Prp8 in a Reduced Spliceosome Lacks a Conserved Toggle that Correlates with Splicing Complexity across Diverse Taxa

E.L. Garside [1], T.A. Whelan [2], M.R. Stark [3], S.D. Rader [3], N.M. Fast [2] and A.M. MacMillan [1]

*1 - **Department of Biochemistry,** University of Alberta, Edmonton, AB, Canada T6G 2H7*
*2 - **Biodiversity Research Center and Department of Botany,** University of British Columbia, Vancouver, BC, Canada V6T 1Z4*
*3 - **Department of Chemistry,** University of Northern British Columbia, Prince George, BC, Canada V2N 4Z9*

***Correspondence to A.M. MacMillan:** Department of Biochemistry, 474 Medical Sciences Bldg., University of Alberta, Edmonton, Alberta, Canada T6G 2H7. andrew.macmillan@ualberta.ca.*
https://doi.org/10.1016/j.jmb.2019.04.047
Edited by Anna Pyle

Abstract

Conformational rearrangements are critical to regulating the assembly and activity of the spliceosome. The spliceosomal protein Prp8 undergoes multiple conformational changes during the course of spliceosome assembly, activation, and catalytic activity. Most of these rearrangements of Prp8 involve the disposition of the C-terminal Jab-MPN and RH domains with respect to the core of Prp8. Here we use x-ray structural analysis to show that a previously characterized and highly conserved β-hairpin structure in the RH domain that acts as a toggle in the spliceosome is absent in Prp8 from the reduced spliceosome of the red alga *Cyanidioschyzon merolae*. Using comparative sequence analysis, we show that the presence or absence of this hairpin corresponds to the presence or absence of protein partners that interact with this hairpin as observed by x-ray and cryo-EM studies. The presence of the toggle correlates with increasing intron number suggesting a role in the regulation of splicing.

Introduction

Eukaryotes contain a split-gene structure in which coding exon sequences are interrupted by non-coding intron sequences. Intron excision and exon ligation take place through two sequential transesterifications catalyzed by the spliceosome (reviewed in Ref. [1]). This large RNA–protein assembly consists of the U1, U2, U4/U6, and U5 small nuclear ribonucleoprotein particles (snRNPs), each containing a unique snRNA and associated proteins, that assemble on the pre-mRNA substrate. In the first step of splicing, a conserved adenosine within the intron, selected by virtue of base pairing between the U2 snRNA and intron branch sequence, performs a nucleophilic displacement at the 5′ splice site yielding the free 5′ exon and lariat intermediate. In the second step, attack of the 5′ exon at the 3′ splice site produces the ligated exons and lariat intron as products.

Spliceosome assembly is a multi-step process that proceeds, following snRNP biogenesis, through discrete complexes and involves significant structural and conformational rearrangements [2,3]. Recruitment of the U4/U6•U5 tri-snRNP to a complex containing pre-mRNA associated with the U1 and U2 snRNPs involves the displacement of the former at the 5′ splice site by the U6 snRNA to yield the B complex. Unwinding of a U4/U6 snRNA duplex within the B complex, mediated by the Brr2 helicase, allows formation of a U6 internal stem loop and U2/U6 snRNA structure to produce the pre-catalytic activated spliceosome referred to as B^{act}. Together these U6 and U2/U6 snRNA structures comprise the active site of the spliceosome that catalyzes the two splicing transesterifications. Rearrangement leads to the B* complex, which catalyzes the first step and results in the C complex. Further remodeling yields the C* complex where the second step occurs followed by a transition to the post-catalytic P complex [4–6].

The highly conserved U5 snRNP protein Prp8 (61% identity between *Saccharomyces cerevisiae* and humans [7]) plays key roles in the regulation of

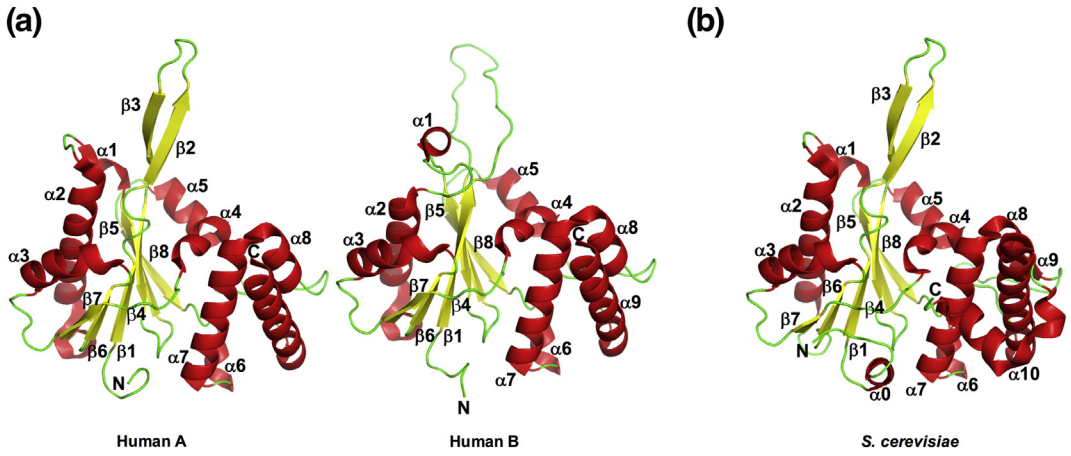

Fig. 1. X-ray structural analysis of Prp8 RH domain. (A) Ribbon diagram of the human RH domain (PDB 4JK7) showing closed (left) and open (right) conformations. (B) Ribbon diagram of the *S. cerevisiae* (PDB 3E9O) RH domain. α-Helices and β-strands are colored red and yellow, respectively.

spliceosome assembly and transitions through the two catalytic steps while remaining intimately associated with the spliceosomal core [8–10]. Prp8 contains reverse transcriptase-like, endonuclease-like, and RNase H-like (RH) domains [11] believed to be derived from an ancestral group II intron maturase [12,13] as well as a C-terminal Jab-MPN domain that regulates the activity of Brr2 [14–16].

A large number of mutant *prp8* alleles related to spliceosome activation and the two transesterification steps have been characterized in yeast (reviewed in Ref. [17]). A clustering of these alleles led us and others to structurally and functionally characterize the domain of Prp8 corresponding to the RH fold ([18–20]; Fig. 1A, B). Sequence analysis was unable to predict the RH fold due to a 17-amino-acid insertion between the first and second β-strands of the RH domain. Crystallographically, we observed two conformations of the RH domain: in the closed conformation, the insertion corresponds to an anti-parallel β-hairpin; while in the second, open conformation, the hairpin is disrupted to yield a loop translated ~45° back with respect to the closed structure ([21]; Fig. 1A). A hydrated magnesium ion observed in the second conformation, bound in a site conserved from RNase H, possibly acts to stabilize the open structure.

X-ray structural analyses of Prp8 RH domain mutants affecting either step of splicing, or U4/U6 snRNA unwinding, as characterized in yeast and yeast splicing extracts, revealed a stabilization or destabilization of one of the two RH domain conformations [21]. This transition between two conformations corresponded to favoring the first step of splicing (at the expense of the second) or the second step of splicing (at the expense of the first), as shown in both *in vitro* and yeast reporter splicing assays, suggesting a switch mechanism to regulate splicing. Recently, it has been proposed that the two conformations of the RH domain represent a toggle mechanism related to proofreading, where the closed conformation favors accurate but inefficient splicing and the open conformation favors inaccurate but efficient splicing [22]. This toggle may relate to different conformations of the U2 snRNA, involving stem IIc (catalytically active) and the alternate, mutually exclusive stem IIa (catalytically inactive; [23]). Deletion of the 17 amino acids comprising the RH domain hairpin is lethal in *S. cerevisiae* [20,22].

The best-characterized spliceosomes are those from humans and budding yeast. Although these spliceosomes are highly conserved, there are significant differences between them. In order to more fully understand the function of the RH domain of Prp8, we have examined it structurally in an organism distantly related to budding yeast and humans and also carried out a comparative bioinformatic analysis across diverse eukaryotic taxa. As a starting point, we chose Prp8 from the red alga *Cyanidioschyzon merolae* due to evidence of a considerably less complex spliceosome in this organism compared to either humans or yeast [24]. The genome of *C. merolae* contains only 27 introns in a genome with 4803 genes [25]. A computational and biochemical analysis revealed the presence of a greatly reduced spliceosome containing only 40 core proteins and completely lacking the U1 snRNA and its associated proteins [24]. Here we report our x-ray structural analysis of the RH domain of *C. merolae* Prp8 (CmPrp8) showing that while the RNase H fold is maintained, the highly conserved β-hairpin is absent. We have extended this analysis across diverse taxa showing that the presence or absence of the hairpin corresponds with intron number in agreement with the observation that among red algal genomes spliceosome complexity is associated with a greater number of introns [26]. We also show that

Fig. 2. Curated alignment of Prp8 β-hairpin insert from 25 organisms. The 17-amino-acid insertion is highly conserved (bold, colored by similarity to the human sequence). *C. merolae* and 2 microsporidia (*E. cuniculi,* and *N. bombycis*) lack the insert completely, and *G. intestinalis* and the kinetoplastid *T. cruzi* lack a conserved sequence. Two conserved Asp residues are highlighted (bold, black). Taxa are grouped by evolutionary relationships: purple, Obazoa: microsporidia; orange, excavata; black, cryptista; maroon, amoebozoa; green, archaeplastida; yellow, SAR (Stramenopiles, Alveolates, and Rhizaria); blue, Obazoa: animals; red, Obazoa: fungi. Alignment begins at the start of the RH domain.

the loss of this hairpin corresponds with the predicted absence of factors that associate with the RH hairpin that act at multiple steps of spliceosome assembly or progression supporting its role as a regulatory feature in complex spliceosomes.

Results

Identification of the Prp8 RH domain of *C. merolae* and comparison across taxa

Protein components of the *C. merolae* (CM) splicing machinery were previously identified by searching the National Center for Biotechnology Information database using sequences of two or more species to retrieve the *C. merolae* homolog [24]. This analysis confirmed the existence of a CM Prp8 ortholog. In examining the alignment of CM and human Prp8 sequences, we noted a gap of 13 amino acids corresponding to the β-hairpin insertion in the human and yeast RH domains. To determine if this loss was unique to CM, we sought to compare Prp8 sequences across diverse eukaryotic taxa. As a core component of the spliceosome, Prp8 is present in every known intron-containing taxon. Eukaryotic diversity separates phylogenetically into major groups [27], although thorough analyses of the spliceosome have occurred in relatively few eukaryotic taxa [28]. We analyzed Prp8 sequences from 36 taxa that represent the major groups of eukaryotic diversity discussed in Burki *et al.* [27]; see Materials and Methods; curated selection of taxa in Fig. 2; all analyzed sequences in Supplemental Fig. 1). We also increased sampling to represent taxa known to have reduced genomes and low intron densities to assess whether the amino acid gap seen in CM is a shared feature of lineages that have undergone spliceosomal reduction. Additional fungal taxa were also included (see Materials and Methods). Although all of these exhibited the expected high degree of sequence conservation, the apparent deletion of the β-hairpin sequence was observed in two microsporidian taxa (*Encephalitozoon cuniculi* and *Nosema bombycis*) with low intron densities (Table 1). In addition, *Giardia intestinalis* and the kinetoplastid *Trypanosoma cruzi* showed a highly divergent insertion that is not predicted to form a hairpin. Like *C. merolae*, these organisms have few annotated introns

Table 1. Comparative genome summary of taxa lacking the conserved hairpin along with *S. cerevisiae* and human[a]

Species	Genome size (Mb)	Gene number	Introns
Cyanidioschyzon merolae	16.5	4803	27
Encephalitozoon cuniculi	2.3	1981	37
Nosema bombycis	15.7	4488	167
Trypanosoma cruzi	41.5	19,607	3
Giardia intestinalis	11.2	6502	2
Trichomonas vaginalis	176.4	59,679	66
Saccharomyces cerevisiae	12.1	5983	253
Homo sapiens	2939.6	29,399	272,667

[a] Genome size, gene number, and introns in taxa that lack the hairpin interaction as well as *S. cerevisiae* and *H. sapiens*. Intron number estimated by extracting exon and coding sequences whose sequence overlaps with annotated mRNA, then subtracting CDS number from exon number,

(Table 1); however, they have been shown to undergo extensive trans-splicing, where the acceptor and donor RNAs are encoded on distinct pre-mRNAs and are brought together in the spliceosome [29–33].

Structural analysis of the Prp8 RH domain of *C. merolae*

In order to investigate the role of the RH domain of Prp8 during splicing, we decided to analyze the structure of the domain in the minimal spliceosome represented by *C. merolae*. We amplified and cloned a sequence representing the CmPRP8 RH domain from a full-length CmPrp8 clone derived from genomic DNA. We expressed, purified, and crystallized CmPRP8 RH and determined its structure to 2.75 Å resolution using x-ray diffraction data (Fig. 3A; Table 2). The crystals of CmPRP8 RH contain two monomers in the asymmetric unit representing essentially identical (rmsd 0.7 Å) conformations of the RH domain. The structure of the CM protein, as observed in the human [19,21] and yeast homologs [18,20] is bipartite containing an N-terminal RNase H fold (CmPrp8 1852–1989) as well as a C-terminal five helix cluster (CmPrp8 1993–2059; Figs. 1 and 3A). As predicted by an initial sequence analysis, the CM protein lacks the 17-amino-acid insertion interrupting the RH fold within human and *S. cerevisiae* Prp8. Instead, it is replaced by a well-ordered type II β-turn (Figs. 3B, 4A). Both monomers in the crystal lattice represent the closed conformation of the RH domain (rmsd 3.3 Å with respect to the human closed conformation lacking the insertion). The residues that coordinate a metal ion in the open conformation of the human protein (D1781, D1782, T1864, and Q1894; Fig. 4B) are partially conserved in CmRH (D1862, D1863, A1932, and Q1964, respectively). The two consecutive aspartates are just N-terminal to the β-hairpin insertion and almost universally conserved (Fig. 2; cmAsp1862/cmAsp1863 corresponding to hAsp1781/hAsp1782 and scAsp1853/scAsp1854). Most species also feature a conserved threonine following the two

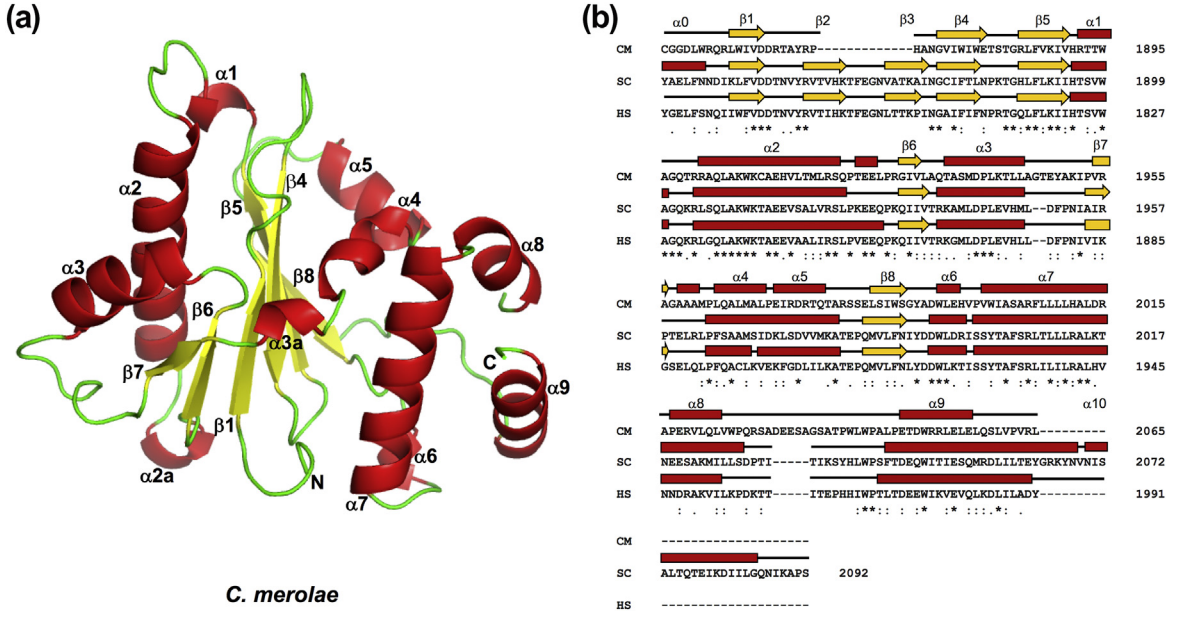

Fig. 3. X-ray structural analysis of *C. merolae* Prp8 RH domain. (A) Ribbon diagram of the CM RH domain. α-Helices and β-strands are colored red and yellow, respectively. (B) Sequence alignment and secondary structure (α-helices as red rectangles and β-strands as yellow arrows) of the *C. merolae* (CM), *S. cerevisiae* (SC), and human (HS) Prp8 RH domain.

Table 2. Data collection and refinement statistics[a]

cmPrp8 RH	
Data collection	
Space group	*P*1211
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)
$I/\sigma I$	9.7 (5.1)
Completeness (%)	99.54 (99.39)
Redundancy	3.3 (3.3)
Refinement	
Resolution (Å)	36.42–2.75 (2.85–2.75)
Unique reflections	29,496 (2313)
R_{work}/R_{free}	0.2529/0.2911 (0.3137/0.3845)
No. atoms	
Protein	3123
Ligand	8
B-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

[a] 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 with chain A of PDB 4JK7 as the search model.

aspartate residues. However, in CM (and three of the four organisms predicted to lack the hairpin), this position is occupied by a basic residue (arginine or lysine; Fig. 2). In CM, the side chain of this arginine (cmArg1864) hydrogen bonds with cmAsp1862 (hAsp1781; Fig. 4C). In humans, a different residue, hArg1865, hydrogen bonds with hAsp1782 in both open and closed conformations; a corresponding interaction is observed in the *S. cerevisiae* and *Schizosaccharomyces pombe* (closed conformation) structures. The positively charged terminus of the cmArg1864 side-chain occupies the same position as the Mg^{2+} observed in the open conformation of the human RH domain (Fig. 4C). However, the positively charged terminus of the hArg1865 side chain is 4 and 7 Å away from that site in the closed and open conformations, respectively. Thus, cmArg1864 may serve as a surrogate for the bound magnesium ion.

Analysis of hairpin interacting factors

Because deletion of the RH domain hairpin is lethal in *S. cerevisiae*, we examined x-ray and cryo-EM structures of yeast and human Prp8 within pre-spliceosomal and spliceosomal complexes. We noted three distinct sets of interactions between spliceosomal proteins and the RH domain hairpin as well as a role for the hairpin in separating spliceosomal RNA elements in the post-catalytic spliceosome.

In the x-ray structure of a large fragment of yeast Prp8, a β-strand of the C-terminal region of the U5 snRNP-associated assembly factor Aar2 inserts into Prp8 forming a continuous β-sheet spanning the hairpin, Aar2, and the Jab-MPN domain ([11,34]; Fig. 5A). In the yeast tri-snRNP, as well as the yeast and human B complex structures, a portion of the protein Snu66 forms the third strand of a β-sheet through its interactions with the RH hairpin (Fig. 5B; [35–38]). Snu66 dissociates from the spliceosome with the U4 snRNP and is not seen in the subsequent B^{act}, C, or other complexes.

In the structure of the post-second step C* complex, the RH β-hairpin is inserted between the U2/branch sequence duplex on one side and the duplex formed by U6 and the 5′ end of the intron on the other ([39–42]; Fig. 5C). The nearby catalytic core of the U6 snRNA is oriented into the active site of the spliceosome, which is nestled in the core of Prp8.

Finally, in the intron lariat spliceosome complex purified from *S. pombe*, a helical bundle from the C terminus of the Cwf19 protein bridges the RH domain

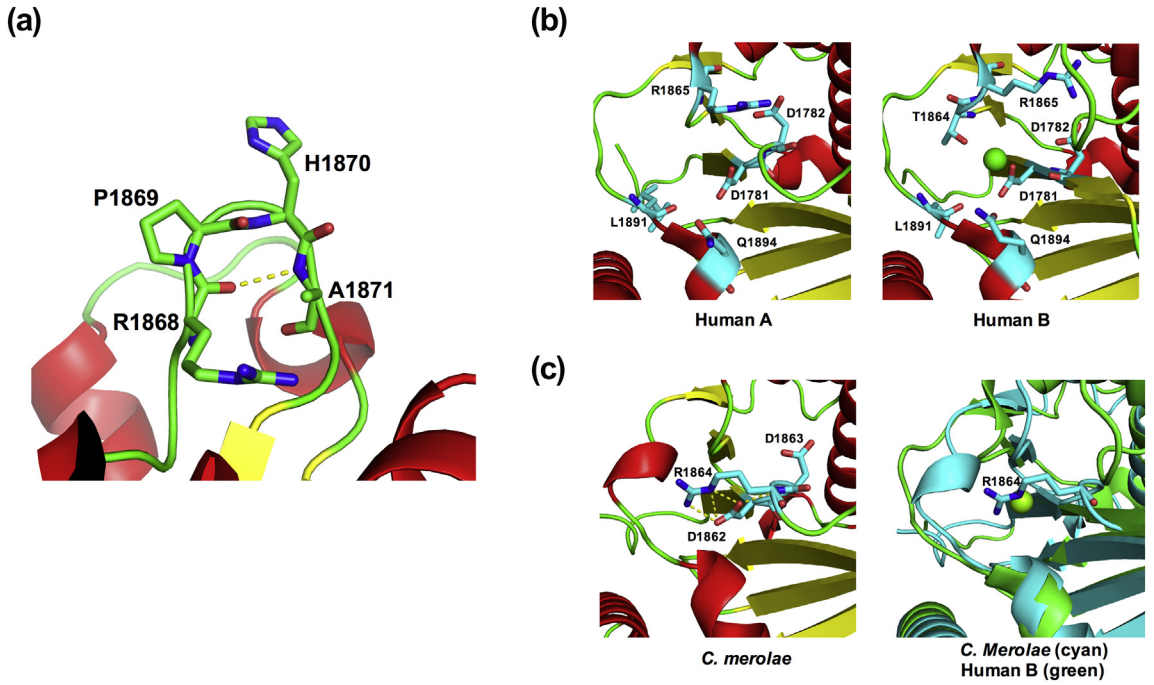

Fig. 4. Details of the *C. merolae* Prp8 RH domain. (A) Detail of RH domain structure highlighting replacement of the RH hairpin insert with a type II β-turn. (B) Comparison of closed and open conformations of human RH domain. Arg1865 interacts with Asp1782 blocking the metal binding site in the closed confirmation (left). Displacement of Arg1865 with Mg^{2+} bound in the open conformation (right). (C) Interaction between Arg1864 and Asp1862 in the *C. merolae* RH domain (left). Alignment of the *C. merolae* and human open structure superimposes *C. merolae* Arg1864 with Mg^{2+} bound in the human structure (right).

to the catalytic core of the spliceosome including an interaction with the RH hairpin ([43]; Fig. 5D). This region of the protein is conserved in the human spliceosome, and the human ortholog of Cwf19 has been reported in mass spectrometric characterizations of human spliceosomal complexes [44].

Conservation of Prp8 and hairpin interacting factors

We performed a search through the taxa we had analyzed with respect to the conservation of the RH domain in order to characterize the conservation of spliceosomal proteins interacting with the hairpin. Using two conserved regions in *S. cerevisiae* Aar2, one in the N-terminal domain (aa 4–46) and one in the C-terminal domain (aa 220–250; Supplemental Fig. 2), we failed to identify an Aar2 homolog in the six taxa where the RH hairpin is missing or the sequence is not conserved. Of the 28 taxa with a conserved hairpin, we failed to find Aar2 candidates in four (*Amanita muscaria*, *Chondrus crispus*, *Entamoeba histolytica*, *Trichomonas vaginalis*; Table 3). This latter observation must be qualified by the relatively low conservation of Aar2 across eukaryotes. Nevertheless, there is a correlation between the presence of the hairpin and of an Aar2 homolog.

We identified Snu66 candidates in 18 of 28 taxa (Table 3; Supplemental Fig. 3). Despite the divergence of Snu66, there appear to be two regions of conservation: the first is near the C-terminus (*S. cerevisiae* aa 527–570), and the other is near the N-terminus (*S. cerevisiae* aa 38–55). This predicted presence of Snu66 again corresponds to the presence of the RH hairpin.

While the *S. pombe* splicing factor associated with Prp8, Cwf19, has been proposed to be Drn1 in *S. cerevisiae,* our BLAST searches did not identify it as a homolog of Cwf19 ([45]; Supplementary Data). We were able to identify candidate Cwf19 homologs in 17 taxa, all of which also have an Aar2 candidate, except *A. muscaria*. Cwf19 is divergent in sequence, with conservation near the C-terminus (*S. pombe* aa 426–582; Supplemental Fig. 4). Once again, the predicted presence of a Cwf19 homolog correlates well with the presence of the hairpin within the RH domain.

Discussion

Recent studies have provided detailed static representations of multiple steps in spliceosome assembly, activation, and catalysis (reviewed in Ref. [46]). Comparison of these spliceosomes on a structural level will yield insights into the mechanism

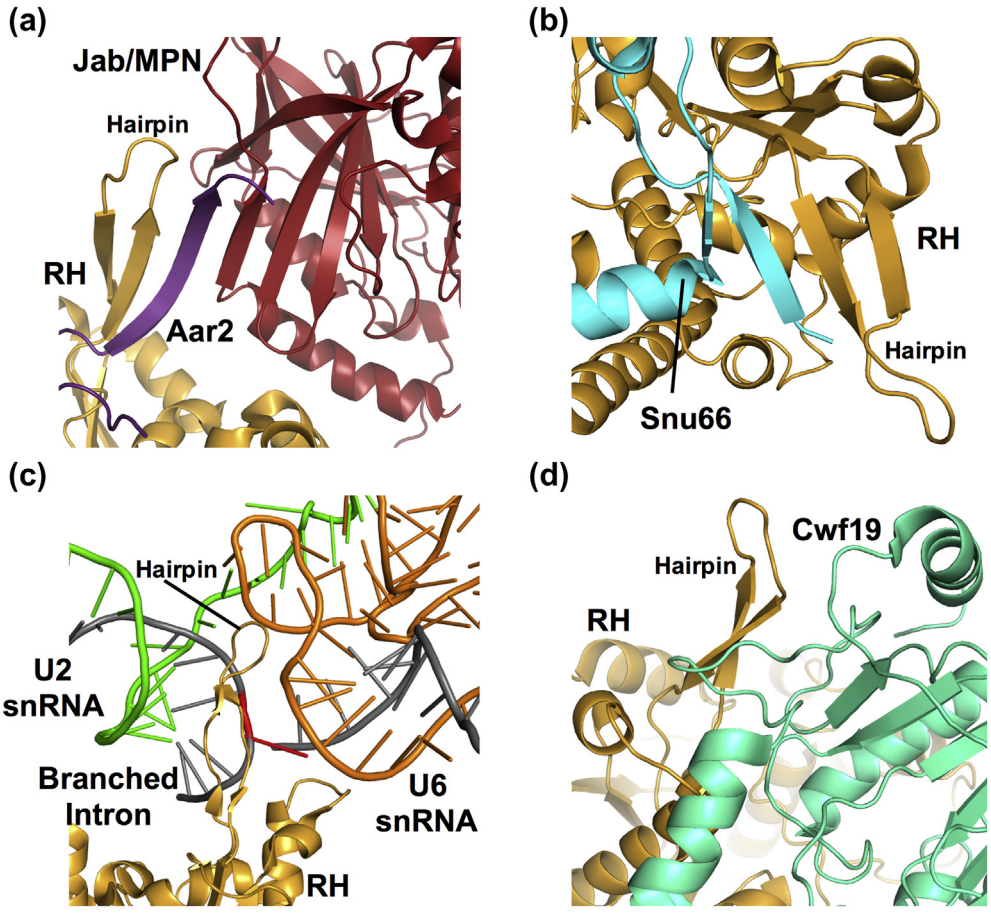

Fig. 5. Structural analysis of hairpin interacting factors of the RH domain β-hairpin. (A) Detail of the extended β-sheet formed by the RH domain hairpin (gold), U5 snRNP assembly factor Aar2 (purple), and the Prp8 Jab/MPN domain (red; PDB 4I43). (B) Detail of β-sheet formed through the interaction of the tri-snRNP protein Snu66 (cyan) and the RH domain hairpin (gold; PDB 5ZWO). (C) RH β-hairpin at the core of the spliceosome. The β-hairpin (light orange) inserts into the minor groove of the intron/U2 snRNA duplex (gray and green, respectively; branch adenosine in red) and separates it from the intron/U6 snRNA duplex (PDB 5WSG). (D). Bridging interaction of Cwf19 (mint) between the RH domain (gold) and the Prp8 core (blue; PDB 3JB9).

of splicing. However, given the high conservation of the spliceosomal machinery, sequence comparison across species is also a useful approach to dissecting spliceosome function in light of these structures. These comparisons should inform our understanding of both constitutive and regulated splicing events.

Prp8 is a highly conserved protein that serves as a scaffold for the snRNA catalytic core of the spliceosome. Structural analyses of Prp8 from a pre-U5 snRNP assembly complex and through the splicing cycle reveal little change in the disposition of the Prp8 core, while the C-terminal Jab-MPN and RH domains undergo significant movement during the spliceosomal cycle. The Jab-MPN domain is connected to the RH domain by 70 amino acids and is not visible in the cryo-EM structures of post-first step spliceosomes suggesting considerable movement [5,6,39–43,47–50]. The RH domain undergoes a localized but significant circular translation through the spliceosomal cycle [46]. The interactions with the RH domain described here in the Prp8/Aar2 complex, the B complex (with Snu 66), the C* complex (with the branch duplex RNA), and the intron lariat spliceosome (with Cwf19) essentially lock the RH domain in a specific conformation and presumably relate to regulation of its orientation in pre-spliceosomal and spliceosomal complexes. These interactions are lacking in the reduced spliceosome of *C. merolae* and other organisms with reduced splicing complexity. Although the spliceosomes of budding yeast and humans are highly conserved overall, there are significant differences in spliceosomal factors including the SF3B component SF3B6 (present in fission but not budding yeast; [51]). The association of the RH domain with regulation of the U4/U6 to U2/U6 transitions [8], the promotion of either the first or second steps [21], and the accuracy

Table 3. Presence of Aar2, Snu66, and Cwf19 across taxa[a,b]

Species	Hairpin	Aar2	Snu66	Cwf19
Cyanidioschyzon merolae	–	–	–	–
Encephalitozoon cuniculi	–	–	–	–
Nosema bombycis	–	–	–	–
Trypanosoma cruzi	Not conserved	–	–	–
Giardia intestinalis	Not conserved	–	–	–
Trichomonas vaginalis	+	–	–	–
Naegleria gruberi	+	+	+	+
Guillardia theta	+	+	–	+
Phaeodactylum tricornutum	+	+	+	+
Phytophthora infestans	+	+	+	+
Paramecium tetraurelia	+	+	+	+
Plasmodium falciparum	+	+	+	+
Chondrus crispus	+	–	+	–
Galdieria sulphuraria	+	+	–	+
Chlamydomonas reinhardtii	+	+	+	+
Arabidopsis thaliana	+	+	+	+
Entamoeba histolytica	+	–	+	–
Dictyostelium discoideum	+	+	+	+
Caenorhabditis elegans	+	+	+	+
Homo sapiens	+	+	+	+
Rozella allomycis	+	+	+	+
Batrachochytrium dendrobatidis	+	+	+	+
Amanita muscaria	+	+	+	+
Schizosaccharomyces pombe	+	+	+	+
Saccharomyces cerevisiae	+	+	+	–

[a] Spliceosomal proteins Aar2, Snu66, and Cwf19 identified with reciprocal BLAST hits to genes from *S. cerevisiae* and *H. sapiens*.
[b] Referenced to curated alignment of Fig. 2; see also Supplementary Figs. 1–4.

of splicing [22] suggest a regulatory role for this domain that appears to be absent in the reduced spliceosome of *C. merolae* and presumably other organisms lacking the RH hairpin and its interacting factors. Splicing of the few introns in *C. merolae* varies considerably in efficiency ranging from 10% to 90% [24]; low efficiency of splicing may be related to the lack of the hairpin toggle associated with a balance between efficiency and accuracy.

Cross-linking and analyses of cDNAs has suggested a possible interaction between U1 snRNA and Prp8, which could be part of the transition in formation of the spliceosomal B complex [52]. This raises the possibility that, in the absence of U1 snRNP, there could be compensating changes in CM Prp8. An issue with identifying additional shared indels in hairpin-lacking taxa is that these taxa are also the most divergent of Prp8 sequences. This makes it even more challenging to find indels because the alignment is often poor for these specific taxa of interest making identifying homologous regions that could be indels almost impossible. The identification of CM Prp8 as lacking the hairpin was mostly made possible by a combination of the previous genetic and structural studies in other organisms. Although we did not find any further sequence features that unite taxa without hairpin sequences to the exclusion of other eukaryotic Prp8 sequences, this does not exclude the possibility that other compensatory adaptations do exist. Such adaptations are clearly of key interest with respect to our understanding of splicing in reduced systems, but advances will require further analyses of non-model systems and physical studies of Prp8 in these or other divergent taxa.

Although the correlation between intron number or the predicted spliceosome complexity and the presence or absence of the hairpin toggle is clear, the role of this structure and that of the RH domain in splicing regulation is not. As noted above, deletion of the hairpin in *S. cerevisiae* (residues 1860–1874, replaced with a GlySer linker) has been reported to be lethal, although the integrity of the overall domain structure and stability was not compromised [20]. When examining the sequences immediately surrounding the deleted hairpin region (Fig. 2), we do not observe strong sequence similarity in those taxa that lack the hairpin. This suggests that the observed lethality of the yeast Prp8 mutant may not be strictly due to the absence of the hairpin. There is strong sequence conservation of the hairpin in those taxa that contain it (with the exception of *G. intestinalis* and *T. cruzi*), even though none of the interactions described here depend on residue-specific recognition. The observation that mutations favoring or disfavoring the hairpin confer a phenotype, in the presence of intron mutations [17,21] suggests that conservation of residues may be important in terms of governing the stability and regulatory role of the hairpin. It appears that the hairpin, through interactions with associated factors, locks the position of the RH domain through snRNP and spliceosome assembly and then post-catalytically. The movements of the RH domain through the catalytic cycle, specifically in the sequential transitions from B^{act} to C*, are significant. Given genetic and biochemical evidence, these transitions must involve the RH domain hairpin, where present, in a yet to be determined fashion.

Materials and Methods

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, which was used to transform *Escherichia coli*. Expression of the maltose binding protein-tagged protein in LB medium supplemented with 2 g/L dextrose was followed by centrifugation, resuspension of the cell pellet in lysis buffer [50 mM Tris

(pH 8), 500 mM NaCl, 5 mM BME] and lysis 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 purified by size exclusion (SD75) chromatography. Concentrated fractions were cleaved overnight with TEV protease to remove the maltose binding protein tag and re-purified on the SD75 column.

Crystallization

Crystals were grown at 23 °C using the hanging drop vapour diffusion technique 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 $MgCl_2$, 10%–12% PEG 8000]. Crystals were cryo-protected in precipitant with the addition of 20% (v/v) glycerol and frozen in liquid nitrogen.

Data collection and processing

Data were collected at beamline CMCF-ID of the Canadian Light Source, University of Saskatchewan, Saskatoon, Canada. Data were processed and scaled using the HKL2000 package [53].

Model building and refinement

The structure was solved using molecular replacement in Phenix Phaser [54] with monomer A of PDB 4JK7 lacking residues 1786–1801 as a search model. Refinement in Phenix Refine alternated with manual model building in Coot [55] to complete and refine the model. Refinement statistics are summarized in Table 2.

Sequence alignment and comparison

We retrieved genome and annotation data from the National Center for Biotechnology Information database for 27 taxa sampled from the major eukaryotic groups discussed in Burki *et al.* [27] These taxa sorted into their major group highlighted include the following: Excavates: *G. intestinalis, Leishmania donovani, Naegleria gruberi, T vaginalis, T cruzi*; Amoebozoa: *Dictyostelium discoideum, E. histolytica*; Obazoa: *Caenorhabditis elegans, Drosophila melanogaster, Homo sapiens, S. cerevisiae, S. pombe, A. muscaria, Batrachochytrium dendrobatidis, Rozella allomycis*, *E. cuniculi, N. bombycis*; Archaeplastida: *Chlamydomonas reinhardtii, Arabidopsis thaliana, C. crispus, C. merolae, Galdieria sulphuraria*; Cryptista: *Guillardia theta*; Stramenopiles, Alveolata and Rhizaria (SAR): *Phaeodactylum tricornutum, Phytophthora infestans, Paramecium tetraurelia, Plasmodium falciparum* (for taxon information/accessions and BLAST scores please, see Supplementary Data). We increased sampling on taxa with reduced genomes and low intron densities similar to CM. We also analyzed additional fungal taxa to improve the resolution of our cwf19 analysis. We performed reciprocal BLASTP analyses ($e > 1 \times 10^{-5}$ cutoff) to recover homologs of Prp8, Aar2, Cwf19, and Snu66. Alignments of each were performed using MUSCLE and further refined manually.

Coordinates

Protein Data Bank: Coordinates for the CM RH domain have been deposited under accession code 6NQI.

Supplementary material includes alignments of Prp8, Aar2, Snu66, and Cwf19 (Supplementary Figs. 1–4) and details of BLAST analysis (Supplementary Data). Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2019.04.047.

Acknowledgments

This work was supported by operating grants to A. M.M. from the Natural Sciences and Engineering Research Council of Canada and from the Canadian Institutes of Health Research as well as Natural Sciences and Engineering Research Council of Canada operating grants to S.D.R. and N.M.F. E.L. G. was supported by a doctoral fellowship from Alberta Innovates Health Solutions.

Received 4 February 2019;
Received in revised form 27 April 2019;
Available online 9 May 2019

Keywords:
pre-mRNA splicing;
spliceosome assembly;
splicing regulation;
Prp8

Abbreviations used:
snRNP, small nuclear ribonucleoprotein particle; RH, RNase H-like; CM, *C. merolae*.

References

[1] C.L. Will, R. Luhrmann, Spliceosome structure and function, Cold Spring Harb Perspect Biol 3 (7) (2011), a003707.
[2] D.A. Brow, Allosteric cascade of spliceosome activation, Annu Rev Genet 36 (2002) 333–360.
[3] A.A. Hoskins, L.J. Friedman, S.S. Gallagher, D.J. Crawford, E.G. Anderson, R. Wombacher, ... M.J. Moore, Ordered and dynamic assembly of single spliceosomes, Science 331

(6022) (2011) 1289–1295, https://doi.org/10.1126/science.1198830.
[4] R. Bai, C. Yan, R. Wan, J. Lei, Y. Shi, Structure of the post-catalytic spliceosome from *Saccharomyces cerevisiae*, Cell 171 (7) (2017) 1589–1598.
[5] S. Liu, X. Li, L. Zhang, J. Jiang, R.C. Hill, Y. Cui, ... R. Zhao, Structure of the yeast spliceosomal postcatalytic P complex, Science 358 (6368) (2017) 1278–1283.
[6] M.E. Wilkinson, S.M. Fica, W.P. Galej, C.M. Norman, A.J. Newman, K. Nagai, Postcatalytic spliceosome structure reveals mechanism of 3′-splice site selection, Science 358 (6368) (2017) 1283–1288.
[7] P.E. Hodges, S.P. Jackson, J.D. Brown, J.D. Beggs, Extraordinary sequence conservation of the PRP8 splicing factor, Yeast 11 (4) (1995) 337–342.
[8] A.N. Kuhn, D.A. Brow, Suppressors of a cold-sensitive mutation in yeast U4 RNA define five domains in the splicing factor Prp8 that influence spliceosome activation, Genetics 155 (4) (2000) 1667–1682.
[9] L. Liu, C.C. Query, M.M. Konasrska, Opposing classes of prp8 alleles modulate the transition between the catalytic steps of pre-mRNA splicing, Nat Struct Mol Biol 14 (6) (2007) 519–526.
[10] C.C. Query, M.M. Konarska, Suppression of multiple substrate mutations by spliceosomal prp8 alleles suggests functional correlations with ribosomal ambiguity mutants, Mol Cell 14 (3) (2004) 343–354.
[11] W.P. Galej, C. Oubridge, A.J. Newman, K. Nagai, Crystal structure of Prp8 reveals active site cavity of the spliceosome, Nature 493 (7434) (2013) 638–643.
[12] G. Qu, P.S. Kaushal, J. Wang, H. Shigematsu, C.L. Piazza, R.K. Agrawal, ... H.W. Wang, Structure of a group II intron in complex with its reverse transcriptase, Nat Struct Mol Biol 23 (6) (2016) 549–557.
[13] C. Zhao, A.M. Pyle, Crystal structures of a group II intron maturase reveal a missing link in spliceosome evolution, Nat Struct Mol Biol 23 (6) (2016) 558–565.
[14] C. Maeder, A.K. Kutach, C. Guthrie, ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C terminus of Prp8, Nat Struct Mol Biol 16 (1) (2009) 42–48.
[15] M. Mayerle, C. Guthrie, Prp8 retinitis pigmentosa mutants cause defects in the transition between the catalytic steps of splicing, RNA 22 (5) (2016) 793–809.
[16] S. Mozaffari-Jovin, T. Wandersleben, K.F. Santos, C.L. Will, R. Luhrmann, M.C. Wahl, Inhibition of RNA helicase Brr2 by the C-terminal tail of the spliceosomal protein Prp8, Science 341 (6141) (2013) 80–84.
[17] R.J. Grainger, J.D. Beggs, Prp8 protein: at the heart of the spliceosome, RNA 11 (5) (2005) 533–557.
[18] V. Pena, A. Rozov, P. Fabrizio, R. Luhrmann, M.C. Wahl, Structure and function of an RNase H domain at the heart of the spliceosome, EMBO J 27 (21) (2008) 2929–2940.
[19] D.B. Ritchie, M.J. Schellenberg, E.M. Gesner, S.A. Raithatha, D.T. Stuart, A.M. MacMillan, Structural elucidation of a PRP8 core domain from the heart of the spliceosome, Nat Struct Mol Biol 15 (11) (2008) 1199–1205.4.
[20] K. Yang, L. Zhang, T. Xu, A. Heroux, R. Zhao, Crystal structure of the beta-finger domain of Prp8 reveals analogy to ribosomal proteins, Proc Natl Acad Sci USA 105 (37) (2008) 13817–13822.
[21] M.J. Schellenberg, T. Wu, D.B. Ritchie, S. Fica, J.P. Staley, K.A. Atta, ... A.M. MacMillan, A conformational switch in PRP8 mediates metal ion coordination that promotes pre-mRNA exon ligation, Nat Struct Mol Biol 20 (6) (2013) 728–734.
[22] M. Mayerle, M. Raghavan, S. Ledoux, A. Price, N. Stepankiw, …. Hadjivassiliou Hm, J. Abelson, Structural toggle in the RNase H domain of Prp8 helps balance splicing fidelity and catalytic efficiency, Proc Natl Acad Sci USA 114 (18) (2017) 4739–4744.
[23] M.L. Rodgers, U.S. Tretbar, A. Dehaven, A.A. Alwan, G. Luo, H.M. Mast, A.A. Hoskins, Conformational dynamics of stem II of the U2 snRNA, RNA 22 (2) (2016) 225–236.
[24] M.R. Stark, E.A. Dunn, W.S. Dunn, C.J. Grisdale, A.R. Daniele, M.R. Halsead, ... S.D. Rader, Dramatically reduced spliceosome in *Cyanidioschyzon merolae*, Proc Natl Acad Sci USA 112 (11) (2015) E1191–E1200.
[25] M. Matsuzaki, O. Misumi, T. Shin-I, S. Maruyama, M. Takahara, S.Y. Miyagishima, T. Kuroiwa, Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D, Nature 428 (6983) (2004) 653–657.
[26] H. Qui, A.W. Rossoni, A.P.M. Weber, H.S. Yoon, D. Bhattacharya, Unexpected conservation of the RNA splicing apparatus in the highly streamlined genome of *Galdieria sulphuraria*, BMC Evol Biol 18 (1) (2018) 41.
[27] Burki F, Kaplan M, Tikhonenkov DV, Zlatogursky V, Minh BQ, Radaykina LV, Smirnov A, Mylnikov AP, & Keeling PJ. (2016). Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proc Biol Sci, 283 (1823),* pii: 20152802
[28] A.J. Hudson, M.R. Stark, N.M. Fast, A.G. Russell, S.D. Rader, Splicing diversity revealed by reduced spliceosomes in *C. merolae* and other organisms, RNA Biol 12 (11) (2015) 1–8.
[29] B. Cuypers, M.A. Domagalska, P. Meysman, G. Muylder, M. Vanaerschot, H. Imamura, F. Dumetz, T.W. Verdonckt, P.J. Myler, G. Ramasamy, et al., Multiplexed spliced-leader sequencing: a high-throughput, selective method for RNA-seq in trypanosomatids, Sci Rep 7 (1) (2017) 3725.
[30] A. Gunzl, The pre-mRNA splicing machinery of trypanosomes: complex or simplified? Eukaryot Cell 9 (8) (2010) 1159–1170.
[31] R. Kamikawa, Y. Inagaki, T. Hashimoto, A novel spliceosome-mediated trans-splicing can change our view on genome complexity of the divergent eukaryote *Giardia intestinalis*, Biophys Rev 3 (4) (2011) 193–197.
[32] S.W. Roy, Genomic and transcriptomic analysis reveals spliced leader trans-splicing in cryptomonads, Genome Biol Evol 9 (3) (2017) 468–473.
[33] S. Vanacova, W. Yan, J.M. Carlton, P.J. Johnson, Spliceosomal introns in the deep-branching eukaryote *Trichomonas vaginalis*, Proc Natl Acad Sci USA 102 (12) (2005) 4430–4435.
[34] G. Weber, V.F. Cristao, K.F. Santos, S.M. Jovin, A.C. Heroven, N. Holton, R. Luhrmann, J.D. Beggs, M.C. Wahl, Structural basis for dual roles of Aar2p in U5 snRNP assembly, Genes Dev 27 (5) (2013) 525–540.
[35] R. Bai, R. Wan, C. Yan, J. Lei, Y. Shi, Structures of the fully assembled *Saccharomyces cerevisiae* spliceosome before splicing activation, Science 360 (6396) (2018) 1423–1429.
[36] K. Bertram, D.E. Agafonov, O. Dybkov, D. Haselbach, M.N. Leelaram, C.L. Will, ... H. Stark, Cryo-EM structure of a pre-catalytic human spliceosome primed for activation, Cell 170 (4) (2017) 701–713.
[37] T.H.D. Nguyen, W.P. Galej, X.C. Bai, C. Oubridge, A. Newman, ... K. Nagai, Cryo-EM structure of the yeast U4/

U6.U5 tri-snRNP at 3.7 Å resolution, Nature 530 (7590) (2016) 298–302.
[38] C. Plaschka, P.C. Lin, K. Nagai, Structure of a pre-catalytic spliceosome, Nature 546 (7660) (2017) 617–621.
[39] K. Bertram, D.E. Agafonov, W.T. Liu, O. Dybkov, C.L. Will, K. Hartmuth, ... R. Lurhmann, Cryo-EM structure of a human spliceosome activated for step 2 of splicing, Nature 542 (7641) (2017) 318–323.
[40] S.M. Fica, C. Oubridge, W.P. Galej, M.E. Wilkinson, X.C. Bai, A.J. Newman, K. Nagai, Structure of a spliceosome remodelled for exon ligation, Nature 542 (7641) (2017) 377–380.
[41] C. Yan, R. Wan, R. Bai, G. Huang, Y. Shi, Structure of a yeast step II catalytically activated spliceosome, Science 355 (6321) (2017) 149–155.
[42] X. Zhang, C. Yan, J. Hang, L.I. Finci, J. Lei, Y. Shi, An atomic structure of the human spliceosome, Cell 169 (5) (2017) 918–929.
[43] C. Yan, J. Hang, R. Wan, M. Huang, C.C. Wong, Y. Shi, Structure of a yeast spliceosome at 3.6 Å resolution, Science 349 (6253) (2015) 1182–1191.
[44] J. Rappsilber, U. Ryder, A.I. Lamond, M. Mann, Large-scale proteomic analysis of the human spliceosome, Genome Res 12 (8) (2002) 1231–1245.
[45] Garrey SM, Katolik A, Prekeris M, Li X, York K, Bernards S, … & Hesselberth JR. (2014). A homolog of lariat-debarnching enzyme modulates turnover of branched RNA. *RNA,* 20 *(8),* 1337-1348.
[46] S.M. Fica, K. Nagai, Cryo-electron microscopy snapshots of the spliceosome: structural insights into a dynamic ribonucleoprotein machine, Nat Struct Mol Biol 24 (10) (2017) 791–799.
[47] W.P. Galej, M.E. Wilkinson, S.M. Fica, C. Oubridge, A.J. Newman, K. Nagai, Cryo-EM structure of the spliceosome immediately after branching, Nature 537 (7619) (2016) 197–201.
[48] R. Wan, C. Yan, R. Bai, G. Huang, Y. Shi, Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution, Science 353 (6302) (2016) 895–904.
[49] R. Wan, C. Yan, Lei J. Bai, Y. Shi, Structure of an intron lariat spliceosome from *Saccharomyces cerevisiae*, *Cell* 171 (1) (2017) 120–132.
[50] X. Zhan, C. Yan, X. Zhang, J. Lei, Y. Shi, Structure of a human catalytic step I spliceosome, Science 359 (6375) (2018) 537–545.
[51] A. Dziembowski, A.P. Ventura, B. Rutz, F. Caspary, C. Faux, F. Halgand, ... B. Seraphin, Proteomic analysis identifies a new complex required for nuclear pre-mRNA retention and splicing, EMBO J 23 (24) (2004) 4847–4856.
[52] X. Li, W. Zhang, T. Xu, J. Ramsey, L. Zhang, R. Hill, K.C. Hansen, J.R. Hesselberth, R. Zhao, Comprehensive in vivo RNA-binding site analyses reveal a role of Prp8 in spliceosomal assembly, Nucleic Acids Res 41 (6) (2013) 3805–3818.
[53] Z. Otwinowski, W. Minor, Processing of x-ray diffraction data collected in oscillation mode, Methods Enzymol 276 (1997) 307–326.
[54] P.D. Adams, P.V. Afonine, G. Bunkoczi, V.B. Chen, I.W. Davis, N. Echols, ... P.H. Zwart, PHENIX: a comprehensive Python-based system for macromolecular structure solution, Acta Crystallogr D Biol Crystallogr 66 (Pt 2) (2010) 213–221.
[55] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of Coot, Acta Crystallogr D Biol Crystallogr 66 (Pt 2) (2010) 486–501.