant was then phenol/chloroform extracted and ethanol precipitated. Pelleted RNA was resuspended in RNase-free water and stored at -20°C .

3-12c. Radiolabelling

RNA was 5' end labeled with ^{32}P γ -ATP. Transcribed RNAs were CIAPed before labeling to remove the 5' phosphate according to the manufacturer's instructions, phenol/chloroform extracted and ethanol precipitated. Kinase reactions were carried out in 1x forward reaction buffer and 1 μ L hot ATP per 20 μ L reaction. Ellman's reagent (DTNB) was added to 12.5 mM to reprotect modified RNA. Kinased RNA was run on an acrylamide denaturing gel and visualized by autoradiography. Hot RNA was cut from the gel and extracted overnight in 400 μ L 0.3 M sodium acetate and 100 μ L phenol. The RNAs were filtered through spin columns to remove gel, and the supernatant was then phenol/chloroform extracted and ethanol precipitated. The RNA pellet was resuspended in RNase-free water and stored at -20°C .

3-12d. Electrophoretic mobility shift assays (EMSAs)

EMSAs were carried out by incubating increasing concentrations of protein with radiolabeled RNA in 10 mM HEPES pH 7.9, 60 mM KCl, 2 mM MgCl_2 , and

0.1 mM EDTA. The reactions were then run on a 6% 80:1 acrylamide:bisacrylamide native gel in 1x Tris glycine. The gel was dried in a gel drier, and exposed overnight to a storage phosphor screen. The phosphor screen was imaged on a Typhoon scanner.

3-12e. RNA synthesis

Modified RNA was synthesized in accordance with Schellenberg et al. (2011) and McCarthy et al. (2017).

RNAs were synthesized on an ABI 394 synthesizer using standard protocols with 2'-TBDMS protection modified to incorporate the thiol tether during synthesis. Following the first detritylation step, the RNA synthesis column was removed from the machine, and the following steps were performed manually using a syringe to introduce reagents to the column: (1) wash with 1 mL of 50:50 acetonitrile:pyridine; (2) coupling with 28 mg of the appropriate 5'-trityl 3'-H phosphonate monomer (thymidine, deoxyguanosine, or deoxyadenosine; ChemGenes) dissolved in 1 mL of 50:50 acetonitrile:pyridine and 40 μ L of pivaloyl chloride (Sigma) for 60 sec; (3) wash with 1 mL of 50:50 acetonitrile:pyridine; (4) wash with 5 mL of acetonitrile; (5) oxidation for 1 h with 900 μ L 50:25:25 carbon tetrachloride:pyridine:cystamine disulfide (freshly dried over $MgSO_4$ in dichloromethane); (6) wash with 5 mL of pyridine; (7) wash with 5 mL 50:50 acetonitrile:pyridine; and finally, (8) wash with 5 mL of acetonitrile. During the reaction steps, a second syringe was attached to the column to allow the reagents to

be pushed back and forth across the solid support. The column was returned to the ABI 394 synthesizer and capped with acetic anhydride followed by detritylation, and the synthesis was continued using standard 2'-TBDMS chemistry. The oligonucleotides were deprotected according to the manufacturer's instructions with the added steps of addition of 0.1 M β -mercaptoethanol during base deprotection. The 2'-TBDMS were removed with 1 M tetrabutylammonium fluoride in tetrahydrofuran (room temperature, 24 hr in the dark). These reactions were quenched in 50 mL 0.1 M TEAB, with 5 mM DTT. Following desalting (Waters C-18 Sep Pak), crude oligonucleotides were then modified with 5,5-dithiobis-(2-nitrobenzoic) acid (Ellman's reagent) prior to separation by denaturing PAGE (20%, 19:1 acrylamide:bisacrylamide, 8 M urea) in TBE running buffer. The band corresponding to the full-length oligonucleotide containing the thiol modification was identified, excised, and extracted from the gel slice overnight in 400 μ L 0.3 M sodium acetate and 100 μ L phenol. Following this, the RNAs were filtered through spin columns to remove gel, phenol/chloroform extracted and ethanol precipitated. The precipitated RNAs were resuspended in RNase free water and quantified by UV absorbance at 260 nm.

3-12f. Cross-linking and Reduction time courses

R28C/C25N and K29C/C25N were dialyzed overnight into U1C-1 buffer (10 mM Tris pH8, 60 mM KCl, 1 mM BME), then into U1C-0.1 buffer (10 mM Tris pH8, 60 mM KCl and 0.1 mM BME) for an hour at 4 °C. Proteins were cross-

linked to modified RNA in 60 μ L U1C-0.1 buffer with 0.9-1.8 nMol U1C and RNA at room temperature for 15 minutes, then overnight at 4°C. Cross-links were reduced in BME and quenched at 0-60 minutes with 5 mM iodoacetamide. The protein was then run on an 18% 200:1 acrylamide:bisacrylamide SDS gel in non-reducing SDS dye and either silver stained or colloidal Coomassie stained. The gels were visualized with a Li-Cor scanner and quantified with ImageQuant.

3-12g. RNA ligation

200 pMol middle piece were kinased as described above. The modified middle piece was protected with DTNB as above. To the kinase reaction were added 300 pMol 5' piece, 100 pMol transcribed 3' piece that had been CIAPed and kinased, 100 pMol DNA bridge and 1 μ L 10 mg/mL glycogen. The RNA mix was phenol/chloroform extracted and ethanol precipitated together. The pellet was resuspended in 5 μ L 10x ligase buffer and 33.5 μ L RNase free water. The mixture was heated to 70°C for 5 minutes and cooled slowly to room temperature to anneal the RNAs with the DNA bridge. 1 μ L 1 mM ATP, 0.5 μ L RNase out and 10 μ L ligase were added and the mixture incubated at room temperature for 6 hours to overnight. The reaction was run on a 10% denaturing acrylamide gel, and visualized by autoradiography. The ligated product was cut out of the gel and incubated overnight in 400 μ L 0.3 M sodium acetate and 100 μ L phenol. The ligated RNA was phenol/chloroform extracted and ethanol precipitated. The pellet was resuspended in RNase free water and stored at -20°C.

3-12h. Splicing

Splicing reactions were carried out under standard HeLa splicing conditions. Labeled RNA was mixed with 2 mM, 20 mM KCl, 1 mM ATP, 5 mM creatine phosphate, 2% RNase out and 40% HeLa nuclear extract, then incubated at 30°C for 1-60 minutes. 10 µL of reaction was quenched in 3.8 µL stop solution (0.06% SDS, 50 mM EDTA and 50 mM Tris pH 8) and 0.2 µL proteinase K at 65°C for 5 minutes. The RNA was then phenol/chloroform extracted and ethanol precipitated. The pellet was resuspended in 5 µL RNA dye (1x TBE, 10 M Urea, xylene cyanol and bromophenol blue) and run on a 15% acrylamide gel in 1x TBE. The gel was exposed to a storage phosphor screen and visualized with a Typhoon scanner.

3-12i. U1 snRNP depletion

To prepare the extract for depletion 400 µL HeLa nuclear extract was thawed quickly at 30°C then dialyzed 3 x 40 minutes into MD 0.6 buffer (20 mM HEPES pH 7.9, 10% glycerol, 0.6 M KCl, 0.2 mM EDTA, 0.5 mM DTT, and 0.5 mM PMSF). After dialysis the volume was increased to 600 µL with MD 0.6 buffer. 4 nM biotinylated anti U1 snRNA oligo, 1.5 mM ATP, 5 mM creatine phosphate and 0.05% NP 40 was added, then incubated at 30°C for 45 minutes.

To preblock streptavidin beads 2 x 300 µL beads were spun down at 3700 rpm for 1 minute and the supernatant removed. Beads were preblocked in 600 µL

blocking buffer (20 mM HEPES pH 7.9, 0.01% NP 40, 250 mM KCl, 50 μ g/mL glycogen, 0.5 mg/mL BSA, 50 μ g/mL tRNA and 0.5 mM DTT) at 4°C for 15 minutes. Beads were centrifuged as before and supernatant removed. Beads were washed 4 x 600 μ L in wash buffer (20 mM HEPES pH 7.9, 0.01% NP 40, 250 mM KCl and 0.5 mM DTT), spinning 4000 rpm for 20 seconds.

For antisense affinity depletion of U1 snRNP the nuclear extract was added to first set of beads and incubated rotating at 4°C for no longer than 45 minutes. Beads were spun down at 4000 rpm for 1 minute and the supernatant transferred to the second set of beads. Incubation and depletion were repeated. Beads were removed by centrifuging 2 x 4000 rpm for 1 minute. The extract was concentrated back to 400 μ L and dialyzed into buffer D (20 mM HEPES pH 7.9, 0.2 mM EDTA, 100 mM KCl, 0.5 mM DTT and 20% glycerol) 3 x 40 minutes at 4°C. Aliquoted extract was snap frozen in liquid nitrogen and stored at -80°C.

3-12j. Binding reconstituted U1 snRNP to pre-mRNA

1 μ L recombinant U1 snRNP (13.3 μ M) was incubated with 0.5 μ L hot PIP per reaction at 4°C for 15 minutes. The U1 snRNP mixed with PIP was then added to splicing reaction with U1 snRNP depleted extract and splicing performed as above.

Chapter 4
The Interaction of Sap49 with Sap145 and RNA

4-1. SF3b

Splicing factor 3 B (SF3b) is a seven protein subcomplex of the U2 snRNP that dissociates under high salt conditions (Figure 4-1; Behrens et al., 1993). SF3b recognizes and sequesters the branch adenosine until the spliceosome is correctly assembled, to prevent premature splicing at incorrect splice sites (Lardelli et al., 2010). Prp2 releases SF3b from the spliceosome in the transition from the B^{act} to B* complex immediately prior to the first step of splicing (Warkocki et al., 2009).

Human SF3b contains seven proteins, including Sap49, Sap145, Sap155, Sap130, SF3b10, p14 and PHF5A (Golas et al., 2003). *S. cerevisiae* SF3b is composed of six proteins: Hsh49 (Sap49), Cus1 (Sap145), Hsh155 (Sap155), Rse1 (Sap130), Ysf3 (SF3b10), and Rds3 (PHF5A). There is no p14 homologue in *S. cerevisiae* (Dziembowski et al., 2004).

Sap49 contains two RNA recognition motifs (RRMs; discussed below. Champion-Arnaud & Reed, 1994). The structure of Sap145, a series of short α -helices connected by long loops, was not known prior to the publication of the *S. cerevisiae* B complex structure (Plaschka et al., 2017). Sap155 contains 20 or 22 C-terminal HEAT repeats, which are composed of a pair of antiparallel α -helices (Wang et al., 1998; Cretu et al., 2016). The HEAT repeats stack into a curve that encircles SF3b. Sap130 is composed of three β -propeller domains followed by an α -helical C-terminal domain. SF3b10 is a small protein composed of three α -helices, and PHF5A folds as a triple knot (Cretu et al., 2016). p14 consists of an RRM that binds to Sap155 (Schellenberg et al., 2006), and cross-links directly to

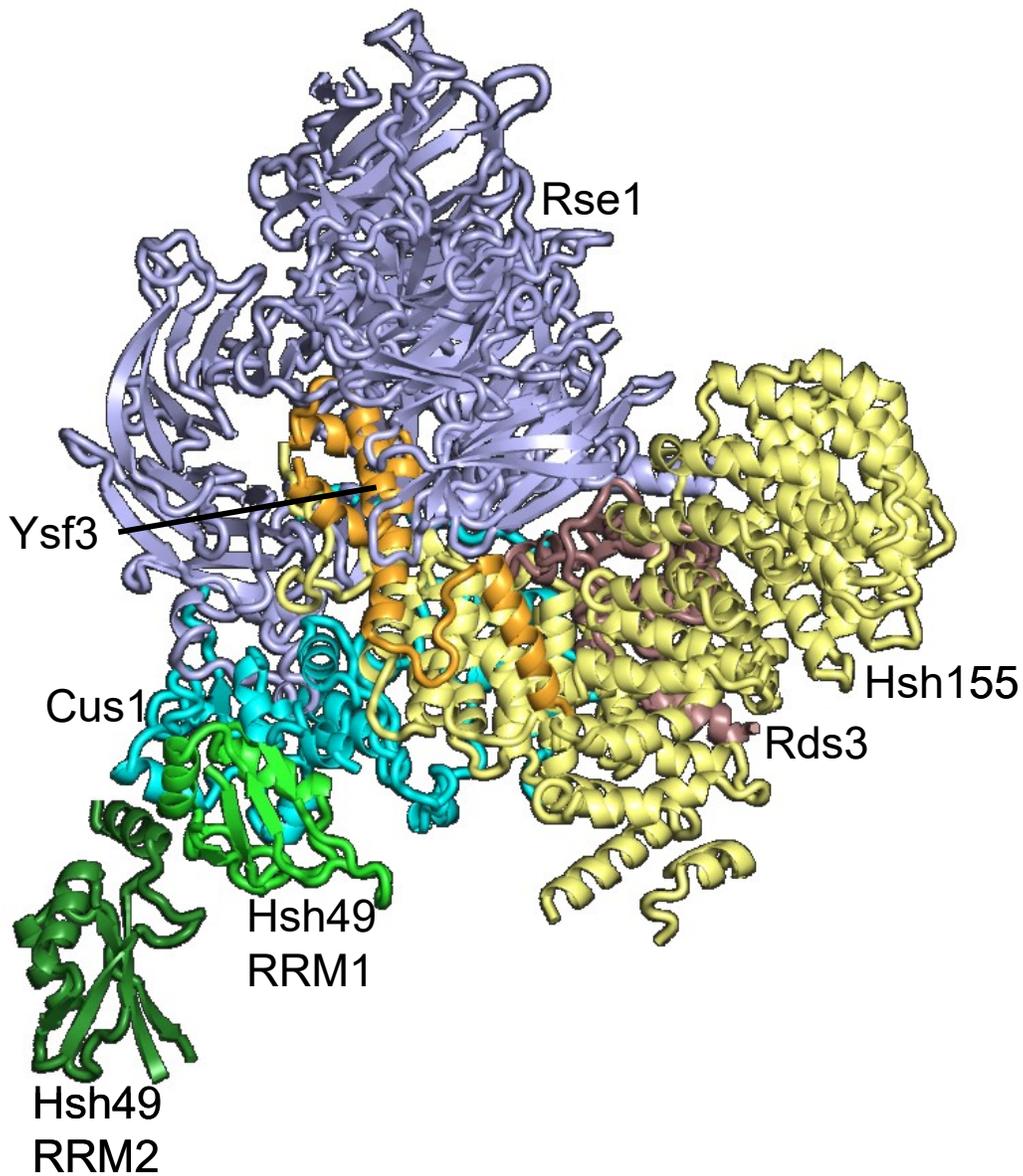


Figure 4-1. Structure of SF3b. Structure of the U2 snRNP subcomplex SF3b from the *S. cerevisiae* A complex (PDB 6G90) is composed of six proteins: Hsh155 (yellow), Rse1 (blue), Rds3 (brown), Ysf3 (orange) Cus1 (teal) and Hsh49 (green).

the branch adenosine (MacMillan et al., 1994).

A low resolution cryo-EM structure of the SF3b particle (Golas et al., 2003) placed the HEAT repeats of Sap155 in an S-shape around the complex with p14

located within the center of SF3b. The two RRM of Sap49 were modeled at the edge of the structure. At the time, the structures of the other SF3b components were not known, and were not placed into the cryo-EM envelope.

More recently, the crystal structure of the human SF3b core, consisting of Sap155, Sap130, SF3b10, and PHF5A, showed extensive contacts between the four proteins (Cretu et al., 2016). Sap155 surrounds the particle, with its termini anchored by PHF5A. The C-terminus of Sap155 is bound to Sap130 in a cleft between β -propeller domains that also binds SF3b10. This arrangement was confirmed in a high resolution cryo-EM structure consisting of the same four SF3b components (Finci et al., 2018).

The *S. cerevisiae* A complex reveals the structure of SF3b within the U2 snRNP. Alignment of the human SF3b core with SF3b from the *S. cerevisiae* A complex reveals a similar structure for the protein components, although their relative orientations are different. PHF5A bridges the two ends of Sap155, as does its homologue Rds3. However, while PHF5A sits at the edge of the circle formed by Sap155, Rds3 is located towards the center.

4-2. Sap49 and Sap145

The essential gene Sap49 (Hsh49 in *S. cerevisiae*) encodes a small protein component of the SF3b complex featuring tandem N-terminal RRMs (RRM1: 15-86; RRM2: 102-174; Figure 4-2A). RRMs are composed of a β 1- α 1- β 2- β 3- α 2- β 4 secondary structure folded into a four-stranded β -sheet buttressed by the two helices

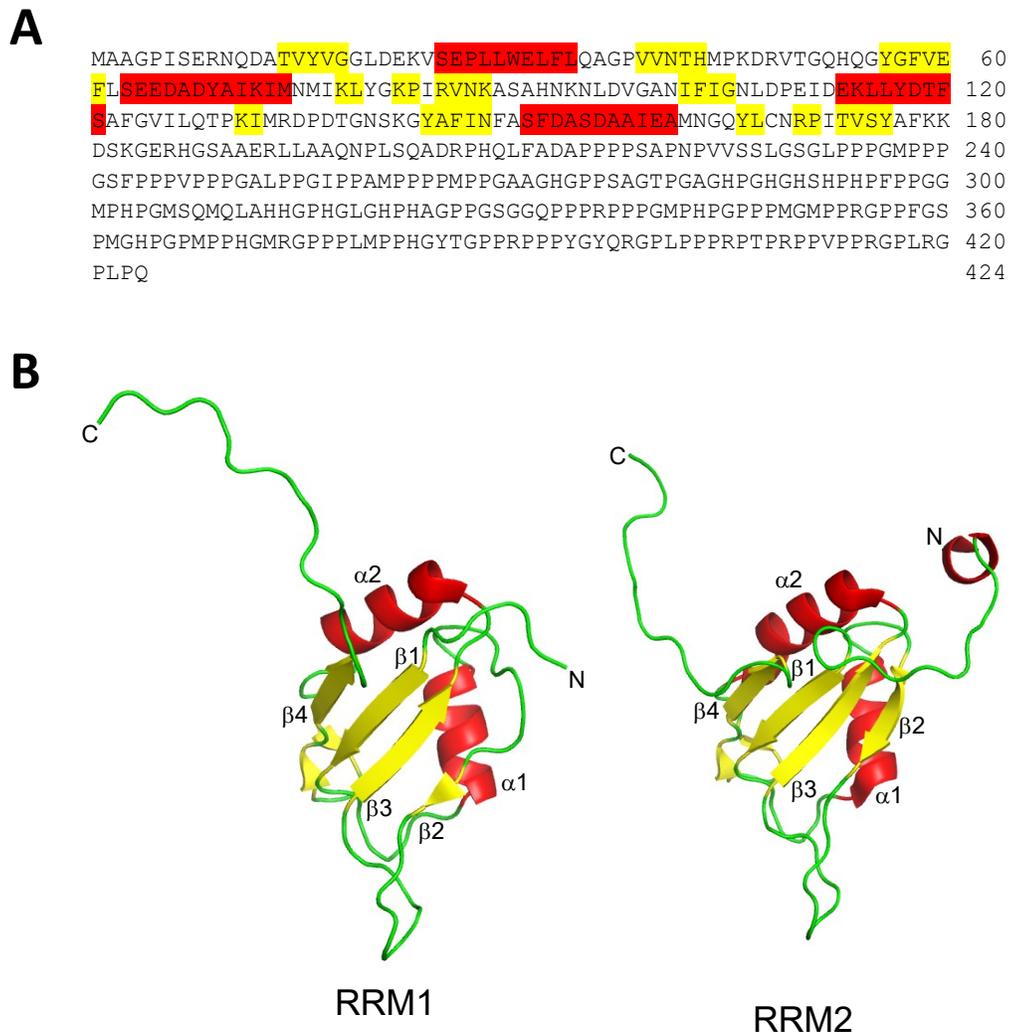


Figure 4-2. Structure of the human Sap49. A. Sequence of Sap49. Sap49 (Hsh49 in *S. cerevisiae*) is composed of two RNA recognition motifs (RRMs) consisting of a four-stranded β sheet and two α helices with RRMs highlighted by secondary structure (yellow β -strand; red α -helix) followed by a long tail predicted to be unstructured. B. NMR structures of the two RRMs of Sap49 (RRM1: PDB 5GVQ; RRM2: PDB 1X5T).

against one side of the β -sheet (Figure 4-2B). Two conserved motifs, RNP1 and RNP2 in $\beta 3$ and $\beta 1$ respectively, contain aromatic residues on the face of the β -sheet that stack with RNA bases upon binding of single-stranded RNA (Oubridge et al., 1994). While the β -sheet binds RNA, the α -helices may participate in protein-

S. cerevisiae	-----MNYSDSGNTVYVGNIDPRITKEQLYELFIQINPV-----	35
S. pombe	-----MSIREDRNQDATIYLGNDLDEKVTDSILFELCLQAGPV-----	37
C. elegans	MPKYFFLRSEMSAGPIVERNQDATIYVGGLEDEKVESILWELMVQAGPV-----	49
D. melanogaster	-----MAAGPIAERNQDATIYAGGLDDKVSETLLWELFVQAGFVKGSGHTCLVPS	50
H. sapiens	-----MAAGPIISERNQDATVYVGGLEDEKVSEPLLWELFLQAGPV-----	39
X. laevis	-----MAAGPIISERNQDATVYVGGLEDEKVSEPLLWELFLQAGPV-----	39
 * : * * . * : : . . * : * * . * . *	
S. cerevisiae	-----LRIKYPKDKVLQAYQGYAFIEFYNQGDAQYAIKIMNNTVRLYDRLIKVRQVT	87
S. pombe	-----VNIHIPRDRVRNNSHNGFGFCEFLHEQDVEYACQILN-QVKLFGKPIRVNRAS	88
C. elegans	-----VSNMMPKDRVTANHQGFVFEFMGEEDADYAIKILN-MIKLYGKPIKVNKAS	100
D. melanogaster	SSLIVMLAVNVHMPKDRVTQMHQGYGFVEFLSEEDADYAIKIMN-MIKLYGKPIRVNKAS	109
H. sapiens	-----VNTHMPKDRVTGQHQGYGFVEFLSEEDADYAIKIMN-MIKLYGKPIRVNKAS	90
X. laevis	-----VNTHMPKDRVTGQHQGYGFVEFLSEEDADYAIKIMN-MIKLYGKPIRVNKAS	90
	: : * : * * : : * : * * : * : * . : * * : : * : * . * : * : *	
S. cerevisiae	NSTGTTNLPNSISKDMLPIAKLFIKLNLAIDSQDLVKIFNKFGKLIPEPEIFYLSN-G	146
S. pombe	QDRGV-----NTLIGANLFLVGNLDPVDERVLYDTFSALGQLVKAPQVARD-ENG	137
C. elegans	AHE-K-----NMDVGANIFVGNLDPVEDEKLLYDTFSAFGVILQVPKIMRDVDSG	149
D. melanogaster	AHQ-K-----NLDVGANIFIGNLDVEVDEKLLYDTFSAFGVILQTPKIMRDPETG	158
H. sapiens	AHN-K-----NLDVGANIFIGNLDPEIDEKLLYDTFSAFGVILQTPKIMRDPDTG	139
X. laevis	AHN-K-----NLDVGANIFIGNLDPEIDEKLLYDTFSAFGVILQTPKIMRDPDTG	139
	* : * : * * * : * . * . * . * . * : : * : * : *	
S. cerevisiae	-KLKCAVYFEDFEKADLAIKSLNNQLVANNRITVDYAFKENGKGNKYGDDVDRLLNKE	205
S. pombe	RSKGYGFVSYDSFETADAAIEAMNNQFLMNKPI TVSYAFKREGKGERH-GDIAERKLA	196
C. elegans	TSKGFAPINFASFEASDTALEAMNGQFLCNRAITVSYAFKRDGKGERH-GTAAERMLAAQ	208
D. melanogaster	KSKSFAPINFASFEASDAAMDAMNGQYLCNRPISVSYAFKDKHGERH-GSAAERLLAAQ	217
H. sapiens	NSKGYAFINFASFDASDAAIEMNGOYLCNRPITVSYAFKDKSGERH-GSAAERLLAAQ	198
X. laevis	NSKGYAFINYSFDASDAAIEMNGOYLCNRPITVSYAFKDKSGERH-GSAAERLLAAQ	198
	. . : : . * : * * : : * : * . * . * . * . * : : * : * . * : * *	
S. cerevisiae	ALKHNMLK-----	213
S. pombe	AKKNKVAVTPQS--TLPPGFSPATPAPSAANTPATIAA-----TSIPPVNPV--	243
C. elegans	NPLFPKDRPHQVFSVDVPLGV--PANTPLAMPVGHAAIAAHATG----RPGYQPPPLMGMA	262
D. melanogaster	NPSTHADRPHQLFADAPVQT--MMFQ-----MPGQIPAQMPGQMPPPMMAPPVVPV	269
H. sapiens	NPLSQADRPHQLFADAPPPP--SAPN-----PVVS----SLGSLGPPPGMPPPGSFPFP	246
X. laevis	NPLSQADRPHQLFADAPPPP--SVPA-----VITS----LT-SAVA-AGI----PTFPF	240
S. cerevisiae	-----	213
S. pombe	---LVGATTAVPP-LS---IPNV-----	259
C. elegans	QSGYQGQYPPVPPPPSVTPMPPMPPPTPGMTPRPPPPSSGMWPPPPPPPGRT----	317
D. melanogaster	SNNNMGL-APPPVVPQAPFPATIPP-----PPLPMT-----	302
H. sapiens	-----VPPPGALPPGIPPMAPP-----PPMPFGAAGHGPPSAGTPGAGHPGHG	289
X. laevis	-----VPPPGAMPPIPPSMPP-----PPMSPVTGA-----GQATAA	272
S. cerevisiae	-----	213
S. pombe	----LPFTA-AQHFFGMPAMPMMN-----VPMGGGAPLVPPPPGMVMASPS	303
C. elegans	-----PGPPGMPGMPPPP-----PPSRFGPPGGMGMP--PPPPGMRYPGGMP	358
D. melanogaster	-----GGQPPL-P--PAMGIPPPPR	319
H. sapiens	HSHPHFPFGMPHPGMSQMLAHHGPHGLGHPHAGPPGSGGQPPP-RPPPGMPHPGPPP	348
X. laevis	PQVLPFQSTAM-HPGM-QMQIPH-----PS-----IPG-TRPPGMGPPGPPP	312
S. cerevisiae	-----	213
S. pombe	AAA---T-----IPGA-----PVMPNIPFYQTIANAQNGYSQQQRR-----	335
C. elegans	PPPPRYPSAG----PG-----MYPPPP-SRPPAPPS---GHGMIPPPPPP	397
D. melanogaster	MMQPNAWAPPGMPAP-PPRPPTNWRPVPVFPPTPY-ARPYQPDGYQY-----	366
H. sapiens	MGMPPRPPFGSPMGHPGMPHPGMRGPPPLMPPHY-TGPPRPPPYQYRGLPFPPT	407
X. laevis	MGMPPRAPPPFGA-MGHPGMP--PGMR-PPPLMP--PY-NGPPRPPPYQYRQAPLPLPRPA	365
S. cerevisiae	-----	213
S. pombe	-----	335
C. elegans	S-----	398
D. melanogaster	-----	366
H. sapiens	PRPPVPRGRLRGLPQ	424
X. laevis	P-----PRPPMLPMTQ	377

Figure 4-3. Sequence alignment of Sap49 from six species. The human RRM8s are highlighted in green. *S. cerevisiae* lacks the N-terminal tail.

protein interactions. Relative to *S. cerevisiae*, metazoan Sap49 has a long proline- and glycine-rich C-terminal extension made up of amino acids not predicted to be structured (Champion-Arnaud & Reed, 1994; Figure 4-3).

Sap49 is required for efficient splicing, and can be cross-linked to RNA 25 nucleotides upstream of the branch adenosine in the A complex (Gozani et al., 1996). It was not shown to cross-link to either exon or the 5' end of the intron (Champion-Arnaud & Reed, 1994). Mutations to the conserved aromatics in the RNP motif of the RRM's shown to be essential for RNA binding in other RRM's are lethal in *S. cerevisiae*. Tandemly expressed Hsh49, each bearing a single mutation in either RRM1 or RRM2 does not rescue Hsh49 knockout in cells, suggesting that both RRM's of a single molecule of Hsh49 are required for viability (Igel et al., 1998).

Sap145 (Cus1 in *S. cerevisiae*), a 97 kDa protein required for the stable association of U2 snRNP with the spliceosome, was initially identified as a protein that suppressed cold sensitive U2 snRNA mutations (Wells et al., 1996). It is divided into two domains of unknown structure and function based on sequence comparison with the Pfam database. Cus1 residues 121-392 are essential for viability in *S. cerevisiae* (Pauling et al., 2000).

A C-terminal region of Sap145 interacts with the first RRM of Sap49 (Champion-Arnaud & Reed, 1994): it pulls down, and is pulled down by, Sap49 even in the absence of RNA (Igel et al., 1998). While mutations to the RNP's of

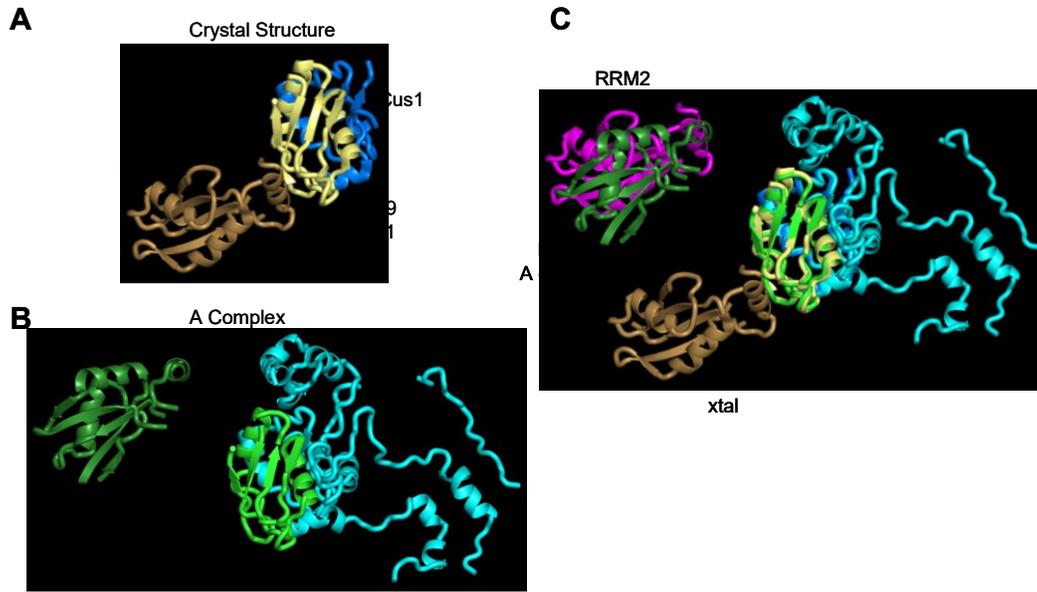


Figure 4-4. Interaction of Hsh49 with Cus1. A. Crystal structure of Hsh49 (yellow and tan) with Cus1 290-368 (blue; PDB 5LSB). B. The interaction of Hsh49 (green) and Cus1 from the A complex (PDB 6G90) is similar to that from the crystal structure. C. Overlay of Hsh49/Cus1 crystal structure (RRM1 yellow, RRM2 brown) with that of the A complex (RRMs green) via RRM1 shows a displacement of RRM2. An RRM2 from a symmetry mate (magenta) overlays more closely with RRM2 from the A complex.

Sap49 are lethal in *S. cerevisiae*, they do not impair binding to Sap145 *in vitro*.

Sap145 also cross-links to RNA about 15 nucleotides upstream of the branch region (Gozani et al., 1996).

4-3. Structures of Sap49, Hsh49, and Cus1

The structures of both of human Sap49 RRM1 and RRM2 were solved by NMR (PDB 5GVQ and 1X5T; Kuwasako et al., 2017). They show a canonical RRM fold and a hydrophobic region between the α -helices (Figure 4-2B). Titration of Sap49 RRM1 with a fragment of Sap145 spanning residues 598-631 produced changes in the chemical shifts of residues on helix α 1, indicative of protein-protein interaction.

NMR of the Sap145 fragment alone shows it to be unstructured; however, it becomes structured upon interaction with Sap49 RRM1 (Kuwasako et al., 2017). The Sap145 fragment 609-621 is suggested to contain the residues that interact with Sap49. Leu28, Leu32, and Tyr80 of Sap49 are thought to form part of the interface between Sap49 and Sap145. Individual mutation of each residue to alanine, while not changing the structure of the RRM, abolished Sap49 pull down with Sap145.

Van Roon et al. (2017) solved the crystal structures of a fragment of Cus1 interacting with either RRM1 or the full length Hsh49, consisting of the two canonical RRMs (Figure 4-4A). Unfortunately, the linker between the RRMs was not visualized, and it is impossible to tell which RRM1 is covalently linked to which RRM2 within the three Hsh49 molecules in the asymmetric unit. The linker between the RRMs is long enough to permit domain exchange between molecules of Hsh49. The α -helical C-terminal extension of RRM2 is observed to fold back and interact with both RRMs, likely due to crystal packing. The Cus1 fragment (290-368) features two short α -helices, and two short β -strands. The C-terminal 18 amino acids are not visualized. The α -helical faces of both Hsh49 RRMs make hydrophobic interactions with Cus1. Mutation of these hydrophobic contacts to alanine disrupts complex formation.

Both RRMs of Hsh49 are visible in the *S. cerevisiae* A complex (Plaschka et al., 2018; Figure 4-4B). They are further apart than predicted by the crystal structure. Upon alignment via RRM1, one of the symmetry-related RRM2s, as opposed to the closest RRM2, is positioned in approximately the same place as

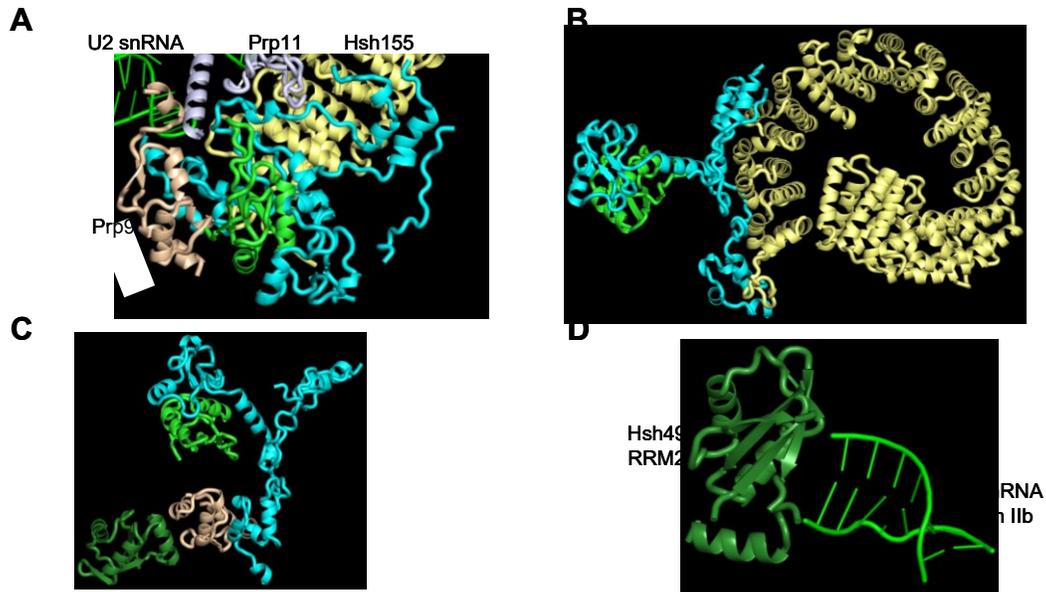


Figure 4-5. Cus1 and Hsh49 in the A complex. A. Cus1 (cyan) interacts with Hsh155 (yellow), Prp11 (light blue), Prp9 (tan) and Hsh49 (green). B. The N-terminal region of Cus1 lies along the HEAT repeats of Hsh155. C. Prp9 binds between the RRM1 and RRM2 of Hsh49. D. Hsh49 RRM2 may interact with the loop of U2 snRNA stem IIb.

RRM2 from the A complex (Figure 4-4C). This superposition may be a coincidental result of crystal packing. Alignment of Hsh49 via RRM1 aligns the crystallized fragment of Cus1 very closely to the A complex Cus1.

In the A complex, the visualized Cus1 fragment (125-376) is predominantly α -helical, and forms two regions that interact with a number of proteins (Plaschka et al., 2018; Figure 4-5A). The N-terminal region, 125-274, forms a flat sheet that binds the HEAT repeats of Hsh155 (Figure 4-5B). The N-terminus is positioned to interact with the loop of U2 snRNA stem IIa, Prp9, and the α -helical extension of the third β propeller of Rse1. The other half of the Cus1 N-terminal region interacts with Prp11. The C-terminal region is perpendicular to the N-terminal region and encircles Hsh49 RRM1. Hsh49 RRM1 binds a fragment of the intron upstream of

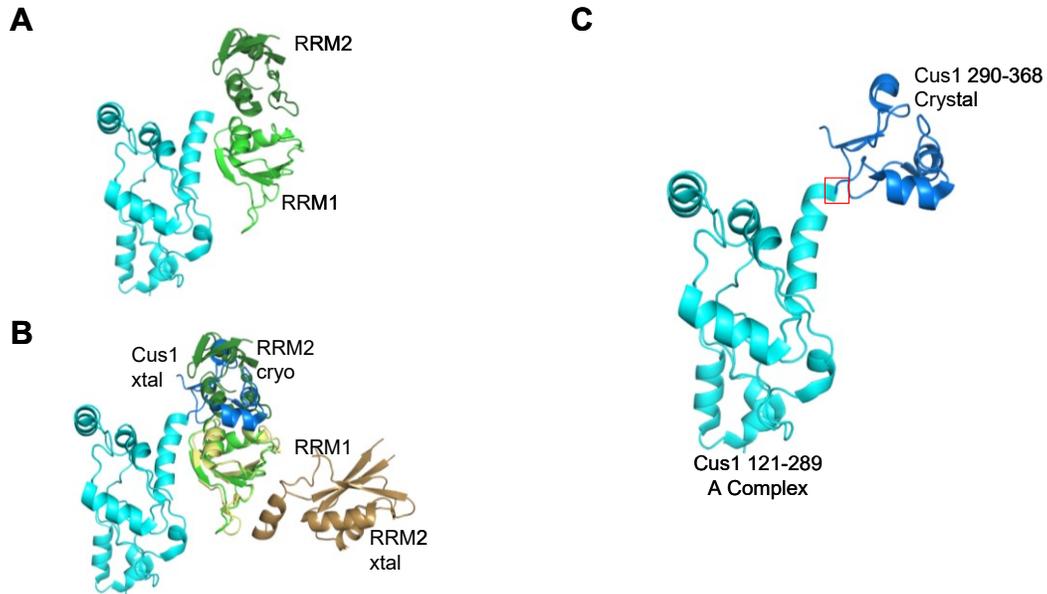


Figure 4-6. Hsh49 and Cus1 in the B^{act} structure. A. Hsh49 (green) and Cus1 (teal) in the B^{act} structure (PDB 5GM6). B. Overlay of the crystal structure of Hsh49 and Cus1 (PDB 5LSB) with the B^{act} structure. When aligned via RRM1 (green B^{act}; yellow crystal), the Cus1 fragment (blue) superimposes with RRM2 (green) in the cryo-EM structure. The crystal RRM2 (tan) is rotated down. C. Alignment of the crystal structure with the B^{act} cryo-EM structure positions the crystallized Cus1 fragment (290-368; teal) at the C-terminus of the cryo-EM Cus1 (131-289; cyan) as highlighted.

the BPS, consistent with cross-linking studies. Hsh49 RRM2 is not tightly bound to the rest of the U2 snRNP. It contacts the small, C-terminal domain of Prp9 opposite to Cus1 (Figure 4-5C). RRM2 is positioned such that it may interact with the loop of U2 snRNA stem IIb; however, there is no density observed for this loop (Figure 4-5D). The backbones of SF3b from the A complex aligns very closely to those of SF3b in the B and B^{act} complexes, suggesting that the particle is very rigid.

The major difference between the B and B^{act} complexes is the placement of the second RRM of Hsh49 (Figure 4-6A). van Roon et al. (2017) dispute the placement of this RRM and of Cus1 in the B^{act} structure published by the Shi lab (Yan et al. 2016). When RRM1 from their crystal structure is aligned to RRM1 in

the cryo-EM map the Cus1 fragment is positioned into, and fits more closely, density that was modeled as Hsh49 RRM2 (Figure 4-6B; van Roon et al., 2017). This rearrangement links the N-terminus of the Cus1 fragment (290-368) to the C-terminus of the modeled Cus1 (131-289; Figure 4-6C). RRM2 may then be placed in unassigned density beside RRM1 near stem IIb of U2 snRNP, consistent with their published structure of the A complex (Plaschka et al., 2018). If the original modeling of the B^{act} RRM2 is correct, that places both RRMs of Hsh49 adjacent to each other. Yan et al. (2016) also model intron RNA bound to both RRMs, which is inconsistent with studies conducted by van Roon et al. (2017) that show RRM2 does not bind RNA.

4-4. Nager Syndrome

Nager syndrome is a genetic disorder belonging to the group of acrofacial dysostoses (Lansinger & Rayan, 2015). It is characterized by abnormalities of bone growth in the face and hands, resulting in pre-axial upper limb anomalies, and mandibulofacial dysostosis (Figure 4-7). Abnormalities in the middle and external ear can lead to deafness. 61% of patients have mutations in Sap49 (Bernier et al., 2012). Most of the Sap49 mutations seen in Nager syndrome are frameshift mutations (Table 4-1), although mutations to the initiating methionine as well as nonsense mutations have been reported. Nager syndrome is thought to be caused by haplo-insufficiency of Sap49 resulting from these disruptions in half of the expressed protein. In addition to its role in splicing, Sap49 has also been shown to

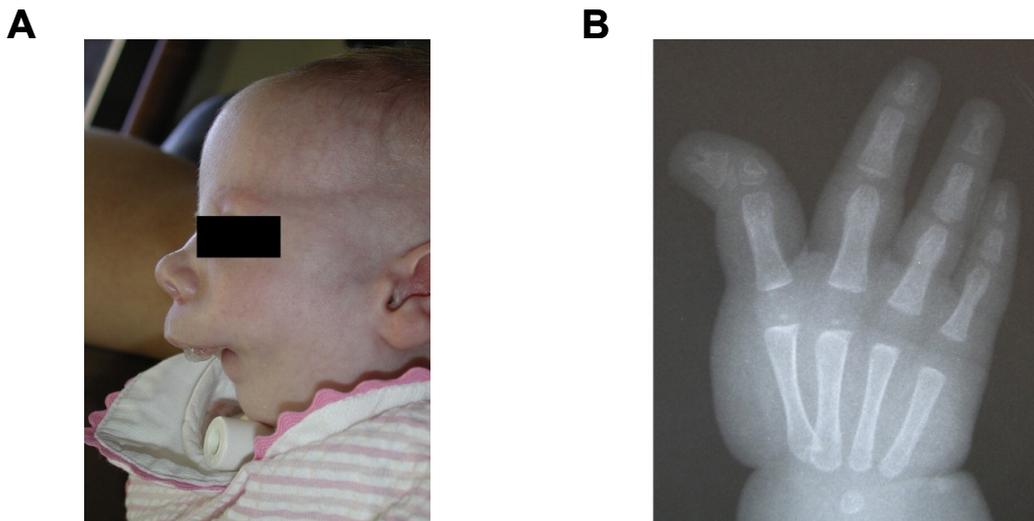


Figure 4-7. Phenotype of Nager syndrome. Nager syndrome is associated with facial abnormalities (A) and upper limb anomalies, including an absence of thumbs (B). Images previously published by Lansinger & Rayan (2015), reprinted with permission from Elsevier.

	Initiator Met	Non-sense	Frame-shift	Splicing	Total positive cases	Total study cases	Percent positive cases
Bernier et al., 2012	5	2	17	1	25	41	61%
Czeschik et al., 2013	0	4	3	0	7	12	58%
McPherson et al., 2014	0	0	0	1	1	1	100%
Petit et al., 2014	2	3	5	3	13	18	72%
Total cases	7	9	25	5	46	72	64%

Table 4-1. Sap49 mutations in Nager syndrome patients. 64% of patients with Nager syndrome have a mutation within Sap49.

interact with bone morphogenic protein receptor IA (BMPR-IA; Nishanian & Waldman, 2004; Watanabe et al., 2007). The results of mutant Sap49 may be due to disruptions of this signaling pathway rather than, or in addition to, its effects on

splicing.

The similar, but more severe, Rodriguez syndrome features the symptoms of Nagar syndrome in addition to lower limb anomalies, 11 ribs, and occasionally anomalies in internal organs (McPherson et al., 2014). Whole genome sequencing of the one surviving individual with Rodriguez syndrome showed a mutation in the coding region of Sap49. The identified G287D mutation is located in the long C-terminal tail with no predicted structure. McPherson et al. (2014) suggest this is the causative mutation, as other observed mutations to the gene occurred within introns. If Rodriguez syndrome is due to mutated Sap49 it may be a severe form of Nagar syndrome, rather than its own disease.

A recently identified form of mild Nagar syndrome, due to the Sap49 I84R mutation, is the first described case of Nagar syndrome arising from a missense mutation rather than a more dramatic frameshift or nonsense mutation. Ile84 is the residue C-terminal to the final β -strand in Sap49 RRM1, and points into the protein core. Mutation from the hydrophobic isoleucine to the large, positively charged arginine is expected to disrupt the structure of the RRM.

4-5. Sap49 Depletion in Frogs

The African clawed frog (*Xenopus laevis*) Sap49 is 377 amino acids, and exhibits the same basic architecture as other Sap49s, namely two N-terminal RRMs and a C-terminal tail. The RRMs of human and *Xenopus* share almost 100% sequence identity. Overall, the proteins are about 80% identical (Figure 4-3).

A Western blot for Sap49 showed no protein following injection of 40 ng antisense morpholino that blocks the translation of Sap49 (Devotta et al., 2016). Embryos injected with 10-40 ng morpholino did not survive past stage 45, while embryos injected with less morpholino (2 or 5 ng) showed reduced craniofacial structures, analogous to the structures seen in human patients with Nager syndrome. As Sap49 is an essential gene, it is not surprising that abolition of the protein is fatal. Nager syndrome is due to haplo-insufficiency, which may be mirrored in the phenotype appearing at low levels of Sap49 knock down (Devotta et al., 2016).

Results

4-6. Yeast Mutants

Not all Sap49 mutations result in disease. The ExAC database, collected from 60,706 individuals without severe pediatric disease, shows 45 missense mutations in Sap49, five of which are within the RRM1s of Sap49: V24A, L28P, N41S, A122T, and D153V (Lek et al., 2016). V24A and N41S are expected to be conservative mutations. We decided to test the effects of L28P and D153V, along with the Nager syndrome mutant I84R, on Sap49. The mutation L28A abolishes interaction of Sap49 with a fragment of Sap145 in pull down assays (Kuwasako et al., 2017). As Leu28 is located in $\alpha 1$ of RRM1 its mutation to proline is predicted to disrupt the helix. Mutation of the solvent exposed Asp153 to valine might interfere with protein-protein interactions.

As Sap49 is an essential gene, a plasmid shuffle strain (courtesy of M. Ares;

Igel et al., 1998) was used to test the effects of our mutations in *S. cerevisiae*. The endogenous Hsh49 gene is interrupted by the His3 gene, and Hsh49 is supplemented on a Ura3 plasmid. Cells that contain the Ura3 plasmid cannot survive on media containing 5-fluoroorotic acid (5-FOA), as they convert it to the toxic 5-flourouracil, which inhibits thymidylate synthase and disrupts DNA replication. Hsh49 is then expressed on a second plasmid, and is shuffled into the cells to provide the necessary Hsh49, allowing cells to drop the Ura3 plasmid. After recovery in media containing uracil the cells are plated on 5-FOA. The cells will grow if the Hsh49 on the second plasmid is able to rescue the loss of the wild type Hsh49 with the Ura3 plasmid. If the shuffled Hsh49 is nonfunctional, cells die either from lack of Hsh49 after dropping the Ura3 plasmid, or from production of 5-flourouracil due to the retention of the Ura3 plasmid. Cells that carry both plasmids may also be selected for on media lacking 5-FOA and uracil, producing a pseudo-diploid genotype.

We performed the plasmid shuffle with wild type, I81R, Q24P, and E159V, (I84R, L28P and D153V in humans respectively) generating both shuffled and pseudo-diploid strains. All the pseudo-diploid strains grew, suggesting that none of the mutants have a dominant negative effect. This is consistent with Sap49 haplo-insufficiency in Nager syndrome.

Wild type Hsh49, Q24P and E159V are able to rescue a deletion of Hsh49 upon plasmid shuffle. I81R was not. This result is consistent with the phenotypes seen in humans: L28P and D153V do not cause disease, while I84R does. We then

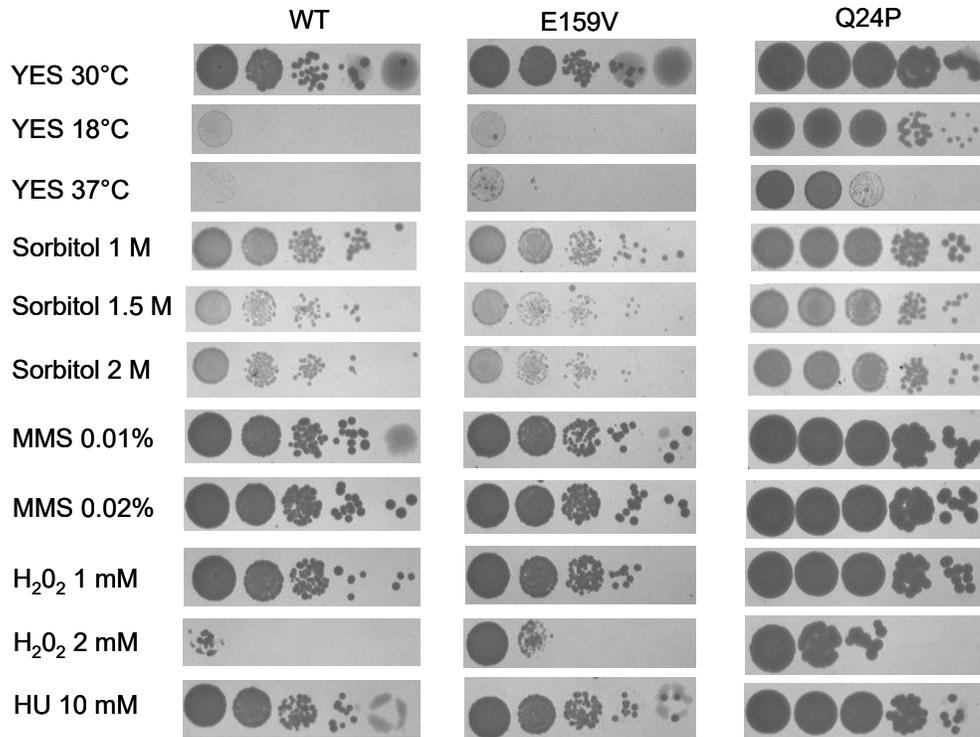


Figure 4-8. Growth tests of Hsh49 mutants. Wild type, E159V, and Q24P Hsh49 was plasmid shuffled into Hsh49 knockout cells. The resulting strains were spotted onto stress plates in 1/10 dilutions and cell growth tested. MMS – methyl methanesulfonate; HU – hydroxyurea.

tested the Q24P and E159V mutants for specific phenotypes by comparing their growth under a number of stress conditions: heat (37°C), cold (18°C), oxidative (1.8 and 2.0 M hydrogen peroxide), replication stress (10 mM hydroxyurea), DNA damage (0.01 and 0.02% methyl methanesulfonate) and osmotic (1.5 M sorbitol). E159V grew about the same as wild type. Surprisingly, Q24P grew the same as or better than wild type (Figure 4-8).

4-7. Protein Expression

To test the effects of Sap49 mutation on protein expression, we cloned,

expressed and purified both RRM1s (1-184) of Sap49 fused to a maltose binding protein (MBP) tag followed by a TEV cleavable linker. We designed a wild type construct and the L28P, D153V, or I84R mutants. We were able to purify the wild type, L28P, and D153V proteins. However, I84R did not stably express, and was proteolyzed either in the cells or during the early stages of purification (Figure 4-9). The MBP tag, which has a stabilizing effect on other proteins, was not sufficient to rescue Sap49 I84R. The instability of the I84R mutant provides an explanation for the lethality of I81R in *S. cerevisiae* and the Nager syndrome phenotype.

The L28P mutation, predicted to disrupt helix α 1, does not notably destabilize the protein. This construct behaved the same as wild type throughout the purification. We also tested the ability of L28P to pull down a fragment of Sap145 (Figure 4-10A) by co-expressing the His-tagged Sap49 L28P RRM1s with an untagged fragment of Sap145 (589-712). Both the wild type and L28P pulled down the Sap145 fragment, suggesting that the Leu28 interaction abolished by the L28A mutation is maintained in the proline mutant (Figure 4-10B).

4-8. Binding of Sap49 to RNA

An interaction between Hsh49 and U2 snRNA or pre-mRNA in the absence of Cus1 is demonstrated in some, but not all, experiments. Igel et al. (1998) showed that recombinantly expressed Hsh49 is able to bind both U2 snRNA and pre-mRNA. However, for van Roon et al. (2017) the Cus1 fragment was required for Hsh49 to shift an RNA mimicking the 5' region of U2 snRNA in an electrophoretic

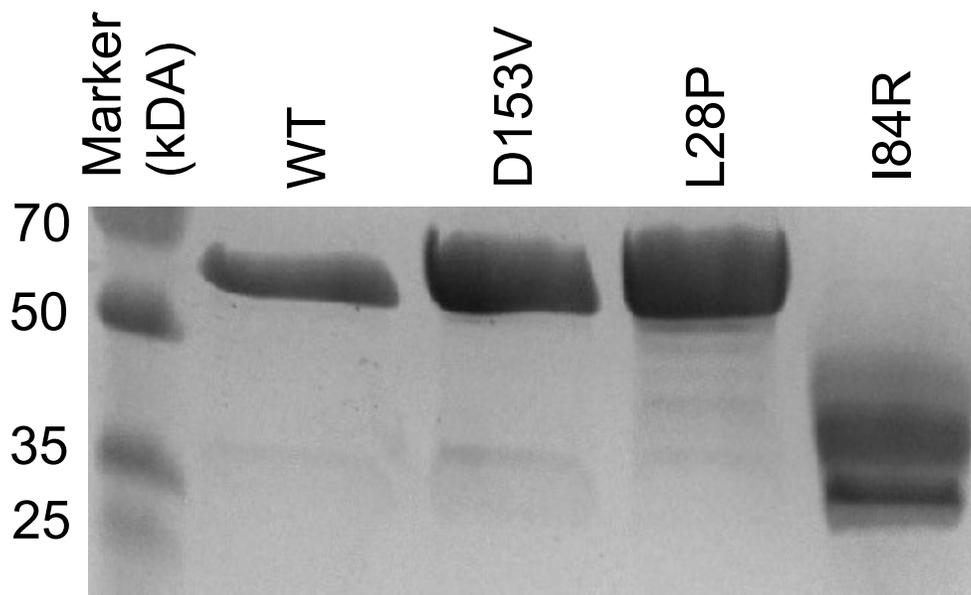


Figure 4-9. Expression of Sap49. MBP-tagged Sap49 (1-184) was stably expressed for the wild type and D153V and L28P mutants. I84R protein was not stably expressed. Proteins run on SDS-PAGE gel following elution from amylose column.

mobility shift assay (EMSA). In fluorescence anisotropy studies the addition of Cus1 increased the binding of Hsh49 to RNA by an order of magnitude (van Roon et al., 2017). Truncation studies show that only RRM1 is required for binding the U2 snRNA 5' piece, as RRM1 shifted the RNA with low affinity (33 μ M) and RRM2 did not shift the RNA at all. Addition of the Cus1 fragment to RRM1 increased affinity for the RNA (0.9 μ M). Deletion of the eight C-terminal residues of the Cus1 fragment decreased the affinity by twofold. Unfortunately, this region is not ordered in the crystal structure. In the recent structure of the A complex residues 354-360 are not seen. Residues 361-368 interact with Rse1, and do not contact the intron, the U2 snRNA or Hsh49.

We performed EMSAs of wild type and L28P Sap49 with U2 snRNA.

A

```
Kuwasako  ETRLKEKKPGDLSDELRLISLGMVGPNAHKVPPP----- 631
Hybrid    ETRLKEKKPGDLSDELRLISLGMVGPNAHKVPPPWLIAMQRYGPPPSYPNLKIPGLNS 655
van Roon  -----RPGRISQELRAIMNLPEG-----QLPPWCMKMKDIGLPTGYPDLKIAGLNW 335

Kuwasako  -----
Hybrid    PIPESCSFGYHAGGWGKPPVDETGKPLYGDVFGTNAAEFQTKTEEEEIDRTPWGELE 712
van Roon  DITN-----LKGDVYGKIIIPNHHSRP--KKQGRNYFGALI 368
```

B

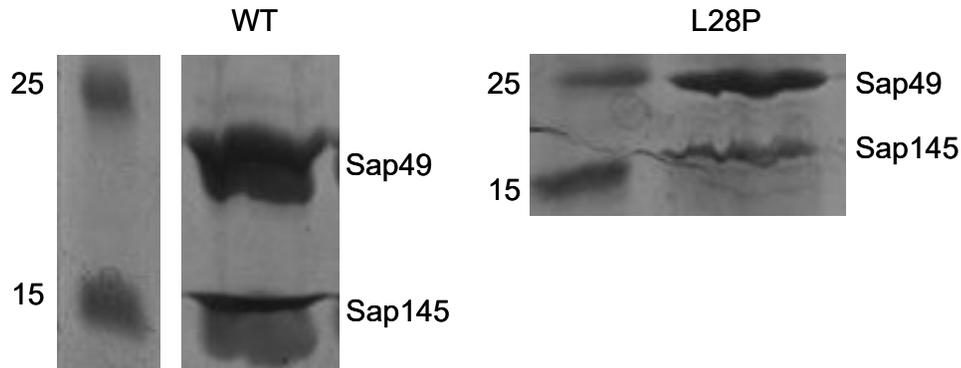


Figure 4-10. Sap49 pulls down Sap145. A. Design of the Sap145 construct is based on the Kuwasako et al. (2017) fragment (Sap145 598-631) and the van Roon et al. (2017) fragment (Cus1 290-368). B. Wild type Sap45 1-184 (left) and L28P (right) His-tagged Sap49 are able to pull down a fragment of Sap145 (589-712).

Sap49 bound RNA only in the presence of the Sap145 fragment. The wild type Sap49 shifts both the full length and the 5' end of U2 snRNA (Figure 4-11A). The presence or absence of Mg^{2+} had no effect on binding (Figure 4-11B). Addition of competitive tRNA abolished binding, suggesting the interaction between protein and RNA is non-specific. L28P bound the 5' end of U2 snRNA at similar concentrations to wild type (Figure 4-11C).

4-9. Discussion

Nager syndrome is thought to be caused by haplo-insufficiency of Sap49 (Bernier et al., 2012). The newly discovered mutant associated with Nager syndrome, Sap49 I84R, likely destabilizes the protein, as evidenced by our inability

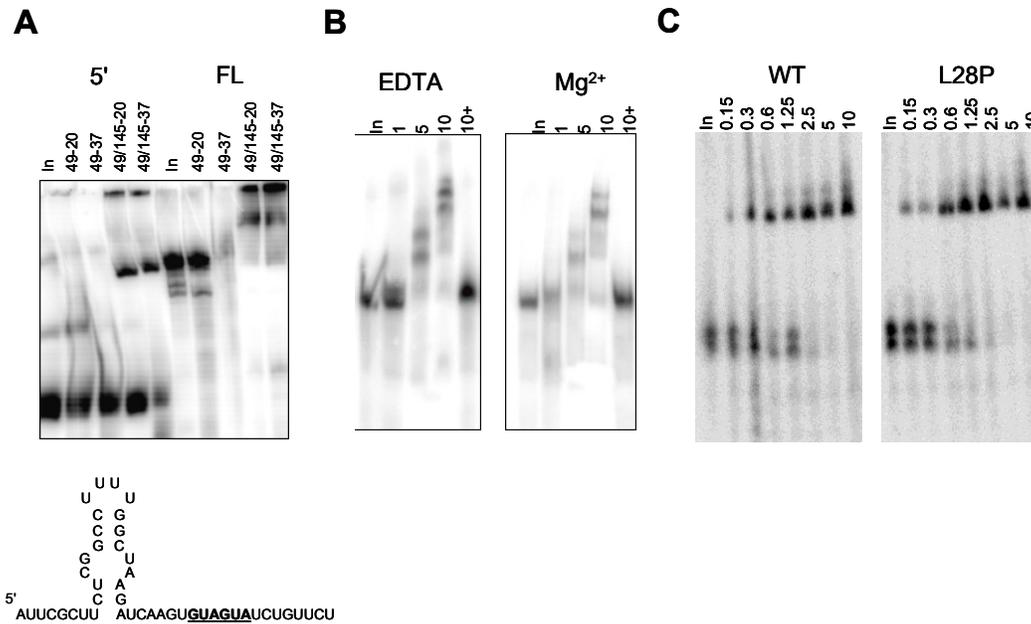


Figure 4-11. Sap49 EMSAs. A. 10 μM Sap49 binds both full length (FL) U2 snRNA and a 5' piece only in the presence of the Sap145 fragment (top) at both 20°C and 37°C. Bottom, sequence of the U2 snRNA 5' piece. Branch binding sequence is underlined. B. The presence of EDTA (left) or Mg^{2+} (right) does not influence the binding of the Sap49/145 dimer to the 5' piece at 1, 5, or 10 μM . The addition of 500 ng tRNA abolishes binding even at 10 μM Sap49/Sap145 (10+). C. Hsh49 L28P dimer (right) binds the 5' piece comparably to the wild type dimer (left) 0.15-10 μM Sap49/Sap145 dimer.

to produce a stable, recombinant I84R Sap49. While Hsh49 I81R alone is not viable, the mutation does not have a dominant negative effect, as seen in the pseudo-diploid yeast growth. Sap49 has been shown to interact with bone morphogenic protein receptor IA, and it has been proposed that disruption of this interaction is the cause of Nager syndrome (Watanabe et al., 2007). Overexpression of Sap49 was seen to inhibit osteogenesis of C2C12 cells, and chondrocytic differentiation of ATDC5 cells (Watanabe et al., 2007). *S. cerevisiae* lacks bone morphogenic protein and its receptors, but still shows an inviable phenotype for the I81R mutation, likely due to splicing defects. In addition, haplo-insufficiency of the

splicing protein Snu114 is associated with Guion-Almeida mandibulofacial dysostosis, whose phenotype is similar to that of Nager syndrome (Lines et al., 2012). These results suggest that the loss of Sap49 also has implications for splicing in Nager syndrome.

Hsh49 Q24P does not have the predicted deleterious effect. The resulting protein is able to compensate for the knockout of Hsh49, and actually grows better than wild type. We expect the proline substitution in helix α 1 to destabilize this helix, but this mutation has no adverse effect. Perhaps the proline is accommodated, or the N-terminus is not required to be strictly α -helical. In both the A and B complex structures, this region of the helix is not observed to interact with anything; the mutation and potential disruption may be accommodated.

Creation of temperature and cold sensitive mutants has been used to identify splicing proteins and study splicing in *S. cerevisiae*. Temperature sensitive mutants usually result from a destabilizing mutant, such that vital interactions are only stable at lower temperatures. The Prp8 mutation G2347D is temperature sensitive, as it disrupts the interaction between Prp8 and Brr2 (Jamieson et al., 1991). Cold sensitive mutants are usually caused by a hyper-stabilization such that higher temperatures are required to melt the interaction. A cold sensitive mutant of U4 snRNA (U4-cs1) extends the base-pairing at the end of stem 1 between U4 and U6 snRNAs, and suppressor mutants disrupt the extended base-pairing (Li & Brow, 1996). Mutations in U2 snRNA that stabilize alternative conformations cause cold sensitivity that can be overcome by mutations in Cus1 (Wells et al., 1996). Given

that L28P grows better than wild type and E159V at both higher and lower temperatures, it might be stabilizing one interaction and destabilizing another. Hsh49 E159V has no effect. The growth similarities between the wild type and mutant Hsh49 is consistent with a lack of disease phenotype in humans.

We tried to crystallize the wild type Sap49 RRMs, but were unsuccessful. We are now working to crystallize the Q24P Hsh49-Cus1 fragment based on van Roon et al. (2017) to determine the effect on the structure of the Q24P mutation.

Various programs predict the effects of missense mutations, including SIFT (Sim et al., 2012), Polyphen 2 (Adzhobei et al., 2010), Mutation Taster (Schwarz et al., 2014) and Align GVGD (Mathe et al., 2006; Tavtigian et al., 2006). Each algorithm predicted I84R to be disruptive, in agreement with our modeling and experimental evidence. SIFT and Polyphen 2 predicted, as did we, that L28P would be deleterious. In contrast, biochemical results show Sap49 L28P is stable, interacts with Sap145, and binds RNA equivalent to wild type. Additionally, Hsh49 Q24P is viable in *S. cerevisiae* and confers an unexpected growth advantage. SIFT predicted Sap49 D153V would be tolerated; Polyphen 2 predicted a disruption. We did not see an effect due to the D153V mutation on either protein stability or yeast viability. Protein prediction software is often a good starting point, especially for drastic changes in side chain size or charge in the interior of the protein, such as the isoleucine to arginine mutation. However, such programs may be less reliable for more subtle mutations, such as L28P, or mutations on the surface of proteins, such as D153V.

4-10. Materials and Methods

4-10a. *S. cerevisiae* mutants

S. cerevisiae with the Hsh49 gene interrupted by the His3 gene and carrying the Hsh49 on a Ura3 plasmid (generous gift, M. Ares USCS) were transformed with plasmids containing the Leu2 marker and either wild type or mutant (I81R, Q24P, or E159V) Hsh49 and plated on -His -Leu plates. Pseudo-diploid strains were streaked onto fresh -His -Leu, and individual colonies grown in liquid culture were then frozen at -80°C in 20% glycerol. To create plasmid shuffle strains, cells were restreaked onto plates containing uracil and 5-fluoroorotic acid (5-FOA) to ensure loss of the wild type Ura3 plasmid. Plasmid shuffle was verified by total DNA extraction and sequencing following PCR of the Hsh49 gene. Shuffled cells were grown in -His -Leu media then frozen at -80°C in 20% glycerol.

4-10b. *S. cerevisiae* growth tests

S. cerevisiae cells in which the wild type Hsh49/Ura3 plasmid had been dropped and replaced by a wild type, Q24P or E159V Hsh49 Leu2 plasmid were grown in liquid -His -Leu dropout media at 30°C shaking overnight. In the morning, cells were diluted to OD₆₀₀=0.2 and allowed to grow at 30°C for an additional three hours. Cells were then diluted back down to OD₆₀₀=0.2 and serially diluted 1/10 four times. 10 µL of cells was plated on YES plates with either no additive; or 1, 1.5, or 2 M sorbitol; 0.01, or 0.02% methyl methanesulfonate; 1, or 2 mM hydrogen peroxide; or 10 mM hydroxyurea. Cells were grown at 30°C for

and photographed after three days. Additional YES plates were grown at 18°C or 37°C.

4-10c. Expression of MBP-tagged Sap49

DNA for human Sap49 residues 1-184 wild type, I84R, L28P or E153V was cloned into pMAL to give a Sap49 with an N-terminal maltose binding protein (MBP) tag and TEV cleavable linker. Plasmids were transformed into *E. coli* Rosetta cells, which were grown in LB supplemented with 2.5 g/L dextrose at 37°C until the OD₆₀₀=0.8. Protein expression was induced with 1 mM IPTG at 18°C overnight. Cells were harvested by centrifugation and frozen at -20°C for 30 minutes. The cell pellets were resuspended in amylose lysis buffer (20 mM Tris pH 8, 100 mM NaCl, and 5 mM β-mercaptoethanol (BME)) supplemented with PMSF and lysozyme on ice. Cells were then lysed by sonication. Lysate was cleared by centrifugation and run over an amylose column. Protein was washed with amylose lysis buffer and eluted with amylose elution buffer (amylose lysis buffer with 20 mM maltose). Fractions were pooled and concentrated, then run on a Superdex 75 column in standard gel filtration buffer (SGFB; 20 mM Tris pH 8, 100 mM NaCl, 5 mM BME). Fractions were then concentrated and stored at 4 °C. Proteins were run on a 16% 19:1 SDS PAGE gel and visualized by Coomassie staining.

4-10d. Co-expression of His-tagged Sap49 and the Sap145 fragment

DNA for human Sap49 1-184 wild type, L28P or E153V was cloned into the first, His-tagged site of pACYC. DNA for Sap145 589-712 was cloned into the second, untagged site of pACYC containing Sap49. Plasmids were transformed into *E. coli* BL-21 cells. Cells were grown in LB at 37°C until the OD₆₀₀=0.8. Protein expression was induced with 1 mM IPTG overnight at 37°C. Cells were harvested by centrifugation and frozen at -20°C for 30 minutes. Cells were resuspended in 49/145 lysis buffer (20 mM Tris pH 7.5, 500 mM NaCl, 500 mM Urea, 25 mM imidazole, 5 mM BME) on ice. Cells were lysed by sonication and lysate cleared by centrifugation. Cleared lysate was run over a Ni-NTA column, washed with 49/145 lysis buffer and eluted with 45/145 elution buffer (49/145 lysis buffer with 200 mM imidazole). Protein was pooled and concentrated, then run on a Superdex 200 column in GFB (20 mM Tris pH 7.5, 200 mM NaCl, 5 mM BME). Fractions were pooled, concentrated, and stored at 4 °C. Proteins were run on a 16% 19:1 SDS PAGE gel and visualized by Coomassie staining.

4-10e. Hot RNA transcription

RNA was transcribed from double-stranded PCR templates of the full length U2 snRNA or the 5' end. Each 25 µL reaction contained 1 x transcription buffer, 0.5 mM each riboNTP, 1 µL PCR template, and 1 µL T7 RNAP at 37 °C for 3 hours.

Following transcription, RNA was run on a 15-20% (depending on product

size) 19:1 acrylamide:bisacrylamide denaturing gel with 8 M urea in 1x TBE. RNA was visualized by autoradiography and sliced from the gel. Gel slices were incubated, shaking overnight in 400 μ L 0.3 M sodium acetate and 100 μ L phenol. The gel slices were spun through a spin filter and the supernatant was then phenol/chloroform extracted and ethanol precipitated. Pelleted RNA was resuspended in RNase-free water and stored at -20°C.

4-10f. Electrophoretic Mobility Shift Assays (EMSAs)

All EMSAs were performed in 20 mM HEPES pH 7.5, 125 mM KCl, and 5 mM DTT in 10 μ L reactions and incubated for 30 minutes, then run on 6% 80:1 acrylamide:bisacrylamide native gels in 0.5x TBE at room temperature.

To test the effect of temperature, mixtures of 375 nM hot U2 snRNA full length or 5' piece with 10 μ M Sap49 alone or Sap49/Sap145 dimer were incubated at room temperature or 37°C before being run on the native gel.

To test the effect of Mg²⁺ mixtures of 375 nM hot 5' piece were incubated with 1, 5 or 10 μ M Sap49/Sap145 dimer with either 5 mM EDTA or 5 mM MgCl₂ at room temperature with and without 500 ng tRNA before being run on the native gel.

To compare wild type and L28P, 15 nM hot 5' piece was incubated with 0.1, 0.3, 0.6, 1.2, 2.5, 5 and 10 μ M wild type or L28P Sap49/Sap145 dimer at room temperature before being run on the native gel.

Chapter 5
The U4 snRNP Protein Snu13 from *C. merolae*

Adapted from Black et al. (2016).

5-1. Snu13

Snu13p is a small, highly conserved protein initially identified in fractionation of the *S. cerevisiae* and human U4/U6.U5 tri-snRNPs (Nottrott et al., 1999; Stevens & Abelson, 1999). It is an essential gene, and homologues have been found in a number of organisms, including human (15.5K), *S. pombe* (Snu13), *C. merolae* (Snu13) and others. It belongs to family of proteins based on the archaeal protein L7Ae (Kuhn et al., 2002b).

The fold common to the proteins in the L7Ae family is five α -helices alternating with four β -strands (Hamma & Ferré-D'Amaré, 2004). The β -strands form a central β -sheet sandwiched between the α -helices: three on one side ($\alpha 1$ $\alpha 4$ and $\alpha 5$) and two ($\alpha 2$ and $\alpha 3$) on the other (Figure 5-1A). This globular fold is highly conserved from archaea to humans (Figure 5-1B). Crystal structures of Snu13 alone and bound to RNA show that the protein's fold does not change upon RNA binding (Vidovic et al., 2000; Oruganti et al., 2005; Dobbyn et al., 2007; Liu et al., 2011).

Snu13 binds the RNA kink-turn (K-turn) motif, which is composed of a long stem loop interrupted by an asymmetric (5+2) bulge (Figure 5-1C; Nottrott et al., 1999). The bulge bends the flanking stems relative to each other and features one nucleotide (usually U) protruding from the internal loop (Figure 5-2A, 5-2B). Snu13 binds this motif in the 5' stem loop of U4 snRNA (Figure 5-2B; Vidovic et al., 2000). The binding of Prp31 to the U4 snRNP is dependent on the presence of Snu13 (Liu et al., 2011). In humans, but not in *S. cerevisiae*, Snu13 is required for

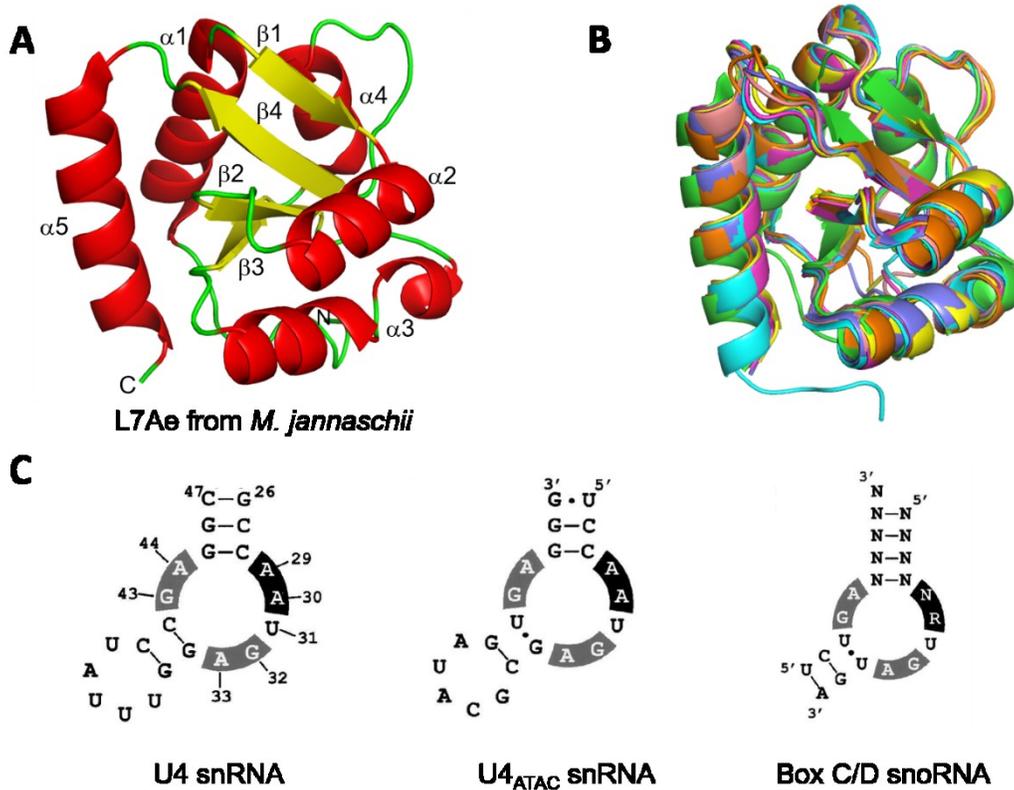


Figure 5-1. Conservation of Snu13 and K-turn motifs. A. Structure of L7Ae from *M. jannaschii* (PDB 1RA4) exhibits the conserved fold of the L7Ae family of proteins that includes Snu13. B. Overlay of Snu13 structures from *S. cerevisiae* (cyan, PDB 2ALE; magenta, PDB 5GAN; yellow, PDB 5NRL) and human (salmon, PDB 2OZB; blue, PDB 3SIU; orange, PDB 5O9Z) with L7Ae (green, PDB 1RA4). C. Snu13 binds a K-turn motif consisting of a 5+2 bulge within an RNA stem found in U4 snRNA, U4_{ATAC} snRNA, and box C/D snoRNAs. Image previously published in Vidovic et al., 2000, used with permission from Elsevier.

the binding of Prp3 and Prp4 to the U4 snRNP (Nottrott et al., 2002; Liu et al., 2015). Prp6, a U5 snRNP component, binds upon formation of the U4/U6 U5 tri-snRNP to stabilize the particle (Galisson & Legrain, 1993). Snu13 serves as a nucleation factor for the formation of U4 and U4_{atac} snRNPs either by stabilizing the RNA structure, or acting as a binding site for other proteins (Liu et al., 2011).

Small nucleolar RNPs (snoRNPs) catalyze modifications to RNA, including cleavage, 2'-O-methylation (box C/D snoRNPs) and pseudo-uridylation (box

H/ACA snoRNPs; reviewed in Watkins & Bohnsack, 2012). These snoRNPs base-pair with their targets, such as ribosomal RNA, to select the site of modification. Box C/D snoRNAs also feature a K-turn motif to which Snu13 binds in the formation of these snoRNPs (Oruganti et al., 2005). Interactions of Snu13 bound to snoRNA with Nop5 in *S. cerevisiae*, or Nop58 and Nop56 in higher eukaryotes, recruits Fibrillarin, which catalyzes 2'-O-methylation (Aittaleb et al., 2003). As with the U4 snRNP, Snu13 acts as a nucleating factor for protein association with the snoRNP.

5-2. Snu13 and U4 snRNA

The crystal structure of the U4 snRNA 5' stem loop bound to Snu13 reveals the architecture of the K-turn (Figure 5-2C; Vidovic et al., 2000). Nucleotides 26-47 of human U4 snRNA fold into two helical stems separated by an asymmetrical 5+2 internal loop consisting of residues 29-33 and 43-44 respectively. The internal loop adopts a complex fold, which kinks the RNA stem. Four of the nucleotides form G-A base-pairs to extend stem 2. The remaining three nucleotides are unpaired. U31 is flipped out of the loop. A29 and A30, unpaired, stack on the base-paired G45-C28 and A44 respectively. Stem 1 ends in three G-C base-pairs. Stem 2, preceded by the two G-A base-pairs, is composed of two G-C base-pairs, and capped with a UUUAU pentaloop. The fold of the internal loop is further stabilized by a network of hydrogen bonds. This structure is seen in both the tri-snRNP and the B complex (Nguyen et al., 2016; Plaschka et al. 2017).

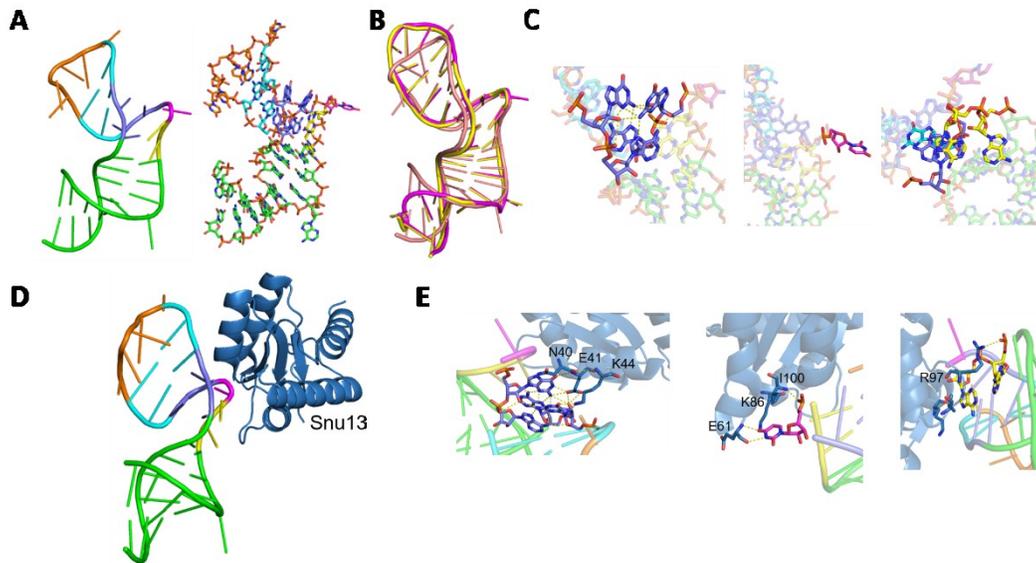


Figure 5-2. Snu13 binds the K-turn of U4 snRNA. A. Secondary structure of U4 snRNA 5' stem loop (PDB 2OZB). Stem 1 in green, single-stranded adenosines in yellow, bulged uridine in magenta, G-A base-pairs in blue, stem 2 in cyan, and the capping pentaloop in orange. Left, cartoon model; right, sticks. B. The K-turn structure is maintained in complex with Snu13 (PDB 2OZB, pink), and the *S. cerevisiae* tri-snRNP (PDB 5GAN, magenta) and B complex (PDB 5NRL, yellow). C. Close up of K-turn structural features. Left, two G-A base-pairs form to extend stem 2; middle, U31 is flipped out of the loop; right, two single-stranded adenosines stack on the G-A base-pairs. D. Snu13 binds the K-turn. E. Close up of interactions between Snu13 and the K-turn. Left, Asn40, Glu41, and Lys44 hydrogen bond to the guanosines involved in the G-A base-pairs; middle, U31 is recognized by Glu61, Lys86, and Ile100; right, the phosphate backbone of the single-stranded adenosines hydrogen bonds to Arg97.

Snu13 binds the K-turn of the U4 snRNA 5' stem loop in a sequence specific manner (Figure 5-2D, 5-2E). The RNA binding face of Snu13 is composed of the helices $\alpha 2$ and $\alpha 4$, β -strand $\beta 1$, and three loops $\beta 1$ - $\alpha 2$, $\beta 2$ - $\alpha 3$ and $\alpha 4$ - $\beta 4$ (Vidovic et al., 2000; Liu et al., 2011). Snu13 interacts primarily with the internal, asymmetric loop in U4 snRNA. In humans, a pocket formed by Glu61, Ile65, Lys86 and Ile100 binds the flipped out U31. This sequence specific interaction is abolished upon mutation of U31. The two guanosines involved in the G-A base-pairs are bound by Asn40 and Glu41 in the $\beta 1$ - $\alpha 2$ loop and by Lys44 in helix $\alpha 2$. The adenosines are

not recognized in a sequence specific manner: the unpaired bases A29 and A30 pack against Arg97 in loops α 4- β 4 and Lys37 (β 1- α 2) and Val95 (α 4- β 4) respectively. Lys44 hydrogen bonds with the phosphate backbone of U47; Arg97 with A35.

5-3. Snu13 and Prp31

Snu13 was crystallized in a complex with U4 snRNA and Prp31 (Figure 5-3A; Liu et al., 2011). The structure of Snu13 is the same in both bound and free forms. Snu13 is required for the binding of Prp31 to U4 snRNA. Additionally, in humans, U4 snRNA is required for the binding of Prp31 to Snu13 (Nottrott et al., 2002).

The fragment of Prp31 crystallized, residues 78-333, features eleven α -helices separated by short linkers. Snu13 binds to the C-terminal helix bundle (215-331) composed of the final three helices. A long helix (180-214) links the C-terminal helices with five short N-terminal helices (124-173) and the longer helix α 1 (88-119; Figure 5-3B). Prp31 binds snRNA via a domain homologous to that in Nop56/58, which binds snoRNA (Liu et al., 2007b). NMR data show Prp31 binds helices α 2 and α 3 of Snu13 (Li et al., 2009). In humans Snu13 Asn40 hydrogen bonds to Prp31 Arg304.

The U4 snRNA is sandwiched between Snu13 and Prp31. Prp31 stabilizes the pentaloop that caps stem 2 and interacts with stem 2 in a sequence independent manner (Figure 5-3C).

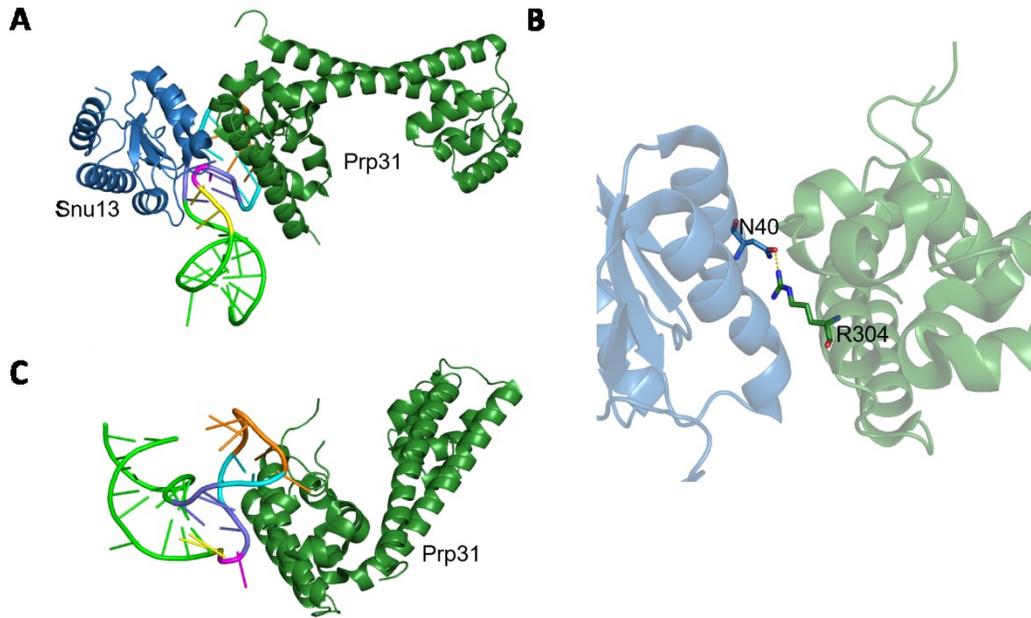


Figure 5-3. Interactions of Snu13 with Prp31. A. Prp31 binds both Snu13 and the U4 snRNA (PDB 3SIU). B. In the interface between Snu13 and Prp31. Asn40 of Snu13 interacts with Arg304 of Prp31. C. The C-terminal helices of Prp31 stabilize stem 2 and the pentaloop.

5-4. Snu13 in the Tri-snRNP and B Complex

The recent structure of the U4/U6.U5 tri-snRNP shows *S. cerevisiae* Snu13 bound to U4 snRNA in the center of four proteins: Prp3, Prp4, Prp6 and Prp31 (Figure 5-4; Nguyen et al., 2016). In humans, Snu13 mediates the binding of Prp31, Prp3, and Prp4 to U4 snRNA (Agafonov et al., 2016).

The structure of the U4 snRNA 5+2 K-turn is very similar between the human crystal structure and the *S. cerevisiae* tri-snRNP. The minor difference between these K-turn motifs is the loop capping stem 2: in humans it is a pentaloop, while in *S. cerevisiae* it is a tetraloop. In addition to the interactions seen in the crystal structure, Snu13 binds stem II of the U4/U6 snRNA duplex.

Prp31 exhibits a different conformation in the tri-snRNP relative to the

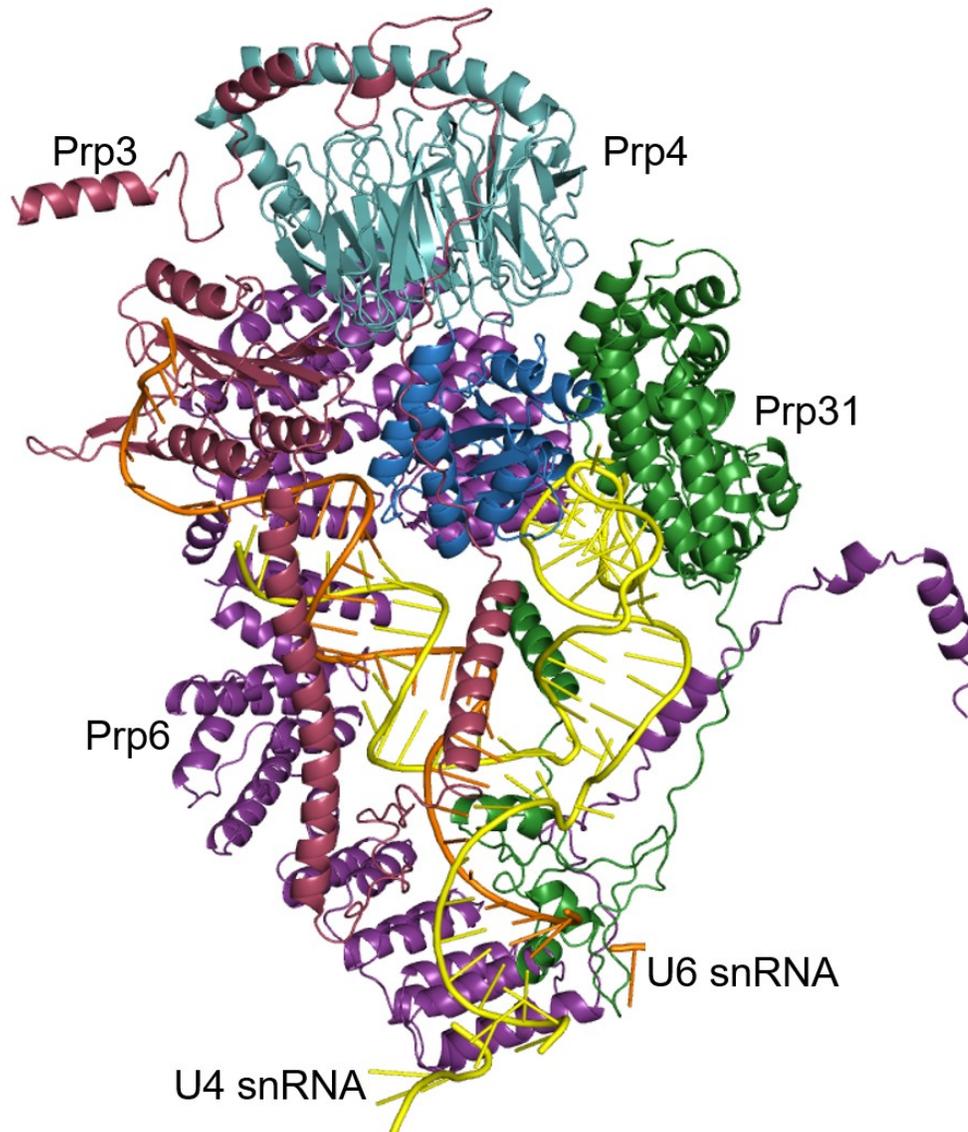


Figure 5-4. Snu13 at the heart of the U4/U6 di-snRNP. In the *S. cerevisiae* tri-snRNP (PDB 5GAN), Snu13 (blue) binds the 5' stem loop of U4 and interacts with Prp31 (green), Prp3 (raspberry), Prp4 (teal), and Prp6 (purple).

human crystal structure of Snu13 and the U4 5' stem loop. The binding region for U4 snRNA and Snu13, Prp31 residues 210-331, remains the same. However, there is rearrangement in the N-terminal region. These helices crystallized extended from

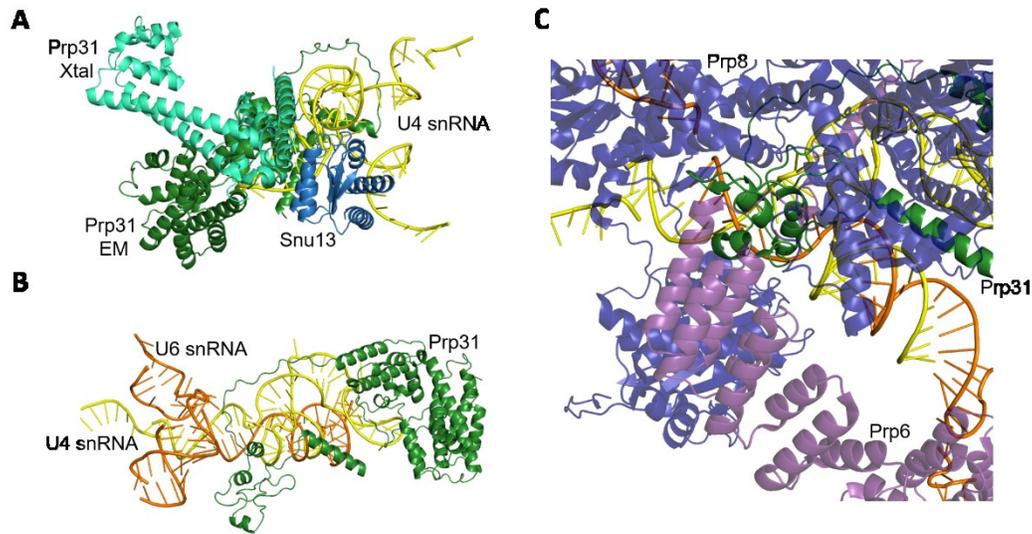


Figure 5-5. Rearrangement of the N-terminus of Prp31 upon incorporation into the tri-snRNP. A. The N-terminal helices of Prp31 rotate toward Snu13 in the *S. cerevisiae* tri-snRNP (PDB 5GAN) relative to the crystal structure (PDB 2OZB). B. The long linker of Prp31 connecting the C-terminus to the center of the protein lies along the U4/U6 snRNA duplex. C. In the tri-snRNP, the termini of Prp31 interact with Prp8 and Prp6.

the Snu13-interacting domain, but in the cryo-EM are rotated back toward Snu13 (Figure 5-5A). They interact with the NT domain of Prp8 and the C-terminus of Prp6. The C-terminus of Prp31, not crystallized but visible in the tri-snRNP structure, features five α -helices separated by long linkers. The first helix is bundled with the crystallized helices alongside Snu13. The linker lies atop the phosphate backbone of stem 2 of the K-turn. The second new helix is inserted into the major groove of the U4/U6 snRNA duplex (Figure 5-5B). The third and fourth helices interact with Prp6, while the final helix interacts with the RT domain of Prp8 (Figure 5-5C).

Prp6, a protein from the U5 snRNP, binds both ends of Prp31 in the tri-snRNP. Its N-terminus extends toward Snu13, running anti-parallel along the C-

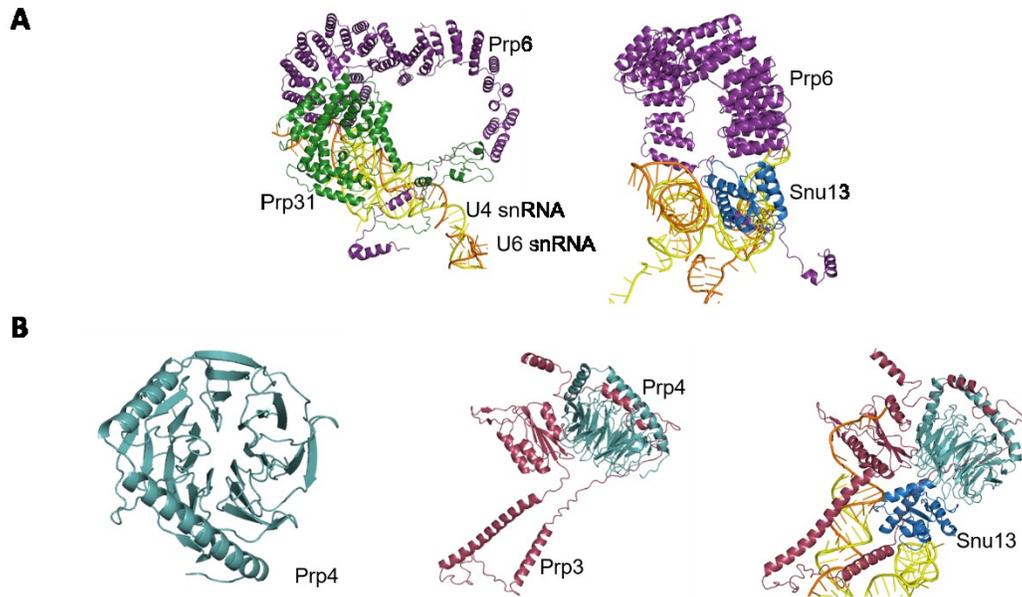


Figure 5-6. Prp6, Prp3, and Prp4 in the tri-snRNP. A. In the *S. cerevisiae* tri-snRNP (PDB 5GAN) Prp6 contacts Prp31 while interacting with the U4/U6 snRNA duplex (left). Right, Prp6 also contacts Snu13. B. Prp4 and Prp3 bound to the di-snRNP. Prp4 has a WD40 fold (left). Center, Prp3 interacts with Prp4. Right, Prp3 also binds the U4/U6 snRNA duplex, and Prp4 sits atop Snu13.

terminus of Prp31, with its backbone potentially interacting with the bulged residues between stem I of the U4/U6 snRNAs and the U4 5' stem loop (Figure 5-6A).

Prp4 contains seven four-stranded β -sheets arranged into the WD40 domain preceded by two α -helices. The helices of Prp4 interact with the three N-terminal helices of Prp3. A long linker of Prp3 winds around to connect these helices with its helix α 4 that sits between the U4/U6 snRNA duplex from the U4 5' stem loop. The long helix α 5 lies alongside the U4/U6 snRNA duplex. The 3' single-stranded region of U6 snRNA is wrapped around the C-terminal bundle of three α -helices sandwiching a four-stranded β -sheet (Figure 5-6B). Prp3 and Prp4 do not make close contacts with Snu13, consistent with their binding to the U4/U6 snRNA

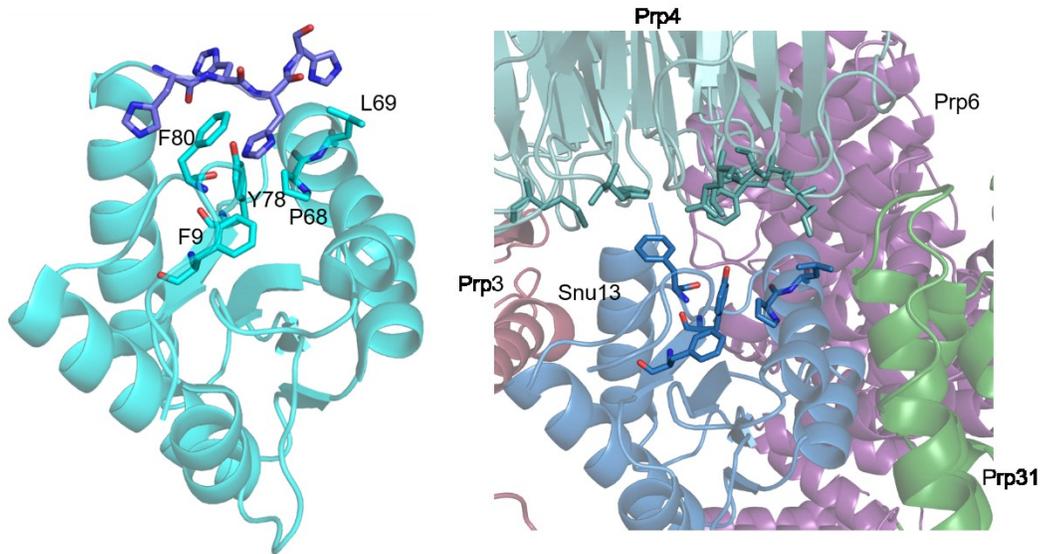


Figure 5-7. The conserved hydrophobic binding pocket of Snu13. Left, Snu13 residues Phe9, Phe80, Tyr78, Leu69, and Pro68 form a hydrophobic binding pocket that in the *S. cerevisiae* structure bound the C-terminal His tag tail (blue, PDB 2ALE). Right, the hydrophobic pocket is empty in the *S. cerevisiae* tri-snRNP (PDB 5GAN). Prp4 in teal, Prp31 in green, Prp6 in purple, Prp3 in raspberry, Snu13 in blue.

duplex being independent of Snu13 in *S. cerevisiae*. However, the requirement of Snu13 for Prp3 and Prp4 binding in humans could be due to a stabilization of this duplex by Snu13.

Dobbyn et al. (2007) crystallized *S. cerevisiae* Snu13 with a C-terminal His tag. The six histidines from the tag folded back up onto Snu13 and bind to a hydrophobic pocket composed of Phe9, Phe80, Pro68, Leu69 and Tyr78. However, this pocket does not appear to be occupied in either the *S. cerevisiae* tri-snRNP or the B complex. It is oriented toward Prp4, along with the C-terminus of Snu13 (Figure 5-7).

The U4 snRNP in the B complex is very similar to that in the tri-snRNP. There are no major rearrangements between the tri-snRNP and B complex, so the

interactions in the tri-snRNP are maintained in the B complex. Minor differences between the complexes include visualization of the N-terminal helices of Prp3, the N-terminal helices of Prp4 that bind Prp3, and the N-terminus of Prp6 that interacts with Prp8.

Results

5-5. *C. merolae* Snu13 Structure

C. merolae is a unicellular red alga with a minimal spliceosome that inhabits geothermal vents at temperatures up to 56°C (Ferris et al., 2005). As such, it may provide a promising source of thermostable spliceosomal proteins to study crystallographically. In addition, its reduced spliceosome may give insights into the essential core of pre-mRNA splicing (Stark et al., 2015).

We expressed, purified, crystallized and solved the structure of *C. merolae* Snu13 (Figure 5-8A; Table 5-1). Like the structures of all other published Snu13s, *C. merolae* Snu13 is composed of five α -helices and four β -strands (Suryadi et al., 2005; Dobbyn et al., 2007; Liu et al., 2007b; Liu et al., 2011). The β -strands fold together to form a mixed β -sheet at the center of the protein with three helices on one face and two on the other. This fold is highly conserved among the solved structures from other organisms, including the *S. cerevisiae* and humans with a backbone RMSD of 0.9 and 1 respectively (Figure 5-8B).

The RNA binding residues at the main interface are also conserved in *C. merolae*, suggesting a similar mechanism of RNA binding (Figures 5-8C, 5-9).

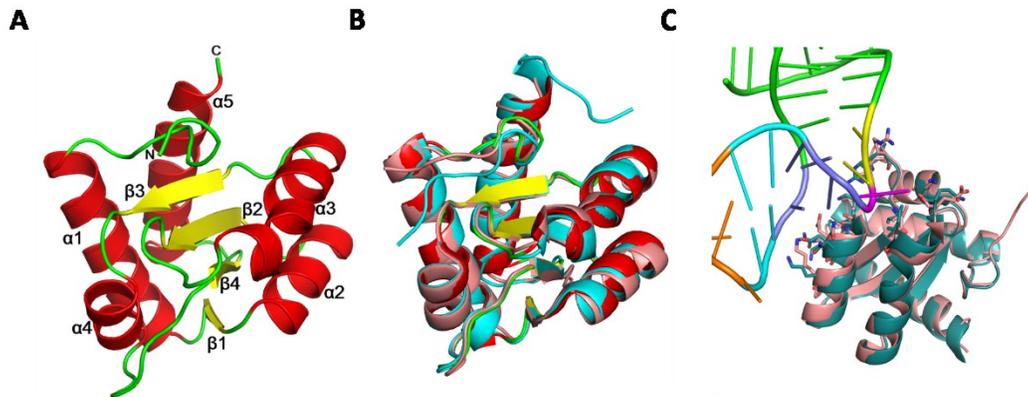


Figure 5-8. Structure of *C. merolae* Snu13. A. Structure of *C. merolae* Snu13 (PDB 3EWR). B. *C. merolae* Snu13 exhibits the highly conserved fold of the L7Ae family. *C. merolae* coloured by secondary structure; human (PDB 2OZB) in pink; *S. cerevisiae* (PDB 2ALE) in cyan. C. The conserved RNA binding residues are shown as sticks. *C. merolae* (teal) and human (pink).

However, residues at the tip of helix $\alpha 1$ that interact with the backbone of U4 snRNA in stem II and with U6 snRNA are not conserved. Despite the lack of Prp4, Prp6, and Prp31, their binding sites on Snu13 are conserved.

C. merolae Snu13 maintains the putative protein binding pocket first described in the *S. cerevisiae* Snu13 (Dobbyn et al., 2007). Based on the lack of Prp4 and Prp6, it was predicted that in *C. merolae* Snu13 this pocket would bind Prp3. However, the recent cryo-EM structure of both the tri-snRNP and B complex show Prp3 binding elsewhere, alongside helices $\alpha 1$ and $\alpha 4$. This does not appear to be a protein binding pocket in either the tri-snRNP or B complex.

5-6. *C. merolae* Snu13 and RNA

The primary sequence of U4 snRNA is not highly conserved; however, the basic architecture of the U4 snRNA is conserved (Figure 5-10A). Complementarity

Data collection	
Space group	P 21 21 21
Cell dimensions	
a, b, c (Å)	30.33, 57.58, 65.38
α , β , γ	90, 90, 90
Wavelength	1.1271
Resolution (Å)	2.35
R _{meas}	0.117
I / σ I	21.36 (11.85)
Completeness (%)	99.77 (99.39)
Redundancy	7.7 (7.6)
Refinement	
Resolution (Å)	43.21 - 2.35
No. reflections	158,498
<i>R</i> _{work} / <i>R</i> _{free}	0.1662 / 0.2200 (0.178/0.227)
No. atoms	
Protein	933
Water	42
<i>B</i> factors	
Protein	49.5
Ligand	40.0
R.M.S. deviations	
Bond lengths (Å)	0.016
Bond angles (°)	1.52
Ramachandran favored (%)	97.5
Ramachandran allowed (%)	2.5
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	2.97
Clashscore	4.68

Table 5-1. Data collection and refinement statistics for *C. merolae* Snu13. Data were collected from a single crystal. Values in parentheses are for the highest resolution shell (2.43-2.35 Å). Molecular replacement was performed with PDB 2ALE as the search model. Structure deposited as PDB 5EWR.

between U4 and U6 snRNA to allow base-pairing is conserved, as is the U4 5' stem loop interrupted by a 5+2 bulge that forms a K-turn. The predicted structures of human and *S. cerevisiae* U4 snRNA have been confirmed by the cryo-EM structures of their tri-snRNPs and B complexes. While the actual structure of *C.*

```

C. merolae      MEPLSSTEAPSRPMDVVV TAPNQADPRAYPFAPADLVVEILDVQQASHYKQIKKGLNEVL 60
D. melanogaster -----MTEEVNPKAFPLADAQLTAKIMNLLQQALNYNQLRKGANEAT 42
H. sapiens     -----MTEADVNPKAYPLADAHLTKKLLDLVQQSCNYKQLRKGANEAT 43
S. cerevisiae  -----MSAPNPKAFPLADAAL TQQILDVVQQAANLRQLKKGANEAT 41
C. elegans     -----MADDGVNPKAFPLADTNLSQKLMDLVQQAMNYKQLKKGANEAT 43
S. pombe       -----MSVNPKAFPLADSGLTQQILDLVQQASHYKQLRKGANEAT 40
                :*:***:* : * :****:***. . *::** **

C. merolae      KSMNRGLAEFVVLAA DTQPLEILLSAPLVAEDKAVPYVFPVSKAALGRACGVS RPVIACA 120
D. melanogaster KTLNRGLADIVVLAGDAEPIEILLHLPLLCEDKNVPYV FVRSKQALGRACGVS RPVIACS 102
H. sapiens      KTLNRGISEFIVMAADAEPLEIILH LPLLCEDKNVPYV FVRSKQALGRACGVSRPVIACS 103
S. cerevisiae   KTLNRGISEFIIMAADCEPIEILLHLPLLCEDKNVPYV FVRSVALGRACGVS RPVI AAS 101
C. elegans      KTLNRGISEIIVMAADAEPLEILLHLPLLCEDKNVPYV FVRSK AALGRACGVT RPVIAAS 103
S. pombe        KTLNRGISEFIVMAADTEPIEILLHLPLLCEDKNVPYV FVPSKAALGRACGVS RPVISAS 100
                *:***:****:*. * :***:* **:.*** ***** *: *****:****:..

C. merolae      VLRADMSQLRNQITALRTKIEQLL 145
D. melanogaster VTTNEGSQ LKSQITSIQQEIERLLV 127
H. sapiens      VTIKEGSQ LKQIQSIQQS IERLLV 128
S. cerevisiae   ITTNDASAIKTQIYAVKD KIE TLLI 126
C. elegans      ITQNEGSQ LKSQIQKIKEDVEKLLI 128
S. pombe        ITTNEASD LLPQIQAIKLAIEKLLI 125
                : : * : ** :: :* **

```

Figure 5-9. Sequence alignment of Snu13 from six organisms. RNA-binding residues are bold and underlined.

merolae U4 snRNA is unknown, analysis of its sequence predicts the same structural features of other U4 snRNAs: base-pairing with U6 snRNA, and the 5' stem loop interrupted by a K-turn (Stark et al., 2015).

We also tested the affinity of *C. merolae* Snu13 for U4 snRNA using fluorescence polarization. Snu13 bound U4 snRNA with an affinity of 160 nM (Figure 5-10B). This binding is specific for the K-turn structure, as Snu13 had a much lower affinity (16 μ M) for U6 snRNA, which lacks this feature. The binding affinities observed with *C. merolae* Snu13 are similar to those observed with Snu13 proteins from other organisms binding a variety of K-turn motifs, including those found in snoRNAs (Figure 5-10C).

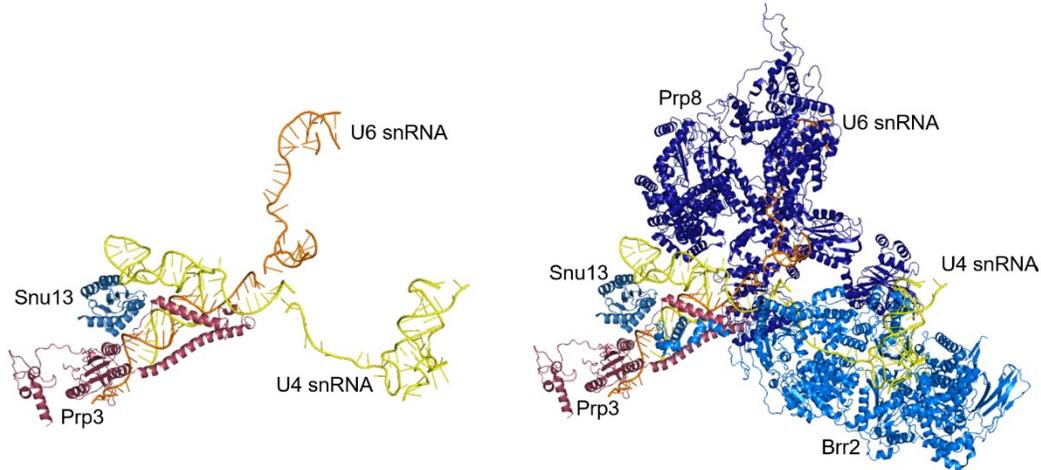


Figure 5-11. Partial potential *C. merolae* tri-snRNP. *C. merolae* lacks homologues for Prp4, Prp6, and Prp31. The resulting di-snRNP, left, would be composed of U4/U6 snRNA, Snu13, and Prp3. Upon binding to the U5 snRNP to create the tri-snRNP Prp8 could bind the end of U6 snRNA and Brr2 the end of U4 snRNA, right.

nucleating factor upon binding of the K-turn motif. Following Snu13 binding, other snoRNP factors associate. In humans, these factors include Nop56, Nop58 and Fibrillarin (Nop1p in *S. cerevisiae*). A BLAST search shows *C. merolae* has annotated orthologues of Nop56, Nop58 and Fibrillarin, suggesting an active box C/D snoRNP. Given the lack of its binding partner, Prp31, the high conservation of Snu13 may be due to selective pressures on its interaction with box C/D snoRNP. The unused binding pocket observed by Dobbyn et al. (2007) may be involved in protein-protein interactions within the snoRNP. Due to the similarity in structure to the U4 snRNA 5' stem loop, the interaction with the K-turn of box C/D snoRNA could conserve residues required to interact with U4 snRNA.

5-8. Materials and Methods

5-8a. Cloning, Expression, and Protein Purification

Snu13 from *C. merolae* genomic DNA was amplified via PCR and cloned into pMCGS7 for expression with a TEV-cleavable N-terminal His tag. The protein was expressed in *E. coli* and purified on Ni-NTA resin before and after cleavage with TEV protease. Cleaved Snu13 was run on a Mono S column and concentrated for further study.

5-8b. Crystallization and Structure Determination

Snu13 was crystallized at 10 mg/ml using the hanging drop method in 31% PEG 3350 and 100 mM sodium acetate pH 4.4. Crystals were cryo-protected with 20% glycerol, looped, and frozen in liquid nitrogen. Diffraction data were collected at the Stanford Synchrotron Radiation Light Source, and processed with HKL2000 (Otwinowski & Minor, 1997). Phenix was used for molecular replacement with PDB 2ALE (Adams et al., 2010). Structure refinement was done with Phenix refine and manually in COOT (Emsley et al., 2010).

5-8c. Fluorescence Polarization

5 nM of RNA derived from either U4 snRNA or U6 snRNA and 3' end labeled with Fluorescein was incubated with increasing concentrations (0-100 μ M) of Snu13. Anisotropy was measured with a Synergy 2 Multi-Mode reader.

Chapter 6
The RH domain of *C. merolae* Prp8

6-1. Splicing in *C. merolae*

The unicellular red alga *Cyanidioschyzon merolae* is an acidophilic organism that can grow at temperatures up to 56 °C (Matsuzaki et al., 2004). Its genome size is comparable to that of *Saccharomyces cerevisiae* both in chromosome length and number of genes. However, unlike *S. cerevisiae*, which contains 296 introns in 5,404 genes (5%; Neueglise et al., 2011) *C. merolae* has only 27 introns in a total of 26 genes within a genome of 4,803 genes (0.5%; Matsuzaki et al., 2004). The diminished number of introns is associated with a reduced number of splicing associated proteins, and splice sites are highly conserved (Stark et al., 2015). Investigating splicing in this organism may lead to a greater understanding of the core set of essential spliceosomal proteins, as less essential proteins are more likely to be lost during evolution. Additionally, the elevated growth temperature of *C. merolae* may render its thermostable proteins more amenable to crystallization.

Stark et al., (2015) identified four of the five spliceosomal snRNAs (U2, U4, U5 and U6 snRNA) in *C. merolae* by conserved secondary structural features. In addition to no candidate for U1 snRNA being found, no U1 snRNP proteins were identified: *C. merolae* lacks U1C, U1A and U1-70k, as well as the seven additional proteins present in the *S. cerevisiae* U1 snRNP (Luc7, Nam8, Prp39, Prp40, Prp42, Snu56, and Snu71) which led to the conclusion that U1 snRNP is entirely absent in this organism. Proteins that associate with the assembled U1 snRNP, such as Prp28 which catalyzes the release of U1 snRNP from the spliceosome, are also absent.

There is precedence for U1-independent splicing *in vitro* (Crispino et al., 1994), and in some naturally occurring transcripts (Crispino et al., 1996; Fukumura et al., 2009). *C. merolae* exhibits extended base-pairing between the 5' SS and the U6 snRNA, which has been shown to increase splicing efficiency in the absence of U1 snRNP (Crispino & Sharp, 1995).

The 43 splicing proteins identified in *C. merolae* include: the Sm and Lsm cores, the SF3a and SF3b (minus p14) complexes, the U4/U6 di-snRNP proteins Snu13 and Prp3, four of the U5 snRNP proteins (Prp8, Brr2, Snu114 and Dib1), Cef1 of the nineteen complex, first step factor Yju2 (but not Cwc25), second step factor Prp22, and disassembly helicase Prp43 (Figure 6-1A). An additional 20 splicing adjacent proteins, such as the exon junction complex, were also identified (Stark et al., 2015).

The splice sites of *C. merolae* are highly conserved. The consensus 5' splice site (5' SS) is GUAAGU (Figure 6-1B; Matsuzaki et al., 2004), as compared to GUAUGU in *S. cerevisiae* (Spingola et al., 1999), and GURAGU (R is a purine) in humans (Zhang, 1998). Twenty-one introns use GUAAGU, one uses GCAAGU (difference from the consensus 5' SS bolded), and two introns use GUAAGC. Both introns with GUAGGU are found in the only gene containing two introns. All introns retain base-pairing with the U6 snRNA ACUGAG sequence and the U5 snRNA sequence GUCUGC. There is no degeneracy within the *C. merolae* branch region: 100% of introns have the branch site ACUA**A**CC (branch adenosine bold and underlined; Figure 6-2B) that forms an imperfect duplex with GGUAGU in the

U2 snRNA (Matsuzaki et al., 2004). The AG dinucleotide at the 3' splice site (3' SS) is also completely conserved. Unlike in *S. cerevisiae* and human spliceosomes, complementarity between U5 snRNA loop 1 and both exons, rather than just the 5' exon, is predicted to hold the splice sites in place for the second step of splicing (Stark et al., 2015).

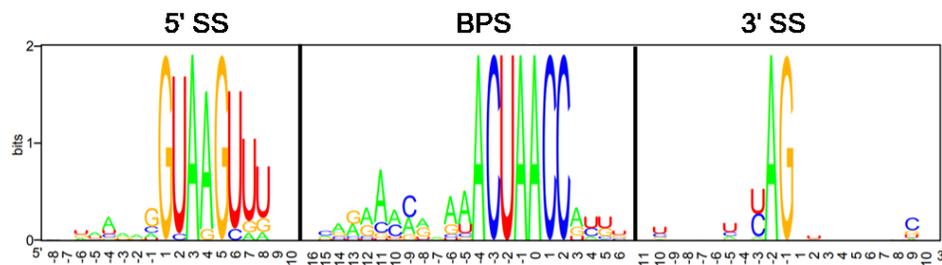
The identified introns in *C. merolae* are spliced, as shown by RT-PCR (Figure 6-1C; Stark et al., 2015). Primers that bind to either exon produce two bands upon RT-PCR: the unspliced pre-mRNA, and the spliced mRNA. *C. merolae* has a much higher ratio of unspliced RNA:spliced RNA when compared with other organisms. Only five of 27 introns (18.5%) are removed in more than 50% of their transcripts. In *S. cerevisiae*, 70% of introns were shown to be spliced at levels above 80%, and 85% of introns splice above 50% as shown by RNA-seq (Grisdale et al., 2013). Introns in *Candida albicans* were spliced at similar rates (80% of introns were spliced above 80%, and 90% spliced above 50%; Grisdale et al., 2013).

To determine intron length I compared the annotated mRNA sequence with the *C. merolae* genome (Table 6-1). Branch site annotation was done by hand. Intron length in *C. merolae* ranges from 70 to 429 bases, with an average length of 208 bases over 26 introns. Five introns are less than 100 bases, eight are between 104-184 bases, and eight introns are between 234-276 bases. There are no introns 185-233 bases long. Four introns are between 301-371, and one intron is over 400 bases. The 27th intron has been excluded from this analysis because it is predicted to 1,309 bases long. RT-PCR performed with primers for this intron expected an

A

Complex	Proteins
SM Core	SmB, SmD1, SmD2, SmD3, SmE, SmF, SmG
U1 snRNP	none
U2 snRNP	Prp9, Prp11, Prp21, Hsh155, Cus1, Hsh49, Rse1, Rds3, Prp5
U2 Related	Prp43, Mud2/U2AF ⁶⁵
U5 snRNP	Prp8, Brr2, Snu114, Dib1
U4/U6 snRNP	Prp3, Snu13
U6 Lsm Core	Lsm1, Lsm2, Lsm3, Lsm4, Lsm5, Lsm6, Lsm7
NTC	Cef1, Prp46, Bud31
Complex A	SF1, Sub2
Complex B	Prp38
Complex B ^{act}	Yju2
Second Step	Prp16, Prp22
Exon Junction	Fal1, Yra1, Thoc2
SR Proteins	Rsp31, SRSF2
hnRNPs	hnRNP H3
Other	Dbr1

B



C

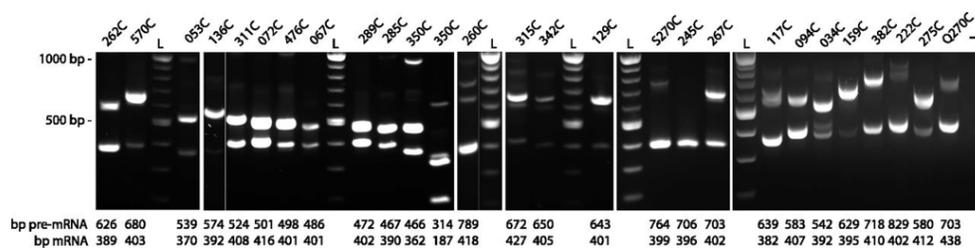


Figure 6-1. Splicing in the red alga *C. merolae*. A. List of splicing proteins found in *C. merolae*. Based on Stark et al., 2015. B. Conservation of the *C. merolae* 5' splice site (5' SS), branch point sequence (BPS), and 3' splice site (3' SS); adapted from Matsuzaki et al. (2004). Images generated with WebLogo. C. RT-PCR of *C. merolae* exons to show spliced (lower band) or unspliced (upper band) mRNA. Image originally published in Stark et al. (2015). Permission not required for use.

unspliced amplicon of 706 bases. The average distance between the branch adenosine and the 3' SS is 28 bases, with the shortest at 8 bases and the longest at

Gene	Length	BPS-3' SS
289	70	15
285	77	16
72	85	27
67	85	26
476	98	46
350b	104	28
311	116	60
350a	127	26
34	150	53
275	168	21
53	169	8
94	176	19
136	182	37
159	234	19
129	242	15
315	245	27
342	245	9
117	257	24
Q270	265	27
262	267	27
570	276	28
267	301	27
382	308	33
S270	366	27
260	371	20
222	429	61
245	1309	23

Table 6-1. *C. merolae* introns. Length and BPS to 3' SS distance of the 27 introns in *C. merolae*. Gene numbers correspond to those in Figure 6-1C.

27 bases. Twenty introns have a distance less than 30 bases. *C. merolae* has no predicted poly-pyrimidine tract. Based on visually sorting intron removal by comparing levels of spliced and unspliced RNA, it appears that longer introns are spliced more efficiently than shorter introns. There does not appear to be a

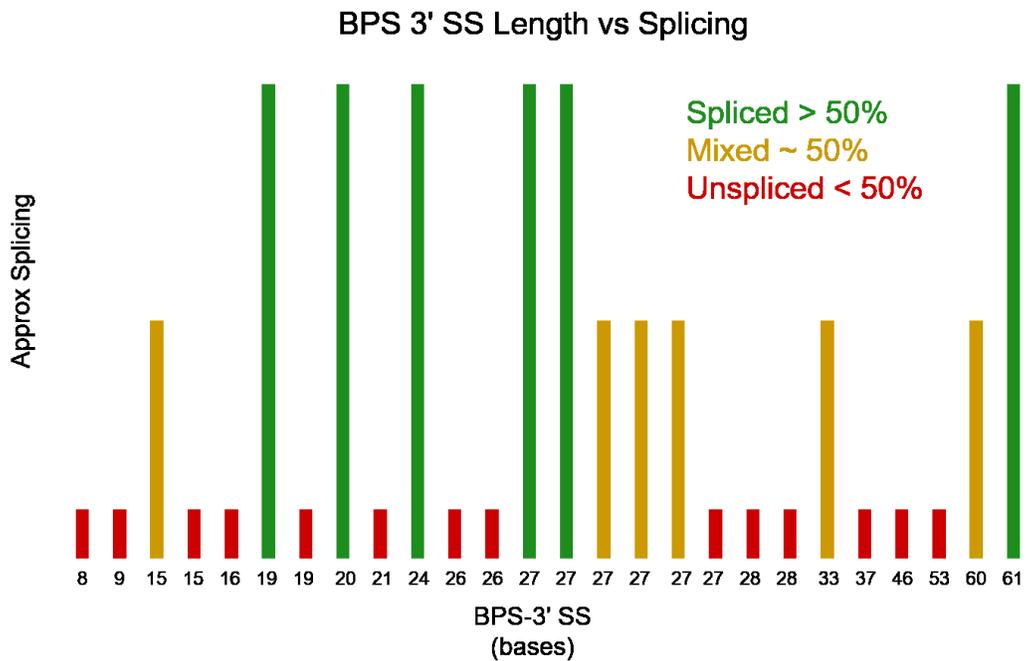
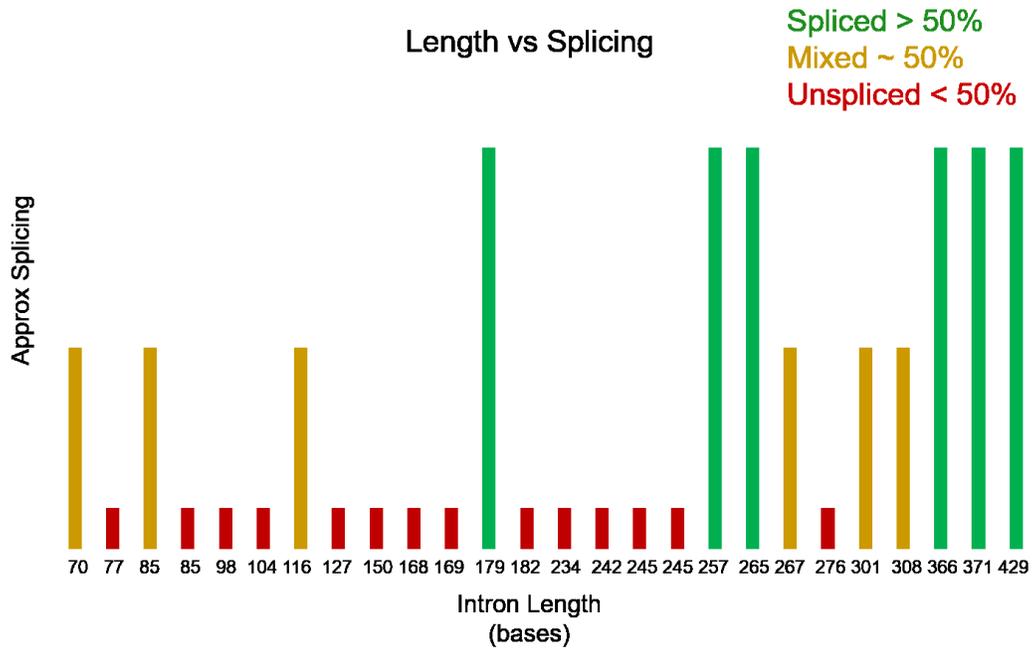


Figure 6-2. *C. merolae* Intron Removal Levels. Estimations of *C. merolae* splicing from Figure 6-1C. Splicing has been approximated by eye to spliced (>50% splicing) mixed spliced and unspliced (~50% splicing) and unspliced (<50% splicing) and compared to intron length (top) and distance between the BPS and 3' SS (bottom).

relationship between BPS to 3' SS distance and intron removal (Figure 6-2)

6-2. The RNase H domain of Prp8

Prp8 is a large, highly conserved, protein at the center of the spliceosome composed of six domains: the N-terminal domain (NT); the core, consisting of the reverse transcriptase and thumb domain (RT), the linker region, and the endonuclease domain (EN); and the C-terminal region featuring the RNase H domain (RH), and the Jab/MPN domain (Galej et al., 2013).

RNase H is a protein that cleaves the RNA in RNA/DNA hybrids (reviewed in Crouch & Dirksen, 1982). In cells it removes Okazaki fragments following DNA replication. During retroviral infection RNase H degrades the RNA genome following synthesis of the DNA copy by the reverse transcriptase. The structure of RNase H consists of a five-stranded β -sheet and at least two α -helices. Conserved acidic residues coordinate a Mg^{2+} ion that catalyzes RNA cleavage (Figure 6-3A; Yang et al., 1990).

The structure of the Prp8 RH domain was published simultaneously by three independent groups, and shows an N-terminal RNase H fold followed by a C-terminal cluster of five α -helices (Figure 6-3B; Pena et al., 2008; Ritchie et al., 2008; Yang et al., 2008). A 17 amino acid insertion between β 1 and β 2 precluded prediction of the domain structure from the primary sequence. A number of alleles that suppress mutations in the splice sites and BPS, as well as mutations within U4 and U6 snRNA, are located in this insertion (Schellenberg et al., 2013). Deletion

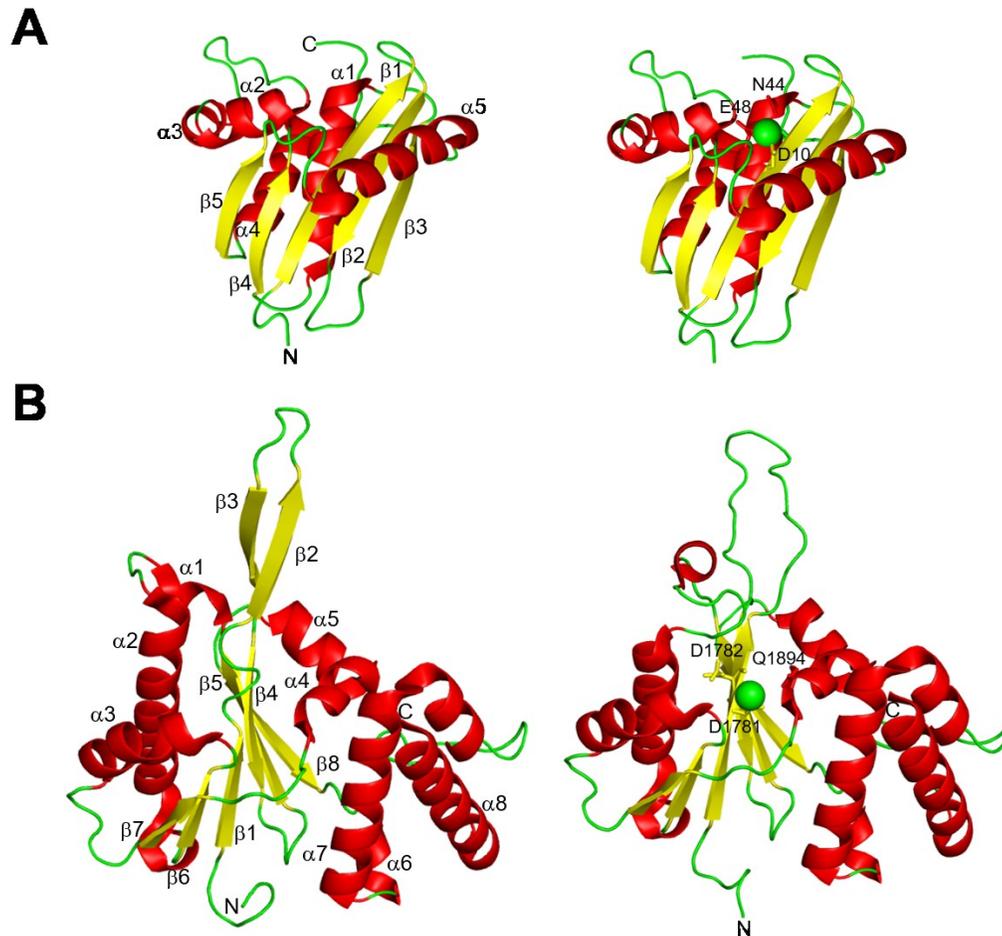


Figure 6-3. RH domains of RNase H and Prp8. A. Structure of RNase H (PDB 1RNH) with the 5-stranded β -sheet buttressed by two α -helices and two additional helices. Right, a metal ion is coordinated by three conserved residues: Asp10, Asn44, and Glu48. B. The human RH domain from Prp8 (PDB 4JK7) features a 17 amino acid insert between β 1 and β 2, and crystallized in two monomers. In monomer A (left) the insertion is folded into a β -hairpin, while in monomer B (right), it forms a loop. A metal ion is observed bound in monomer B, coordinated by Asp1781, Asp1782, and Gln1894.

of these 17 amino acids is lethal in *S. cerevisiae* (Mayerle et al., 2017).

The crystal structure from our lab modelled two conformations for the insertion: a β -hairpin (observed in other structures; monomer A), and a disordered region that in subsequent structures was resolved into a loop (monomer B). The

metal binding site of RNase H is also conserved in the Prp8 RH domain, and a bound metal ion was observed in monomer B (Schellenberg et al., 2013).

6-3. The RH domain of Prp8 in the spliceosome

In the cytosol, the U5 snRNP contains the U5 snRNA, the Sm core, Snu114, Prp8 and the U5 snRNP assembly factor Aar2 (Weber et al., 2011). Upon import to the nucleus, Aar2 is replaced with the mutually exclusive Brr2 (Boon et al., 2007). As Prp8 did not crystallize alone, the C-terminal two-thirds of *S. cerevisiae* Prp8 was co-crystallized with Aar2 (Galej et al., 2013). Examination of the resulting structure shows a β -strand from Aar2 linking the β -hairpin of the RH domain with a β -sheet in the Jab/MPN domain into one long β -sheet (Figure 6-4). As the C-terminal RH and Jab/MPN domain do not contact the rest of Prp8, Aar2 may have been required to homogenize the position of the C-terminus of Prp8 to allow for crystallization.

The assembled U5 snRNP binds the U4/U6 di-snRNP to form the U4/U6.U5 tri-snRNP. Alignment of the core of Prp8 from the crystal structure to the tri-snRNP shows a repositioning of the RH and Jab/MPN domains (Nguyen et al., 2016). The Jab/MPN domain bound to Brr2 is rotated down to be in line with the EN domain and the rest of the core of Prp8. There is a long, 58 amino acid linker between the RH and Jab/MPN domains to facilitate this movement. The RH domain flips down 180° so that the β -hairpin points down towards the Jab/MPN domain, instead of up towards the NT domain. The hairpin is modeled as a bent

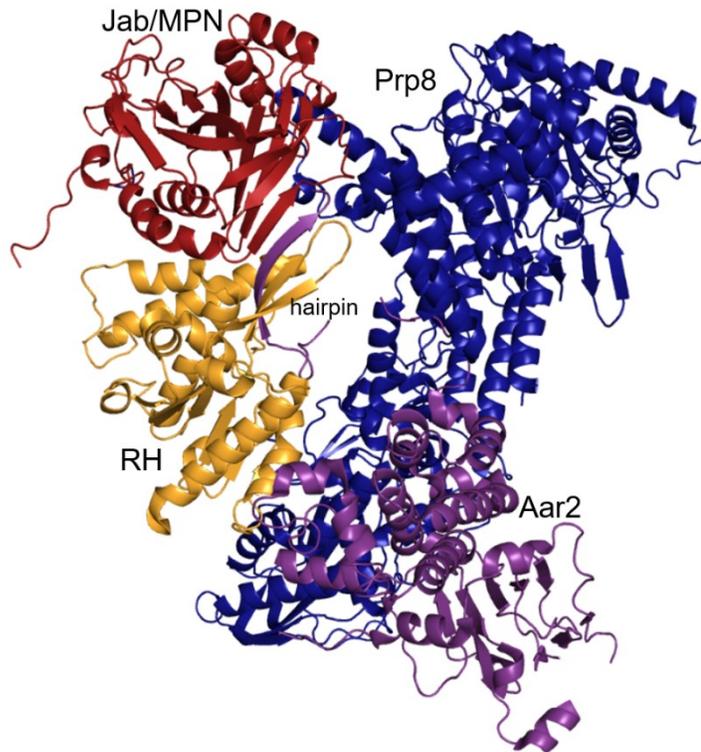


Figure 6-4. Crystal structure of *S. cerevisiae* Prp8 with Aar2. *S. cerevisiae* Prp8 residues 885-2413 was crystallized with the U5 snRNP assembly protein Aar2 (PDB 4I43). A β -sheet is formed between the RH β -hairpin, the β -sheet of the Jab/MPN domain and a β -strand from Aar2.

loop unique to this structure. A small segment of the hairpin interacts with an unidentified peptide that, based on alignment with the B complex structure, is likely Snu66 (Figure 6-5A).

The RH domain is located at the edge of the tri-snRNP. The two α -helices that are part of the RNase H fold interact with a small helix of Prp31 and helices at the N-terminus of Prp6. The C-terminal helical bundle of the RH domain interacts with a long helix from Prp3, which links the RH domain with Prp4, Snu13, and the U4/U6 snRNA duplex. A helix from the unidentified chain that is likely Snu66 inserts into the cleft formed between the N-terminal RNase H fold and the C-

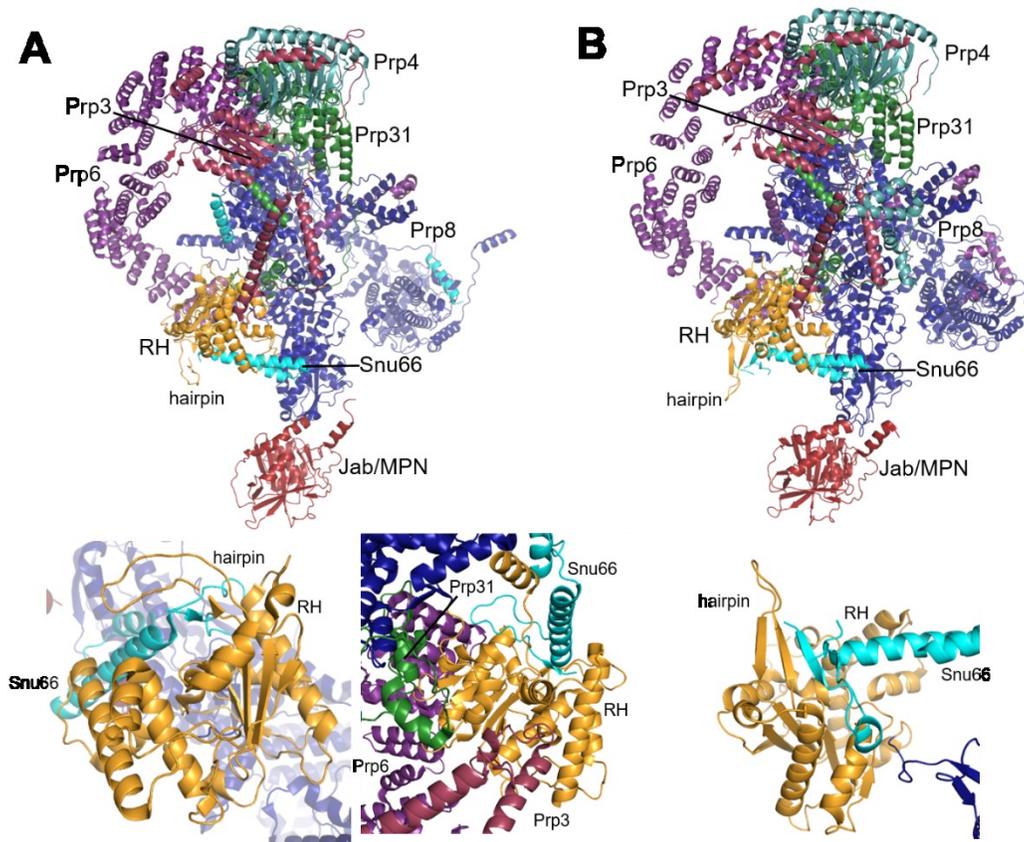


Figure 6-5. Interactions of the RH domain of Prp8 in the tri-snRNP and the spliceosomal B complex. A. Prp8 (blue) in the tri-snRNP (top; PDB 5GAN) interacts with Snu66 (cyan), Prp6 (purple), Prp3 (raspberry), and Prp31 (green). Prp3, Prp6, and Prp31 link Prp8 with Prp4 (teal). The RH β -hairpin (orange) interacts with Snu66 (bottom left). Bottom right, a close up view of the interactions of Prp3, Prp6, Prp31 and Snu66 with the RH domain. B. The interactions of the RH domain of Prp8 in the spliceosomal B complex (PDB 5ZWM) are the same as those in the tri-snRNP. Bottom, the β -hairpin forms a longer β -sheet with a β -strand of Snu66.

terminal helical bundle (Figure 6-5A).

Incorporation of the tri-snRNP into the A complex, giving rise to the B complex, does not change the positioning of Prp8 or the proteins it interacts with (Plaschka et al., 2017; Bai et al., 2018). The 17 amino acid insertion that was a bent loop in the tri-snRNP is again a hairpin that forms a three-stranded β -sheet with a β -strand of Snu66 (Figure 6-5B).

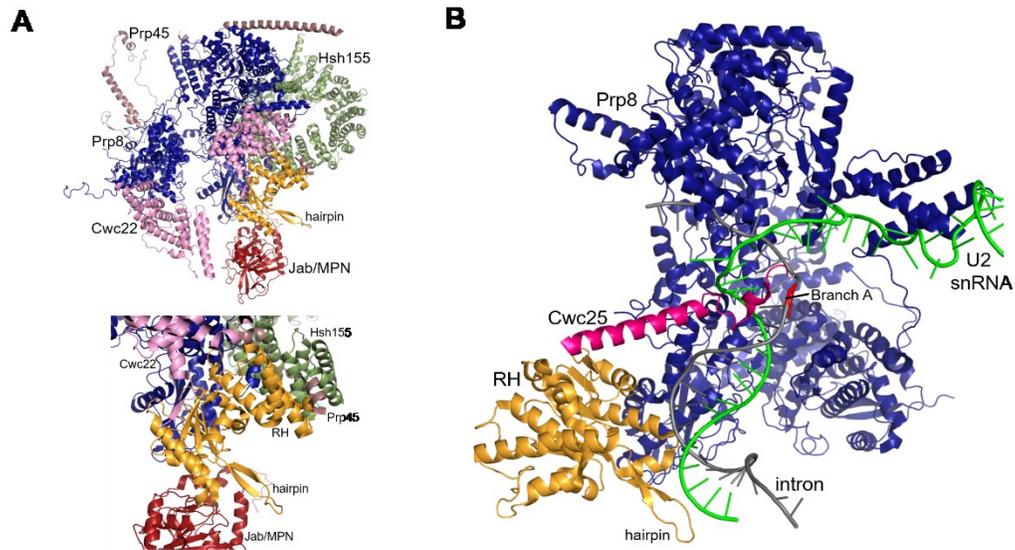


Figure 6-6. Prp8 RH interactions in the B^{act} and C spliceosomal complexes. A. The RH domain loses all its interacting proteins in the transition from the B to B^{act} complex (top; PDB 5GM6) and the loss of the U4 snRNP. It gains new interactions with Prp45 (mauve), Cwc22 (pink) and Hsh155 (green). Bottom, a close up view of the RH domain and its interactions with Prp45, Hsh155, and Cwc22. B. Interactions of the RH domain in the spliceosomal C complex (PDB 5GMK). During the transition from the B^{act} to the C complex (via B*) the spliceosome loses Hsh155, along with the rest of SF3, and the first step of splicing occurs. In the C complex Cwc25 (magenta) stabilizes the BPS/U2 snRNA duplex (grey and green, respectively) and interacts with the RH domain. The β -hairpin lies parallel to this RNA duplex.

Prp8 undergoes a conformational change between the B and B^{act} complexes (Yan et al., 2016). While the core remains constant, the NT and Jab/MPN domains shift slightly. The RH domain is displaced around the core of Prp8 towards the Jab/MPN domain, which has also moved towards the RH domain, to sit in a pocket formed by the EN and Jab/MPN domains. In the tri-snRNP and the B complex, the RH domain interacts with components of the U4 snRNP that are dissociated from the spliceosome in the transition to B^{act}, including Prp3, Prp4, Prp31 and Snu66. As such, the RH domain loses all its previous interacting partners. It forms new interactions with Hsh155, Prp45, and Cwc22. The C-terminal helices of the RH

domain interact with the HEAT repeats of Hsh155 about halfway around the protein from the interaction of Hsh155 with the branch duplex. The C-terminal tail of Prp45 bridges the C-terminal helices of the RH domain with the linker region. Cwc22 adds a short β -strand to the β -sheet at the center of the RNase H fold. The β -hairpin has no observed interactions in this complex (Figure 6-6A).

The RH domain continues to rotate around the core of Prp8 in the transition from the B^{act} to the C complex and the first step of splicing (Wan et al., 2016). It also twists so that the β -hairpin, which was facing away from the NT domain, is now parallel to the NT domain. Hsh155 is released from the spliceosome with the rest of SF3b during the transition from the B^{act} to the B* complex and the first step of splicing. The movement of the RH domain disrupts its interactions with Prp45 and Cwc22, whose positions do not change. The rearrangement of the U2 snRNA moves the U2 Sm core into proximity with the C-terminal helices of the RH domain. The N-terminus of the first step factor Cwc25 fragment lies in the groove of the intron/U2 snRNA duplex, while the C-terminus is positioned near the C-terminal helix of the RH domain. The intron/U2 snRNA duplex lies alongside Prp8, with regions in the two helices of the RNase H fold and the β -hairpin positioned to hydrogen bond with the RNA. The β -hairpin is beside the duplex (Figure 6-6B).

During the transition from the C to the C* complex the RH domain flips so that the β -hairpin points up towards the NT domain (Figure 6-7A; Yan et al., 2017). The hairpin is modeled mostly as a loop, with small patches of β -strand. The movement places the hairpin in the midst of the U2 and U6 snRNAs. The β -hairpin

is in the minor groove of this duplex underneath the branched adenosine. The duplex between the 5' end of the intron and U6 snRNA is on the other side of the β -hairpin from the BPS/U2 snRNA duplex. The top of the β -hairpin is perpendicular to, and may contact, the long helix of Cef1 that helps maintain the structure of the U6 snRNA (Figure 6-7B). Prp17 interacts with the 5' end of the BPS/U2 snRNA duplex and one side of the RH domain. The helices of Prp18 are packed at the bottom of the RNase H fold (Figure 6-7C).

The RH domain and its interactions in the post-catalytic spliceosome structure (P, following exon ligation; Liu et al., 2017; Wilkinson et al., 2017) is very similar to that in the C* structure. Slu7, which was not visualized in the C* complex, is bound to the RH domain. Two helices bind the C-terminal helical bundle of the RH domain, while the rest of the protein binds the bottom of the RNase H fold, with a loop extending up along the two helices and pointing towards the intron/U6 snRNA duplex (Figure 6-8A).

Dissociation of the ligated exons from the P complex gives rise to the intron lariat spliceosome (ILS). The RH domain rotates away from the interactions it had in the P complex (Wan et al., 2017). It is now at the edge of the spliceosome: Slu7 and Prp17 are not visible, nor are there any other factors observed interacting with the RH domain (Figure 6-8B).

When the core domains of Prp8 from each step in the splicing cycle (the B, B^{act}, C, C* and P complexes) are overlaid, there is a circular movement of the RH domain over the course of splicing: the position of the RH domain is between the

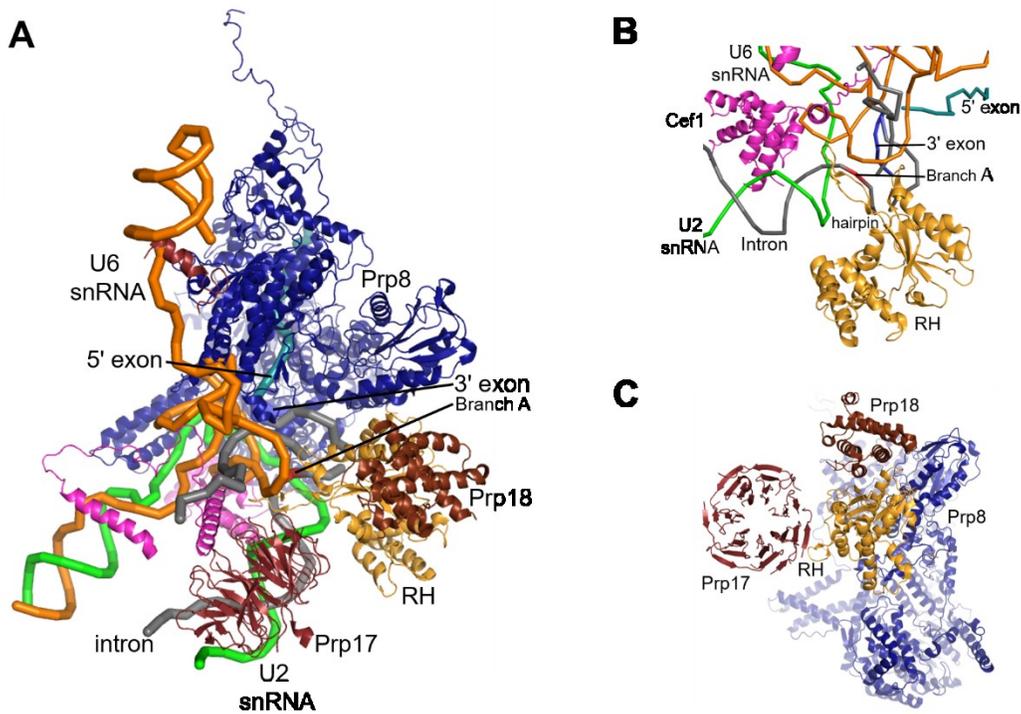


Figure 6-7. Interactions of the Prp8 RH domain in the C* complex. In the C* complex (PDB 5WSG) the RH β -hairpin (orange) inserts into the minor groove of the BPS/U2 snRNA duplex (grey and green, respectively). U6 snRNA is on the other side of the β -hairpin. The tip of the hairpin may contact the long helix of Cef1 (magenta). Second step factors, Prp17 (maroon) and Prp18 (brown) also interact with the RH domain. The RH domain, and the β -hairpin in particular, may help stabilize the RNA core for the catalysis of exon ligation. B. Close up view of the β -hairpin with Cef1, the BPS/U2 snRNA duplex, and the U6 snRNA. C. rotated view of the RH domain with Prp17 and Prp18.

location of the RH domain in the complex before and the complex after (Fica & Nagai, 2017). Following splicing, in the C* and P complexes, the RH domain is beside where it began in the tri-snRNP and B complexes (Figure 6-9).

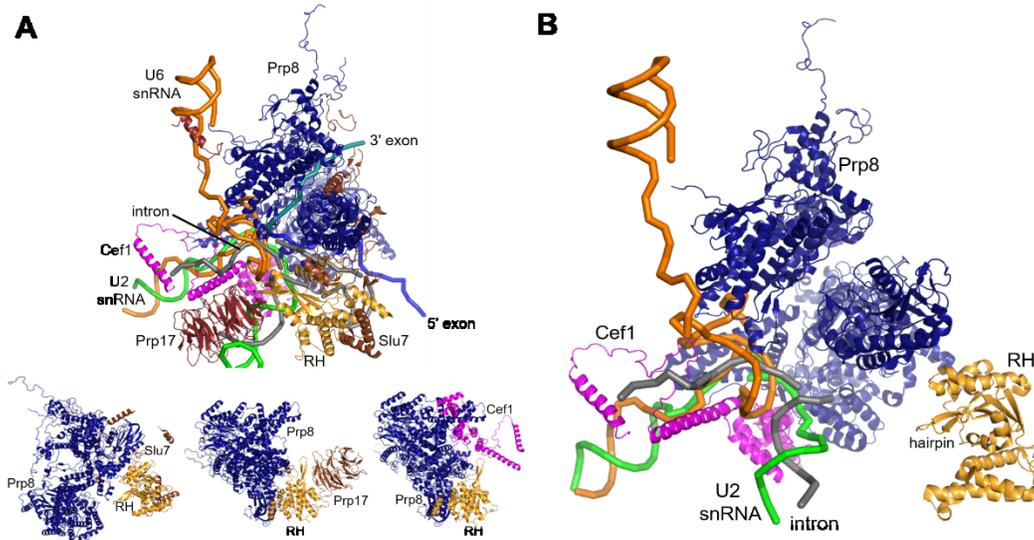


Figure 6-8. Prp8 RH domain interactions in the P and ILS complexes. A. Interactions of the Prp8 RH domain in the P and ILS complexes. A. In the P (post-catalytic) spliceosome (PDB 6BK8), the RH domain interacts with Slu7 (brown), Prp17 (maroon), and Cef1 (magenta). The interaction of the hairpin with the RNA as seen in the C* complex is maintained. Bottom, Prp8 with Slu7 (left), Prp17 (middle) or Cef1 (right). B. During the release of the ligated exons and the transition to the ILS complex (PDB 5Y88), the RH domain of Prp8 rotates away from the RNA core and is no longer interacting with other proteins in the spliceosome.

Results

6-4. The RH domain of *C. merolae* Prp8

The RH domain is highly conserved: 69% between humans and *S. cerevisiae*, compared to 61% conservation of Prp8 as a whole (Hodges et al., 1995). Prp8 is also conserved in *C. merolae*, with 31% sequence conservation overall, and 33% identity for the RH domain with humans. However, sequence alignment predicted a lack of the β -hairpin insert within the RH domain (Figure 6-10). Homologues for proteins that interact with the hairpin insert in *S. cerevisiae*, such as Aar2 and Snu66, have not been identified in *C. merolae* (Stark et al., 2015).

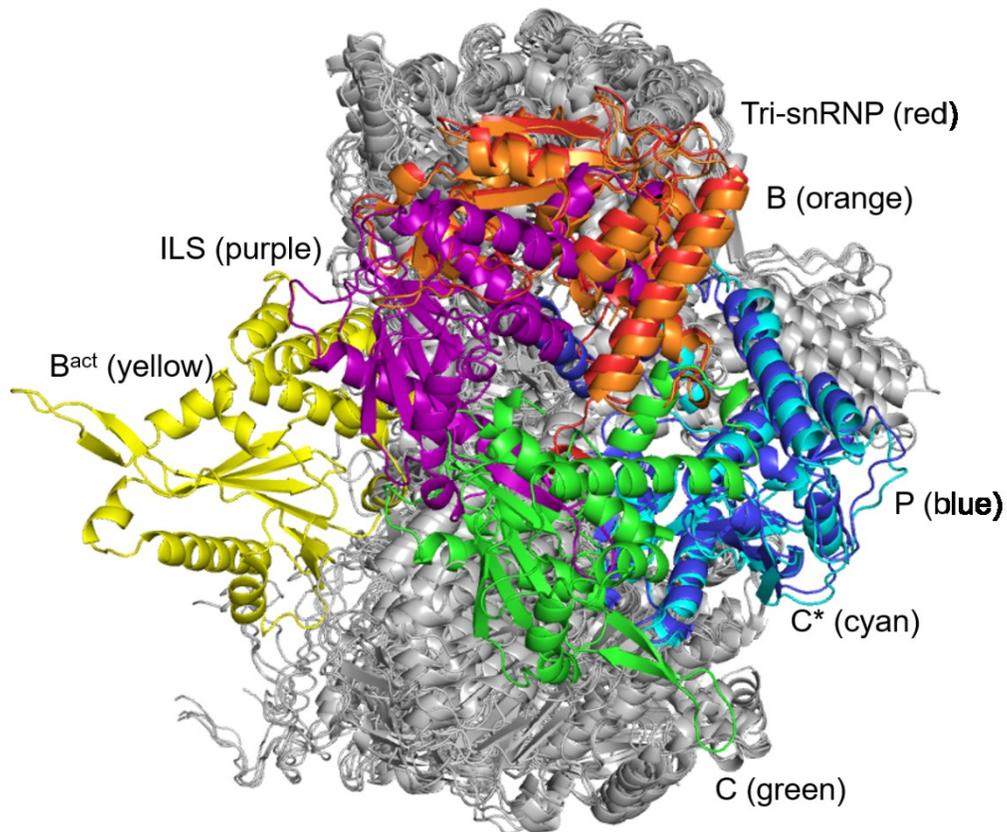


Figure 6-9. Location of the RH domain relative to the Prp8 core during the splicing cycle. The RH domain rotates around the bottom of Prp8 during splicing (grey). Complex assembly proceeds from the tri-snRNP (red) to the B (orange), B^{act} (yellow), C (green), C* (cyan) and P (blue) complexes. It then flips over to its position in the ILS (purple) before being recycled back into the tri-snRNP.

In order to confirm the predicted absence of the β -hairpin insert, I cloned, expressed and purified the RH domain of *C. merolae* consisting of residues 1848-2065. It expressed cleanly, and I was able to crystallize it, then solve the structure to 2.75 Å resolution using the human RH domain as a search model for molecular replacement (PDB 4JK7; Table 6-2).

The *C. merolae* RH domain crystallized with two monomers in the asymmetric unit. It adopts the RNase H fold followed by five C-terminal helices

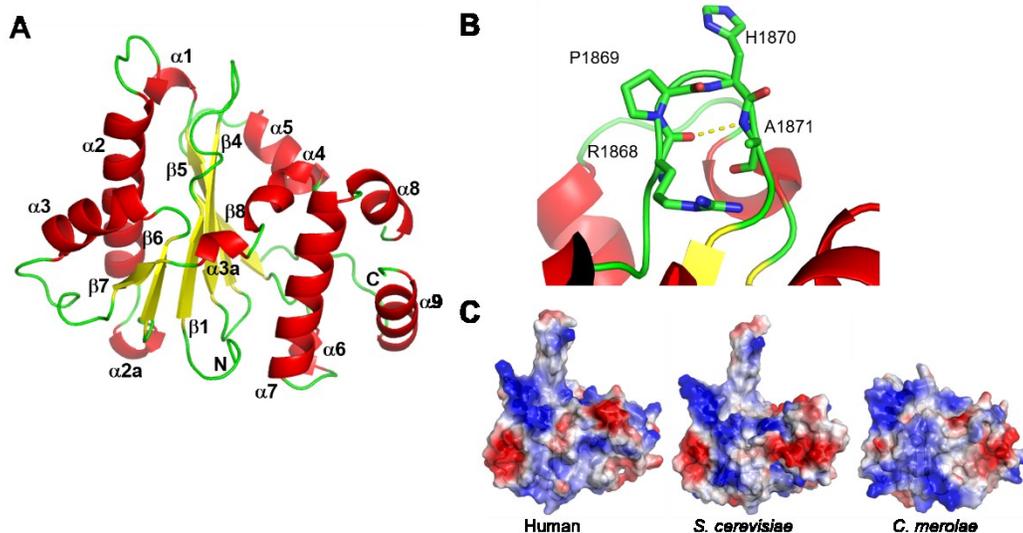


Figure 6-11. Crystal structure of the *C. merolae* Prp8 RH domain. A. Structure of the *C. merolae* RH domain with colours and numbering consistent with the alignment in Figure 6-10. B. The 17 amino acid β -hairpin in other organisms is replaced by RPHA (1868-1871) in a β -turn. C. Electrostatic representation of human (left), *S. cerevisiae* (middle) and *C. merolae* (right) show a conserved basic patch beside the hairpin. *C. merolae* lacks the acidic patch on the N-terminal helices present in the human and *S. cerevisiae* RH domains.

The surface charge distribution shows the C-terminal helical bundle is acidic, while the N-terminal helices in the RNase H fold are basic. In both the human and *S. cerevisiae* RH domains the negative patch at the C-terminus is smaller than that in *C. merolae*. The N-terminal helices in human and *S. cerevisiae* are also acidic, versus basic in the *C. merolae* structure. The basic patch alongside the β -hairpin in both human and *S. cerevisiae* is present in *C. merolae* even though the β -hairpin itself is absent (Figure 6-11C). This region binds the intron in the C, C* and P spliceosomes (PDB 5GMK; 5WSG; 6BK8 respectively), an interaction that may be conserved in *C. merolae* despite the lack of β -hairpin.

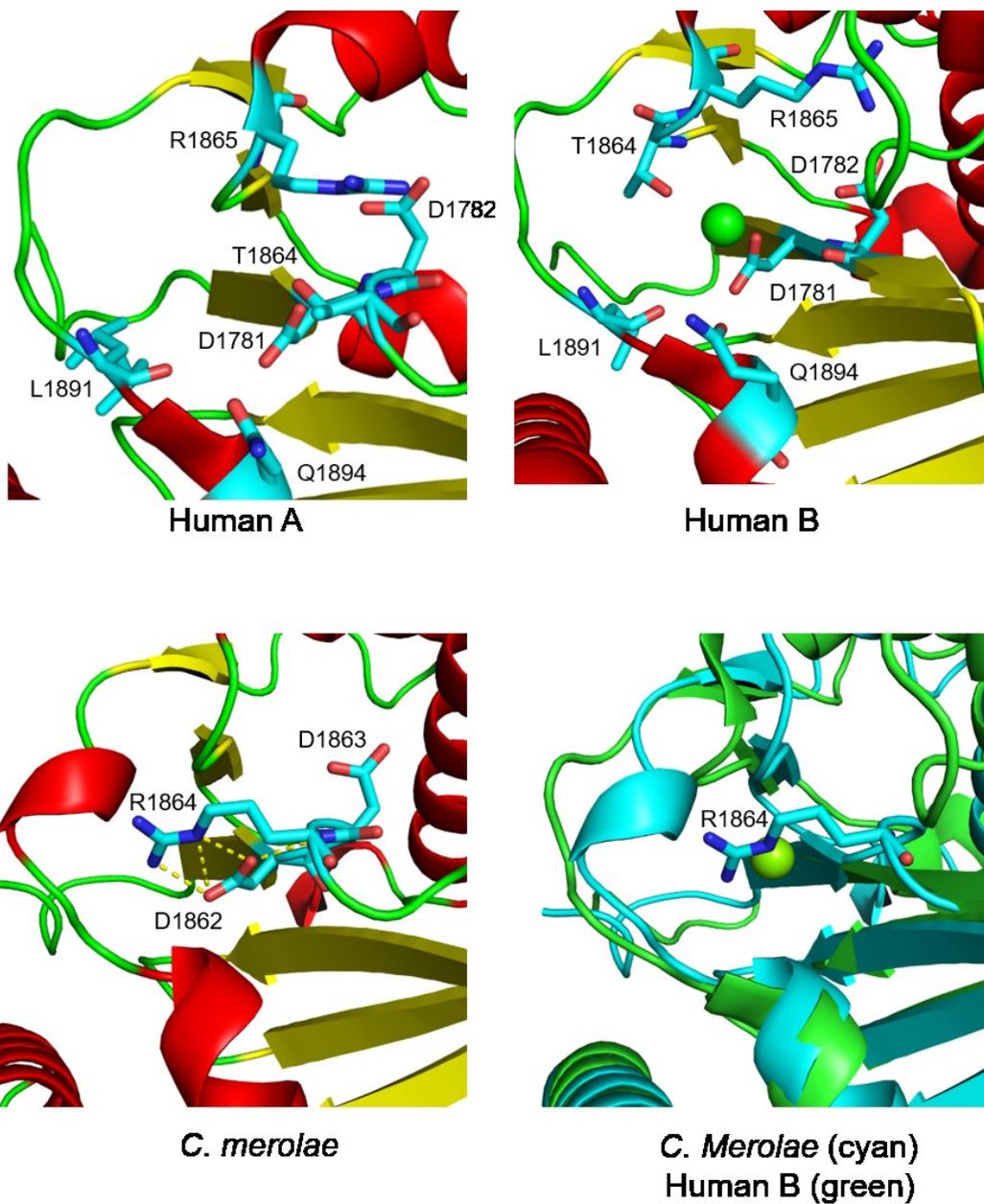


Figure 6-12. Metal binding by Prp8. Top, the metal binding site of human Prp8 (PDB 4JK7) consists of Asp1781, Asp1782, Gln1894 and the backbone of Leu1891. The side chain of Arg1865 blocks the metal binding site in monomer A (left), but is rotated out of the site in monomer B (right), allowing a metal to bind. *C. merolae* has a partially conserved metal binding site (Asp1862, Asp1863, and Gln1964; bottom left). Arg1864 hydrogen bonds with Asp1862 and occupies the metal binding site. Bottom right, alignment of the *C. merolae* structure with human monomer B shows *C. merolae*'s Arg1864 overlapping with the bound metal, suggesting the arginine and metal may play similar roles.

In monomer B of the human RH structure a metal is observed coordinated by Asp1781, Asp1782, Gln1894, Thr1864, and the backbone of Leu1891 (Schellenberg et al., 2013). This metal is not observed in the *S. cerevisiae* structures (all of which are in monomer A), and the equivalent of Gln1894 is Ser1966 (discussed in the next chapter). The metal binding site is similarly conserved in *C. merolae*, with both aspartates (Asp1862 and Asp1863) and the glutamine (Gln1964) present, but the threonine is replaced by an alanine (Ala1932). However, the metal binding site is occupied by the sidechain of Arg1864 (Thr1783 in humans). Whatever role the metal in the human RH domain is playing may be performed by Arg1864 in *C. merolae* (Figure 6-12).

6-5. Conservation of the β -hairpin in other species

C. merolae is not the only organism predicted to lack a β -hairpin insertion in the RH domain (Figure 6-13). Sequence alignment does not predict a β -hairpin in a number of microsporidia species, including *Encephalitozoon cuniculi*, an intracellular parasite that infects humans with suppressed immunity, and the first microsporidian genome to be completely sequenced (Katinka et al., 2001). Microsporidia have some of the smallest eukaryotic genomes, such as the 2.9 Mb genome of *E. cuniculi* that encodes fewer than 2,000 proteins. Only about 30 spliceosomal proteins have been identified in *E. cuniculi*, including Prp8 and Brr2, along with 34 introns (Katinka et al., 2001). As in *C. merolae*, the splice sites and BPS are highly conserved (Grisdale et al., 2013). The consensus sequence of the

Data collection	
Space group	P 1 21 1
Cell dimensions	
a, b, c (Å)	56.516, 67.508, 58.859
α , β , γ	90 101.64 90
Wavelength (Å)	0.97949
Resolution (Å)	2.75
R _{meas}	0.109 (0.581)
I / σ I	9.7
Completeness (%)	99.54 (99.39)
Redundancy	3.3 (3.3)
Refinement	
Resolution (Å)	36.42 - 2.75 (2.85 - 2.75)
No. reflections	376,466
<i>R</i> _{work} / <i>R</i> _{free}	0.2529 / 0.2911 (0.3137/0.3845)
No. atoms	
Protein	3123
Ligand	8
<i>B</i> factors	
Protein	73.77
Ligand	54.51
R.M.S. deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.15
Ramachandran favored (%)	97.85
Ramachandran allowed (%)	2.15
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	3.73
Clashscore	14.61

Table 6-2. Data collection and refinement statistics for the RH domain of *C. merolae* Prp8. Data were collected from a single crystal. Values in parentheses are for the highest resolution shell (2.85-2.75 Å). Molecular replacement was performed using PDB 4JK7 chain A as the search model.

5' SS is GUAAGU, and the AG at the 3' SS is completely conserved. The BPS is UAAYUU (Y=pyrimidine). These introns do not contain a PPT, and the distance between the end of the BPS and the 3' SS is no longer than three nucleotides.

Further annotation of the *E. cuniculi* genome is required, as there was no U1 snRNA found, despite homologues of U1A and U1C being annotated (Katinka et al., 2001). Homologues of both U2AF subunits were identified (*C. merolae* encodes for the large subunit of U2AF, but not the small). Homologues for Sap49 and Sap145 have been identified, but not for the other components of SF3b, including Sap155.

BLAST searches for homologues of Aar2 and Snu66 did not identify these proteins in *E. cuniculi* or other organisms with low conservation of the β -hairpin. These organisms also have fewer annotated introns than organisms that possess a highly conserved β -hairpin and the proteins Aar2 and Snu66 (Table 6-3). This suggests a decrease in intron number is accompanied by a simplification of the spliceosome similar to that observed in the red algae. (Qui et al., 2018).

A second group of organisms conserves the length of the insert, but not its sequence. These organisms, including *G. lamblia* and *T. cruzi*, have few *cis*-introns but all undergo some form of *trans*-splicing (Vanacova et al., 2005; Gunzl, 2010; Kamikawa et al., 2011; Cuypers et al., 2017; Roy, 2017). In contrast to *cis*-splicing, where the ligated exons are transcribed on the same piece of pre-mRNA, *trans*-splicing features exons transcribed on different pre-mRNA transcripts that are then

V. culicis	-----L-SN-----S IT KYEEYMFNARTGVL	1449
N. bombycis	VSSPKEIFHDS---LIVEDRLL FTS -----NRSLLILDPE SGRK	1564
E. cuniculi	VNSNGDLFTSG---LLVDVKAL L RK-----EKTLFVLD PASGNL	1748
O. colligata	VRSSGDLFSSG---MIIDSRCL G RK-----QRAVFAFD PNGNGL	1616
C. merolae	VGCGGDLWRQ--RLWIVDDRTAY R P-----HANGVIWI WETSTGRL	1885
G. lamblia	PSSIGDLFTGK--VLI V DDSLAY NFRMLNRDD TRASRV I INGFISIFNP QTGRL	1732
G. theta	ILNIDDFKKK--CL I VDD S CLSN H L E L Q N L Q K N V IN V H S G F L F I F NP I NG L I	1636
L. donovani	DTNIAELFSSGMR T WIVDD S ATY V T S E Q P T E D G G K K F K S EN G AV L V F E P V S G S L	1892
T. cruzi	VTNIAELFSEGM R TWIVDD S ATY V T S E Q P T A E G G R K F R S EN G AV L L F E P T T G Q L	1873
C. albicans	SNNYAELFNND-TQ L FVDD T NVY R V T V H K T F E G N L A T K P I N G C I F I L N P K S G Q L	1894
S. cerevisiae	SSNYAELFNND-I K L F VDD T NVY R V T V H K T F E G N V A T K A I N G C I F T L N P K T G H L	1889
M. daphniae	SANYGELFAGG-T T W F VDD S AVY R V T I H K T F E G N L T T R P I N G A L I I F N P R T G Q L	1867
T. vaginalis	STNFGELFGNK-I T W I V E D K H V R V T I H K T F E G N L T T S P V N G G V F I M N P A T G Q L	1807
P. falciparum	TQNYNELFSSQ-T I W F VDD T NVY R V T I H K T F E G N L T T K P I N G A I F I L N P K T G Q L	2532
E. histolytica	ITNYGELFT N Q-I I W F VDD S N I R V T T H K T F E G N H I T K P L N G C I F I F N P R T G G V	1760
C. crispus	SQNYGELFG N Q-V I W F VDD T NVY R V T A H K T F D G N H V T K P I N G A V L I F N P R T G Q L	1899
P. tricornutum	SQNYGELF S NQ-V I W F VDD T NVY R V T I H K T F E G N L T T K P I N G A I F I F N P R T G Q L	1818
P. tetraurelia	TQNYAELF S NQ-I I W F VDD T NVY R V T I H K T F E G N L T T K P I N G A I I I F N P K T G Q L	1799
T. thermophila	SSNFSELF S NQ-T I W F VDD T NVY R V T I H K T F E G N L T T K P I N G A I F I F N P R T G E L	1834
S. pombe	SSNYAELF S NQ-I Q L F VDD T NVY R V T I H K T F E G N L T T K P I N G A I F I F N P R T G Q L	1841
N. gruberi	SQNYGELF S NQ-T I W F VDD S DVY R V I H Q T S E G N S T S K P V N G A I I I F N P K T G Q L	1554
A. muscaria	SQNYSELF S NQ-I I W F VDD T NVY R V T I H K T F E G N L T T K P I N G A I F I F N P R S G Q L	1817
G. sulphuraria	SQNYGELF G NQ-I I W F VDD T NVY R V T I H K T F E G N L V T K P I N G A I F I F N P R T G Q L	1842
D. discoideum	SQNFGE L F S NK-I M W F VDD S NVY R V T I H K T F E G N L T T K P I N G A I F I F N P R T G Q L	1805
R. allomyces	SQNYGELF S NQ-V I W F VDD T NVY R V T I H K T F E G N L T T K P I N G A I F I F N P R T G Q L	1831
A. macrogynus	SQNYAELF S NQ-T I W F VDD T NVY R V T I H K T F E G N L T T K P I N G A I F I L N P R T G Q L	2087
C. reinhardtii	SQNYGELF S NQ-T V W F VDD T NVY R V T I H K T F E G N L T T K P I N G A I F I F N P R T G Q L	1829
A. thaliana	SQNYGE I F S NQ-I I W F VDD T NVY R V T I H K T F E G N L T T K P I N G A I F I F N P R T G Q L	1841
T. gondii	SQNYGELF S NQ-T I W F VDD T NVY R V T I H K T F E G N L T T K P V N G A I F I F N P R T G Q L	2019
P. infestans	SQNYGELF S NQ-I I W F VDD T NVY R V T I H K T F E G N L T T K P I N G A I F I F N P R T G Q L	1799
B. dendrobatidis	SQNYGELF S AQ-I I W F VDD T QVY R V T I H K T F E G N L T T K P I N G A L I I F N P R T G Q L	1851
R. delemar	STNYGELF S NQ-I I W F VDD T NVY R V T I H K T F E G N L T T K P I N G A I F I F N P R T G Q L	1834
C. elegans	SQNYGELF S NQ-I I W F VDD T NVY R V T I H K T F E G N L T T K P I N G A I F I F N P R T G Q L	1810
D. melanogaster	SQNYGELF S NQ-I I W F VDD T NVY R V T I H K T F E G N L T T K P I N G A I F I F N P R T G Q L	1877
H. sapiens	SQNYGELF S NQ-I I W F VDD T NVY R V T I H K T F E G N L T T K P I N G A I F I F N P R T G Q L	1817

Figure 6-13. Sequence alignment of the Prp8 β -hairpin region from 35 taxa. Organisms besides *C. merolae* also lack a β -hairpin insertion. Some, *V. culicis*, *N. bombycis*, *E. cuniculi*, and *O. colligata*, like *C. merolae*, lack the insert completely, while other organisms, *G. lamblia*, *G. theta*, *L. donovani*, and *T. cruzi*, retain the hairpin's length, but do not conserve its sequence. Hairpin sequence in bold, and coloured with respect to conservation of the human sequence.

spliced together (Lasda & Blumenthal, 2011). Both cis- and trans-splicing are catalyzed by the spliceosome. The most common form of trans-splicing is spliced leader trans-splicing (SLTS), where a standard, small RNA sequence is spliced onto the 5' end of the cell's mRNA. Conservation of the splice sites varies from highly conserved (*T. vaginalis*, with a stretch of 13 conserved residues between the BPS and 3' SS; Vanacova et al., 2005) to less conserved, as in *G. lamblia* (Roy et al.,

Species	Genome size (MB)	Gene number	Introns	Hairpin identity (% human)	Aar2	Snu66
<i>C. merolae</i> ¹	16.5	4803	27	absent	No	No
<i>E. cuniculi</i> ²	2.3	1981	37	absent	No	No
<i>O. colligata</i> ³	2.3	1820	30	absent	No	No
<i>N. bombycis</i> ³	15.7	4488	10	absent	No	No
<i>V. culicis</i> ³	6.1	2773	7	absent	No	No
<i>G. lamblia</i> ⁴	11.2	6502	19*	6	No	No
<i>G. theta</i> ⁵	0.55	464	17*	6	Yes	No
<i>L. donovani</i> ⁶	32.2	8032	9*	18	No	No
<i>T. cruzi</i> ⁷	41.5	13643	unknown*	18	No	No
<i>T. vaginalis</i> ⁸	176.4	59679	66	71	No	No
<i>S. cerevisiae</i> ⁹	12.1	5404	253	76	Yes	Yes
<i>G. sulphuraria</i> ¹⁰	13.7	6623	13630	94	Yes	No
<i>S. pombe</i> ¹¹	12.5	4929	4730	100	Yes	Yes
<i>C. elegans</i> ¹²	97	16260	108151*	100	Yes	Yes
<i>D. melanogaster</i> ¹³	~180	13601	>41000*	100	Yes	Yes
<i>H. sapiens</i> ¹⁴	2939.6	29399	272667*	100	Yes	Yes

Table 6-3. Intron number, hairpin conservation, and presence of Aar2 and Snu66 in 16 species.

Hairpin conservation is associated with an increased number of introns, and the presence of Aar2 and Snu66. * indicates documented trans-splicing. References: ¹Matsuzaki et al. (2004). ²Grisdale et al. (2013). ³Calculated by subtracting CDS number from annotated exons, from NCBI. ⁴Roy et al. (2012). ⁵Roy (2017). ⁶Lamontagne & Papadopoulou (1999). ⁷Araujo & Teixeira (2011). ¹⁰Vanacova et al. (2005). ⁹Neuveglise et al. (2011). ¹⁰Schonknecht et al. (2013). ¹¹Wood et al. (2002). ¹²Spieth et al. (2014); Blumenthal (2012). ¹³Adams et al. (2000); McManus et al. (2010). ¹⁴Sakharkar et al. (2004); Lei et al. (2016).

2012). Other organisms, such as *C. elegans*, have both cis-introns, and SLTS (Zorio et al., 1994). It is possible that, with the reduction of selective pressures from cis-splicing, the role of the RH insert has diverged in organisms that undergo predominantly trans-splicing.

6-6. Discussion

The crystal structure of *C. merolae* Prp8 RH domain confirmed the prediction that the 17 amino acid β -hairpin is absent. The protein maintains the RNase H fold and C-terminal helices seen in other structures of the RH domain.

An arginine is bound by an aspartate in a way that mimics the binding of a metal ion by an aspartate in the human structure. These similarities, along with the conservation of the Snu13 structure (discussed in chapter 5) suggest that *C. merolae* could provide a more minimal spliceosome for study with insights into human splicing. Genetic manipulation tools, such as those developed by Fujiwara et al. (2017) can be used to study the role of various splicing proteins within the organism, and shed light on the most essential components of the spliceosome.

The β -hairpin insertion in the Prp8 RH domain has been shown genetically and biochemically to regulate splice site proof reading and progression of the splicing cycle (reviewed in Grainger & Beggs, 2005). Structural analysis shows it at the heart of the C* complex, stabilizing the RNA core of the spliceosome (Yan et al., 2017). The lack of so highly conserved an insert in *C. merolae* and other organisms hints at its role in splicing. The concurrent absence of proteins that interact with this hairpin in these organisms suggests a network of protein interactions based on the hairpin.

Red algal genomes with fewer spliceosomal proteins also feature fewer introns (Qui et al., 2018). Additionally, hairpin deletion appears to occur in organisms with already reduced spliceosomes and few introns. Introns from these organisms also have highly conserved splice sites and BPS. While the splice site and BPS conservation of *S. cerevisiae* is high, relative to human, it is degenerate relative to *C. merolae* and *E. cuniculi*. Mutations in the β -hairpin are able to overcome defects in splicing due to mutated intron sequences, suggesting a

regulatory or proofreading role. Perhaps this function is not needed in organisms with such highly conserved splice site and BPS sequences. The conservation of the insert, but not its sequence, suggests this structure may play a similar but distinct role in trans-splicing, and that spliceosomes that predominantly splice in trans may differ from the cis-spliceosome in other ways. These data, taken together, can shed more light on the role of the hairpin both within the RH domain, Prp8, and the spliceosome.

6-7. Materials and Methods

6-7a. Identification, Cloning, and Expression of the Prp8 RH domain from *C. merolae*.

A cDNA corresponding to amino acids 1848-2065 of *C. merolae* Prp8 was cloned into the pMAL expression vector and transformed into *E. coli* Rosetta cells. The maltose binding protein (MBP) tagged RH domain was expressed in LB medium supplemented with 2g/L dextrose. The cells were pelleted by centrifugation, resuspended in amylose lysis buffer (50 mM Tris pH 8, 500 mM NaCl, 5 mM BME) and lysed by sonication. The cleared lysate was run over an amylose column and protein was eluted in the lysis buffer supplemented with 20 mM maltose. Fractions were concentrated and run over an SD75 column in GFB (20 mM Tris pH 8, 300 mM NaCl, 5 mM BME). Concentrated fractions were cleaved overnight with TEV protease to remove the MBP tag and re-purified on the SD75 column in GFB buffer.

6-7b. Crystallization

Crystals were grown at 23°C using hanging drop vapour diffusion by mixing 1 µL of 10 mg/mL protein solution (in GFB) with 1 µL precipitant (100 mM Tris pH 8.5, 100-150 mM MgCl₂, 10-12% PEG 8000). Crystals were cryo-protected in precipitant with the addition of 20% (v/v) glycerol and frozen in liquid nitrogen.

6-7c. Data Collection and Processing

Data were collected at beamline CMCF-ID of the Canadian Light Source, University of Saskatchewan, Saskatoon, Canada. Data were then processed and scaled using the HKL2000 package (Otwinowski & Minor, 1997).

6-7d. Model Building and Refinement

The structure was solved using molecular replacement with Phenix Phaser (Adams et al., 2010) using monomer A of PDB 4JK7 lacking residues 1786-1801 as a search model. Refinement in Phenix Refine alternated with manual model building in Coot (Emsley et al., 2010) to complete and refine the model. Refinement statistics are summarized in Table 6-2.

Chapter 7
Metal Binding by the Human Prp8 RH Domain

7-1. The Active Sites of the Spliceosome

Biochemical evidence suggests that the spliceosome requires individual active sites for each step of splicing. The substrates and products of the steps of splicing are structurally distinct: in the first step of splicing, the nucleophile is the 2' hydroxyl on the branch adenosine and produces a branched intron; while in the second step it is a 3' hydroxyl on a free 5' exon and results in a linear mRNA. Additionally, splicing can be stalled between the two steps, suggesting a transition. Substitution of the 2' hydroxyl at the 5' SS with a 2' O-methyl (-OCH₃) has no effect on the first step, but impairs the second step of splicing, indicating different interactions with this hydroxyl group (Moore & Sharp, 1992). The spliceosome is also stereoselective: a pro-R_p sulfur substitution of the non-bridging oxygen at both splice sites is inhibitory, while the pro-S_p sulfur substitution is tolerated for both steps (Moore & Sharp, 1993). The transesterification reactions, intron branching and exon ligation, are both S_N2 reactions that result in inversion of stereochemistry – the product of splicing with the pro-S_p sulfur substitution is the R_p stereocenter. If the second step of splicing was chemically the exact reverse of the first step then the stereochemistry would be S_p to R_p at the 5' SS in the first step, the R_p branched intron would leave the common active site, and be replaced by the R_p 3' SS. As the pro-R_p sulfur substitution at the 3' SS is inhibitory the second step is therefore not simply the reverse of the first step.

The spliceosome could accommodate the individual steps of splicing in one of two ways: two physically distinct active sites, with the free 3' end of the 5' exon

being moved into the second active site of the spliceosome and joined by the 3' SS; or, a structural rearrangement of the first step active site following branching to facilitate positioning of the 3' SS. The recent cryo-EM structures reveal that the spliceosome possesses a single active site that is rearranged between the steps of splicing (reviewed in Fica & Nagai, 2017). The branched lariat structure is moved out of the active site, and is replaced by the 3' SS. The bound metal ions switch roles: metal 1, which had activated the nucleophile for the first step, now becomes metal 2 and stabilizes the transition state.

Query and Konarska (2004) proposed an equilibrium between two conformations of the spliceosome: one would favour the first step of splicing, and the other would favour the second. A balance between the two conformations leads to productive splicing (Figure 7-1). A spliceosome more inclined towards the first step generates a lariat-exon intermediate, but little spliced mRNA. A second step spliceosome has difficulty with the first step, but any intermediate free 5' exon generated quickly undergoes the second step to produce ligated mRNA. Different proteins within the spliceosome assist in the transition between these two spliceosomal conformations, such as the helicases Prp16 and Prp22.

Prp16 is an RNA helicase that acts after the first step of splicing to rearrange the spliceosome for the second step (Burgess et al., 1990). It has been proposed that mutating Prp16 to slow its activity overcomes defects in the pre-mRNA that inhibit splicing, such as mutation of the branch site adenosine to cytidine (BSC), by increasing the amount of time available for the first step of splicing to occur prior

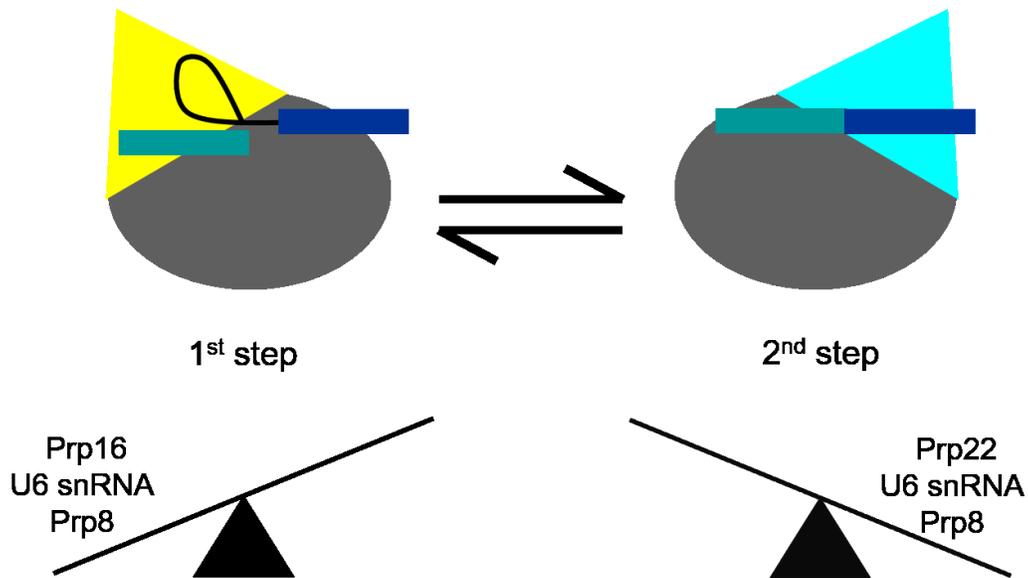


Figure 7-1. Equilibrium between the two steps of splicing. The spliceosome is in equilibrium between first and second step conformations (top). Mutations in spliceosome components can shift the equilibrium towards the first step (bottom right), or second step (bottom left).

to triggering spliceosome disassembly due to an incorrect substrate (Burgess & Guthrie, 1993). Additionally, mutations in Prp16 have been used to stall the spliceosome between steps in order to elucidate details of spliceosome assembly and activation (Schwer & Guthrie, 1991; Fica et al., 2014).

Prp22 is the second step equivalent of Prp16 (Mayas et al., 2006). It releases the ligated exons from the spliceosome and allows for recycling of the snRNPs in the next round of splicing. Mutations in Prp22 that slow its action are able to overcome defects that inhibit the second step of splicing, such as a branch site guanosine (BSG), likely by delaying spliceosome disassembly and allowing more time for the less optimal substrate to splice (Schneider et al., 2004). Like Prp16, Prp22 mutations have also been used to study the splicing cycle (Bai et al., 2017; Liu et al., 2017).

Alleles of other spliceosomal components, such as the U6 snRNA, are able to overcome defects in both steps of splicing. U6 snRNA G71A and G71U allow splicing of substrates with a non-adenosine branch (BSG and BSC; McPheeters, 1996).

In the tri-snRNP and early stages of spliceosome assembly U6 snRNA U57 is base-paired with U4 snRNA G62. Upon dissociation of the U4 snRNP, U6 U57 base-pairs with U2 snRNA A27, and remains base-paired for the duration of the splicing cycle. However, depending on the mutation, base substitutions of the U6 snRNA U57 have opposite effects: U57C is able to overcome a block of the first step of splicing, such as the one presented by BSC, but exacerbates blocks to the second step, as in BSG; U57A is able to splice the BSG substrate and bypass the second step block, but use of BSC in the first step worsens the splicing defect (McPheeters, 1996). U6 snRNA U57C is thus called a first step allele, in that it can overcome defects of the first step of splicing at the expense of the second step. U6 snRNA U57A is a second step allele, with the ability to overcome blocks at the second step, but causes an increased defect at the first step. Improving splicing of one step at the expense of the other is further evidence of the equilibrium between the two steps of splicing.

7-2. Prp8 and the Spliceosomal Equilibrium

Another protein that can overcome defects in the two steps of splicing is Prp8. A number of alleles throughout the protein are able to suppress defects in the

5' and 3' splice sites as well as in the branch site, with many alleles that suppress BSC and BSG clustering on the RH β -hairpin. In *S. cerevisiae* (y) yV1860D, yT1865K, yA1871E, and yT1872E, suppress defects in the first step of splicing caused by BSC (first step alleles), while yT1861P, yV1862Y, yH1863E, yN1869D, and yV1870N suppress defects in the second step caused by BSG (second step alleles; reviewed in Grainger & Beggs, 2005; Figure 7-2A).

The human Prp8 RH domain crystallized in two conformations: monomer A and monomer B (Figure 7-2B). In monomer A, the 17 amino acid insertion in the RH domain forms a β -hairpin. In monomer B, this hairpin is disrupted into loop, and a metal ion is observed bound in the traditional Mg^{2+} binding site of the RNase H domain (Ritchie et al., 2008). First step alleles stabilize monomer A and the β -hairpin, while second step alleles stabilize the loop in monomer B, or destabilize the β -hairpin in monomer A (Figure 7-2C; Schellenberg et al., 2013).

The canonical RNase H metal ion binding site binds a metal ion only in monomer B of the Prp8 RH domain. The transition from monomer A to monomer B includes the disruption of the β -hairpin, and its translocation back by $\sim 45^\circ$, the extension of strand $\beta 1$ and the movement of the loop connecting helices $\alpha 1$ and $\alpha 2$. These rearrangements move the metal coordinating (human; h) hAsp1781 and hAsp1782 closer to each other, and displace hThr1783 out of the metal binding site. hArg1865, which also partially blocks the metal binding site, is flipped up to hydrogen bond with the backbone of hLeu1836. The hR1865A mutant increases the occupancy of the metal bound in monomer B, but does not allow for metal

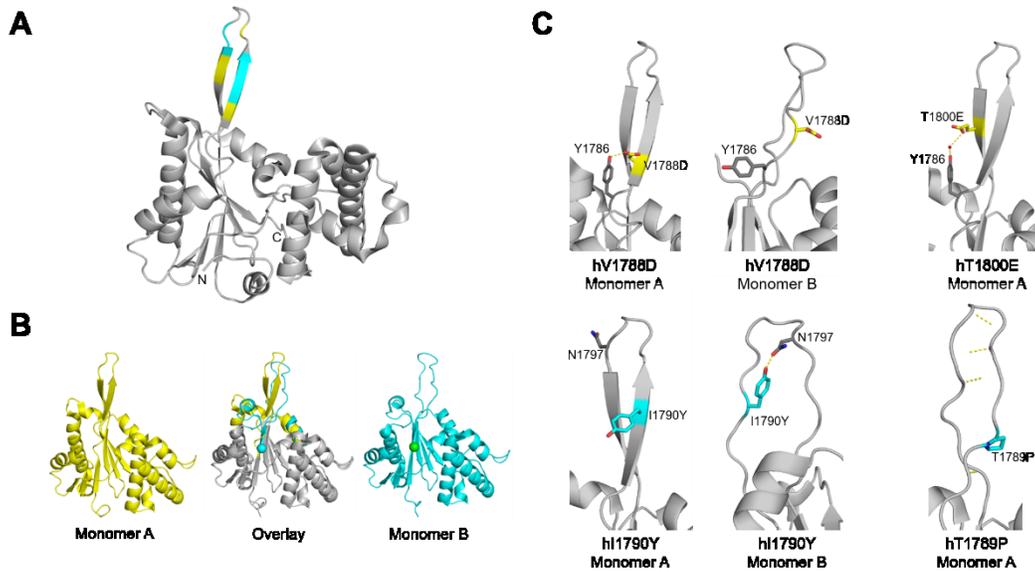


Figure 7-2. Prp8 alleles in the β -hairpin stabilize one of two monomers. A. A number of mutations of Prp8 that influence splicing map to the β -hairpin. B. Prp8 crystallized with two conformations (monomers A and B) in the asymmetric unit (PDB 4JK7). C. Mutations that stabilize the β -hairpin (monomer A; top) are able to overcome defects in the first step of splicing, while mutations that stabilize the loop form of the hairpin (monomer B; bottom) or disrupt the hairpin in monomer A are able to overcome defects in the second step of splicing. hV1788D forms an additional hydrogen bond with hY1786 in monomer A, but not in monomer B. hT1800E forms a water-mediated hydrogen bond with hY1786 in monomer A. The loop in monomer B is not visible. hI1790Y forms a hydrogen bond with hN1797 in monomer B, but not in monomer A. hT1789P disrupts the β -hairpin of monomer A. The loop in monomer B is not visible.

binding to monomer A. Mutation of the metal binding yAsp1853 (hAsp1781) to cysteine reduced second step splicing and can be rescued by co-mutation with the second step allele yV1862Y, suggesting the bound metal stabilizes monomer B during the second step of splicing.

Results

7-3. Structures of hT1783A/S hR1865A Double Mutants

hThr1783 is highly conserved in organisms that possess the 17 amino acid insert in the RH domain. We hypothesized it plays a regulatory role in splicing by blocking the metal binding site in monomer A and relocating to allow metal binding in monomer B.

In monomer A, hThr1783 is held in place by hydrogen bonding to hAsp1781; in monomer B it hydrogen bonds to the backbone carbonyl of hAsn1784 (Figure 7-3A). To investigate the role of the bound metal ion in monomer B we mutated the highly conserved hThr1783 (yThr1855) to serine or alanine. We then crystallized and solved the structure for the human RH domain with the hT1783A/S hR1865A double mutants. Both structures have a resolution of 1.6 Å.

The Prp8 RH structure with the hT1783S hR1865A double mutation (Table 7-1) is very similar to the wild type structure: monomer A aligns with an RMSD of 0.36 Å, and monomer B with an RMSD of 0.4 Å (Figure 7-3B). In monomer A, the sidechain of the serine is rotated such that the hydroxyl is pointing in a different direction than the hydroxyl on the threonine. However, this change does not interrupt the hydrogen bonding patterns of the hydroxyl. hSer1783 hydrogen bonds to the sidechain of hAsp1781. In monomer B the hydroxyl of the serine is oriented similarly to that of the threonine. However, the serine hydroxyl is shifted up from the threonine by 1 Å due to slight changes in the backbone. The shift does not interfere with hydrogen bonding, and the serine interacts with the backbone

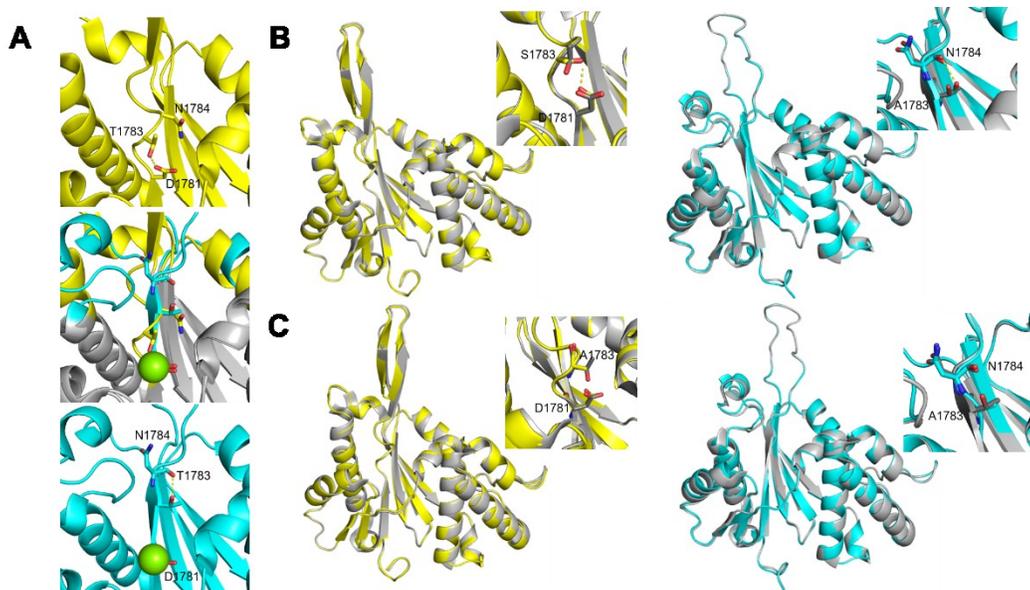


Figure 7-3. The hT1783S/A hR1865A double mutants of the Prp8 RH domain. A. In monomer A hThr1783 hydrogen bonds to hAsn1781 (yellow; top); upon the transition to monomer B (cyan; bottom) it hydrogen bonds to hAsn1784. Alignment of monomer A and monomer B shows hThr1783 blocks the metal binding site in monomer A, and its repositioning in monomer B allows a metal to bind (middle). B. Structure of double mutant hT1783S hR1865A Prp8. As with the hT1983A mutant, both monomers align closely to wild type (monomer A yellow, left; monomer B cyan, right; wild type Prp8 grey). hT1783S forms the same hydrogen bonds in both monomers as hThr1783. C. Structure of hT1783A hR1865A Prp8. Both monomers align very closely to wild type (monomer A yellow, left; monomer B cyan, right; wild type Prp8 grey). The mutation to alanine removes the hydrogen bonds to hAsp1781 and hAsn1784.

carbonyl of hAsn1784. A metal is observed in monomer B of this structure bound to the RNase H metal binding site. There do not appear to be any changes in the structure that would cause a first step or second step phenotype for this mutant.

The hT1783A hR1865A double mutant structures (Table 7-1) also closely align to those of wild type: RMSD of 0.6 Å for monomer A; and RMSD of 0.8 Å for monomer B (Figure 7-3C). In alignment of monomer A the alanine is positioned directly atop the wild type threonine. The lack of hydroxyl means hT1783A does not hydrogen bond with hAsp1781. In monomer B the backbone of the mutant is

not as extended as that in the wild type, but maintains the hydrogen bonding pattern of the β -sheet with hAla1806 in strand β 4. Monomer B features a bound metal.

Prediction of the phenotype is more complex with the yT1855A mutant than the yT1855S: both monomers lose a hydrogen bond. If the loss of the hydrogen bond causes a slight destabilization of monomer B relative to monomer A, we could observe a change in splicing with the BSG substrate. The first step could proceed normally, but a less stable second step conformation may be unable to overcome the defect, leading hT1783A to appear as a first step allele with this substrate. The effects of yT1855A on the BSC substrate are harder to predict. yT1855A might destabilize both monomers, making the transition between them easier. In this case, BSC might show worse splicing than wild type, as neither conformation is stable enough to overcome the defects caused by BSC. It is also possible that the lost hydrogen bonds do not play a significant role in the equilibrium between the two states, leading to a wild type phenotype.

7-4. Phenotypes of yT1855A/S

We also produced *S. cerevisiae* strains with the wild type endogenous Prp8 replaced by yT1855S or yT1855A (hThr1783) Prp8 on a plasmid. The CUP1 gene, a non-essential gene that allows for growth on media containing copper, has also been deleted. These strains contain an ACT1-CUP1 reporter plasmid, with the CUP1 gene split by the actin (ACT) intron. The branch point sequence of the intron may be wild type (BSA), guanosine (BSG) or cytidine (BSC). Growth on

increasing concentrations of copper indicates a higher level of splicing to remove the ACT intron and produce functional CUP1 protein (Lesser & Guthrie, 1993). BSG causes a defect in the second step of splicing. Second step alleles are able to overcome this defect, splice this intron more efficiently, and will grow at higher concentrations of copper than cells with a wild type Prp8. First step alleles exacerbate this second step defect and are more sensitive to copper concentration (Liu et al., 2007c). Wild type intron (BSA) is spliced highly efficiently by all alleles. yE1960K, a strong first step allele, was used as a control strain (Umen & Guthrie, 1995b).

yT1855S and yT1855A both have a wild type phenotype. They grow at the same concentrations of copper as the strain with wild type Prp8 (Figure 7-4A). This result was unexpected, and suggests that either yThr1855 does not regulate the transition between the steps of splicing, or that the both monomers are equally destabilized and the equilibrium between them is maintained.

Primer extension is another way to track splicing in yeast cells. After the total RNA is extracted, a radio-labeled reverse primer binds to the 3' exon and reverse transcriptase synthesizes a complementary DNA fragment that corresponds to the full length pre-mRNA (RNA not spliced) or the ligated exons (RNA has been fully spliced). Because the reverse transcriptase is blocked by the branch of the lariat intron, a third species corresponding to the 3' exon and distance between the 3' SS and BPS is created (RNA has undergone the first step of splicing, but not the second). The reaction is then run on a gel, quantified, and the efficiency of the two

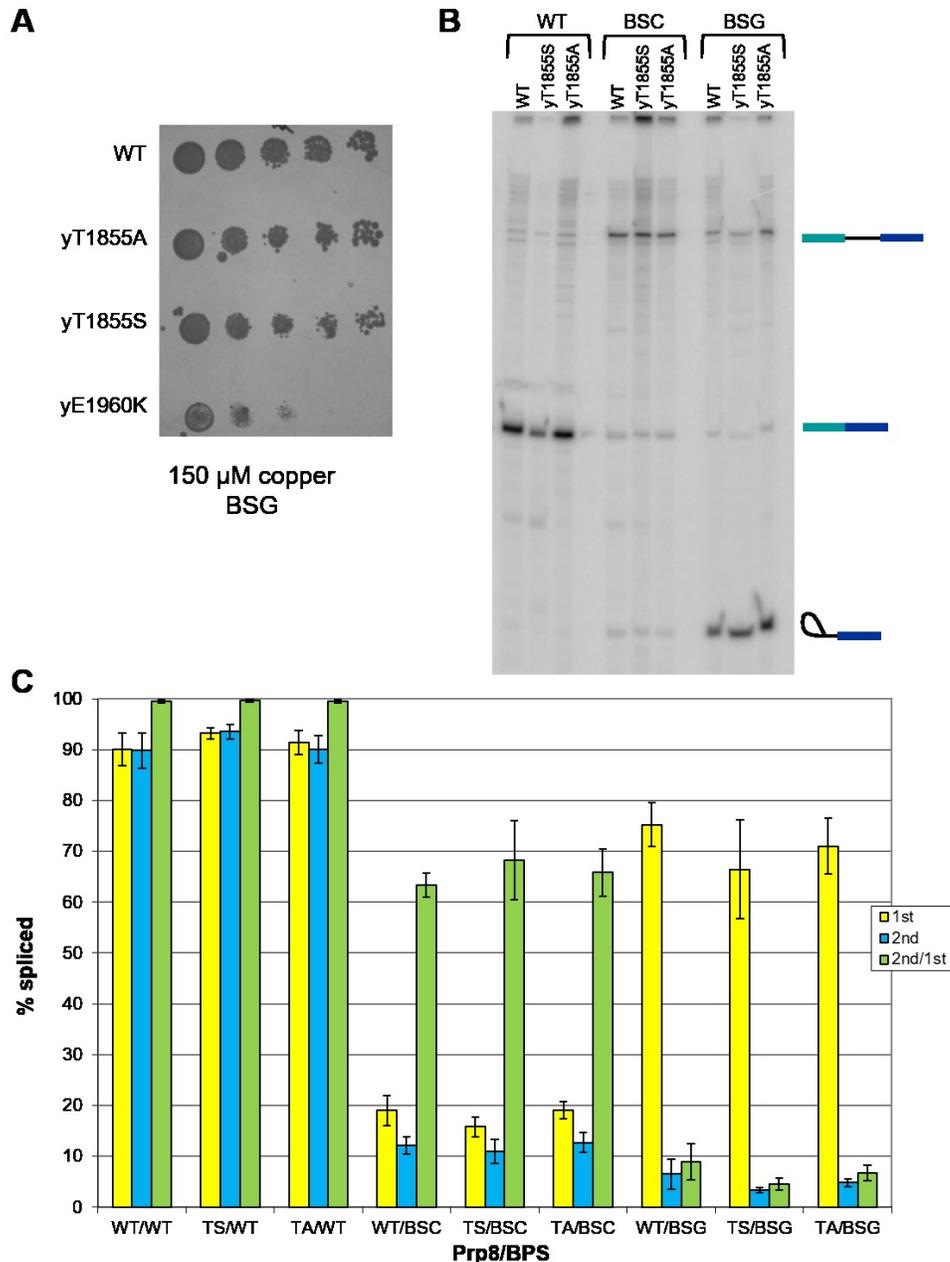


Figure 7-4. Phenotype of the yT1855A and yT1855S mutants is wild type. A. Spot tests of *S. cerevisiae* expressing only wild type or mutant Prp8 on 150 μ M copper with the ACT1-CUP1 construct featuring a branch site G (BSG) in its intron. Increased growth on copper is associated with an increase in the removal of the ACT1 intron from the CUP1 gene. Growth of yT1855A and yT1855S Prp8 is equivalent to wild type. yE1960K, a strong first step allele, shows worse growth and serves as a control. B. Primer extension of the ACT1-CUP1 RNA with wild type (WT), branch site C (BSC) or branch site G (BSG) introns grown in wild type, yT1855S or yT1855A cells. Products of each step of splicing are visualized. C. Graph of primer extension showing percent splicing levels of total RNA for the first step (yellow) and second step (blue). Second step efficiency (green) is calculated by dividing the amount second step product by the amount of RNA that underwent the first step.

steps of splicing can be compared. A wild type intron will be mostly spliced and show a strong band for ligated exons with weaker bands for unspliced pre-mRNA and the lariat intron. Introns blocked at the second step of splicing will show an accumulation of lariat intron relative to spliced mRNA. Introns blocked at the first step of splicing will show very little lariat intron, as the few pieces of RNA that manage the first step of splicing quickly undergo the second step. Primer extension of a first step allele with a BSC intron shows an accumulation of lariat intron and a decrease in the amount of full length mRNA. BSC with a second step allele leads to low levels of splicing, as the first step is inefficient, but with a higher ratio of ligated exons to branched intron. The second step allele is able to stabilize the second step of splicing for each intron that is able to undergo the first step. A first step allele splicing a BSG intron will show an accumulation of lariat intron, while a second step allele will show an increase in the amount of ligated exons (Query & Konarska, 2004).

Primer extension of RNA extracted from cells containing wild type or mutant Prp8 and wild type, BSC or BSG ACT1-CUP1 does not show clear first or second step phenotypes for the yT1855A and yT1855S mutants (Figure 7-4B). The Prp8 mutants have no effect on the wild type intron. Splicing looks equivalent to wild type Prp8 levels: both steps of splicing are very effective, and almost all of the introns that undergo the first step of splicing also undergo the second step of splicing (Figure 7-4C). This result is expected, as the Prp8 mutations do not affect splicing of an intron with a branch adenosine.

yT1855S and yT1855A both appear to have a wild type phenotype with the BSC intron (Figure 7-4B). They may have a weak first step phenotype with the BSG intron, as seen by the slight decrease in the amount of second step product relative to wild type (Figure 7-4C). However, this is not supported by the splicing seen with the BSC intron. When compared with the primer extension results from known strong first and second step alleles, the effects of the yT1855S and yT1855A mutations are very mild (Schellenberg et al., 2013). Our primer extension data are consistent with the results of the copper growth tests: yT1855S and yT1855A have a wild type phenotype, and yThr1855 is unlikely to play a regulatory role within the *S. cerevisiae* spliceosome.

7-5. Structures of hQ1894E/S hR1865A Double Mutants

We also crystallized the hQ1894E and hQ1894S hR1865R double mutants to a resolution of 1.48 and 2.1 Å respectively (Table 7-1). hGln1894 coordinates the metal ion in monomer B via a water molecule, and is ySer1966 in *S. cerevisiae* (Figure 7-5A).

The hQ1894E/hR1865A Prp8 RH structure is very similar to that of wild type: monomer A aligns with an RMSD of 0.6 Å, and monomer B with an RMSD of 0.8 Å. hGlu1894 occupies the same position in both monomers as hGln1894 (Figure 7-5B). The rest of the metal binding side chains are also in the same position as their wild type counterparts in both monomer A and monomer B. The metal ion is present in monomer B, in an equivalent position to the wild type (Figure

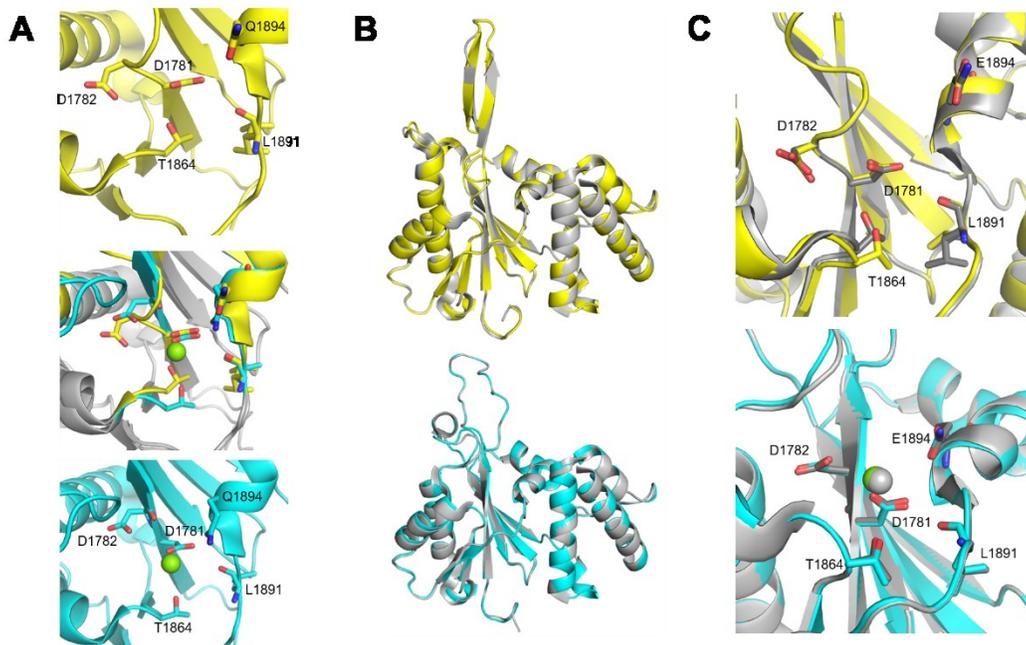


Figure 7-5. The hQ1894E hR1865A double mutant of the RH domain of Prp8. A. Metal binding site of the Prp8 RH domain. There is no metal bound in monomer A (top). Rearrangement to monomer B allows a metal to bind (bottom). Overlay of monomer A (yellow) with monomer B (cyan) shows the difference in the metal binding site between monomers (middle). B. Overall structure of hQ1894E hR1865A Prp8 RH domain. Both monomer A (yellow; top) and monomer B (cyan; bottom) align closely to the wild type RH domain (grey; PDB 4JK7). C. There is no change to the metal binding site upon hQ1894E mutation. There is no metal bound in monomer A (yellow; top), but there is a metal in monomer B (cyan; bottom).

7-5C). The metal occupancy in hQ1894E is 64%, higher than the 46% occupancy of wild type, possibly due to the negative charge of glutamate substituted for the polar side chain of glutamine.

hQ1894S/hR1865A also aligns closely to the wild type: monomer A with an RMSD of 0.7 Å and monomer B at 0.8 Å (Figure 7-6A). The metal binding sites of the mutants also align closely with those of the wild type (Figure 7-6B). The serine is not long enough to hydrogen bond to the water that coordinates the metal ion, and there is no clear metal ion bound to monomer B in this structure

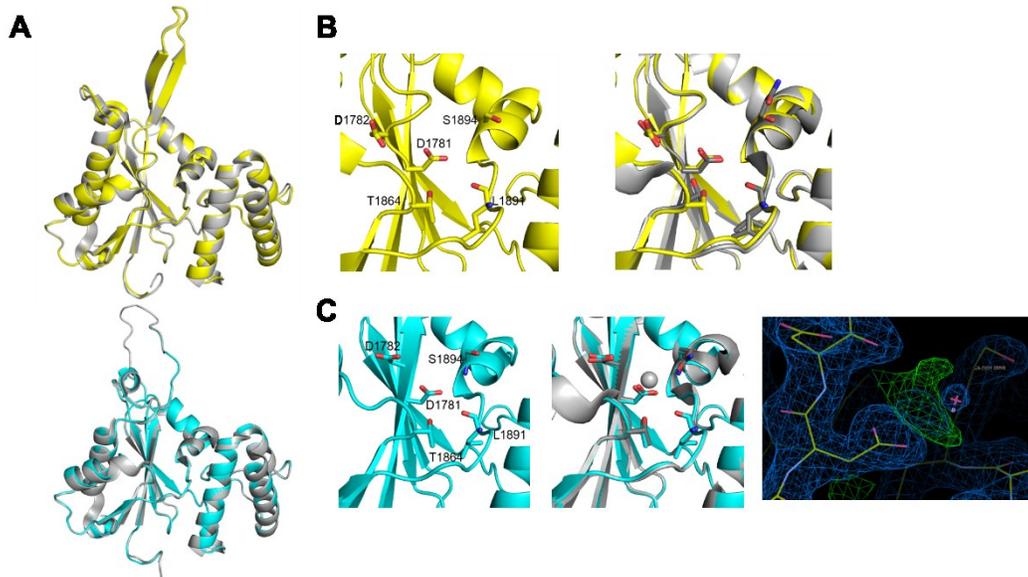


Figure 7-6. The hQ1894S hR1865A double mutant of the Prp8 RH domain. A. Structure of the hQ1894S hR1865A Prp8 RH domain. Both monomers align very closely to wild type (monomer A top, yellow; monomer B bottom, cyan; wild type grey; PDB 4JK7). B. hQ1894S mutation does not affect the metal binding site in monomer A. Metal binding site of hQ1894S shown with (left) and without (right) the wild type RH domain. C. hQ1894S does not change the arrangement of the metal binding site in monomer B (left). However, there is no clear metal bound in the mutant (wild type RH and bound metal in grey; middle). There is density for something in the metal binding site, but it is not clearly a bound metal (right).

(Figure 7-6C). The blob of electron density that might be a bound metal is not sufficiently defined to confidently place a metal in the map.

We were not able to create the yS1966E or yS1966Q plasmid to generate mutant strains.

7-6. Discussion

Crystal structures of the *S. cerevisiae* Prp8 RH domain show a single monomer corresponding to monomer A, even in crystals with multiple copies of the RH domain in the asymmetric unit (Pena et al., 2008; Yang et al., 2008).

Data collection	hT1783A	hT1783S	hQ1894E	hQ1894S
	hR1865A	hR1865A	hR1865A	hR1865A
Space group	P 21 21 21			
Cell dimensions				
<i>a</i>	77.935	76.146	76.1	76.189
<i>b</i>	75.943	77.743	77.748	78.058
<i>c</i> (Å)	93.315	93.126	93.09	93.186
α, β, γ	90 90 90	90 90 90	90 90 90	90 90 90
Wavelength	0.97949	0.97949	0.97949	0.9749
Resolution (Å)	1.60	1.60	1.48	2.10
R _{meas}	0.092 (0.631)	0.103 (0.684)	0.104 (0.920)	0.160 (0.622)
I / σ I	8.7	11.7	11.7	8.0
Completeness (%)	99.89 (99.99)	99.83 (99.99)	99.81 (98.34)	99.18 (100.00)
Redundancy	7.2 (7.3)	7.2 (7.2)	7.4 (7.4)	6.8 (6.8)
Refinement				
Resolution (Å)	40.03-1.60 (1.66-1.60)	39.72-1.60 (1.66-1.60)	46.96-1.48 (1.53-1.48)	40.01-2.10 (2.18-2.10)
Reflections	2,246,726	1,987,375	1,323,866	2,543,382
<i>R</i> _{work}	0.1728 (0.2268)	0.1751 (0.2329)	0.1569 (0.2018)	0.2079 (0.2273)
<i>R</i> _{free}	0.1920 (0.2556)	0.1925 (0.2585)	0.1807 (0.2893)	0.2511 (0.2953)
No. atoms	4,024	4,087	4,137	3,728
Protein	3,564	3,566	3,567	3,483
Ligand	6	8	2	0
Solvent	454	513	568	245
<i>B</i> factor	34.59	31.81	29.29	40.36
Protein	33.59	30.65	27.85	40.12
Ligand	36.01	40.47	30.62	-
Solvent	42.49	39.76	38.32	43.71
R.M.S. deviations				
Bond lengths (Å)	0.007	0.007	0.014	0.007
Bond angles (°)	1.08	1.09	1.45	1.19
Ramachandran favored (%)	97.93	98.39	97.01	96.93
Ramachandran allowed (%)	1.84	1.61	2.53	2.83
Ramachandran outliers (%)	0.23	0.00	0.46	0.24
Rotamer outliers (%)	0.00	1.01	0.25	2.57
Clashscore	2.20	2.48	4.26	3.52

Table 7-1. Data collection and refinement statistics for four human Prp8 RH domain mutants. Data were collected from single crystals. Values in parentheses are for the highest resolution shell. Molecular replacement was performed with PDB 4JK7 as the search model.

Furthermore, monomer B is not observed in any of the cryo-EM spliceosomal structures. This begs the question: does the RH domain adopt the conformation of monomer B at any point during the splicing cycle? Alternatively, is monomer B simply a crystallographic artefact? The correlation between stabilization of monomer B or destabilization of monomer A with a second step phenotype suggests that monomer B does play a role in splicing. An overlay of the available RH domains shows the RH domain is very rigid, with the exception of the flexible β -hairpin, suggesting the possibility of conformational change during the splicing cycle (Figure 7-7). Perhaps the RH domain and the β -hairpin adopt monomer B transiently during the splicing cycle on a time-scale that prevents its observation by cryo-EM.

In *C. merolae* the residue following the two aspartates involved in metal binding is cmArg1864, rather than hThr1783 or yThr1855 in humans and *S. cerevisiae* respectively (Figure 7-8). In the human Prp8 RH monomer A, this threonine blocks the metal binding site and is shifted away from this site in monomer B. In the *C. merolae* structure, which lacks the β -hairpin insert, the arginine remains hydrogen bonded to cmAsp1862 in the metal binding site. The equivalent aspartate in humans, hAsp1781, makes inner sphere contacts with the magnesium in monomer B, suggesting that the positive charge of cmArg1864 might be the functional equivalent of the magnesium ion (Figure 7-8).

S. cerevisiae lacks one of the metal binding side chains: hGln1894 is ySer1966. To test metal binding in the absence of a full set of ligands we created and crystallized the human hQ1894S mutant. We did not observe a coordinated

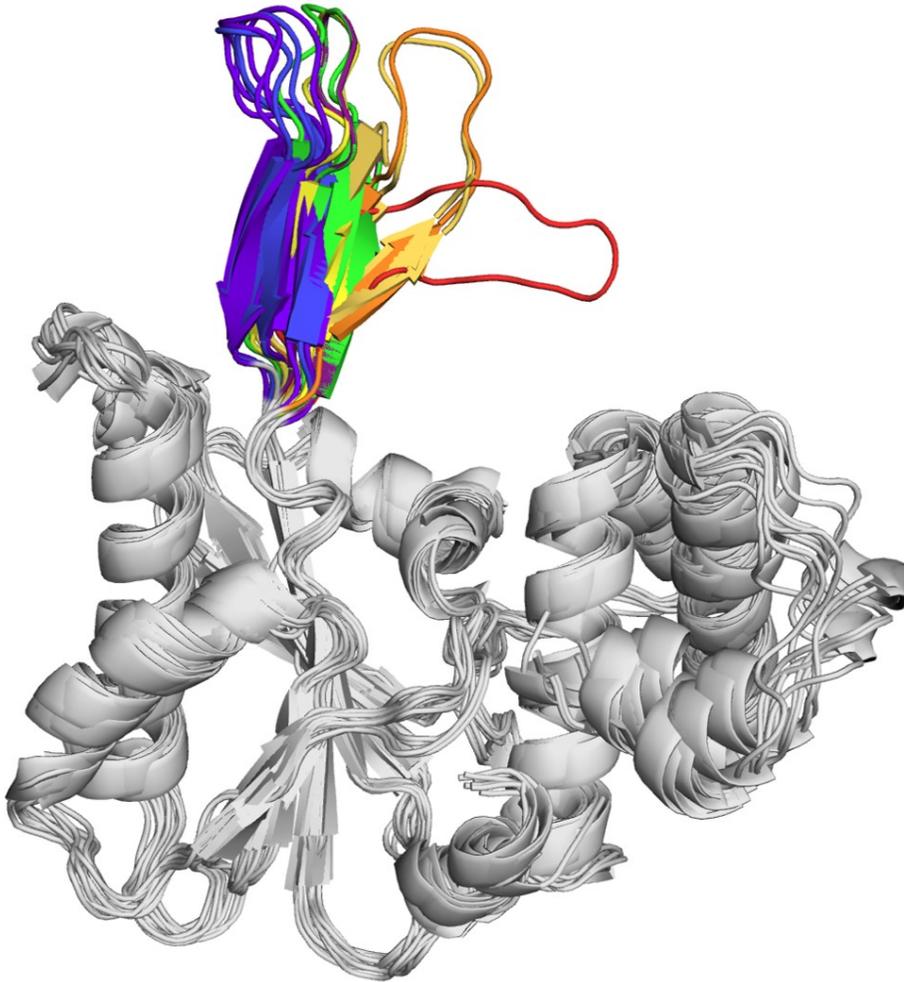


Figure 7-7. β -hairpin flexibility. Alignment of the Prp8 RH domain from the *S. cerevisiae* spliceosomes reveals flexibility of the β -hairpin. Throughout the splicing cycle the rest of the Prp8 RH domain is rigid. The β -hairpin is coloured by its progression through the splicing cycle: tri-snRNP (red; PDB 5GAN), pre-B (orange; PDB 5ZWM), B (yellow orange; PDB 5ZWO, 5NRL), B^{act} (yellow; PDB 5GM6), C (green; PDB 5GMK, 5LJ3), C* (blue; PDB 5MQ0, 5WSG), P (indigo; PDB 6EXN, 6BK8, 5YLZ), and ILS (purple; PDB 5Y88).

metal ion in that structure, suggesting the need for a full complement of ligands to observe metal binding. This suggests that the RH domain of *S. cerevisiae* Prp8 may not have a metal ion bound, or that the metal ion observed in the human monomer

<i>C. merolae</i>	RLWIV DDR FA	1866	LPRGTVLAQ T ASMDPLK T LLAGTEYAK I PVRAGAAAMP I Q A LMALP	1970
<i>E. cuniculi</i>	--LIV D V K AL	1731	LNKRSTIVVPE S MI D AMEN F --I I DHPS I SV R TCL S AP P S N VM G I--	1820
<i>O. colligata</i>	--M I L D S R CL	1599	LNRRSVSV T DD M DAVEN V --I V D P SV V RS L CS G I P FS N VM M L--	1688
<i>N. bombycis</i>	--LIV EDR LL	1547	-----I E V L GG Y --Q V D Y PK I N L K I LE N E I PK Y K I GG Y V I	1599
<i>V. culicis</i>	-----	1433	-----M D R S E G V K -----S K C R K I F F P A L R T L L	1501
<i>G. intestinalis</i>	-V I I V DD S A	1700	R P T V V V C S R Q S I D P I R T L L--S V L N P P I V V R G T S L H W Q I Q N I R S M V I	1817
<i>T. cruzi</i>	RTWIV DD S AT	1841	R P G K I I V TR S R F R Q T L H N M L--V L D Y P N I I G Q S D L N L S I P M V L R H S	1957
<i>I. donovani</i>	RTWIV DD S AT	1860	R P S K I I V TR S R F R Q T L H N M L--V L D Y P N I I G Q S D L N L S I P M V L R H S	1976
<i>N. gruberi</i>	TIW FV DD S IV	1522	Q P K Q I I V TR K G L D P L E--V H --L L D F P N I V I K G S E L Q L P I Q S I L K I D	1637
<i>T. vaginalis</i>	TIW FV EDK IV	1775	Q P K Q V I C T S E L L D P V Q --S Y --L S E F P N T V V K G S D M D L P I P A F M K I P	1890
<i>G. theta</i>	IIW FV DD T IV	1763	Q P K Q I I V TR K G M L D P L E --V H --L L D F P N I V I K G S D L Q L P I Q A S L K I E	1878
<i>E. histolytica</i>	IIW FV DH S NI	1728	Q P K Q I I V SL R G M L D P L E --T H --M I D F P N I L L K G S E I A L P F G E I M Q M E	1843
<i>D. discoideum</i>	IIW FV DD S IV	1773	Q P K Q I I V TR K G M M D P L E --V H --L L D F P N I V I Q G S E L Q L P I Q A C L K V E	1888
<i>C. crispus</i>	VIW FV DD T IV	1867	Q P K Q L I V IR K G M L D P L E --T H --C L D F P N V I I K G S D L Q L P I Q T L L K I E	1982
<i>G. sulphuraria</i>	IIW FV DD T IV	1810	Q P K Q I I V TR K G M L D P L E --V H --L L D F P N I V I K G S D L Q L P I Q A L K I E	1925
<i>A. thaliana</i>	IIW FV DD T IV	1809	Q P K Q I I V TR K G M L D P L E --V H --L L D F P N I V I K G S E L Q L P I Q A C L K I E	1924
<i>Z. mays</i>	IIW FV DD T IV	1816	Q P K Q I I V TR K G M L D P L E --V H --L L D F P N I V I K G S E L Q L P I Q A C L K I E	1931
<i>C. reinhardtii</i>	TVW FV DD T IV	1797	Q P K R I I V TR K G M L D P L E --V H --L L D F P N I V I T G S E L Q L P I Q A A I K L E	1912
<i>T. gondii</i>	TIW FV DD T IV	1987	Q P K Q L I V TR K G M L D P L E --V H --L L D F P N I V I K G S E L N L P I Q A I M K V E	2102
<i>P. falciparum</i>	TIW FV DD T IV	2482	Q P K Q I I V TR K G M L D P L E --V H --L L D F P N I I I K G T E L N L P I Q A L L K L N	2597
<i>P. tetraurelia</i>	IIW FV DD T IV	1766	Q P R Q I I V TR K G L L D P L E --V H --L L D F P N I V I K G S E L S L P I Q A I L K I E	1871
<i>T. thermophila</i>	TIW FV DD T IV	1802	Q P R Q I I V TR K G M L D P M E --V H --L L D F P N I V I K G S E M Q L P I Q S C L K V E	1917
<i>P. tricornutum</i>	VIW FV DD T IV	1786	Q P K Q I I V TR K G M L D P L E --V H --C L D F P N I V I K G S E L Q L P I Q A A L K V E	1901
<i>P. infestans</i>	IIW FV DD T IV	1767	Q P K Q I I V TR K G M L D P L E --V H --L L D F P N I V I K G S E L Q L P I Q A C L K V E	1882
<i>C. elegans</i>	IIW FV DD T IV	1778	Q P R Q I I V TR K A M L D P L E --V H --L L D F P N I V I K G S E L M L P I Q I M K V E	1893
<i>D. melanogaster</i>	IIW FV DD T IV	1845	Q P K Q I I V TR K G M L D P L E --V H --L L D F P N I V I K G S E L Q L P I Q A C L K V E	1960
<i>H. sapiens</i>	IIW FV DD T IV	1785	Q P K Q I I V TR K D M L D P L E --V H --L L D F P N I V I K G S E L Q L P I Q A C L K V E	1900
<i>M. daphniae</i>	TIW FV DD S AV	1835	Q P G T V I V TR K G M L D P L E --V H --L L D F P N I V I K G S E L A I P F A A L L K I E A	1951
<i>R. allomyces</i>	VIW FV DD T IV	1799	Q P K Q I I V TR K G M L D P L E --V H --L L D F P N I I I K G S E L Q L P I Q A C L K I E	1914
<i>B. dendrobatidis</i>	IIW FV DD T IV	1819	Q P K Q I I V TR K G M L D P L E --V H --C L D F P N I V I K G S E L Q L P I Q A C I K I E	1934
<i>A. macrogynus</i>	TIW FV DD T IV	1815	Q P N Q I I V SR K G M L D P L E --V H --L V S F P N C L I R G A E L Q L P I Q A V M K I E	1930
<i>R. delemar</i>	IIW FV DD T IV	1802	Q P K Q I I V TR K G M L D P L E --V H --L L D F P N I V I K G S E L Q L P I Q A C L K V E	1917
<i>A. muscaria</i>	IIW FV DD T IV	1785	Q P K Q V I V TR K G M L D P L E --V H --L L D F P N I V I K G S E L Q L P I Q A C M K M E	1900
<i>C. albicans</i>	T Q L F V DD T IV	1862	Q P K Q L I V SR R G M M D P L E --V H --M L D F P N I S I R P S E L M L P F A A V M K I D	1977
<i>S. pombe</i>	T Q L F V DD T IV	1809	Q P R Q I I V TR K G M L D P L E --V H --L L D F P N I T I K G S E L Q L P I Q A I I K L D	1924
<i>S. cerevisiae</i>	I K L F V DD T IV	1857	Q P K Q I I V TR K A M L D P L E --V H --M L D F P N I A I R P E L R L P S A A M S I D	1972

Figure 7-8. Conservation of the metal binding site and hR1865. The residues involved in metal binding (hAsp1781, hAsp1782, hThr1864, hGln1894; green) are highly conserved. hArg1865 (blue) is also highly conserved. Absence of arginine at the equivalent of hArg1865 is coupled with an arginine or lysine at the equivalent of cmArg1864, and may be playing a similar role to hArg1865. hThr1873, mutated in this study, is shown in red.

B is an opportunistic crystallographic artefact rather than biologically relevant.

There is a rotamer for the side chain of hArg1865 in monomer B that occupies the metal binding site and hydrogen bonds with hAsp1781. However, this requires disrupting the hydrogen bonds with hAsp1782. If monomer A, with hArg1865

hydrogen bonded to hAsp1872, is the predominant form of the RH domain in solution it is possible that it is easier for a metal to bind upon formation of monomer B than for hArg1865 to rearrange its hydrogen bonds.

If the RNase H metal binding site of the RH domain is occupied by an arginine instead of a metal the lack of coordination by the side chain of ySer1966 would be irrelevant. Similarly, the *C. merolae* cmAla1932 (equivalent of hThr1864, another metal ligand) is accommodated. Of the 36 taxa surveyed previously (see chapter 6), two (*G. intestinalis* and *E. histolytica*) lack both the positive residue following the two aspartates, and an arginine at the equivalent to hArg1865. *C. merolae*, *O. colligate*, and *N. bombycis* lack a conserved arginine at the equivalent of hArg1865, but have the sequence DDR (cm1862-cm1864), DSR and EDR respectively. *E. cuniculi* and *T. vaginalis* have a lysine, in DVK and EDK respectively. In the *C. merolae* structure cmArg1864 hydrogen bonds to cmAsp1862, bypassing cmAsp1863. If these other organisms are similar, the positive charge could interact with the first aspartate, rendering the identity of the intervening amino acid irrelevant. The other 29 taxa have an arginine at the equivalent of hArg1865. Such high conservation of an arginine available to hydrogen bond to an aspartate suggests it might be relevant for splicing.

Recently, a second interpretation of the two RH conformations based on similarities with the group II intron was published (Mayerle et al., 2017). The group II intron adopts an intermediate, transitional structure between the catalytic structures of the first and second steps (Marcia & Pyle, 2012). Mayerle et al. (2017)

propose that in the spliceosome, transitional alleles of Prp8 stabilize monomer A and are associated with low efficiency high fidelity splicing. Catalytic alleles of

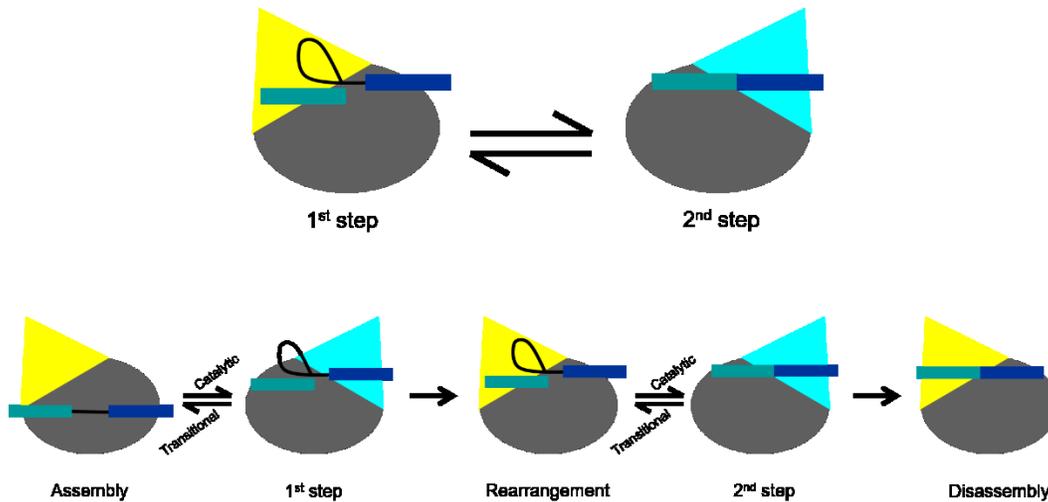


Figure 7-9. Model of the catalytic and transitional alleles of Prp8 the RH domain. Top, model of the two monomers of the Prp8 RH domain influencing the equilibrium between the first and second steps. Bottom, new evidence suggests that the Prp8 RH domain toggles between a transitional, high fidelity, low efficiency state (associated with monomer A) and a catalytic, low fidelity, high efficiency state (monomer B) during the splicing cycle, as proposed by Mayerle et al. (2017). The transitional form dominates during spliceosomal assembly and rearrangement, then adopts the catalytic conformation during the two chemical steps of splicing.

Prp8 stabilize monomer B and are associated with a high efficiency low fidelity spliceosome. The β -hairpin insert thus acts as a structural toggle for spliceosomal cycling between the catalytic and transitional states during splicing. Monomer A is present during spliceosome assembly, and is toggled to monomer B for the first catalytic step of splicing. Upon branching, the RH domain reverts to monomer A. It then toggles to monomer B for exon ligation, and back to monomer A following the second step (Figure 7-9).

Using the ACT1-CUP1 reporter, alleles were sorted by their ability to rescue splicing and allow growth on copper. Alleles that grew worse than wild-type Prp8

were sorted into the high fidelity, low efficiency (monomer A, previously first step) category, while alleles that grew better than wild-type Prp8 were considered high efficiency low fidelity (monomer B, previously second step). These experiments were performed with other intron mutants in addition to the BSC and BSG, and give a wider view of the effects on splicing of the Prp8 alleles. Sequencing of introns from strains containing these alleles showed that the division between the low efficiency high fidelity, with use of fewer aberrant splice sites, and high efficiency low fidelity, with use of more aberrant splice sites, is maintained in the cell (Mayerle et al., 2017). The toggle model is the next iteration of the first and second step equilibrium hypothesis. First step alleles, the low efficiency high fidelity alleles, appear first step because both steps are impaired, and the few spliceosomes that manage to perform the first step are unable to struggle through the second step. Likewise second step alleles, the low fidelity high efficiency alleles associated with monomer B, are able to quickly splice both steps, leading to an accumulation of more mRNA and less lariat intermediate.

This toggling is similar to ribosomal mutants that stabilize one form of the ribosome over the other: the closed catalytic form, or the open transitional proofreading form (Ogle et al., 2002). Prp8 is proposed to work in concert with other structural toggles in the spliceosome, including the U2 snRNA stem II, which switches between stem IIa, appearing similar to the Prp8 high fidelity alleles, and stem IIc, appearing similar to Prp8 high efficiency alleles (Hilliker et al., 2007; Perriman & Ares, 2007).

The two monomers observed in the human crystal structure appear to be functionally relevant. The disposition of the 17 amino acid insert acts as a toggle between two states in the spliceosome, either as an equilibrium between the first and second steps of splicing, or between a highly efficient or a highly proofreading spliceosome. Additional work is needed to determine the exact role of these amino acids. The metal seen bound in monomer B of the human RH domain may play a structural role in stabilizing monomer B in splicing, that might be functionally replaced by an arginine.

7-7. Materials and Methods

7-7a. Protein Expression and Purification

A cDNA encoding for the human Prp8 RH domain, amino acids 1769-1990 was cloned into the pMAL plasmid. Mutants were generated by overlapping PCR and confirmed by sequencing. The plasmids were then transformed into *E. coli* Rosetta cells and grown in LB supplemented with 2 g/L dextrose. Expression of MBP-tagged Prp8 RH was induced with 1 mM IPTG overnight at 16°C. Cells were harvested by centrifugation, and lysed in amylose lysis buffer (50 mM Tris pH 8, 100 mM NaCl, 5 mM BME, with 20 mg/L PMSF and 20 mg/L lysozyme) and lysed by sonication. Lysate was cleared by centrifugation, and run over an amylose column. Protein was eluted in lysis buffer with 20 mM maltose. Fractions were concentrated and run over the SD 200 column, then concentrated for TEV cleavage to remove the MBP tag. Cleaved protein was run over the anion exchange column and protein-containing fractions were concentrated and dialyzed into 10 mM Tris pH 8, 0.1 mM EDTA, and 5 mM DTT.

7-7b. Crystallization

Hanging drop crystals were grown at 23°C by mixing 1 μ L 10 mg/mL purified protein with 1 μ L reservoir solution (10-14% PEG 4000, 100 mM Tris pH

8, and 300 mM MgCl₂). Crystals were cryo-protected in reservoir solution containing 15% glycerol and frozen in liquid nitrogen.

7-7c. Data Collection and Processing

Data were collected at the Canadian Light Source (University of Saskatchewan, Saskatoon, Canada) on beamline CMCF-ID. Data were processed and scaled with the HKL 2000 package (Otwinowski & Minor, 1997).

7-7d. Model Building and Refinement

The structure was solved via molecular replacement using monomer A of PDB 4JK7 in Phenix Phaser (Adams et al., 2010). Model building and refinement was carried out by Phenix Refine (Adams et al., 2010) alternated with manual model building in COOT (Emsley et al., 2010). Refinement statistics are summarized in Table 7-1.

7-7e. Creation of Mutant *S. cerevisiae* Strains

PCR of mutant Prp8 was used to create plasmids via gap repair of plasmid pJU186 (Umen and Guthrie, 1995b) containing a HIS selectable marker in *S. cerevisiae* cells. Plasmids were then purified and confirmed by sequencing. Mutant plasmids were then transformed into *S. cerevisiae* strain JDY8.06 (*ura3-52, leu2-3,-112, ade2, his3-A1, trpl-289, prp8:LEU2*) containing plasmid pJU169 (*PRP8, URA3, CEN, ARS*), courtesy of Richard Grainger and Jean Beggs (University of

Edinburgh, UK). This strain has a chromosomal deletion of Prp8 rescued by Prp8 on a counter-selectable URA3 plasmid (Brown & Beggs, 1992). Following transformation with the mutant Prp8 pJU186 plasmid, cells were grown in media lacking histidine and leucine at 30°C for 16 hours, then streaked onto plates lacking histidine and leucine but containing 5-fluoroorotic acid (5-FOA; Boeke et al., 1987) to select for cells containing the mutant Prp8 on the HIS plasmid that had also dropped wild type Prp8 on the URA3 plasmid in a plasmid shuffle. Total DNA was extracted from single colonies growing on the selection plates (Qiagen DNeasy kit) and mutant Prp8 confirmed by sequencing.

7-7f. Copper Resistant Strains

Copper resistant strains were generated by transforming the mutant Prp8 plasmids with the ACT1-CUP1 (*LEU*) plasmid containing wild type, branch site G (BSG) or branch site C (BSC) intron into *S. cerevisiae* strain yJU75 (MATa, *ade2*, *cup1Døura3*, *his3*, *lys2*, *prp8DøLYS2*, *trp1*) containing the pJU169 plasmid (Umen & Guthrie, 1996; courtesy of Jonathan Staley). Plasmid shuffle was then performed as described above. DNA was then extracted (Qiagen DNeasy kit), and Prp8 mutations verified by sequencing.

7-7g. Growth Assays

Spot tests were performed by growing copper resistant strains in –LEU

-HIS media overnight at 30 °C. Cells were then diluted to OD₆₀₀=0.2 and grown at 30 °C for an additional three hours. Cells were again diluted to OD₆₀₀=0.2, and serially diluted 1/10 times. 10 µL of each dilution was then spotted on SDC plates lacking leucine and histidine, but containing 0-0.25 mM CuSO₄ (Lesser & Guthrie, 1993) and grown at 30°C for three days.

7-7h. Primer extension

Cells containing mutant Prp8 and an ACT1-CUP1 reporter plasmid were grown in -HIS -LEU media overnight at 30 °C, then diluted to 5 mL at OD₆₀₀=0.2. The cells were then grown at 30°C for an additional 6 hours. RNA was extracted using the Qiagen RNeasy kit.

Primer extensions were performed with the radiolabelled YAC6 primer complementary to the 3' exon of CUP1 (Query & Konarska, 2004) using the RevertAid H minus First Strand cDNA synthesis kit (Fermentas). 12 µL water containing 1 µg total RNA and 2 pMol labelled primer was heated to 70°C for 10 minutes, cooled slowly to 40°C then chilled on ice. 4 µL reaction buffer, 1 U RNase inhibitor, 1 mM dNTP mix and 10 U reverse transcriptase were added, then incubated for 5 minutes at 37°C followed by 55 minutes at 42°C. The reaction was terminated and RNA degraded by treatment with 0.5 M NaOH at 70°C. Synthesized DNA was phenol/chloroform extracted and ethanol precipitated, and run on gels containing 7% polyacrylamide and 8 M urea and exposed to a phosphor storage screen. Screens were scanned by a Typhoon scanner. Quantification was

performed using ImageQuant (Molecular Dynamics) and splicing efficiencies were calculated as follows.

$$1^{\text{st}} \text{ step efficiency} = \frac{\text{lariat intermediate} + \text{spliced mRNA}}{\text{lariat intermediate} + \text{spliced mRNA} + \text{pre-mRNA}}$$

$$2^{\text{nd}} \text{ step efficiency} = \frac{\text{spliced mRNA}}{\text{lariat intermediate} + \text{spliced mRNA} + \text{pre-mRNA}}$$

$$2^{\text{nd}} \text{ step}/1^{\text{st}} \text{ step} = \frac{\text{spliced mRNA}}{\text{spliced mRNA} + \text{lariat intermediate}}$$

Chapter 8

Conclusions and Future Directions

8-1. Conclusions and Future Directions

The removal of introns and ligation of exons in pre-mRNA splicing is catalyzed by the extremely complex spliceosome. Recent improvements in cryo-EM technique, as well as years of other structural and biochemical research has increased our understanding of splicing and the spliceosome dramatically since splicing was first described. This thesis explored the structure and function of a number of proteins involved in pre-mRNA splicing concluding: U1C in the U1 snRNP plays a role in recognizing the 5' splice site; a missense mutation in Sap49 that is associated with Nager syndrome is lethal in yeast; the U4 snRNP protein Snu13 is highly conserved from yeast to humans, including in the minimal spliceosome of *C. merolae*; the *C. merolae* Prp8 RH domain lacks a highly conserved 17 amino acid insertion, which may be related to the reduced intron number in this organism and its minimal spliceosome; and, metal binding by the human Prp8 RH domain does not appear to be linked to the equilibrium between the two steps of splicing.

U1C stabilizes the duplex between the 5' SS and the 5' end of the U1 snRNA. In human U1C, Arg28 and Lys29 are in a position to interact with the backbone of the 5' SS. A cross-link between a modified RNA hairpin and the R28C or K29C U1C mutants was designed to mimic this interaction. However, recent structures

of the U1 snRNP interacting with RNA does not show contact between the arginine or lysine and the backbone (Kondo et al., 2015). Pre-mRNA modified with a thiol tether within the 5' SS blocks U1 snRNP association with the spliceosome to prevent splicing. Splicing in the presence of this modification is restored in pre-mRNA sequences shown to splice independently of U1 snRNP, possibly due to steric clashes between the modification and wild type U1C, demonstrating the interaction between U1 snRNP and this region of the 5' SS.

The cysteine mutants of U1C were designed for use with a modified pre-mRNA to stall spliceosome assembly at the E or A complex in order to facilitate structural studies. I was unable to rescue splicing with a reconstituted wild type U1 snRNP added to U1 snRNP depleted extract. Thus, I did not attempt to rescue the modified pre-mRNA with the U1C cysteine mutants, and was not able to stall spliceosome assembly. The recent structures of the human and *S. cerevisiae* U1 snRNPs as well as the *S. cerevisiae* A complex limit the usefulness of this cross-linking approach for structural studies of the U1 snRNP. Despite the results of biochemical experiments, there is as yet no structural evidence of the interaction between U1C Arg28 or Lys29 and the 5' SS. Structural determination of the modified pre-mRNA cross-linked to the U1C cysteine mutants would reconcile these two lines of evidence.

Sap49 is a component of the U2 snRNP SF3b complex. Mutations that decrease protein levels in the cell, including mutations to the initiator methionine and frame shift mutations, are associated with the acrofacial dysostosis Nager

syndrome. We characterized a new mutation associated with Nager syndrome, I84R, that is not stable when recombinantly expressed, and is unable to maintain viability in *S. cerevisiae*. A second mutant discovered in humans but not associated with Nager syndrome, L28P, behaves similarly to wild type in both protein expression, its interaction with a fragment of Sap145 and its ability to bind RNA. Further work is needed to understand how mutations in the universal splicing factor Sap49 cause such a specific, localized phenotype, and why other cell types also expressing mutated Sap49 appear to be unaffected by the mutation. A comparison of mRNA sequencing could be used to determine which, if any, pre-mRNAs are being improperly spliced. A recent paper examining cartilage from an individual with Rodriguez acrofacial dysostosis, also associated with mutations in Sap49, showed defects in splicing and dysregulated gene expression (Marques et al., 2016). Investigation of the intron sequence in mis-spliced mRNAs from individuals with Nager syndrome, in addition to the introns identified by Marques et al. (2016) might reveal the susceptibility of these pre-mRNAs to the Sap49 mutations and improve our understanding of the role of Sap49 in splicing.

The red alga *C. merolae* has a decreased complement of splicing proteins and a total of 27 introns. Crystal structures of the U4 snRNP component Snu13 and the RH domain of the large U5 snRNP protein Prp8 show similar structures to both their human and yeast counterparts, suggesting that study of *C. merolae* splicing proteins is relevant to the human spliceosome. Further investigation into the

spliceosome of *C. merolae* could provide insight into the minimal core of required splicing factors.

The RH domain structure confirmed the sequence-based prediction that *C. merolae* lacks a highly conserved β -hairpin insert. Sequence alignment predicts a similar deletion in other organisms with minimal spliceosomes and few introns, including *E. cuniculi* and *O. colligata*. Structures of Prp8 or its RH domain from these organisms will confirm the lack of this insert. The concomitant absence of other splicing proteins that interact with the RH domain β -hairpin can help clarify its role in splicing.

A second set of organisms, including *G. lamblia*, *G. theta*, and *T. cruzi*, maintains the length of the β -hairpin insert, but not its sequence. These organisms predominantly undergo trans-splicing, where exons from two different pre-mRNA transcripts are ligated together. Trans-splicing is also catalyzed by the spliceosome. Investigation into trans-splicing spliceosomes will reveal further differences between the assembly and regulation of cis- and trans-splicing, including the role of the sequence changes in the β -hairpin.

Investigation into metal binding in the human and *S. cerevisiae* Prp8 RH domains shows that metal binding does not influence the equilibrium between the first step monomer A and the second step monomer B. Prior to the cryo-EM structures of the spliceosome there was some discussion about the nature of the active site of the spliceosome: is it composed of protein or RNA? There was evidence that U6 snRNA was involved, and, similar to the ribosome, an RNA active

site was suggested. However, unlike the ribosome, the spliceosome is predominantly protein. Prp8 cross-links to both splice sites and the branch region in the intron, as well as to U5 and U6 snRNAs, leading to the hypothesis that Prp8 could be involved in spliceosome catalysis by participating in the active site. The metal observed in the RH domain monomer B was thought to be required for the second step of splicing (Schellenberg et al., 2013). Structures of the spliceosome at both steps of splicing show the catalytic metals bound to U6 snRNA, and that splicing is catalyzed by RNA. My experiments with the metal binding site in monomer B of Prp8 supports the conclusion that this metal may be structural, and may not be essential for splicing.

The *C. merolae* Prp8 RH domain structure also featured an arginine bound in a position equivalent to a metal ion observed in the human structure, suggesting a positive charge at this position is somehow relevant. Arg1865 in the human RH domain monomer B could possibly occupy this space. *S. cerevisiae* and *C. merolae* lack the full complement of metal binding residues, suggesting that they do not bind a metal ion, and may instead rely on the positive charge from an arginine. To test this hypothesis, the human T1783R R1865A mutant could be crystallized to mimic the interaction observed in *C. merolae*. Likewise, the *S. cerevisiae* equivalent, the double mutant T1873R R1937T, could be tested for viability and crystallized.

An updated model of the Prp8 RH domain during the splicing cycle suggests that the RH domain toggles between a high fidelity low efficiency conformation (monomer A) that predominates in spliceosome assembly, and a low fidelity high

efficiency conformation that is required for the two transesterification reactions. Monomer B, not seen in any of the published spliceosome structures, may be the short lived catalytic conformation required for the chemical steps of splicing. The U2 snRNA and the U6 ISL are also proposed to toggle between a conformation for spliceosome assembly or rearrangement and a conformation associated with splicing catalysis. Not much is known about the factors that trigger this toggling. The recent cryo-EM structures, while very informative, are static and can only hint at the factors that trigger the toggling of these regulatory elements. Further work on the factors that influence specific spliceosomal conformations is required.

Spliceosomal proteins also interact with non-spliceosomal factors. For example, Snu13 is a component of the C/D snoRNP, and its evolution has been directed by interactions with both the C/D snoRNP and U4 snRNP complexes. Sap49 has been reported to interact with bone morphogenic protein receptor IA (Nishanian & Waldman, 2004; Watanabe et al., 2007). Interference with this interaction, rather than its effects on splicing, may be the mechanism by which mutations in Sap49 cause Nager syndrome. In addition to its role in splicing, U1 snRNP has been shown to be involved in regulating promoter directionality (Almada et al., 2013) and regulating polyadenylation (Kaida et al., 2010). Additional work is required to tease out the other, non-spliceosomal, roles splicing proteins and complexes may be playing in the cell.

With the exception of the B* complex, which was theorized from the spliceosomal B^{act} and C complex structures, spliceosomal structures exist for each

major stage of the splicing cycle. A structure of the B* complex would complete the collection of splicing cycle intermediates. However, it may not be possible to stall the spliceosome at the B* complex, which is formed after the dissociation of SF3a and SF3b but before the first step of splicing occurs.

An increased understanding of the spliceosome will aid the search for treatments of splicing related diseases, such as more productive ways to address mis-splicing. Work is being done to correct splicing errors due to mutated splice sites by competing oligos that block access to incorrect splice sites (Wan & Dreyfuss, 2017). Drugs that bind to mutant SF3b155 are being developed to correct the resulting aberrant splice site usage in the treatment of cancer (Corrionero et al., 2011; Fan et al., 2011; Convertini et al., 2014; Kashyap et al., 2015).

Increasing our understanding of pre-mRNA splicing is vital to understanding gene expression in all eukaryotes.

References

- Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, ... & Venter JC. (2000). The genome sequence of *Drosophila melanogaster*. *Science* 284 (5461), 2185-2195.
- Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, ... & Zwart PH. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*, 66 (Pt 2), 213-221.
- Adzhubei IA, Schmidt P, Peshkin L, Ramensky VE, Gerasimova A, Bork P, ... Sunyaev SR. (2010). A method and server for predicting damaging missense mutations. *Nat Methods*, 7 (4), 248-249. doi:10.1038/nmeth0410-248
- Agafonov DE, Kastner B, Dybkov O, Hofele RV, Liu WT, Urbau H, ... & Stark H. (2016). Molecular architecture of the human U4/U6.U5 tri-snRNP. *Science*, 351 (6280), 1416-1420. doi: 10.1126/science.aad2085
- Aittaleb M, Rashid R, Chen Q, Palmer JR, Daniels CJ, & Li H. (2003). Structure and function of archaeal box C/D snoRNP core proteins. *Nat Struct Biol*, 10 (4), 256-263.
- Almada AE, Wu X, Kriz AJ, Burge CB, & Sharp PA. (2013). Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature*, 499 (7458), 360-363. doi: 10.1038/nature12349
- Anna A, & Monika G. (2018). Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J Appl Genet*, 59 (3), 253-286. doi: 10.1007/s13353-018-0444-7
- Araujo PR, & Teixeira SM. (2011). Regulatory elements involved in the post-transcriptional control of stage-specific gene expression in *Trypanosoma cruzi*: a review. *Mem Inst Oswaldo Cruz*, 106 (3), 257-266.
- Ares M Jr. (2013). Analysis of splicing in vitro using extracts of *Saccharomyces cerevisiae*. *Cold Spring Harb Protoc*, 2013 (10), 982-985.

- Bai R, Yan C, Wan R, Lei J, & Shi Y. (2017). Structure of the post-catalytic spliceosome from *Saccharomyces cerevisiae*. *Cell*, *171* (7), 1589-1598. doi: 10.1016/j.cell.2017.10.038
- Bai R, Wan R, Yan C, Lei J, & Shi Y. (2018). Structures of the fully assembled *Saccharomyces cerevisiae* spliceosome before activation. *Science*, *360* (6396), 1423-1429. doi: 10.1126/science.aau0325
- Barabino SM, Blencowe BJ, Ryder U, Sproat BS, & Lamond AI. (1990). Targeted snRNP depletion reveals an additional role for mammalian U1 snRNP in spliceosome assembly. *Cell*, *63* (2), 293-302.
- Bao P, Will CL, Urlaub H, Boon KL, & Luhrmann R. (2017a). The RES complex is required for efficient transformation of the precatalytic B spliceosome into an activated B^{act} complex. *Genes Dev*, *31* (23-24), 2416-2429. doi: 10.1101/gad.308163.117
- Bao P, Hobartner C, Hartmuth K, & Luhrmann R. (2017b). Yeast Prp2 liberates the 5' splice site and the branch site adenosine for catalysis of pre-mRNA splicing. *RNA*, *23* (12), 1770-1779. doi: 10.1261/rna.063115.117
- Behrens SE, Tyc K, Kastner BM, Reichelt J, & Luhrmann R. (1993). Small nuclear ribonucleoprotein (RNP) U2 contains numerous additional proteins and has a bipartite RNP structure under splicing conditions. *Mol Cell Biol*, *13* (1), 307-319.
- Bellayou H, Hamzi K, Rafai MA, Karkouri M, Slassi I, Azeddoug H, & Nadifi S. (2009). Duchenne and Becker muscular dystrophy: contribution of a molecular and immunohistochemical analysis in diagnosis in Morocco. *J Biomed Biotechnol*, *2009*. doi: 10.1155/2009/325210
- Bennett M, Pinol-Roma S, Staknis D, Dreyfruss G, & Reed R. (1992). Differential binding of heterogeneous nuclear ribonucleoproteins to mRNA precursors prior to spliceosome assembly in vitro. *Mol Cell Biol*, *12* (7), 3165-3175.
- Berget SM. (1995). Exon recognition in vertebrate splicing. *J Biol Chem*, *270* (6), 2411-2414.
- Bernier FP, Caluseriu O, Ng S, Schwartzentruber J, Buckingham KJ, Innes AM, ... & Parboosingh JS. (2012). Haploinsufficiency of SF3B4, a component of the pre-mRNA spliceosomal complex, causes Nager syndrome. *Am J Hum Genet*, *90* (5), 925-933. doi: 10.1016/j.ajhg.2012.04.004

Bertram K, Agafonov DE, Dybkov O, Haselbach D, Leelaram MN, Will CL, ... & Stark H. (2017a). Cryo-EM structure of a pre-catalytic human spliceosome primed for activation. *Cell*, *170* (4), 701-713.e11 doi: 10.1016/j.cell.2017.07.011

Bertram K, Agagonov DE, Liu WT, Dybkov O, Will CL, Hartmuth K, ... & Luhrmann R. (2017b). Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature*, *542* (7641), 318-323. doi: 10.1038/nature21079

Bessonov S, Anokhina M, Krasauskas A, Golas MM, Sander B, Will CL, ... Luhrmann R. (2010). Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. *RNA*, *16* (12), 2384-2403. doi: 10.1261/rna.2456210

Biankin AV, Waddell N, Kassahn KS, Gingras MC, Muthuswamy LB, Johns AL, ... & Grimmond SM. (2012). Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature*, *491* (7424), 399-405. doi: 10.1038/nature11547

Black DL. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem*, *72*, 291-336.

Black CS, Garside EL, MacMillan AM, & Rader SD. (2016). Conserved structure of Snu13 from the highly reduced spliceosome of *Cyanidioschyzon merolae*. *Protein Sci*, *25* (4), 911-916. doi: 10.1002/pro.2894

Blumenthal T. (2012). Trans-splicing and operons in *C. elegans*. *WormBook*, ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.7.1, <http://www.wormbook.org>.

Boeke JD, Trueheart J, Natsoulis G, & Fink GR. (1987). 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. *Methods Enzymol*, *154*, 164-175.

Boise LH, Gonzales-Garcia M, Postema CE, Ding L, Lindsten T, Turka LA, ... & Thompson CB. (1993). Bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death. *Cell*, *74* (4), 597-608.

Boon KL, Grainger RJ, Ehsani P, Barrass JD, Auchynnikava T, Inglehearn CF, & Beggs JD. (2007). prp8 mutations that cause human retinitis pigmentosa lead to a U5 snRNP maturation defect in yeast. *Nat Struct Mol Biol*, *14* (11), 1077-1083.

Bringmann P, & Luhrmann R. (1986). Purification of the individual snRNPs U1, U2, U5 and U4/U6 from HeLa cells and characterization of their protein constituents. *EMBO J*, *5* (13), 3509-3516.

- Brow DA. (2002). Allosteric cascade of spliceosome activation. *Annu Rev Genet*, 36, 333-360.
- Brow DA, & Guthrie C. (1989). Splicing a spliceosomal RNA. *Nature*, 337 (6202), 14-15.
- Brown JD, & Beggs JD. (1992). Roles of PRP8 protein in the assembly of splicing complexes. *EMBO J*, 11 (10), 3721-3729.
- Burgess S, Couto JR, & Guthrie C. (1990). A putative ATP binding protein influences the fidelity of branchpoint recognition in yeast splicing. *Cell*, 60 (5), 705-717.
- Burgess SM, & Guthrie C. (1993). A mechanism to enhance mRNA splicing fidelity: the RNA-dependent ATPase Prp16 governs usage of a discard pathway for aberrant lariat intermediates. *Cell*, 73 (7), 1377-1391.
- Busch A, & Hertel KJ. (2012). Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley Interdiscip Rev RNA*, 3 (1), 1-12. doi: 10.1002/wrna.100
- Caspary F, & Seraphin B. (1998). The yeast U2A'/U2B complex is required for pre-spliceosome formation. *EMBO J*, 17 (21), 6348-6358.
- Cech TR. (1990). Self-splicing of group I introns. *Annu Rev Biochem*, 59, 543-568.
- Chabot B, & Shkreta L. (2016). Defective control of pre-messenger RNA splicing in human disease. *J Cell Biol*, 212 (1), 13-27. doi: 10.1083/jcb.201510032
- Chakarova CF, Hims MM, Bolz H, Abu-Safieh L, Patel RJ, Papaioannou MG, ... & Bhattacharya SS. (2002). Mutations in HPRP3, a third member of pre-mRNA splicing factor genes, implicated in autosomal dominant retinitis pigmentosa. *Hum Mol Genet*, 11 (1), 87-92.
- Champion-Arnaud P, & Reed R. (1994). The prespliceosome components SAP 49 and SAP 145 interact in a complex implicated in tethering U2 snRNP to the branch site. *Genes Dev*, 8 (16), 1974-1983.
- Chan SP, Kao DI, Tsai WY, & Cheng SC. (2003). The Prp19p-associated complex in spliceosome activation. *Science*, 302 (5643), 279-282.
- Chan SP, & Cheng SC. (2005). The Prp19-associated complex is required for specifying interactions of U5 and U6 with pre-mRNA during spliceosome activation. *J Biol Chem*, 289 (35), 31190-31199.

Chen JY, Stands L, Staley JP, Jackups RR Jr, Latus LJ, & Chang TH. (2001). Specific alterations of U1-C protein or U1 small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD box splicing factor. *Mol Cell*, 7 (1), 227-232.

Chen M, & Manley JL. (2009). Mechanisms of alternative splicing regulation: Insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol*, 7 (11), 227-232. doi: 10.1038/nrm2777

Chen W, Shulha HP, Ashar-Patel A, Yan J, Green KM, Query CC, ... Moore MJ. (2014). Endogenous U2.U5.U6 snRNA complexes in *S pombe* are intron lariat spliceosomes. *RNA*, 20 (3), 308-320. doi: 10.1261/rna.040980.113

Cheng SC, & Abelson J. (1987). Spliceosome assembly in yeast. *Genes Dev*, 1 (9), 1014-1027.

Chu CS, Trapnell BC, Curristin S, Cutting GR, & Crystal RG. (1993). Genetic basis of variable exon 9 skipping in cystic fibrosis transmembrane conductance. *Nat Genet*, 3 (2), 151-156.

Convertini P, Shen M, Potter PM, Palacios G, Lagiatti C, de la Grange P, ... & Stamm S. (2014). Sudemycin E influences alternative splicing and changes chromatin modifications. *Nucleic Acids Res*, 42 (8), 4947-4961. doi: 10.1093/nar/gku151

Corrionero A, Minana B, & Valcarcel J. (2011). Reduced fidelity of branch point recognition and alternative splicing induced by the anti-tumor drug spliceostatin A. *Genes Dev*, 25 (5), 445-459. doi: 10.1101/gad.2014311

Cortes JJ, Sontheimer EJ, Seiwert SD, & Steitz JA. (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.

Crawford DJ, Hoskins AA, Friedman LJ, Gelles J, & Moore MJ. (2013). Single-molecule colocalization FRET evidence that spliceosome activation precedes stable approach of 5' splice site and branch site. *Proc Natl Acad Sci USA*, 110 (17), 6783-6788. doi: 10.1073/pnas.1219305110

Cretu C, Schmitzova J, Ponce-Salvatierra A, Dybkov O, De Laurentiis EI, Sharma K, & Pena V. (2016). Molecular architecture of SF3b and structural consequences of its cancer-related mutations. *Mol Cell*, 64 (2), 307-319. doi: 10.1016/j.molcel.2016.08.036

- Crispino JD, Blencowe BJ, & Sharp PA. (1994). Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP. *Science*, *265* (5180), 1866-1869.
- Crispino JD, & Sharp PA. (1995). A U6 snRNA:pre-mRNA interaction can be rate-limiting for U1-independent splicing. *Genes Dev*, *9* (18), 2314-2323.
- Crispino JD, Mermoud JE, Lamond AI, & Sharp PA. (1996). *Cis*-acting elements distinct from the 5' splice site promote U1-independent pre-mRNA splicing. *RNA*, *2*, 664-673.
- Crouch RJ & Dirksen ML. (1982). Ribonucleases H. In SM Linn & RJ Roberts (Eds), *Nucleases* (pp. 211-241). Cold Spring Harbor: Cold Spring Harbor Laboratory.
- Cuypers B, Domagalska MA, Meysman P, Muylder G, Vanaerschot M, Imamura H, ... & Dujardin JC. (2017). Multiplexed spliced-leader sequencing: a high-throughput, selective method for RNA-seq in Trypanosomatids. *Sci Rep*, *7* (1), 3725. doi: 10.1038/s41598-017-03987-0
- Czeschik JC, Voigt C, Alanay Y, Albrecht B, Avci S, Fitzpatrick D, ... & Wieczorek D. (2013). Clinical and mutation data in 12 patients with the clinical diagnosis of Nager syndrome. *Hum Genet*, *132* (8), 885-898. doi: 10.1007/s00439-013-1295-2
- Darman RB, Seiler M, Agrawal AA, Lim KH, Peng S, Aird D, ... Buonamici S. (2015). Cancer-associated SF3B1 hotspot mutations induce cryptic 3' splice site selection through use of a different branch point. *Cell Rep*, *13* (5), 1033-1045. doi: 10.1016/j.celrep.2015.09.053
- Devotta A, Juraver-Geslin H, Gonzales JA, Hong CS, & Saint-Jeannet JP. (2016). Sf3b4-depleted *Xenopus* embryos: a model to study the pathogenesis of craniofacial defects in Nager syndrome. *Dev Biol*, *415* (2), 371-382. doi: 10.1016/j.ydbio.2016.02.010
- Dobbyn HC, McEwan PA, Krause A, Novak-Frazer L, Bella J, & O'Keefe RT. (2007). Analysis of pre-mRNA and pre-rRNA processing factor Snu13p structure and mutants. *Biochem Biophys Res Commun*, *360* (4), 857-862.
- Dole MG, Jasty R, Cooper MJ, Thompson CB, Nunez G, & Castle VP. (1995). Bcl-xL is expressed in neuroblastoma cells and modulates chemotherapy-induced apoptosis. *Cancer Res*, *55* (12), 2576-2582.

Du H, & Rosbash M. (2001). Yeast U1 snRNP-pre-mRNA complex formation without U1snRNA-pre-mRNA base pairing. *RNA*, 7 (1), 133-142.

Du H, & Rosbash M. (2002). The U1 snRNP protein U1C recognizes the 5' splice site in the absence of base pairing. *Nature*, 419 (6902), 86-90.

Dunn EA, & Rader SD. (2014). Pre-mRNA splicing and the spliceosome: assembly, catalysis, and fidelity. In A Sesma & T von der Haar (Eds.), *Fungal RNA biology* (pp 27-57). doi: 10.1007/978-3-319-05687-6_2

Dziembowski A, Ventura AP, Rutz B, Caspary F, Faux C, Halgand F, ... & Seraphin B. (2004). Proteomic analysis identifies a new complex required for nuclear pre-mRNA retention and splicing. *EMBO J*, 23 (24), 4847-4856.

Emsley P, Lohkamp B, Scott WG, & Cowtan K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*, 66 (Pt 2), 486-501.

Faa V, Coiana A, Incani F, Costantino L, Cao A, & Rosatelli MC. (2010). A synonymous mutation in the CFTR gene causes aberrant splicing in an Italian patient affected by a mild form of cystic fibrosis. doi: 10.2353/jmoldx.2010.090126

Fan L, Lagisetti C, Edwards CC, Webb TR, Potter PM, (2011). Sudemycins, novel small molecule analogues of FR901464, induce alternative gene splicing. *ACS Chem Biol*, 6 (6), 582-589. doi: 10.1021/cb100356k

Farkas MH, Grant GR, & Pierce EA. (2012). Transcriptome analyses to investigate the pathogenesis of RNA splicing factor retinitis pigmentosa. *Adv Exp Med Biol*, 723, 519-525. doi: 10.1007/978-1-4614-0631-0_65

Faustino NA, & Cooper TA. (2003). Pre-mRNA splicing and human disease. *Genes Dev*, 17 (4). 419-437.

Ferris MJ, Sheehan KB, Kuhl M, Cooksey K, Wigglesworth-Cooksey B, Harvey R, & Henson JM. (2005). Algal species and light microenvironment in a low-pH, geothermal microbial mat community. *Appl Environ Microbiol*, 71 (11), 7164-7171.

Fica SM, Tuttle N, Novak T, Li NS, Lu J, Koodathingal P. ... & Piccirilli JA. (2013). RNA catalyses nuclear pre-mRNA splicing. *Nature*, 503 (7475), 229-234. doi: 10.1038/nature12734

Fica SM, Mefford MA, Piccirilli JA, & Staley JP. (2014). Evidence for a group II intron-like catalytic triplex in the spliceosome. *Nat Struct Mol Biol*, 21 (5), 494-471. doi: 10.1038/nsmb.2815

Fica SM, Oubridge C, Galej WP, Wilkinson ME, Bai XC, Newman AJ, & Nagai K. (2017). Structure of a spliceosome remodelled for exon ligation. *Nature*, 542 (7641), 377-380. doi: 10.1038/nature21078

Fica SM, & Nagai K. (2017). Cryo-electron microscopy snapshots of the spliceosome: structural insights into a dynamic ribonucleoprotein machine. *Nat Struct Mol Biol*, 24 (10), 791-799. doi: 10.1038/nsmb.3463

Finci LI, Zhang X, Huang X, Zhou Q, Tsai J, Teng T, ... & Larsen NA. (2018). The cryo-EM structure of the SF3b spliceosome complex bound to a splicing modulator reveals a pre-mRNA substrate competitive mechanism of action. *Genes Dev*, 32 (3-4), 309-320. doi: 10.1101/gad.311043.117

Forch P, Puig O, Martinez C, Seraphin B, & Valcarcel J. (2002). The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5' splice sites. *EMBO J*, 21 (24), 6882-6892.

Fortner DM, Troy RG, & Brow DA. (1994). A stem/loop in U6 RNA defines a conformational switch required for pre-mRNA splicing. *Genes Dev*, 8 (2), 221-233.

Freund M, Asang C, Kammler S, Konermann C, Krummheuer J, Hipp M, ... & Schaal H. (2003). A novel approach to describe a U1 snRNA binding site. *Nucleic Acids Res*, 31 (23), 6963-6975.

Freund M, Hicks MJ, Konermann C, Otte M, Hertel KJ, & Schaal H. (2005). Extended base pair complementarity between U1 snRNA and the 5' splice site does not inhibit splicing in higher eukaryotes, but rather increases 5' splice site recognition. *Nucleic Acids Res*, 33, 5112-5119.

Fujiwara T, Ohnuma M, Kuroiwa T, Ohbayashi R, Hirooka S, & Miyagishima SY. (2017). Development of a double nuclear gene-targeting method by two-step transformation based on a newly established chloramphenicol-selection system in the red alga *Cyanidoschyzon merolae*. *Front Plant Sci*, 8, 343. doi: 10.3389/fpls.2017.00343

Fukumura K, Taniguchi I, Sakamoto H, Ohno M, & Inoue K. (2009). U1-independent pre-mRNA splicing contributes to the regulation of alternative splicing. *Nucleic Acids Res*, 37 (6), 1907-1914. doi: 10.1093/nar/gkp050

Galej WP, Oubridge C, Newman AJ, & Nagai K. (2013). Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature*, 493 (7434), 638-343. doi: 10.1038/nature11843

- Galisson F, & Legrain P. (1993). The biochemical defects of prp4-1 and prp6-1 yeast splicing mutants reveal that the PRP6 protein is required for the accumulation of the [U4/U6.U5] tri-snRNP. *Nucleic Acids Res*, 21 (7), 1555-1562.
- Gilbert HF. (1995). Thiol/disulfide exchange equilibria and disulfide bond stability. *Methods Enzymol*, 251, 8-28.
- Golas MM, Sander B, Will CL, Luhrmann R, & Stark H. (2003). Molecular architecture of the multiprotein splicing factor SF3b. *Science*, 300 (5621), 980-984.
- Gottschalk A, Tang J, Puig O, Salgado J, Neubauer G, Colot HV, ... & Fabrizio P. (1998). A comprehensive biochemical and genetic analysis of the yeast U1 snRNP reveals five novel proteins. *RNA*, 4 (4), 372-393.
- Gozani O, Feld R, & Reed R. (1996). Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes Dev*, 10 (2), 233-243.
- Grainger RJ, & Beggs JD. (2005). Prp8 protein: at the heart of the spliceosome. *RNA*, 11 (5), 533-557.
- Grisdale CJ, Bowers LC, Didier ES, & Fast NM. (2013). Transcriptome analysis of the parasite *Encephalitozoon cuniculi*: an in-depth examination of pre-mRNA splicing in a reduced eukaryote. *BMC Genomics*, 14, 207. doi: 10.1186/1471-2164-14-207
- Gunzl A. (2010). The pre-mRNA splicing machinery of trypanosomes: complex or simplified? *Eukaryot Cell*, 9 (8), 1159-1170. doi: 10.1128/EC.00113-10
- Habara Y, Takeshima Y, Awano H, Okizuka Y, Zhang Z, Saiki K, ... & Matsuo M. (2009). In vitro splicing analysis showed that availability of a cryptic splice site is not a determinant for alternative splicing patterns caused by the +1G→A mutations in introns of the dystrophin gene. *J Med Genet*, 46 (8), 542-547. doi: 10.1136/jmg.2008.061259
- Hamma T, Ferré-D'Amaré AR. (2004). Structure of protein L7Ae bound to a K-turn derived from an archaeal box H/ACA snRNA at 1.8 Å resolution. *Structure*, 21 (5), 893-903.
- Hardin JW, Warnasooriya C, Kondo Y, Nagai K, & Rueda D. (2015). Assembly and dynamics of the U4/U6 di-snRNP by single-molecule FRET. *Nucleic Acids Res*, 43 (22), 10961-10974. doi: 10.1093/nar/gkv1011

- Havens MA, & Hastings ML. (2016). Splice-switching antisense oligonucleotides as therapeutic drugs. *Nucleic Acids Res*, 44 (14), 6549-6563. doi: 10.1093/nar/gkw533
- Heinrichs V, Bach M, Winkelmann G, & Luhrmann R. (1990). U1-specific protein C needed for efficient complex formation of U1 snRNP with a 5' splice site. *Science*, 247 (4938), 69-72.
- Hilliker AK, Mefford MA, & Staley JP. (2007). U2 toggles iteratively between the stem IIa and stem IIc conformations to promote pre-mRNA splicing. *Genes Dev*, 21 (7), 821-834.
- Hinterberger M, Pettersson I, & Steitz JA. (1983). Isolation of small nuclear ribonucleoproteins containing U1, U2, U4, U5 and U6 snRNAs. *J Biol Chem*, 258 (4), 2604-2613.
- Hodges PE, Jackson SP, Brown JD, & Beggs JD. (1995). Extraordinary sequence conservation of the PRP8 splicing factor. *Yeast*, 11 (4), 337-342.
- Hogg R, McGrail JC, & O'Keefe RT. (2010). The function of the NineTeen Complex (NTC) in regulating spliceosome conformations and fidelity during pre-mRNA splicing. *Biochem Soc Trans*, 38 (4), 1110-1115. doi: 10.1042/BST0381110
- Hori T, Fukao T, Murase K, Sakaguchi N, Harding CO, & Kondo N. (2013). Molecular basis of two-exon skipping (exons 12 and 13) by c.1248+5g>a in OXCT1 gene: study on intermediates of OXCT1 transcripts in fibroblasts. *Hum Mutat*, 34 (3), 473-480. doi: 10.1002/humu.22258
- Hoskins AA, Friedman LJ, Gallagher SS, Crawford DJ, Anderson EG, Wombacher R, ... & Moore MJ. (2011). Ordered and dynamic assembly of single spliceosomes. *Science*, 331 (6022), 1289-1295. doi: 10.1126/science.1198830
- Huang T, Vilardell J, & Query CC. (2002). Pre-spliceosome formation in *S. pombe* requires a stable complex of SF1-U2AF(59)-U2AF(32). *EMBO J*, 21 (20), 5516-5526.
- Hudson AJ, Stark MR, Fast NM, Russell AG, & Rader SD. (2015). Splicing diversity revealed by reduced spliceosomes in *C. merolae* and other organisms.
- Igel H, Wells S, Perriman R, & Ares M Jr. (1998). Conservation of structure and subunit interactions in yeast homologues of splicing factor 3b (SF3b) subunits. *RNA*, 4 (1), 1-10.

- Jacquier A. (1990). Self-splicing group II and nuclear pre-mRNA introns: how similar are they?. *Trends Biochem Sci*, 15 (9), 351-354.
- Jamieson DJ, Rahe B, Pringle J, & Beggs JD. (1991). A suppressor of a yeast splicing mutant (prp8-1) encodes a putative ATP-dependent RNA helicase. *Nature*, 349 (6311), 715-717.
- Kaida D, Berg MG, Younis I, Kasim M, Singh LN, Wan L, & Dreyfuss G. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature*, 468 (7324), 664-668. doi: 10.1038/nature09479
- Kamikawa R, Inagaki Y, Tokoro M, Roger AJ, & Hashimoto T. (2011). Split introns in the genome of *Giardia intestinalis* are excised by spliceosome-mediated trans-splicing. *Curr Biol*, 21 (4), 311-315. doi: 10.1016/j.cub.2011.01.025
- Kashyap MK, Kumar D, Villa R, La Clair JJ, Benner C, Sasik R, ... & Castro JE. (2015). Targeting the spliceosome in chronic lymphocytic leukaemia with the macrolides FD-895 and pladienolide-B. *Haematologica*, 100 (7), 945-954. doi: 10.3324/haematol.2014.122069
- Katinka MD, Duprat S, Cornillot E, Metenier G, Thomarat F, Prensier G, ... & Vivares CP. (2001). Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature*, 414 (6862), 450-453.
- Kistler AL, & Guthrie C. (2001). Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for sub2, an essential spliceosomal ATPase. *Genes Dev*, 15 (1), 42-49.
- Kondo Y, Oubridge C, van Roon AM, & Nagai K. (2015). Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5' splice site recognition. *Elife*, 4. doi: 10.7554/eLife.04986.
- Konarska MM, Vilardell J, & Query CC. (2006). Repositioning of the reaction intermediate within the catalytic center of the spliceosome. *Mol Cell*, 21 (4), 543-553.
- Koodathingal P, Novak T, Piccirilli JA, & Staley JP. (2010). The DEAH box ATPases Prp16 and Prp43 cooperate to proofread 5' splice site cleavage during pre-mRNA splicing. *Mol Cell*, 39 (3), 385-395. doi: 10.1016/j.molcel.2010.07.014
- Koodathingal P, & Staley JP. (2013). Splicing fidelity: DEAD/H-box ATPases as molecular clocks. *RNA Biol*, 10 (7), 1073-1079. doi: 10.4161/rna.25245

- Kretzner L, Rymond BC, & Rosbash M. (1987). *S. cerevisiae* U1 RNA is large and has limited primary sequence homology to metazoan U1 snRNA. *Cel*, 50 (4), 593-602.
- Kretzner L, Krol A, & Rosbash M. (1990). *Saccharomyces cerevisiae* U1 small nuclear RNA secondary structure contains both universal and yeast-specific domains. *Proc Natl Acad Sci USA*, 87 (2), 851-855.
- Krol A, Westhof E, Bach M, Luhrmann R, Ebel JP, & Carbon P. (1990). Solution structure of human U1 snRNA. Derivation of a possible three-dimensional model. *Nucleic Acids Res*, 18 (13), 3803-3811.
- Kuhn AN, Reichl EM, Brow DA. (2002a). Distinct domains of splicing factor Prp8 mediate different aspects of spliceosome activation. *Proc Natl Acad Sci USA*, 99 (14), 9145-9148.
- Kuhn JF, Tran EJ, & Maxwell ES. (2002b). Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein. *Nucleic Acids Res*, 30 (4), 931-934.
- Kuwasako K, Nameki N, Tsuda K, Takahashi M, Sato A, Tochino N, ... & Muto Y. (2017). Solution structure of the first RNA recognition motif domain of human spliceosomal protein SF3b49 and its mode of interaction with a SF3b145 fragment. *Protein Sci*, 26 (2), 280-291. doi: 10.1002/pro.3080
- Lambowitz AM, & Zimmerly S. (2011). Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb Perspect Biol*, 3 (8), a003616. doi: 10.1101/cshperspect.a003616
- Lamontagne J, & Papadopoulou B. (1999). Developmental regulation of spliced leader RNA gene in *Leishmania donovani* amastigotes is mediated by specific polyadenylation. *J Biol Chem*, 274 (10), 6602-6609.
- Lansinger Y, & Rayan G. (2015). Nager syndrome. *J Hand Surg Am*, 40 (4), 851-854. doi: 10.1016/j.jhsa.2014.10.064
- Lardelli RM, Thompson JX, Yates JR 3rd, & Stevens SW. (2010). Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. *RNA*, 16 (3), 516-528. doi: 10.1261/rna.2030510
- Lasda EL. & Blumenthal T. (2011). Trans-splicing. *Wiley Interdiscip Rev RNA*, 2 (3), 417-434. doi: 10.1002/wrna.71

- Le Hir H, Izaurralde E, Maquat LE, & Moore MJ. (2000). The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J*, 19 (24), 6860-6869.
- Lehalle D, Wieczorek D, Zechi-Ceide RM, Passos-Bueno MR, Lyonnet S, Amiel J, & Gordon CT. (2015). A review of craniofacial disorders caused by spliceosomal defects. *Clin Genet*, 88 (5), 405-415. doi: 10.1111/cge.12596
- Lei Q, Li C, Zuo Z, Huang C, Cheng H, & Zhou R. (2016). Evolutionary insights into RNA trans-splicing in vertebrates. *Genome Biol Evol* 80(3), 562-577. doi: 10.1093/gbe/evw025
- Lek M, Karczewski K, Minikel EV, Samocha KE, Banks E, Fennell T, ... & Exome Aggregation Consortium. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536 (7616), 285-291. doi: 10.1038/nature19057
- Lesser CF, & Guthrie C. (1993), Mutational analysis of pre-mRNA splicing in *Saccharomyces cerevisiae* using a sensitive new reporter gene, CUP1. *Genetics*, 133 (4), 851-863.
- Li Z, & Brow DA. (1996). A spontaneous duplication in U6 spliceosomal RNA uncouples the early and late functions of the ACAGA element in vivo. *RNA*, 2 (9), 879-894.
- Li P, Kirkpatrick J, & Carlomagno T. (2009). An efficient strategy for the determination of the three-dimensional architecture of ribonucleoprotein complexes by the combination of a few easily accessible NMR and biochemical data: intermolecular recognition in a U4 spliceosomal complex. *J Mol Biol*, 388 (2), 283-298. doi: 10.1016/j.jmb.2009.03.001
- Li Z, Li Q, Han L, Tian N, Liang Q, Li Y, ... & Tian Y. (2016). Pro-apoptotic effects of splice-switching oligonucleotides targeting Bcl-x pre-mRNA in human glioma cell lines. *Oncol Rep*, 35 (2), 1013-1019. doi: 10.3892/or.2015.4465
- Li X, Liu S, Jiang J, Zhang L, Espinosa S, Hill RC ... & Zhao R. (2017). CryoEM structure of *Saccharomyces cerevisiae* U1 snRNP offers insight into alternative splicing. *Nat Commun*, 8 (1), 1035. doi: 10.1038/s41467-017-01241-9
- Liang WW, & Cheng SC. (2015). A novel mechanism for Prp5 function in prespliceosome formation and proofreading the branch site sequence. *Genes Dev*, 29 (1), 181-93. doi: 10.1101/gad.253708.114
- Lin PC, & Xu RM. (2012). Structure and assembly of the SF3a splicing factor complex of U2 snRNP. *EMBO J*, 31 (6), 1579-1590. doi: 10.1038/emboj.2012.7

Lines MA, Huang L, Schwartzenruber J, Douglas SL, Lynch DC, Beaulieu C, ...Boycott KM. (2012). Haploinsufficiency of a spliceosomal GTPase encoded by EFTUD2 causes mandibulofacial dysostosis with microcephaly. *Am J Hum Genet*, 90 (2), 369-377. doi: 10.1016/j.ajhg.2011.12.023

Liu YC, Chen HC, Wu NY, Cheng SC. (2007a). A novel splicing factor, Yju2, is associated with NTC and acts after Prp2 in promoting the first catalytic reaction of pre-mRNA splicing. *Mol Cell Biol*, 27 (15), 5403-5413.

Liu S, Li P, Dybkov O, Nottrott S, Hartmuth K, Luhrmann R, ... & Wahl MC. (2007b). Binding of the human Prp31 Nop domain to a composite RNA-protein platform in U4 snRNP. *Science*, 316 (5821), 15-120.

Liu L, Query CC, & Konarska MM. (2007c). 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 S, Ghalei H, Luhrmann R, & Wahl MC. (2011). Structural basis for the dual U4 and U4atac snRNA-binding specificity of spliceosomal protein hPrp31. *RNA*, 17 (9), 1655-1663. doi: 10.1261/rna.2690611

Liu S, Mozaffari-Jovin S, Wollenhaupt J, Santos KF, Theuser M, Dunin-Horkawicz S, ... & Wahl MC. (2015). A composite double-/single-stranded RNA-binding region in protein Prp3 supports tri-snRNP stability and splicing. *Elife*, 4, e07320. doi: 10.7554/eLife.07320

Liu S, Li X, Zhang L, Jiang J, Hill RC, Cui Y, ... & Zhao R. (2017). Structure of the yeast spliceosomal postcatalytic P complex. *Science*, 358 (6368), 1278-1283. doi: 10.1126/science.aar3462

Lykke-Andersen S, & Jensen TH. (2015). Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nat Rev Mol Cell Biol*, 16 (11), 665-677.

MacMillan AM, Query CC, Allerson CR, Chen S, Verdine GL, & Sharp PA. (1994). Dynamic association of proteins with the pre-mRNA branch region. *Genes Dev*, 8 (24), 3008-3020.

Madhani HD, & Guthrie C. (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.

- Maguire SL, Leonidou A, Wai P, Marchio C, Ng CK, Sapino A, ... & Natrajan RC. (2015). SF3B1 mutations constitute a novel therapeutic target in breast cancer. *J Pathol*, 235 (4), 571-580. doi: 10.1002/path.4483
- Maimon A, Mogilevsky M, Shilo A, Golan-Gerstl R, Obiedat A, Ben-Hur V, ... & Karni R. (2014). Mnk2 alternative splicing modulates the p38-MAKP pathway and impacts Ras-induced transformation. *Cell Rep*, 7 (2), 501-513. doi: 10.1016/j.celrep.2014.03.041
- Malcovati L, Papaemmanuil E, Bowen DT, Boulton J, Della Porta MG, Pascutto C, ... & the Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium and of the Associazione Italiana per la Ricerca sul Cancro Gruppo Italiano Malattie Mieloproliferative. (2011). Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. *Blood*, 118 (4), 6239-6246. doi: 10.1182/blood-2011-09-377275
- Marcia M, & Pyle AM. (2012). Visualizing group II intron catalysis through the stages of splicing. *Cell*, 151 (3), 497-507. doi: 10.1016/j.cell.2012.09.033
- Marques F, Tenney J, Duran I, Martin J, Nevarez L, Pogue R, ... & Li B. (2016). Altered mRNA splicing, chondrocyte gene expression and abnormal skeletal development due to SF3B4 mutations in Rodriguez acrofacial dysostosis. *PLoS Genet*, 12 (9), e1006307. doi: 10.1371/journal.pgen.1006307
- Mathe E, Olivier M, Kato S, Ishioka C, Hainaut P, & Tavtigian SV. (2006). Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods. *Nucleic Acids Res*, 34 (5), 1317-1325.
- Matlin AJ, Clark F, & Smith CW. (2005). Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol*, 6 (5), 386-398.
- Matsuzaki M, Misumi O, Shin-I T, Maruyama S, Takahara M, Miyagishima SY, ... Kuroiwa T. (2004). Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D.
- Mayas RM, Haita H, & Staley JP. (2006). Exon ligation is proofread by the DExD/H-box ATPase Prp22p. *Nat Struct Mol Biol*, 13 (6), 482-490.
- Mayerle M, Raghavan M, Ledoux S, Price A, Stepankiw N, Hadjivassiliou H, ... & Abelson J. (2017). Structural toggle in the RNaseH domain of Prp8 helps balance splicing fidelity and catalytic efficiency. *Proc Natl Acad Sci USA*, 114 (8), 4739-4744. doi: 10.1073/pnas.1701462114

- McCarthy P, Garside E, Meschede-Krasa Y, MacMillan A, & Pomeranz Krummel D. (2017). Reversibly constraining spliceosome-substrate complexes by engineering disulfide crosslinks. *Methods*, 125, 25-35.
doi: 10.1016/j.ymeth.2017.06.022
- McGrail JC, Krause A, & O'Keefe RT. (2009). The RNA binding protein Cwc2 interacts directly with the U6 snRNA to link the nineteen complex to the spliceosome during pre-mRNA splicing. *Nucleic Acids Res*, 37 (13), 4205-4217.
doi: 10.1093/nar/gkp341
- McKie AB, McHale JC, Keen TJ, Tarttelin EE, Goliath R, van Lith-Verhoeven JJ, ... & Inglehearn CF. (2001). Mutations in the pre-mRNA splicing factor gene PRPC8 in autosomal dominant retinitis pigmentosa (RP13). *Hum Mol Genet*, 10 (15), 1555-1562.
- McManus CJ, Duff MO, Eipper-Mains J, & Graveley BR. (2010). Global analysis of trans-splicing in *Drosophila*. *Proc Natl Acad Sci USA* 107 (29), 12975-12979.
doi: 10.1073/pnas.1007586107
- McPheeters DS. (1996). Interactions of the yeast U6 RNA with the pre-mRNA branch site. *RNA*, 2 (11), 1110-1123.
- McPherson E, Zaleski C, Ye Z, & Lin S. (2014). Rodriguez syndrome with SF3b4 mutation: a severe form of Nager syndrome? *Am J Med Genet A*, 164A (7), 1841-1845. doi: 10.1002/ajmg.a.36555
- Michaud S, & Reed R. (1991). An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. *Genes Dev*, 5 (12B), 2534-2546.
- Michaud S, & Reed R. (1993). A functional association between the 5' and 3' splice site is established in the earliest prespliceosome complex (E) in mammals. *Genes Dev*, 7 (6), 1008-1020.
- Mohlmann S, Mathew R, Neumann P, Schmitt A, Luhrmann R, & Ficner R. (2014). Structural and functional analysis of the human spliceosomal DEAD-box helicase Prp28. *Acta Crystallogr D Biol Crystallogr*, 70 (Pt6), 1622-1630.
- Moore MJ, & Sharp PA. (1992). Site-specific modification of pre-mRNA: the 2'-hydroxyl groups at the splice sites. *Science*, 256 (5059), 992-997.
- Moore MJ, & Sharp PA. (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature*, 365 (6444), 364-368.

- Mount SM, Pettersson I, Hinterberger M, Karmas A, & Steitz JA. (1983). The U1 small nuclear RNA-protein complex selectively binds a 5' splice site in vitro. *Cell*, 33 (2), 509-518.
- Mozaffari-Jovin S, Wandersleben T, Santos KF, Will CL, Luhrmann R, & Wahl MC. (2013). Inhibition of RNA helicase Brr2 by the C-terminal tail of the spliceosomal protein Prp8. *Science*, 341 (6141), 80-84. doi: 10.1126/science.1237515
- Muto Y, Pomeranz Krummel D, Oubridge C, Hernandez H, Robinson CV, Neuhaus D, & Nagai K. (2004). The structure and biochemical properties of the human spliceosomal protein U1C. *J Mol Biol*, 341 (1), 185-198.
- Nancollis V, Ruckshanthi JP, Frazer LN, & O'Keefe RT. (2013). The U5 snRNA internal loop 1 is a platform for Brr2, Snu114 and Prp8 protein binding during U5 snRNP assembly. *J Cell Biochem*, 114 (12), 2770-2784. doi: 10.1002/jcb.24625
- Nelissen RL, Heinrichs V, Habets WJ, Simons F, Luhrmann R, & van Venrooij WJ. (1991). Zinc finger-like structure in U1-specific protein C is essential for specific binding to U1 snRNP. *Nucleic Acids Res*, 19 (3), 449-454.
- Neueglise C, Marck C, & Gaillardin C. (2011). The intronome of budding yeasts. *C R Biol*, 334 (8-9), 662-670.
- Newby MI, & Greenbaum NL. (2001). A conserved pseudouridine modification in eukaryotic U2 snRNA induces a change in branch-site architecture. *RNA*, 7 (6), 833-845.
- Nguyen TH, Galej WP, Bai XC, Sawa CG, Newman AJ, Scheres SH, & Nagai K. (2015). The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature*, 523 (7558), 47-52. doi: 10.1038/nature14548
- Nguyen THD, Galej WP, Bai XC, Oubridge C, Newman AJ, Scheres SHW, & Nagai K. (2016). Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. *Nature*, 530 (7590), 289-302. doi: 10.1038/nature16940
- Nilsen TW, & Graveley BR. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature*, 463, 457-463. doi: 10.1038/nature08909
- Nishanian TG, & Waldman T. (2004). Interaction of the BMPR-1A tumor suppressor with a developmentally relevant splicing factor. *Biochem Biophys Res Commun*, 323 (1), 91-97.

- Nottrott S, Hartmuth K, Fabrizio P, Urlaub H, Vidovic I, Ficner R, & Luhrmann R. (1999). Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. *EMBO J*, 18 (21), 6119-6133.
- Nottrott S, Urlaub H, & Luhrmann R. (2002). Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. *EMBO J*, 21 (20), 5527-5538.
- Ogle JM, Murphy FV, Tarry MJ, & Ramakrishnan V. (2002). Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell*, 111 (5), 721-732.
- Ohno K, Rahman MA, Nasrin F, & Masuda A. (2015). Decoding abnormal splicing code in human diseases. *J Investig Genomics*, 2 (1), 00016. doi: 10.15406/jig.2015.02.00016
- Oruganti SV, Zhang Y, & Li H. (2005). Structural comparison of yeast snoRNP and spliceosomal protein Snu13p with its homologs. *Biochem Biophys Res Commun*, 333 (2), 550-554.
- Otwinowski Z, & Minor W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol*, 276, 307-326.
- Oubridge C, Ito N, Evans PR, Teo CH, & Nagai K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, 372 (6505), 432-438.
- Oubridge C, Krummel DA, Leung AK, Li J, & Nagai K. (2009). Interpreting a low resolution map of human U1 snRNP using anomalous scatterers. *Structure*, 17 (7), 930-938. doi: 10.1016/j.str.2009.05.009
- Pan Q, Shai O, Lee LJ, Frey BJ, & Blencowe BJ. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet*, 40 (12), 1413-1415. doi: 10.1038/ng.259
- Pandit S, Lynn B, & Rymond BC. (2006). Inhibition of a spliceosome turnover pathway suppresses splicing defects. *Proc Natl Acad Sci USA*, 103 (37), 13700-13705.
- Pauling MH, McPheeters DS, & Ares M Jr. (2000). Functional Cus1p is found with Hsh155p in a multiprotein splicing factor associated with U2 snRNA. *Mol Cell Biol*, 20 (6), 2176-2185.

- Pena V, Rozov A, Fabrizio P, Luhrmann R, & Wahl MC. (2008). Structure and function of an RNase H domain at the heart of the spliceosome. *EMBO J*, 27 (21), 2929-2940. doi: 10.1038/emboj.2008.209
- Perriman RJ, & Ares M Jr. (2007). Rearrangement of competing U2 RNA helices within the spliceosome promotes multiple steps in splicing. *Genes Dev*, 21 (7), 811-820.
- Peters JK, & Toor N. (2015). Group II intron lariat: structural insights into the spliceosome. *RNA Biol*, 12 (9), 913-917. doi: 10.1080/15476286.2015.1066956
- Petit F, Escande F, Jourdain AS, Porchet N, Amiel J, Doray B, ... & Holder-Espinasse M. (2014). Nager syndrome: confirmation of SF3B4 haploinsufficiency as the major cause. *Clin Genet*, 86 (3), 246-251. doi: 10.1111/cge.12259
- Plaschka C, Lin PC, & Nagai K. (2017). Structure of a pre-catalytic spliceosome. *Nature*, 546 (7660), 617-621. doi: 10.1038/nature22799
- Plaschka C, Lin PC, Charenton C, & Nagai K. (2018). Prespliceosome structure provides insights into spliceosome assembly and regulation. *Nature*, 559 (7714), 419-422. doi: 10.1038/s41586-018-0323-8
- Podar M, Perlman PS, & Padgett RA. (1998). The two steps of group II intron self-splicing are mechanistically distinguishable. *RNA*, 4 (8), 890-900.
- Pomeranz Krummel DA, Oubridge C, Leung AKW, Li J, & Nagai K. (2009). Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature*, 458 (7237), 478-480. doi: 10.1038/nature07851
- Price PL, Morderer D, & Rossoll W. (2018). RNP assembly defects in spinal muscular atrophy. *Adv Neurobiol*, 20, 143-171. doi: 10.1007/978-3-319-89689-2_6
- Puig O, Bragado-Nilsson E, Koski T, & Seraphin B. (2007). The U1 snRNP-associated factor Luc7p affects 5' splice site selection in yeast and humans. *Nucleic Acids Res*, 35 (17), 5874-5885.
- Pyle AM. (2016). Group II intron self-splicing. *Annu Rev Biophys*, 45, 183-205. doi: 10.1146/annurev-biophys-062215-011149
- Qu G, Kaushal PS, Wang J, Shigematsu H, Piazza CL, Agrawal RK, ... & Wang HW. (2016). Structure of a group II intron in complex with its reverse transcriptase. *Nat Struct Mol Biol*, 23 (6), 549-557. doi: 10.1038/nsmb.3220

Query CC, Moore MJ, & Sharp PA. (1994). Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev*, 8 (5), 587-597.

Query CC, Strobel SA, & Sharp PA. (1996). Three recognition events at the branch-site adenine. *EMBO J*, 15 (6), 1392-1402.

Query CC, & Konarska MM. (2004). Suppression of multiple substrate mutations by spliceosomal prp8 alleles suggests functional correlations with ribosomal ambiguity mutants. *Mol Cell*, 14 (3), 343-354.

Qui H, Rossoni AW, Weber APM, Yoon HS, & Bhattacharya D. (2018). Unexpected conservation of the RNA splicing apparatus in the highly streamlined genome of *Galdieria sulphuraria*. *BMC Evol Biol*, 18 (1), 41. doi: 10.1186/s12862-018-1161-x

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

Rahman MA, Nasrin F, Masuda A, & Ohno K. (2015). Decoding abnormal splicing code in human diseases. *J Investig Genomics*, 2 (1), 00016. doi: 10.15406/jig.2015.02.00016

Reddy R, Ro-Choi TS, Hemming D, & Busch H. (1974). Primary sequence of U-1 nuclear ribonucleic acid of Novikoff hepatoma ascites cells. *J Biol Chem*, 249 (20), 6486-6494.

Ritchie DB, Schellenberg MJ, Genser EM, Raithatha SA, Stuart DT, & MacMillan AM. (2008). Structural elucidation of a PRP8 core domain from the heart of the spliceosome. *Nat Struct Mol Biol*, 15 (11), 1199-1205. doi: 10.1038/nsmb.1505

Rogers J, & Wall R. (1980). A mechanism for RNA splicing. *Proc Natl Acad Sci USA*, 77 (4), 1877-1879.

Rodgers ML, Tretbar US, Dehaven A, Alwan AA, Luo G, Mast HM, & Hoskins AA. (2016). Conformational dynamics of stem II of the U2 snRNA. *RNA*, 22 (2), 225-236. doi: 10.1261/rna.052233.115

Rossi F, Forne T, Antoine E, Tazi J, Brunel C, & Cathala G. (1996). Involvement of U1 small nuclear ribonucleoproteins (snRNP) in 5' splice site-U1 snRNP interaction. *J Biol Chem*, 271 (39), 23985-23991.

Roy SW, Hudson AJ, Joseph J, Yee J, Russell AG. (2012). Numerous fragmented spliceosomal introns, AT-AC splicing, and an unusual dynein gene expression

pathway in *Giardia lamblia*. *Mol Biol Evol*, 29 (1), 43-49. doi: 10.1093/molbev/msr063

Roy SW. (2017). Genomic and transcriptomic analysis reveals spliced leader trans-splicing in cryptomonads. *Genome Biol Evol*, 9 (3), 468-473. doi: 10.1093/gbe/evx012

Ryan DE, & Abelson J. (2002). The conserved central domain of yeast U6 snRNA: importance of U2-U6 helix Ia in spliceosome assembly. *RNA*, 8 (8), 997-1010.

Sakharkar MK, Chow VT, & Kanguane P. (2004). Distribution of exons and introns in the human genome. *In Silico Biol* 4 (4), 387-393.

Sander B, Golas MM, Makarov EM, Brahm H, Kastner B, Luhrmann R, & Stark H. (2006). Organization of the core spliceosomal components U5 snRNA loop 1 and the U4/U6 di-snRNP within U4/U6.U5 tri-snRNP as revealed by electron cryomicroscopy. *Mol Cell*, 24 (2), 267-278.

Schellenberg MJ, Edwards RA, Ritchie DB, Kent OA, Golas MM, Stark H, ... & MacMillan AM. (2006). Crystal structure of a core spliceosomal protein interface. *Proc Natl Acad Sci USA*, 103 (5), 1266-1271.

Schellenberg M, Dul EL, & MacMillan AM. (2011). Structural model of the p14/SF3b155.branch duplex complex. *RNA*, 17, 155-165. doi: 10.1261/rna.2224411

Schellenberg MJ, Wu T, Ritchie DB, Fica S, Staley JP, Atta K, ... & MacMillan AM. (2013). A conformational switch in PRP8 mediates metal ion coordination that promotes pre-mRNA exon ligation. *Nat Struct Mol Biol*, 20 (6), 728-734. doi: 10.1038/nsmb.2556

Scheper GC, Parra JL, Wilson M, van Kollenburg B, Vertegaal AC, Han ZG, & Proud CG. (2003). The N and C termini of the splice variants of the human mitogen-activated protein kinase-interacting kinase Mnk2 determine activity and localization. *Mol Cell Biol*, 23 (16), 5692-5705.

Schmidt C, Gronborg M, Deckert J, Bessonov S, Conrad T, Luhrmann R, & Urlaub H. (2014). Mass spectrometry-based relative quantification of proteins in precatalytic and catalytically active spliceosomes by metabolic labelling (SILAC), chemical labelling (iTRAQ), and label-free spectral count. *RNA*, 20 (3), 406-420. doi: 10.1261/rna.041244.113

Schneider S, Campodonico E, & Schwer B. (2004). Motifs IV and V in the DEAH box splicing factor Prp22 are important for RNA unwinding, and helicase-defective Prp22 mutants are suppressed by Prp8. *J Biol Chem*, 279 (10), 8617-8626.

Schneider C, Agafonov DE, Schmitzova J, Hartmuth K, Fabrizio P, & Luhrmann R. (2015). Dynamic contacts of U2, RES, Cwc25, Prp8 and Prp45 proteins with the pre-mRNA branch-site and 3' splice site during catalytic activation and step 1 catalysis in yeast spliceosomes. *PLoS Genet*, 11 (9), e1005539 doi: 10.1371/journal.pgen.1005539

Schonknecht G, Chen WH, Ternes CM, Barbier GG, Shrestha RP, Stanke M, ... & Weber AP. (2013). Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. *Science*, 339 (6124), 1207-1210. doi: 10.1126/science.1231707

Schreib CC, Bowman EK, Hernandez CA, Lucas AL, Potts CHS, & Maeder C. (2018) Functional and biochemical characterization of Dab1's role in pre-messenger RNA splicing. *J Mol Biol*, 430 (11), 1640-1651. doi: 10.1016/j.jmb.2018.04.027

Schwarz JM, Cooper DN, Scheulke M, & Seelow D. (2014). MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods*, 11 (4), 361-362. doi: 10.1038/nmeth.2890

Schwer B, & Guthrie C. (1991). PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. *Nature*, 349 (6309), 494-499.

Schwer B, & Guthrie C. (1992). A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. *EMBO J*, 11 (13), 5033-5039.

Schwer, B. (2008). A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. *Mol Cell*, 30 (6), 743-754. doi: 10.1016/j.molcel.2008.05.003

Schwer B, & Shuman S. (2014). Structure-function analysis of the Yhc1 subunit of yeast U1 snRNP and genetic interactions of Yhc1 with Mud2, Nam8, Mud1, Tgs1, U1 snRNA, SmD3 and Prp28. *Nucleic Acids Res*, 42 (7), 4697-4611. doi: 10.1093/nar/gku097

Semlow DR, & Staley JP. (2012). Staying on message: ensuring fidelity in pre-mRNA splicing. *Trends Biochem Sci*, 37 (3), 263-267. doi: 10.1016/j.tibs.2012.04.001

Senapathy P, Shapiro MB, & Harris NL. (1990). Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. *Methods Enzymol* 183, 252-278.

Siliciano PG, & Guthrie C. (1988). 5' splice site selection in yeast: genetic alterations in base-pairing with U1 reveal additional requirements. *Genes Dev*, 2 (10), 1258-1267.

Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res*, 40 (web server issue), W452-W457. doi:10.1093/nar/gks539

Small EC, Leggett SR, Winans AA, & Staley JP. (2006). The EF-G-like GTPase Snu114p regulates spliceosome dynamics mediated by Brr2p, a DExD/H box ATPase. *Mol Cell*, 23 (3), 389-399.

Smith CW, & Valcarcel J. (2000). Alternative pre-mRNA splicing: The logic of combinatorial control. *Trends Biochem Sci*, 25 (8), 381-388.

Smith DJ, Query CC, & Konarska MM. (2008). "Nought may endure but mutability": spliceosome dynamics and the regulation of splicing. *Mol Cell*, 30 (6), 657-666. doi: 10.1016/j.molcel.2008.04.013

Spieth J, Lawson D, Davis P, Williams G, & Howe K. (2014). Overview of gene structure on *C. elegans*. *WormBook*, ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.7.1, <http://www.wormbook.org>.

Spingola M, Grate L, Haussler D, & Ares M Jr. (1999). Genome-wide bioinformatics and molecular analysis of introns in *Saccharomyces cerevisiae*. *RNA*, 5 (2), 221-234.

Staley JP, & Guthrie C. (1999). An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. *Mol Cell*, 3 (1), 55-64.

Stanojevic D, & Verdine GL. (1995). Deconstruction of GCN4/GCRE into a monomeric peptide-DNA complex. *Nat Struct Biol*, 2 (6), 450-457.

Stark MR, Dunn EA, Dunn WS, Grisdale CJ, Daniele AR, Halstead MR, ... Rader SD. (2015). Dramatically reduced spliceosome in *Cyanidioschyzon merolae*. *Proc Natl Acad Sci USA*, 112 (11), E1191-E1120. doi: 10.1073/pnas.1416879112

Steckelberg AL, Altmueller J, Dietrich C, & Gehring NH. (2015). CWC22-dependent pre-mRNA splicing and eIF4A3 binding enables global deposition of

exon junction complexes. *Nucleic Acids Res*, 43 (9), 4687-4700. doi: 10.1093/nar/gkv320

Steitz TA, & Steitz JA. (1993). A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci USA*, 90 (14), 6498-6502.

Stevens SW, & Abelson J. (1999). Purification of the yeast U4/U6.U5 small nuclear ribonucleoprotein particle and identification of its proteins. *Proc Natl Acad Sci USA*, 96 (13), 7226-7231.

Su YL, Chen HC, Tsai RT, Lin PC, & Cheng SC. (2018). Cwc23 is a component of the NTR complex and functions to stabilize Ntr1 and facilitate disassembly of spliceosome intermediates. *Nucleic Acids Res*, 46 (7), 3764-3773. doi: 10.1093/nar/gky052

Suryadi J, Tran EJ, Maxwell ES, & Brown BA 2nd. (2005). The crystal structure of the *Methanocaldococcus jannaschii* multifunctional L7Ae RNA-binding protein reveals an induced-fit interaction with the box C/D RNAs. *Biochemistry*, 44 (28), 6957-6972.

Tarn WY, & Steitz JA. (1994). SR proteins can compensate for the loss of U1 snRNP functions in vitro. *Genes Dev*, 8 (22), 2704-2717.

Tavtigian SV, Deffenbaugh AM, Yin L, Judkins T, School T, Samollow PB, ... & Thomas A. (2006). Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *J Med Genet*, 43 (4), 295-305.

Toor N, Keating KS, Taylor SD, & Pyle AM. (2008). Crystal structure of a self-spliced group II intron. *Science*, 320 (5872), 77-82. doi: 10.1126/science.1153803

Tran HJ, Allen MD, Lowe J, & Bycroft M. (2003). Structure of the Jab1/MPN domain and its implications for proteasome function. *Biochemistry*, 42 (39), 11460-11465.

Tsai RT, Fu RH, Yeh FL, Tseng CK, Lin YC, Huang YH, & Cheng SC. (2005). Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2. *Genes Dev*, 19 (24), 2991-3003.

Tsai RT, Tseng CK, Lee PJ, Chen HC, Fu RH, Chang KJ, .. & Cheng SC. (2007). Dynamic interactions of the Ntr1-Ntr2 with Prp43 and with U5 govern the recruitment of Prp43 to mediate spliceosome disassembly. *Mol Cell Biol*, 27 (23), 8027-8037.

- Tseng CK, Liu HL, & Cheng SC. (2011). DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. *RNA*, 17 (1), 145-154. doi: 10.1261/rna.2459611
- Tseng CK, Chung CS, Chen HC, & Cheng SC. (2017). A central role of Cwc25 in spliceosome dynamics during the catalytic phase of pre-mRNA splicing. *RNA*, 23 (4), 546-556. doi: 10.1261/rna.059204.116
- Turunen JJ, Niemela EH, Verma B, & Frilander MJ. (2013). The significant other: splicing by the minor spliceosome. *Wiley Interdiscip Rev RNA*, 4 (1), 61-76. doi: 10.1002/wrna.1141
- Umen JG, & Guthrie C. (1995a). The second catalytic step of pre-mRNA splicing. *RNA*, 1 (9), 869-885.
- Umen JG, & Guthrie C. (1995b). A novel role for a U5 snRNP protein in 3' splice site selection. *Genes Dev*, 9 (7), 855-868.
- Umen JG, & Guthrie C. (1996). Mutagenesis of the yeast gene PRP8 reveals domains governing the specificity and fidelity of 3' splice site selection. *Genetics*, 143 (2), 723-739.
- Urbanski LM, Leclair N, & Anczukow O. (2018). Alternative-splicing defects in cancer: Splicing regulators and their downstream targets, guiding the way to novel cancer therapeutics. *Wiley Interdiscip Rev RNA*, 9 (4), e1476. doi: 10.1002/wrna.1476
- Van de Water NS, Tan T, May S, Browett PJ, & Harper P. (2004). Factor IX polypyrimidine tract mutation analysis using mRNA from peripheral blood leukocytes. *J Thromb Haemost*, 2 (11), 2073-2075.
- Van Roon AM, Oubridge C, Obayahsi E, Sposito B, Newman AJ, Seraphin B, & Nagai, K. (2017). Crystal structure of U2 snRNP SF3b components: Hsh49p in complex with Cus1p-binding domain. *RNA*, 23 (6), 968-981. doi: 10.1261/rna.059378.116
- Vanacova S, Yan W, Carlton JM, & Johnson PJ. (2005). Spliceosomal introns in the deep-branching eukaryote *Trichomonas vaginalis*. *Proc Natl Acad Sci USA*, 102 (12), 4430-4435.
- Vidovic I, Nottrott S, Hartmuth K, Luhrmann R, & Ficner R. (2000). Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol Cell*, 6 (6), 1331-1342.

- Villa T, & Guthrie C. (2005). The Isylp component of the NineTeen complex interacts with the ATPase Prp16p to regulate the fidelity of pre-mRNA splicing. *Genes Dev*, 19 (16), 1894-1904.
- Vithana EN, Abu-Safieh L, Allen MJ, Carey A, Papaioannou M, Chakarova C, ... & Bhattacharya SS. (2001). A human homolog of yeast pre-mRNA splicing gene, PRP31, underlies autosomal dominant retinitis pigmentosa on chromosome 19q13.4 (RP11). *Mol Cell*, 8 (2), 375-381.
- Vosseberg J, & Snel B. (2017). Domestication of self-splicing introns during eukaryogenesis: the rise of the complex spliceosomal machinery. *Biol Direct*, 12 (1), 30. doi: 10.1186/s13062-017-0201-6
- Wan R, Yan C, Bai R, Huang G, & Shi Y. (2016). Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution. *Science*, 353 (6302), 895-904. doi: 10.1126/science.aag2235
- Wan L, & Dreyfuss G. (2017). Splicing-correcting therapy for SMA. *Cell*, 170 (1), 5. doi: 10.1016/j.cell.2017.06.028
- Wan R, Yan C, Bai R, Lei J, & Shi Y. (2017). Structure of an intron lariat spliceosome from *Saccharomyces cerevisiae*. *Cell*, 171 (1), 120-132. doi: 10.1016/j.cell.2017.08.029
- Wang C, Chua K, Seghezzi W, Lees E, Gozani O, & Reed R. (1998). Phosphorylation of spliceosomal protein SAP 155 coupled with splicing catalysis. *Genes Dev*, 12 (10), 1409-1414.
- Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, ... & Burge CB. (2008). Alternative isoform regulation in human transcriptomes. *Nature*, 456 (7221), 470-476. doi: 10.1038/nature07509
- Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, ... & Wu CJ. (2011). SF3b1 and other novel cancer genes in chronic lymphocytic leukemia. *N Engl J Med*, 365 (26), 2497-2506. doi: 10.1056/NEJMoa1109016
- Ward AJ, & Cooper TA. (2010). The pathobiology of splicing. *J Pathol*, 220 (2), 152-163. doi: 10.1002/path.2649
- Warkocki Z, Odenwalder P, Schmitzova J, Platzmann F, Stark H, Urlaub H, ... & Luhrmann R. (2009). Reconstitution of both steps of *Saccharomyces cerevisiae* splicing with purified spliceosomal components. *Nat Struct Mol Biol*, 16 (12), 1237-1243. doi: 10.1038/nsmb.1729

- Watanabe H, Shionyu M, Kimura T, Kimata K, Watanabe H. (2007). Splicing factor 3b subunit 4 binds BMPR-IA and inhibits osteochondral cell differentiation. *J Biol Chem*, 282 (28), 20728-20738.
- Watkins NJ, & Bohnsack MT. (2012). The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA. *Wiley Interdiscip Rev RNA*, 3 (3), 397-414. doi: 10.1002/wrna.117
- Weber G, Cristao VF, de L Alves F, Santos KF, Holton N, Rappsilber J, ... & Wahl MC. (2011). Mechanism for Aar2p function as a U5 snRNP assembly factor. *Genes Dev*, 25 (15), 1601-1612. doi: 10.1101/gad.635911
- Wells SE, Neville M, Haynes M, Wang J, Igel H, & Ares M Jr. (1996). CUS1, a suppressor of cold-sensitive U2 snRNA mutations, is a novel yeast splicing factor homologous to human SAP 145. *Genes Dev*, 10 (2), 220-232.
- Wiest DK, O'Day CL, & Abelson J. (1996). In vitro studies of the Prp9.Prp11.Prp21 complex indicate a pathway for U2 small nuclear ribonucleoprotein activation. *J Biol Chem*, 271 (52), 33268-33276.
- Wilkinson ME, Fica SM, Galej WP, Norman CM, Newman AJ, & Nagai K. (2017). Postcatalytic spliceosome structure reveals mechanism of 3'-splice site selection. *Science*, 358 (6368), 1283-1288. doi: 10.1126/science.aar3729
- Will CL, Rumpler S, Klein Gunnewiek J, van Venrooij WJ, & Luhrmann R. (1996). In vitro reconstitution of mammalian U1 snRNPs active in splicing: the U1-C protein enhances the formation of early (E) spliceosomal complexes. *Nucleic Acids Res*, 24 (23), 4614-4623.
- Will CL, & Luhrmann R. (2011). Spliceosome structure and function. *Cold Spring Harb Perspect Biol*, 3:a003707. doi: 10.1101/cshperspect.a003707
- Wood V, Gwilliam R, Rajandream MA, Lyne M, Stewart A, Sgouros J, ... & Nurse P. (2002). The genome sequence of *Schizosaccharomyces pombe*. *Nature*, 415, (6874), 871-880.
- Woodward LA, Mabin JW, Gangras P, & Singh G. (2017). The exon junction complex: a lifelong guardian of mRNA fate. *Wiley Interdiscip Rev RNA*, 8 (3). doi: 10.1002/wrna.1411
- Wu S, Li YL, Cheng NY, Wang C, Dong EL, Lu YQ, ... & Chen WJ. (2018). C.835-5>T variant in SMN1 gene causes transcript exclusion of exon 7 and spinal muscular atrophy. *J Mol Neurosci*, 65 (2), 196-202. doi: 10.1007/s12031-018-1079-1

- Wysoczanski P, & Zweckstetter M. (2016). Retention and splicing complex (RES) – the importance of cooperativity. *RNA Biol*, 13 (2), 128-133. doi: 10.1080/15476286.2015.1096484
- Xerri L, Parc P, Brousset P, Schlaifer D, Hassoun J, Reed JC, ... & Birnbaum D. (1996). Predominant expression of the long isoform of Bcl-x (Bcl-xL) in human lymphomas. *Br J Haematol*, 92 (4), 900-906.
- Xu W, Yang X, Hu X, Li S. (2014). Fifty-four novel mutations in the NF1 gene and integrated analyses of the mutations that modulate splicing. *Int J Mol Med*, 34 (1), 53-60. doi: 10.3892/ijmm.2014.1756
- Yan C, Hang J, Wan R, Huang M, Wong CC, & Shi Y. (2015). Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science*, 349 (6253), 1182-1191. doi: 10.1126/science.aac7629
- Yan C, Wan R, Bai R, Huang G, & Shi Y. (2016). Structure of a yeast activated spliceosome at 3.5 Å resolution. *Science*, 353 (6302), 904-911. doi: 10.1126/science.aag0291
- Yan C, Wan R, Bai R, Huang G, & Shi, Y. (2017). Structure of a yeast step II catalytically activated spliceosome. *Science*, 355 (6321), 149-155. doi: 10.1126/science.aak9979
- Yang W, Hendrickson WA, Crouch RJ, & Satow Y. (1990). Structure of ribonuclease H phased at 2 Å resolution by MAD analysis of the selenomethionyl protein. *Science*, 249 (4975), 1398-1405.
- Yang J, Zimmerly S, Perlman PS, & Lambowitz AM. (1996). Efficient integration of an intron RNA into double-stranded DNA by reverse splicing. *Nature*, 381 (6580), 332-335.
- Yang K, Zhang L, Xu T, Heroux A, & Zhao R. (2008). Crystal structure of the beta-finger domain of Prp8 reveals analogy to ribosomal proteins. *Proc Natl Acad Sci USA*, 105 (37), 13817-13822. doi: 10.1073/pnas.0805960105
- Yang F, Wang XY, Zhang ZM, Pu J, Fan YJ, Zhou J, ... & Xu YZ. (2013). Splicing proofreading at 5' splice sites by ATPase Prp28p. *Nucleic Acids Res*, 41 (8), 4660-4670. doi: 10.1093/nar/gkt149
- Zhang MQ. (1998). Statistical features of human exons and their flanking regions. *Hum Mol Genet*, 7 (5), 919-932.

Zhang D, & Rosbash M. (1999). Identification of eight proteins that cross-link to pre-mRNA in the yeast commitment complex. *Genes Dev*, 13 (5), 581-592.

Zhao C, & Pyle AM. (2016). Crystal structures of a group II intron maturase reveal a missing link in spliceosome evolution. *Nat Struct Mol Biol*, 23 (6), 558-565. doi: 10.1038/nsmb.3224

Zhao C, & Pyle AM. (2017). The group II intron maturase: a reverse transcriptase and splicing factor go hand in hand. *Curr Opin Struct Biol*, 47, 30-39. doi: 10.1016/j.sbi.2017.05.002

Zhou Z, & Fu XD. (2013). Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma*, 122 (3), 191-207. doi: 10.1007/s00412-013-0407-z

Zhou L, Hang J, Zhou Y, Wan R, Lu G, ... Shi, Y. (2014). Crystal structures of the Lsm complex bound to the 3' end sequence of U6 small nuclear RNA. *Nature*, 506 (7486), 116-120. doi: 10.1038/nature12803

Zorio DA, Cheng NN, Blumenthal T, & Speith J. (1994). Operons as a common form of chromosomal organization in *C. elegans*. *Nature*, 372 (6503), 270-272.