

# Molecular Mechanisms of pre-mRNA Splicing through Structural Biology of the Spliceosome

Chuangye Yan,<sup>1</sup> Ruixue Wan,<sup>1</sup> and Yigong Shi<sup>1,2</sup>

<sup>1</sup>Beijing Advanced Innovation Center for Structural Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China

<sup>2</sup>Institute of Biology, Westlake Institute for Advanced Study, Westlake University, Hangzhou 310064, Zhejiang Province, China

Correspondence: shi-lab@tsinghua.edu.cn; yancy05@mails.tsinghua.edu.cn; wxruisnow@163.com

## SUMMARY

Precursor messenger RNA (pre-mRNA) splicing is executed by the spliceosome. In the past 3 years, cryoelectron microscopy (cryo-EM) structures have been elucidated for a majority of the yeast spliceosomal complexes and for a few human spliceosomes. During the splicing reaction, the dynamic spliceosome has an immobile core of about 20 protein and RNA components, which are organized around a conserved splicing active site. The divalent metal ions, coordinated by U6 small nuclear RNA (snRNA), catalyze the branching reaction and exon ligation. The spliceosome also contains a mobile but compositionally stable group of about 13 proteins and a portion of U2 snRNA, which facilitate substrate delivery into the splicing active site. The spliceosomal transitions are driven by the RNA-dependent ATPase/helicases, resulting in the recruitment and dissociation of specific splicing factors that enable the reaction. In summary, the spliceosome is a protein-directed metalloribozyme.

## Outline

- 1 Introduction
  - 2 Structural studies of the spliceosome
  - 3 Shared features of the spliceosomes
  - 4 Spliceosome as a metalloribozyme
  - 5 The role of the protein components
  - 6 Recognition of the RNA elements
  - 7 Choreography of the RNA and protein components
  - 8 The human spliceosome
  - 9 Conclusion
- References

## 1 INTRODUCTION

Precursor messenger RNA (pre-mRNA) splicing, discovered 40 years ago (Berget et al. 1977; Chow et al. 1977), is executed by a multi-megadalton, ribonucleoprotein (RNP) complex known as the spliceosome (Brody and Abelson 1985; Frendewey and Keller 1985; Grabowski et al. 1985). The components of the spliceosome constitute five distinct classes: the uridine-rich RNPs (U1, U2, U4, U5, and U6 snRNPs), the NineTeen complex (NTC) and the NTC-related complex (NTR), the splicing factors, the RNA-dependent ATPase/helicases, and other regulatory and/or auxiliary proteins such as kinases/phosphatases and peptidyl prolyl isomerases (PIs). The involvement of snRNPs in pre-mRNA splicing was documented by the early 1980s (Lerner and Steitz 1979; Lerner et al. 1980; Rogers and Wall 1980; Yang et al. 1981) and subsequently confirmed (Mount et al. 1983; Padgett et al. 1983; Kramer et al. 1984; Black et al. 1985; DiMaria et al. 1985; Krainer and Maniatis 1985; Berget and Roberson 1986; Grabowski and Sharp 1986; Pikielny and Rosbash 1986). Only the precatalytic spliceosome (B complex) and the pre-B complex contain all five snRNPs but lack the NTC and NTR. Despite their relatively late discovery (Tarn et al. 1994; Chen et al. 2002; Ohi and Gould 2002; Chan et al. 2003; Chan and Cheng 2005), NTC and NTR are known to be indispensable in structurally shaping the splicing active site (Hang et al. 2015; Yan et al. 2015; Shi 2017b).

The assembled spliceosome sequentially assumes eight different compositional states during each cycle of the splicing reaction: pre-B, B, activated ( $B^{\text{act}}$ ), catalytically activated ( $B^*$ ), step I spliceosome (C), step II activated ( $C^*$ ), postcatalytic (P), and intron lariat spliceosome (ILS) (Fig. 1A). The B complex is preceded by a recently characterized pre-B complex (Boesler et al. 2016), in which the 5'-splice site (5'SS) is still recognized by the U1 snRNP. Each cycle consists of two sequential steps: branching and exon ligation, which proceed in the  $B^*$  and  $C^*$  complexes, respectively. Therefore, the  $B^*$ -to-C and  $C^*$ -to-P transitions occur unassisted. In contrast, the B-to- $B^{\text{act}}$ ,  $B^{\text{act}}$ -to- $B^*$ , C-to- $C^*$ , and P-to-ILS transitions are driven by the RNA-dependent ATPase/helicases Brr2, Prp2, Prp16, and Prp22, respectively (Fig. 1A) (Cordin et al. 2012). The disassembly of the ILS complex is mediated by the ATPase/helicase Prp43 (Arenas and Abelson 1997; Tsai et al. 2005).

Both the branching and exon ligation reactions are of the bimolecular nucleophilic substitution ( $S_N2$ ) type transesterification. In a pre-mRNA, two neighboring exons (5'-exon and 3'-exon) are separated by an intron, which contains three conserved sequence elements: the 5'SS, the branch point sequence (BPS), and the 3'-splice site (3'SS) (Fig. 1A).

The branching reaction results in a free 5'-exon and an intron lariat-3'-exon intermediate, which has a 2'-5' phosphodiester bond linking the invariant adenosine nucleotide of the BPS to the guanine nucleotide at the 5'-end of the 5'SS (Grabowski et al. 1984; Padgett et al. 1984; Ruskin et al. 1984). Exon ligation leads to the joining of the 5'-exon and the 3'-exon and the generation of a free intron lariat (Fig. 1A).

The splicing active site of the spliceosome was postulated to contain two divalent metal ions known as M1 and M2 (Steitz and Steitz 1993), which was supported by sulfur-substituted pre-mRNA splicing studies (Sontheimer et al. 1997). In vitro splicing of *Saccharomyces cerevisiae* using U6 small nuclear RNA (snRNA) with sulfur-substituted nonbridging phosphoryl oxygen unambiguously identified the nucleotide U80 as a coordinating ligand for the catalytic metals (Yean et al. 2000). In the presence of thiophilic ions such as  $Mn^{2+}$ , the spliceosome reconstituted with the sulfur-substituted U80(Sp), but not U80(Rp), U6 snRNA allowed the branching reaction to occur (Yean et al. 2000). U6 snRNA was subsequently shown to coordinate two divalent metals known as M1 and M2 (Fica et al. 2013), and structural analysis of the group II intron allowed prediction of conserved features in the active site of the spliceosome (Toor et al. 2008; Keating et al. 2010).

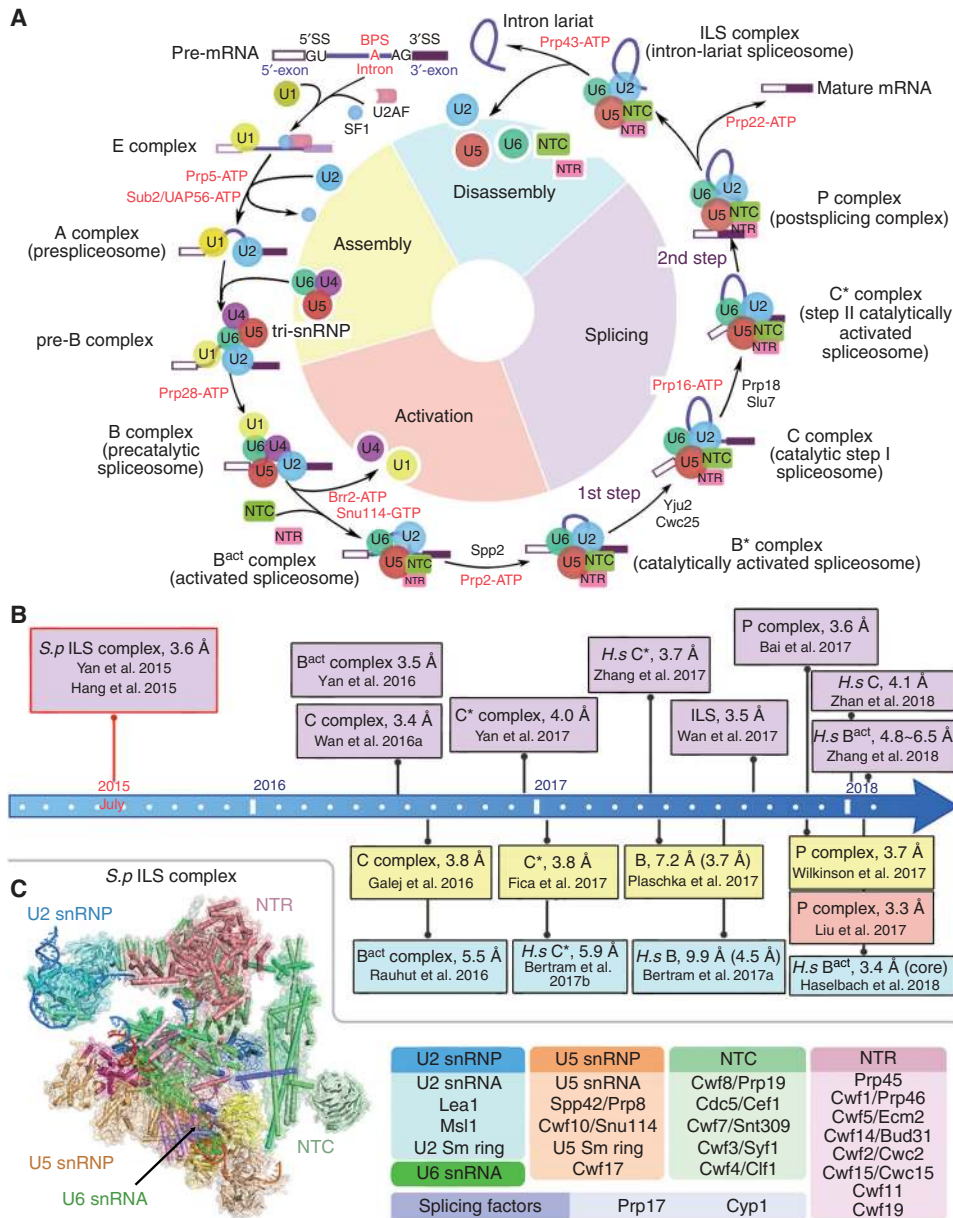
The entire U5 snRNP, the core of U2 snRNP, and U6 snRNA are present in all eight assembled spliceosomal complexes (Fig. 1A). Among all spliceosomal proteins, the central U5 snRNP component Prp8 is most highly conserved and plays a central role by anchoring the splicing active site. The step I splicing factors (Cwc25 and Yju2 in *S. cerevisiae*) only associate with the splicing active site surrounding the branching reaction, whereas the step II factors (Prp18 and Slu7 in *S. cerevisiae*) mainly act to promote exon ligation (Fig. 1A). In contrast, the general splicing factors Cwc21 and Cwc22 are present in the  $B^{\text{act}}$  through P complexes, whereas Cwc24 and Cwc27 are recruited into the  $B^{\text{act}}$  complex but dissociated in the  $B^{\text{act}}$ -to- $B^*$  transition (Ohrt et al. 2012).

## 2 STRUCTURAL STUDIES OF THE SPLICEOSOME

### 2.1 EM-Based Structural Studies before 2015

Mechanistic understanding of pre-mRNA splicing and its regulation requires detailed three-dimensional information on the intact spliceosome. X-ray crystallography allowed structural determination of a number of individual spliceosomal components and subcomplexes, including the U1 snRNP core (Pomeranz Krummel et al. 2009; Weber et al. 2010; Kondo et al. 2015), the U4 snRNP core (Leung et al. 2011), subcomplexes of the U2 and U6 snRNPs (Price et al. 1998; Sickmier et al. 2006; Lin and Xu 2012; Montemayor et al. 2014; Zhou et al. 2014), the exon junc-





**Figure 1.** The precursor messenger RNA (pre-mRNA) splicing cycle and the major structures of the spliceosome. (A) The pre-mRNA splicing cycle. Each cycle includes three phases: assembly and activation of the spliceosome, execution of the splicing reaction, and disassembly of the spliceosome. The pre-B complex represents the first fully assembled spliceosome, in which all five snRNPs are present and the 5' SS is still recognized by U1 snRNP. In addition to the pre-B complex, the assembled spliceosome exists in seven compositionally distinct states: B, B<sup>act</sup>, B\*, C, C\*, P, and intron lariat spliceosome (ILS). The unidirectional conversion of each of these neighboring spliceosomal complexes to the next is driven by the conserved ATPases/helicases Brr2 (B-to-B<sup>act</sup>), Prp2 (B<sup>act</sup>-to-B\*), Prp16 (C-to-C\*), and Prp22 (P-to-ILS). The splicing factors Cwc25 and Yju2 facilitate the branching reaction, whereas Prp18 and Slu7 assist exon ligation. Disassembly of the ILS complex is executed by the ATPase/helicase Prp43. (B) The major cryoelectron microscopy (cryo-EM) structures of the spliceosome since 2015. The time axis is drawn to scale, with each white dot representing a month. The first atomic model of an intact spliceosome was elucidated for the ILS complex from *Saccharomyces pombe* on the basis of an EM density map at an average resolution of 3.6 Å (Hang et al. 2015; Yan et al. 2015). The 18 cryo-EM structures of the spliceosome come from four independent laboratories: nine from the Shi group (shaded light purple), four from the Nagai group (shaded yellow), four from the Luhrmann/Stark group (shaded cyan), and one from the Zhao/Zhou group (shaded light brown). (C) The cryo-EM structure of the ILS complex from *S. pombe* (PDB code 3JB9) (Hang et al. 2015; Yan et al. 2015). The color-coded structure is shown on the left, and the individual components resolved in the structure are tabulated on the right.

tion complex bound to RNA (Andersen et al. 2006), a large fragment of Prp8 bound to Aar2 (Galej et al. 2013), Brr2 bound to a carboxy-terminal fragment of Prp8 (Mozaffari-Jovin et al. 2013; Nguyen et al. 2013), the SF3a complex (Lin and Xu 2012), and the core of the SF3b complex (Cretu et al. 2016). A structure of the retention and splicing complex (RES) was determined by nuclear magnetic resonance (NMR) (Wysoczanski et al. 2014). These structures reveal detailed features of the individual components and their mutual interactions within the colossal spliceosome jigsaw puzzle. Because the intact spliceosome proved to be intractable to crystallization, its structural information comes exclusively from electron microscopy (EM).

Until 2015, EM-based studies, with the highest resolution  $\sim 20$ – $29$  Å, had only allowed a glimpse of the overall appearance of the spliceosome at different splicing stages and tentative assignment of some snRNPs and protein components (Luhrmann and Stark 2009). These EM structures were reported for the A (Furman and Glitz 1995; Behzadnia et al. 2007), B (Boehringer et al. 2004; Deckert et al. 2006; Fabrizio et al. 2009; Wolf et al. 2009), B<sup>act</sup> (Fabrizio et al. 2009; Bessonov et al. 2010), C (Jurica et al. 2004; Ohi et al. 2007; Fabrizio et al. 2009; Golas et al. 2010), P (Ilagan et al. 2013), and ILS (Chen et al. 2014) complexes. In 2015, the 5.9 Å cryoelectron microscopy (cryo-EM) structure of the *S. cerevisiae* U4/U6.U5 tri-snRNP, for the first time, allowed faithful assignment of the protein and RNA components (Nguyen et al. 2015).

## 2.2 EM-Based Structural Studies since 2015

Major advances have been achieved in the mechanistic understanding of the spliceosome structure and function since 2015, beginning with the public release of the first atomic model for an intact spliceosome (Fig. 1B) (Hang et al. 2015; Yan et al. 2015). Thirty-seven protein components and four snRNA molecules of the endogenous *Saccharomyces pombe* ILS complex were atomically modeled from the 3.6 Å cryo-EM reconstruction (Fig. 1C), in which the local resolution in the core regions of the spliceosome reached 2.9 Å (Yan et al. 2015). Most notably, structural insights derived from the *S. pombe* ILS complex are generally applicable to most other spliceosomes from both *S. cerevisiae* and human (Shi 2017a).

Since 2016, at least 17 independent cryo-EM structures of the assembled spliceosome have been reported, reflecting six distinct compositional states (Fig. 1B; Table 1). Among these structures, 11 were reported for the *S. cerevisiae* spliceosomal complexes: B (Plaschka et al. 2017), B<sup>act</sup> (Rauhut et al. 2016; Yan et al. 2016), C (Galej et al. 2016; Wan et al. 2016a), C\* (Fica et al. 2017; Yan et al. 2017), P (Bai et al. 2017; Liu et al. 2017; Wilkinson et al. 2017), and ILS (Wan

et al. 2017); six were reported for the human spliceosomes: B (Bertram et al. 2017a), B<sup>act</sup> (Haselbach et al. 2018; Zhang et al. 2018), C (Zhan et al. 2018), and C\* (Fig. 1B) (Bertram et al. 2017b; Zhang et al. 2017). In addition, three independent structures of the U4/U6.U5 tri-snRNP have been reported for *S. cerevisiae* (Nguyen et al. 2016; Wan et al. 2016b) and for human (Agafonov et al. 2016), and the structure of the *S. cerevisiae* U1 snRNP was elucidated at 3.6 Å resolution (Table 1) (Li et al. 2017).

## 2.3 Quality Assessment of the Recent Cryo-EM Structures

The reported average resolution for the 18 cryo-EM structures of the entire spliceosome ranges between 3.3 Å and 9.9 Å resolution, with the highest limit reaching 2.8 Å (Table 1). Among these structures, only seven were determined at an average overall resolution of 3.6 Å or better, including the *S. pombe* ILS complex at 3.6 Å (Yan et al. 2015), the human B<sup>act</sup> complex at 3.4 Å (Haselbach et al. 2018), and five *S. cerevisiae* complexes: B<sup>act</sup> at 3.5 Å (Yan et al. 2016), C at 3.4 Å (Wan et al. 2016a), P at 3.3 Å (Liu et al. 2017) and 3.6 Å (Bai et al. 2017), and ILS at 3.4 Å (Fig. 1B; Table 1) (Wan et al. 2017). The average overall resolution refers to the entire spliceosome, not just the core or a specific region as evasively claimed in some studies. The local resolution for the core regions of the spliceosome in each of the six structures reaches at least 3.0 Å, which allows accurate assignment of amino acid side chains and placement of RNA nucleotides and metal ions. The ultimate test of the quality of a cryo-EM structure is not the reported resolution, but the quality of the original EM density map. Importantly, side-by-side comparison of the deposited EM density maps shows that the seven structures are indeed of the highest quality among all 18 published structures of the yeast or human spliceosome.

Two general methods were used for the isolation of the spliceosomes. One method involves direct purification of endogenous spliceosomal complexes from the nuclear extract using an affinity-tagged protein component (Ohi et al. 2007; Yan et al. 2015). This method was effectively applied to purification of the yeast, but not human, spliceosomes and allowed determination of seven of the 18 published spliceosome structures (Table 1). The pre-mRNA in these spliceosomes represents a mixture from endogenous genes; its atomic modeling was made possible by the strong sequence conservation of yeast intronic sequences (Spingola et al. 1999; Bon et al. 2003; Kupfer et al. 2004). In contrast, the other method relies on in vitro assembly of the spliceosome on MS2-tagged synthetic pre-mRNA (Jurica et al. 2002; Zhou et al. 2002). This method has been successfully applied to both human and yeast spliceosomes, although



**Table 1.** Cryo-EM studies of the spliceosomes and the spliceosomal complexes

Species	State of the spliceosome	Resolution (Å) (highest limit) <sup>a</sup>	Method of assembly	Method of purification	Date of Epub	Reference(s)	
<i>Schizosaccharomyces pombe</i>	ILS	3.6 (3.0)	Endogenous	Cdc5-TAPs	2015.08.20	Hang et al. 2015; Yan et al. 2015	
<i>Saccharomyces cerevisiae</i>	B <sup>act</sup>	3.5 (2.8)	Endogenous	Cef1-TAPs	2016.07.21	Yan et al. 2016	
		5.8 (4.0)	In vitro assembling	Prp2-G551N; MS2-MBP; Grafix and PMPI cross-linking	2016.08.25	Rauhut et al. 2016	
	C	3.4 (2.9)	Endogenous	Cef1-TAPs	2016.07.21	Wan et al. 2016a	
		3.8 (3.5)	In vitro assembling	MS2-MBP	2016.07.26	Galej et al. 2016	
	C*	4.0 (3.5)	Endogenous	Cef1-TAPs	2016.12.15	Yan et al. 2017	
		3.8 (3.5)	In vitro assembling	Slu7-TAPs; BS3 cross-linking	2017.01.11	Fica et al. 2017	
	B	7.2 (3.5)	In vitro assembling	Brr2-TAPs; MS2-MBP; BS3 cross-linking	2017.05.22	Plaschka et al. 2017	
					2017.09.14	Wan et al. 2017	
		ILS	3.5 (2.9)	Endogenous	Cef1 or Yju2 TAPs	2017.09.14	Wan et al. 2017
			3.6 (3.0)	Endogenous	Prp22-K512A; Cef1-TAPs	2017.11.16	Bai et al. 2017
P	3.3 (3.0)	Endogenous	Prp22-H606A; Cef1-TAPs	2017.11.16	Liu et al. 2017		
	3.8 (3.5)	In vitro assembling	Prp22-S635A; MS2-MBP	2017.11.16	Wilkinson et al. 2017		
<i>Homo sapiens</i>	C*	5.9 (4.5)	In vitro assembling	MS2-MBP; Grafix	2017.01.11	Bertram et al. 2017b	
					2017.05.11	Zhang et al. 2017	
	B	9.9 (3.5)	In vitro assembling	MS2-MBP; Grafix	2017.08.03	Bertram et al. 2017a	
					2018.01.04	Zhan et al. 2018	
	C B <sup>act</sup>	4.1 (3.5)	In vitro assembling	MS2-MBP; Grafix	2018.01.17	Haselbach et al. 2018	
3.4 (3.0)		In vitro assembling	MS2-MBP; Grafix	2018.01.23	Zhang et al. 2018		
Other spliceosomal complexes <i>S. cerevisiae</i>	Tri-snRNP	5.9 (4.5)	Endogenous	Brr2-TAPs	2015.06.24	Nguyen et al. 2015	
		3.8 (3.0)	Endogenous	Prp6-TAPs	2016.01.07	Wan et al. 2016b	
		3.7 (3.0)	Endogenous	Brr2-TAPs	2016.02.01	Nguyen et al. 2016	
	U1 snRNP	3.6 (3.0)	Endogenous	U1A/Mud1-TAPs	2017.10.19	Li et al. 2017	
	Tri-snRNP	7.0 (4.0)	Endogenous	m <sup>3</sup> G-cap antibody; Grafix	2016.02.18	Agafonov et al. 2016	

<sup>a</sup>The resolution shown refers to the average overall resolution of the entire spliceosome or the spliceosomal complex. The resolution in parentheses represents the highest local resolution in the spliceosome or the spliceosomal complex.

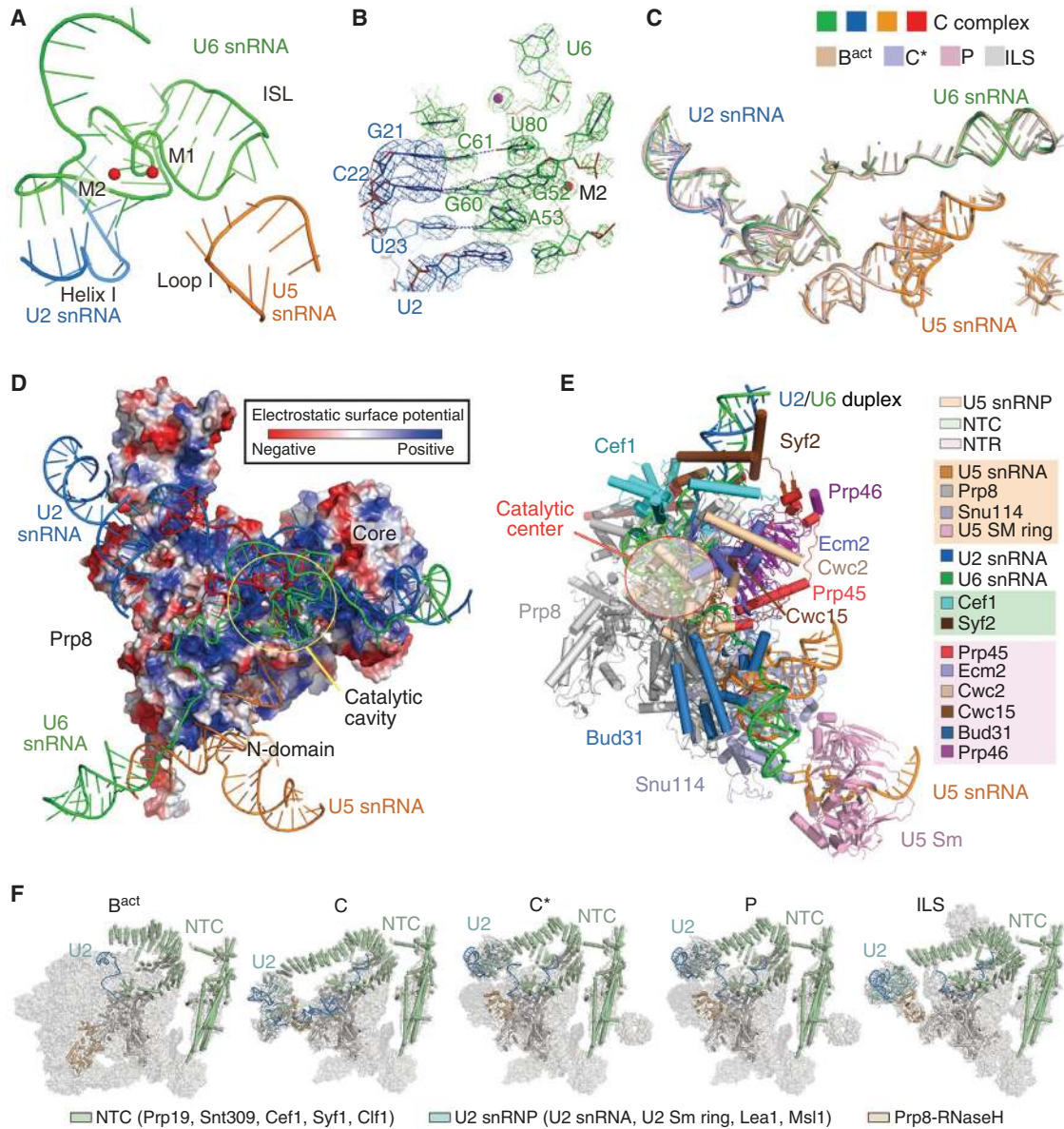
an additional step of chemical cross-linking is required for maintenance of the structural integrity for the human spliceosome (Table 1).

### 3 SHARED FEATURES OF THE SPLICEOSOMES

Except the pre-B and the B\* complexes, six of the eight assembled spliceosomal complexes during each splicing cycle have been structurally characterized in recent cryo-EM studies. Among these, the pre-B and the B complex are the only spliceosomes that do not contain a well-formed splicing active site; they show a major difference compared with the other six spliceosomal complexes. The U4 snRNP remains associated and the NTC and NTR components are yet to be recruited. The other six spliceosomes share a set of strikingly similar structural features that were first observed

in the cryo-EM structure of the *S. pombe* ILS complex (Hang et al. 2015; Yan et al. 2015).

The RNA elements in the splicing active site show a highly conserved conformation in the B<sup>act</sup> through ILS complexes (Fig. 2A). The intramolecular stem loop (ISL) of U6 snRNA forms the center of the splicing active site and coordinates the two catalytic metal ions (M1 and M2). The ISL is stabilized by three conserved structural metal ions. Three consecutive nucleotides of the U2 snRNA from the helix Ib form a catalytic triplex with a duplex of U6 snRNA, which involves five nucleotides preceding the ISL and the key U6 nucleotide U80 in *S. cerevisiae* (Fig. 2B). The catalytic triplex facilitates the splicing reaction. Loop I of U5 snRNA, which recognizes the 5'-exon, is positioned in close proximity to the ISL. Beyond the active site, U5 and U6 snRNAs along their entire lengths and 30 consecutive



**Figure 2.** Shared structural features of the spliceosome during the splicing reaction. (A) The RNA elements at the conserved splicing active site. The active site comprises the intramolecular stem loop (ISL) of U6 small nuclear RNA (snRNA) and associated metal ions, the U2/U6 catalytic triplex, helix I of the U2/U6 duplex, and loop I of U5 snRNA. Although the active site shown here is derived from the *Saccharomyces cerevisiae* C complex (PDB code 5GMK) (Wan et al. 2016a), the overall conformation is identical in the B<sup>act</sup> through intron lariat spliceosome (ILS) complexes. (B) Structure of the conserved catalytic triplex. The original electron microscopy (EM) density map, shown for the *S. cerevisiae* C complex (Wan et al. 2016a), allows unambiguous identification of the metal ions and the nucleobases. (C) U5 and U6 snRNAs along their entire lengths and the 5' 30 nucleotides of U2 snRNA remain largely static in the B<sup>act</sup> through ILS complexes. The RNA elements in the *S. cerevisiae* C complex are color coded; whereas those from the B<sup>act</sup>, C\*, P, and ILS complexes are each displayed in a single color. The PDB codes for the B<sup>act</sup>, C, C\*, P, and ILS complexes are 5GM6, 5GMK, 5WSG, 5YLZ, and 5Y88, respectively. (D) The conserved splicing active site RNA elements are anchored in the catalytic cavity of Prp8. The cavity is formed at the interface between the N-domain and the Prp8 core and is enriched in positively charged residues. (E) Sixteen additional protein components, for their partial or entire lengths, remain largely static in the B<sup>act</sup> through ILS complexes. These proteins, together with Prp8, maintain the rigid conformation of the snRNA elements (U5 and U6 along their entire lengths and the 5' 30 nucleotides of U2 snRNA). These proteins and the snRNAs from the *S. cerevisiae* C complex are color coded here. (F) A compositionally constant group of 13 proteins in the B<sup>act</sup> through ILS complexes. These proteins constitute two groups: nine in the U2 snRNP (the heptameric Sm complex, Lea1, and Msl1) and four in the NTC (Prp19, Snt309, Cef1, Syf1, and Clf1). Each group rigidly maintains its collective structure but is translocated during the two steps of the splicing reaction. These proteins, together with the other 20 rigid components, give rise to the characteristic appearance of the spliceosome.

nucleotides at the 5'-end of U2 snRNA show the same structure in the B<sup>act</sup> through ILS complexes (Fig. 2C). Thus, among the three snRNA molecules, the only mobile region is the sequence downstream from nucleotide 30 of U2 snRNA.

The relatively rigid conformation of the snRNAs in the splicing active site and beyond is maintained by at least 17 protein components, which in turn must share nearly identical structures in the B<sup>act</sup> through ILS complexes (Shi 2017b). Among these proteins, Prp8 plays a pivotal role. A positively charged catalytic cavity, which is formed at the interface between its N-domain and the core (also known as the large domain; Galej et al. 2013), provides approximately half of the needed surface to host the conserved RNA elements at the splicing active site (Fig. 2D) (Hang et al. 2015; Yan et al. 2015). The other half is furnished by segments of the NTC components Cef1 (Cdc5 in *S. pombe*) and Syf2, the NTR components Cwc2 and Ecm2 (Cwf2 and Cwf5 in *S. pombe*), and the various splicing factors. In addition, Snu114, the heptameric Sm complex of the U5 snRNP, and six NTR components (Cwc2, Ecm2, Bud31, Cwc15, Prp45, and Prp46) display the same general conformation in the B<sup>act</sup> through ILS complexes (Fig. 2E).

In addition to these 20 protein and RNA components that remain largely rigid in the B<sup>act</sup> through ILS complexes, 13 additional protein components are included in each of these spliceosomes. These include nine components of the U2 snRNP (Lea1, Msl1, and the heptameric Sm complex), which interact with each other and bind the sequences around the Sm site of U2 snRNA, and four proteins of the NTC (Prp19, Snt309, Syf1, and Clf1), which together with a portion of Cef1 are intertwined to form a stable core of the NTC (Fig. 2F). The relative positioning of the components within the U2 snRNP core or the NTC core remains unchanged in the various spliceosomal complexes;

but the core as a whole undergoes drastic translocation, particularly the U2 snRNP.

Together, the immobile 20 protein and RNA components, along with the compositionally stable but mobile 13 proteins, give rise to a characteristic appearance of the spliceosome that is conserved in the B<sup>act</sup> through ILS complexes (Fig. 2F). In fact, for an experienced structural biologist, visual inspection of the two-dimensional averages of the particles not only allows identification of the spliceosomal complexes but also their specific stages in the splicing reaction. Such a structural hallmark also greatly facilitates atomic modeling of the various spliceosomal complexes. For example, the atomic coordinates of the *S. pombe* ILS complex (Yan et al. 2015) can be conveniently docked into the EM density map of the human C\* complex (Zhang et al. 2017), followed by minor adjustment and residue replacement.

#### 4 SPLICEOSOME AS A METALLORIBOZYME

The available structural evidence proves the notion that the spliceosome is a metalloribozyme. The roles of the two catalytic metal ions M1 and M2 are thought to reciprocate in the two transesterification reactions of each splicing cycle (Steitz and Steitz 1993; Fica et al. 2013). In the branching reaction that occurs in the B\* complex, M2 activates the nucleophile (2'-OH of an adenine nucleotide in the BPS) and M1 stabilizes the leaving group (3'-OH of the 5'-exon). In the exon ligation that proceeds in the C\* complex, M1 activates the nucleophile (3'-OH of the 5'-exon) and M2 stabilizes the leaving group (3'-OH of the intron lariat). In the *S. pombe* ILS complex (Hang et al. 2015; Yan et al. 2015), both M1 and M2 are observed and their locations are consistent with published biochemical characterization (Fica et al. 2013).

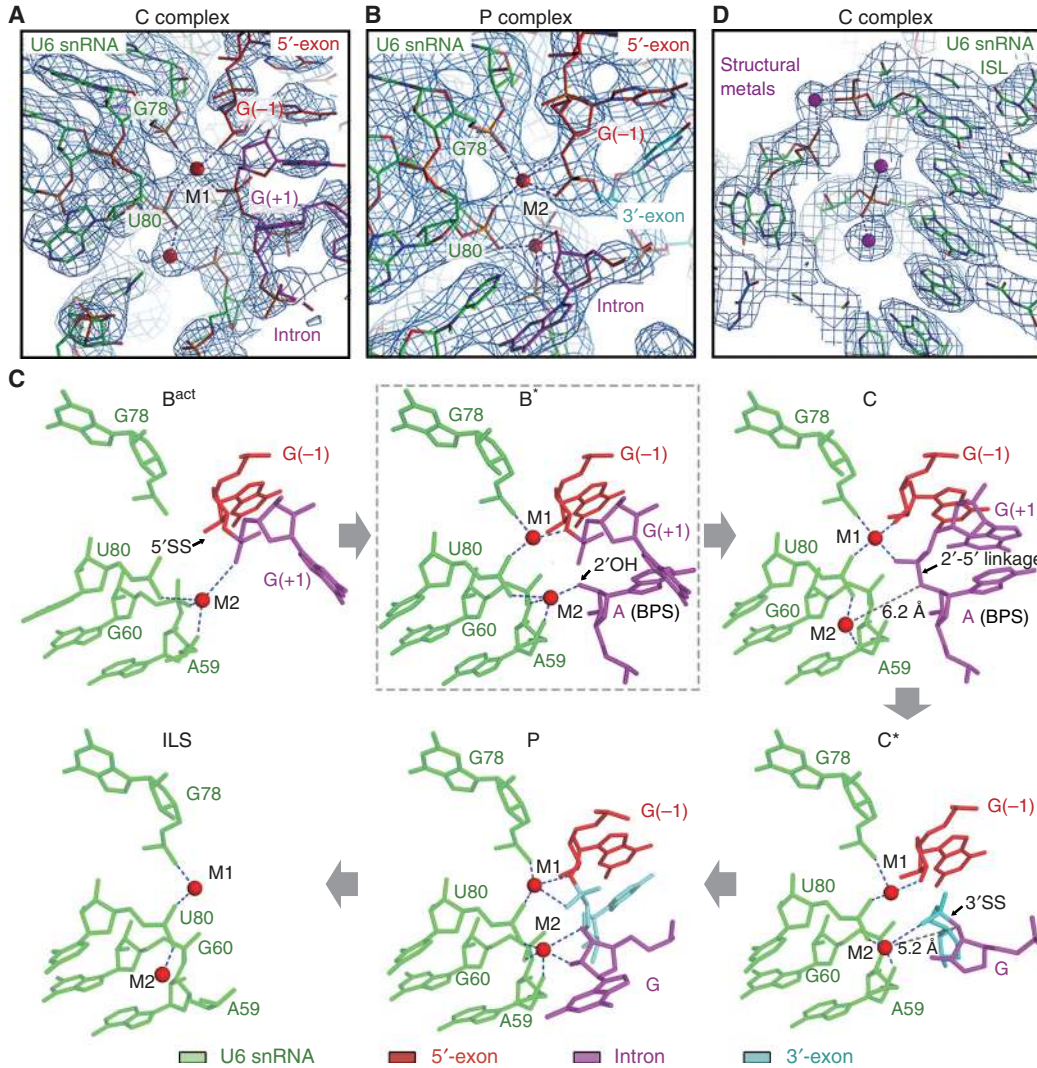
**Table 2.** Coordination of the M1 and M2 metals in different spliceosomal complexes in the yeast of *Saccharomyces cerevisiae*

Ligand	Complex	B <sup>act</sup>		C		C*		P		ILS	
		M1	M2	M1	M2	M1	M2	M1	M2	M1	M2
Identity		–	+	+	+	+	+	+	+	+	+
U6 snRNA	A59: <i>pro-R<sub>p</sub></i>		✓				✓		✓		
	A59: <i>pro-S<sub>p</sub></i>				✓						
	G60: <i>pro-S<sub>p</sub></i>		✓				✓		✓		
	G78: <i>pro-R<sub>p</sub></i>			✓		✓		✓		✓	
	U80: <i>pro-R<sub>p</sub></i>				✓		✓		✓		✓
	U80: <i>pro-S<sub>p</sub></i>			✓		✓		✓		✓	
5'-exon	G-1 at 3' end: 3' OH			✓		✓		✓			
5'SS	G1 at 5' end: <i>pro-R<sub>p</sub></i>			✓		✓		✓			
3'SS	G at 3' end: 2'/3' OH								✓✓		
3'-exon	X at 5' end: <i>pro-R<sub>p</sub></i>							✓			
	Number of ligands		2	4	2	3	3	4	5	2	1

Presence of the M1 and M2 metals is indicated by red and blue check signs, respectively.

The M1 metal ion is uniformly coordinated by two phosphate oxygen atoms from G78 and U80 of U6 snRNA in the spliceosomal C, C\*, P, and ILS complexes (Table 2). The coordination of the M1 metal, as illustrated in the splicing active site center of the *S. cerevisiae* C complex (Wan et al. 2016a), involves two more ligands: a phosphate oxygen atom from the 5'-end of the 5'SS and the 3'-OH

at the 3'-end nucleotide of the 5'-exon (Table 2). These four ligands are arranged in a planar fashion (Fig. 3A). The putatively octahedral coordination suggests M1 to be a divalent ion, likely magnesium ( $Mg^{2+}$ ). In addition to the two uniformly coordinating ligands from U6 snRNA, M1 is bound by the 3'-OH at the 3'-end nucleotide of the 5'-exon in both the C\* (Yan et al. 2017) and P (Bai



**Figure 3.** The spliceosome is a metalloribozyme. (A) Coordination of the catalytic metals M1 and M2. Shown here is metal coordination in the *Saccharomyces cerevisiae* C complex (Wan et al. 2016a). M1 is recognized by four ligands arranged in a planar fashion. The original electron microscopy (EM) density map is shown in panels A, B, and D. PDB code 5GMK and EMDB code EMD-9525 for the C complex; PDB code 5YLZ and EMDB code EMD-6839 for the P complex. (B) Coordination of M1 and M2 in the *S. cerevisiae* P complex (Bai et al. 2017). M2 is bound by five potential ligands. (C) Choreography of the catalytic metals during the splicing reaction. The coordination of M1 and M2 is shown for the spliceosomal B<sup>act</sup>, B\* (predicted), C, C\*, P, and intron lariat spliceosome (ILS) complexes. Elements of the U6 small nuclear RNA (snRNA) are shown in the same orientation in the six complexes to allow meaningful comparison. Metal coordination in the B\* complex is predicted based on the C complex and knowledge of the splicing reaction. (D) The structural metals are defined by an excellent EM density map (shown here for the *S. cerevisiae* C complex; Wan et al. 2016a) and presumably stabilize the intramolecular stem loop (ISL) fold by neutralizing the negative charges of the RNA phosphodiester backbone.



et al. 2017) complexes (Table 2). In the P complex (Bai et al. 2017), M1 is also recognized by a fourth ligand: a phosphate oxygen atom at the 5'-end of the 3'-exon. In the ILS complex (Wan et al. 2017), M1 loses all other ligands and is coordinated only by G78 and U80. Notably, M1 is yet to be loaded into the B<sup>act</sup> complex (Yan et al. 2016).

M2 coordination appears to be more dynamic than M1. The M2 metal ion is stably coordinated by two phosphate oxygen atoms from A59 and U80 of U6 snRNA in the spliceosomal C, C\*, and P complexes (Table 2). The coordination of the M2 metal is exemplified in the active site center of the P complex (Bai et al. 2017; Liu et al. 2017; Wilkinson et al. 2017), in which M2 is bound by five ligands: the 2'-OH and 3'-OH (the leaving group in exon ligation) at the 3'-end of the 3'SS and three phosphate oxygen atoms from the nucleotides A59, G60, and U80 of U6 snRNA (Fig. 3B; Table 2). Although M2 is recognized by only two ligands from A59 and U80 in the *S. cerevisiae* C complex (Wan et al. 2016a), it is coordinated by one more ligand—a phosphate oxygen atom from G60—in the *S. cerevisiae* C\* complex (Table 2) (Fica et al. 2017; Yan et al. 2017). In the *S. cerevisiae* ILS complex (Wan et al. 2017), however, M2 is bound only by one ligand: a phosphate oxygen atom from U80. Although M2 is loaded in the B<sup>act</sup> complex (Yan et al. 2016), it is recognized by the phosphate oxygen atoms of A59 and G60.

This analysis reveals a choreography of the catalytic metals during each cycle of splicing (Fig. 3C). In the B<sup>act</sup> complex (Yan et al. 2016), M2 is loaded in a preactivated fashion and M1 is absent. Both catalytic metals are presumably loaded into the B\* complex, which executes the branching reaction. In the resulting C complex (Wan et al. 2016a), both M1 and M2 are properly coordinated. In the C\* complex, which executes the exon ligation (Fica et al. 2017; Yan et al. 2017), M1 loses one ligand (a phosphate oxygen atom at the 5'-end of the 5'SS) owing to the relocation of the lariat junction and M2 coordination also undergoes rearrangement. In the resulting P complex (Bai et al. 2017; Liu et al. 2017; Wilkinson et al. 2017), M1 gains one more ligand and M2 has two more ligands; the coordination of M1 and M2 appears to be most complete at this stage. Finally, after exon dissociation in the ILS complex (Wan et al. 2017) M1 loses two ligands from the exon, whereas M2 has only one ligand from U80 of U6 snRNA.

In contrast to the changing patterns of the catalytic metals, three structural metal ions are steadily present in the B<sup>act</sup>, C, C\*, P, and ILS complexes as judged by the local features of the EM density map (Fig. 3D). These metals appear to stabilize the delicate fold of the U6 ISL by neutralizing the negative charges of the RNA backbone phosphates. Two of the three metal ions are also present

in the *S. pombe* ILS complex (Hang et al. 2015; Yan et al. 2015).

## 5 THE ROLE OF THE PROTEIN COMPONENTS

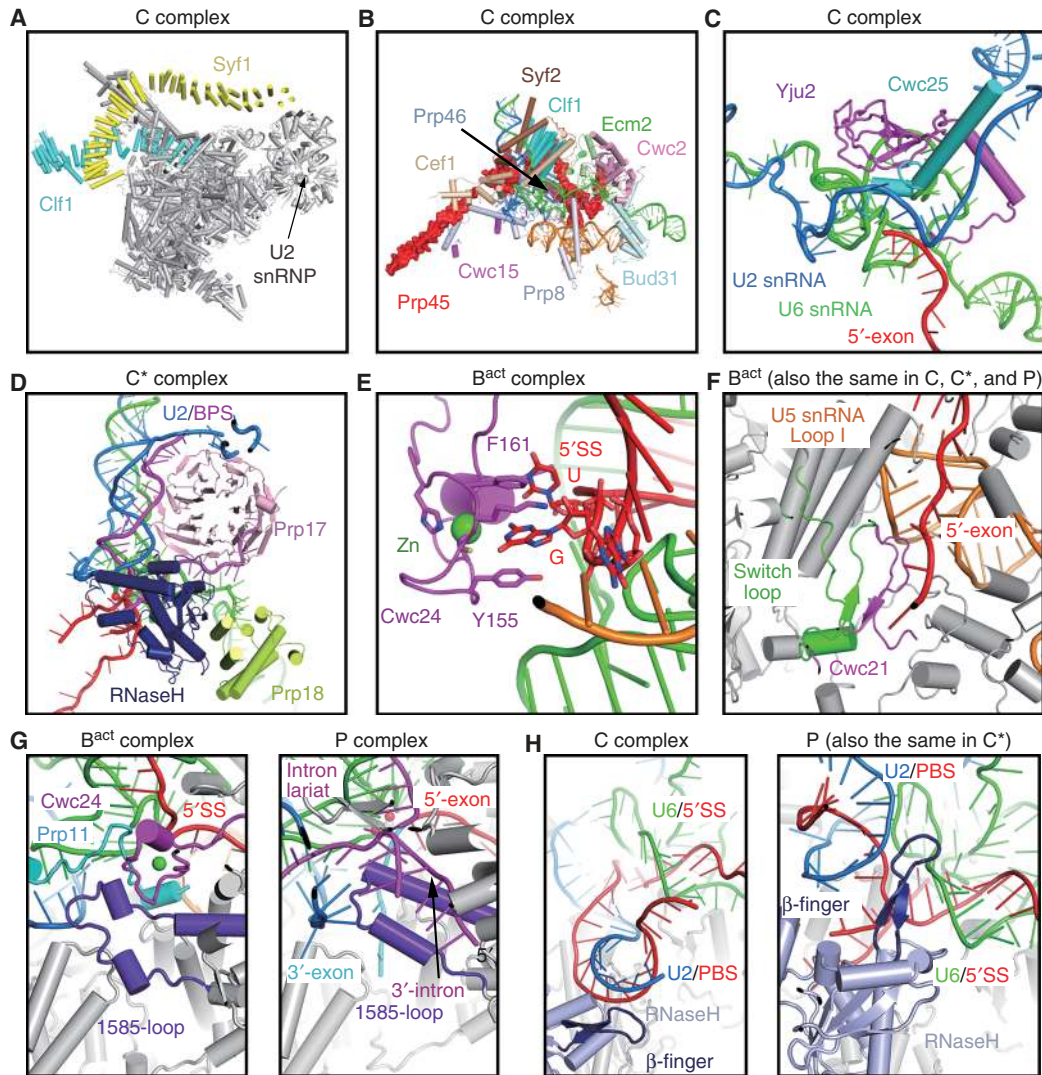
The protein components in each spliceosome can be divided into four distinct groups: the structural proteins, the splicing factors, the RNA-dependent ATPase/helicases, and other regulatory proteins. The structural proteins sustain the splicing active site conformation, support the overall characteristic appearance of the spliceosome, and provide the elasticity that is needed for the splicing reaction. The splicing factors facilitate assembly and activation of the spliceosome and assist the two steps of transesterification. The ATPase/helicases remodel the spliceosomal complexes to allow flux of the splicing factors and other proteins and RNA elements. The other regulatory proteins modulate the splicing of pre-mRNA and will be briefly discussed in a later section.

### 5.1 The Structural Proteins

Both the ribosome and the spliceosome are ribozymes. Unlike the ribosome in which the RNA components constitute the structural scaffold of the entire RNP complex (Ban et al. 2000; Nissen et al. 2000; Schluenzen et al. 2000; Wimberly et al. 2000), the snRNA elements in the spliceosome are flexible and are supported and directed by the surrounding protein components (Shi 2017b). The structural proteins include all protein components of the U2 snRNP core and the U5 snRNP, and virtually all proteins of the NTC and NTR. The structural proteins can be further classified into three distinct categories: the spring-like proteins, the rope-like proteins, and the rest.

The spring-like proteins are exemplified by the superhelical proteins Syf1 and Clf1 (Cwf3 and Cwf4 in *S. pombe*), which can bend and stretch, without being structurally disrupted, to allow large-degree movement of a subcomplex or a region of the spliceosome (Fig. 4A). During the C-to-C\* transition, the U2 snRNP core is translocated by a distance of 60–100 Å, which is mediated by the elastic property of Syf1 (Fica et al. 2017; Yan et al. 2017). The rope-like proteins are represented by Prp45, the carboxy-terminal fragment of Cef1, and Cwc15; these proteins are flexible tethers that glue together multiple components of the spliceosome. For example, the NTR component Prp45 adopts an extended conformation, spans a distance of 150 Å, and simultaneously interacts with nine proteins and two snRNA elements in the spliceosome (Fig. 4B) (Yan et al. 2015). Other than the spring-like and rope-like proteins, each of the remaining protein components has a well-defined conformation and often contains well-folded globular domains.





**Figure 4.** Protein components and catalytic motifs of the spliceosome. (A) The spring-like proteins, exemplified by Syf1 and Clf1 in *Saccharomyces cerevisiae*, allow large conformational rearrangements of the spliceosome while maintaining their structural integrity. The *S. cerevisiae* C complex (Wan et al. 2016a) is used for illustration in panels A–C. Syf1 and Clf1 are colored yellow and cyan, respectively. (B) The rope-like proteins, represented by Prp45 and Cwc15, stabilize spliceosomal organization through direct interactions with multiple proteins and RNA elements. Prp45, colored red and displayed in surface representation, simultaneously interacts with at least nine proteins and two small nuclear RNA (snRNA) elements (U2 and U6). (C) The step I factors Cwc25 and Yju2 constitute an integral part of the splicing active site during the branching reaction. The amino termini of both proteins reach into the center of the active site RNA elements and stabilize their conformations. (D) The step II factors Prp17 and Prp18 stabilize the active site conformation. Shown here is the *S. cerevisiae* C\* complex (Yan et al. 2017). (E) The splicing factor Cwc24 protects the guanine nucleotide at the 5'-end of the 5'SS in the *S. cerevisiae* B<sup>act</sup> complex (Yan et al. 2016). Two aromatic residues Tyr155 and Phe161 sandwich the guanine nucleobase. (F) The splicing factor Cwc21 stabilizes the binding of the 5'-exon to loop I of U5 snRNA. Cwc21 simultaneously interacts with nucleotides of the 5'-exon and the switch loop of Prp8 in the B<sup>act</sup> through P complexes. The interactions, shown here for the *S. cerevisiae* B<sup>act</sup> complex (Yan et al. 2016), are identically preserved in the C, C\*, and P complexes. (G) The 1585-loop of Prp8 in the *S. cerevisiae* B<sup>act</sup> (Yan et al. 2016) (left panel) and P (Bai et al. 2017) (right panel) complexes. (H) The  $\beta$ -finger (from the RNaseH-like domain of Prp8) in the *S. cerevisiae* C (Wan et al. 2016a) (left panel) and P (Bai et al. 2017) (right panel) complexes.

## 5.2 The Splicing Factors

The splicing factors constitute three classes: those that are specific for the branching reaction (step I factors) or the exon ligation (step II factors), and those that are generally required for splicing. The third class is further divided into two categories: those that are required for the early phase of active site formation (Cwc24 and Cwc27) and those that are present throughout both steps of the transesterification (Cwc21, Cwc22).

The step I factors Cwc25 and Yju2 play an indispensable role in stabilizing the active site RNA conformation (Chiu et al. 2009; Warkocki et al. 2009; Schneider et al. 2015), as observed in the cryo-EM structure of the *S. cerevisiae* C complex (Fig. 4C) (Galej et al. 2016; Wan et al. 2016a). Notably, for both Cwc25 and Yju2, the amino-terminal fragment is positioned close to the center of the active site, with conserved residues directly contacting the BPS/U2 duplex. Cwc25 also binds to the helix I of the U2/U6 duplex and the ISL. Yju2 interacts with the 5'-exon, Isy1, and Prp8. These elaborate interactions render Cwc25 and Yju2 integral parts of the splicing active site composite.

The step II factors Prp17, Prp18, and Slu7 play an important role in exon ligation (Jones et al. 1995; Aronova et al. 2007), although they appear to be less critically involved in the splicing active site compared with the step I factors. Indeed, the absence of Prp18 or Slu7 still allowed splicing of some pre-mRNA to proceed when the RNA sequence between the BPS and the 3'SS is shortened (Bryson and Schwer 1996; Zhang and Schwer 1997; James et al. 2002). Nonetheless, these factors are located close to the splicing active site and may facilitate exon ligation through interactions with the protein and RNA elements. In the *S. cerevisiae* C\* (Fica et al. 2017; Yan et al. 2017) and P (Bai et al. 2017; Liu et al. 2017; Wilkinson et al. 2017) complexes, the WD40 domain of Prp17 is located between the BPS/U2 duplex and the 5'SS/U6 duplex (Fig. 4D). Prp18 makes no physical interactions with the RNA elements but directly contacts the RNaseH-like domain of Prp8. The RNaseH-like domain, in turn, interacts with the BPS/U2 duplex in a nearly identical fashion in the spliceosomal C, C\*, and P complexes (Galej et al. 2016; Wan et al. 2016a; Bai et al. 2017; Fica et al. 2017; Liu et al. 2017; Wilkinson et al. 2017; Yan et al. 2017). Both the RNaseH-like domain of Prp8 and the WD40 domain of Prp17 undergo marked translocation in the C-to-C\* transition. Only a small portion of Slu7 was tentatively identified in the *S. cerevisiae* C\* complex (Fica et al. 2017), although the bulk of this protein has been assigned in the human C\* complex (Zhang et al. 2017).

The splicing factor Cwc24, which is recruited into the spliceosome in the B-to-B<sup>act</sup> transition, plays an important

role in the B<sup>act</sup> complex by insulating the 5'-end guanine nucleobase of the 5'SS (Yan et al. 2016). The aromatic side chains of Tyr155 and Phe161 of Cwc24 sandwich the guanine nucleobase (Fig. 4E), whereas Lys160 mediate hydrogen bonds (H-bonds) to the nucleobase and the ribose. Through the action of Prp2, Cwc24 is dissociated in the B<sup>act</sup>-to-B\* transition. The splicing factors Cwc21 and Cwc22 are recruited into the B<sup>act</sup> complex and dissociated in the P-to-ILS transition. The amino-terminal fragment of Cwc21 directly interacts with the 5'-exon to stabilize its binding to loop I of U5 snRNA (Fig. 4F). Cwc21, in turn, is stabilized by the switch loop of Prp8. The 5'-exon-Cwc21-switch loop interactions remain unchanged in the spliceosomal B<sup>act</sup> through P complexes. Cwc22 is located close to Cwc21 and interacts with Prp8, whereas the splicing factor Cwc27 binds both Prp8 and Brr2 and is present only in the B<sup>act</sup> complex.

## 5.3 The Catalytic Motifs of Prp8

Three prominent structural motifs, all from Prp8, appear to facilitate the splicing reaction through close interactions with the active site elements. These motifs include the switch loop and the 1585-loop (also known as  $\alpha$ -finger; Nguyen et al. 2016) of the linker domain and the  $\beta$ -finger (Yang et al. 2008) of the RNaseH-like domain. The switch loop, which comprises an extended  $\beta$ -hairpin-containing structure, stabilizes 5'-exon binding. In all spliceosomes (B<sup>act</sup>, C, C\*, and P) in which the 5'-exon is loaded onto loop I of U5 snRNP, the switch loop is identically arranged (Fig. 4F). Compared with these spliceosomes, the switch loop is flipped by 180° in all structurally characterized complexes in which the interaction of the 5'-exon with U5 loop I is absent, including in the U4/U6.U5 tri-snRNP (Nguyen et al. 2016; Wan et al. 2016b) and in the B (Plaschka et al. 2017) and ILS (Hang et al. 2015; Yan et al. 2015; Wan et al. 2017) complexes.

In the B<sup>act</sup> complex (Yan et al. 2016), the 1585-loop interacts with the U2/U6 duplex and binds the structural elements of Cwc24 and Prp11 that together protect the 5'-end guanine nucleotide of the 5'SS (Fig. 4G, left panel). In the C\* (Fica et al. 2017; Yan et al. 2017) and P (Bai et al. 2017; Liu et al. 2017; Wilkinson et al. 2017) complexes, the tip of the 1585-loop is located close to the 5'-exon and the catalytic metals, contacting the ISL and the lariat junction (Fig. 4G, right panel). In addition, the 1585-loop directly interacts with the pre-mRNA nucleotides of the 3'SS and two nucleobases of the 3'-exon (Bai et al. 2017; Liu et al. 2017; Wilkinson et al. 2017). At the end of the splicing cycle, the 1585-loop is no longer needed and becomes disordered in the ILS complex (Hang et al. 2015; Yan et al. 2015; Wan et al. 2017).

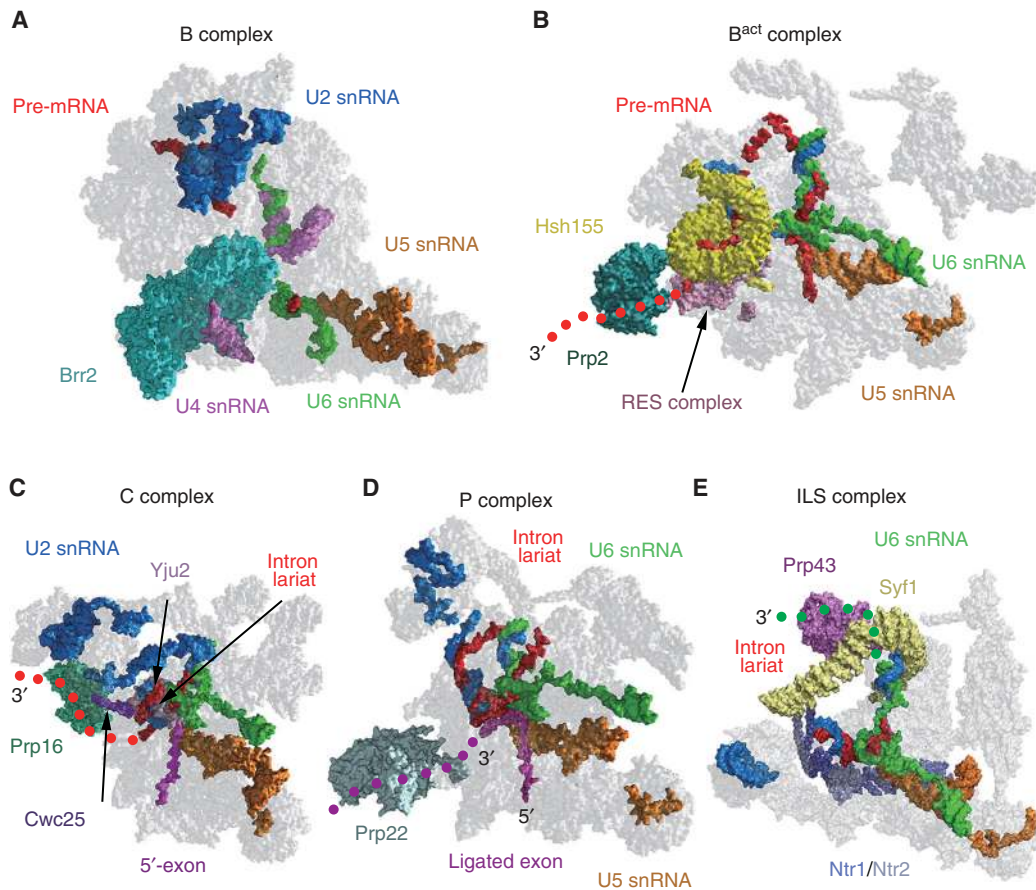
In the B<sup>act</sup> (Rauhut et al. 2016; Yan et al. 2016) and ILS (Hang et al. 2015; Yan et al. 2015; Bai et al. 2017) complexes,



the  $\beta$ -finger remains unengaged, not interacting with any protein or RNA element. In the C (Galej et al. 2016; Wan et al. 2016a) and presumably  $B^*$  complexes, the  $\beta$ -finger interacts with the BPS/U2 duplex (Fig. 4H, left panel). In the  $C^*$  (Fica et al. 2017; Yan et al. 2017) and P (Bai et al. 2017; Liu et al. 2017; Wilkinson et al. 2017) complexes, the  $\beta$ -finger interacts with the lariat junction on the opposite side of the 1585-loop, located between the U2/BPS and U6/5'SS duplexes (Fig. 4H, right panel), and contacts the step II splicing factor Prp17.

#### 5.4 The RNA-Dependent ATPase/Helicases

The cryo-EM structures of the various spliceosomes reveal mechanistic insights into the B-to- $B^{\text{act}}$ ,  $B^{\text{act}}$ -to- $B^*$ , C-to- $C^*$ , and P-to-ILS transitions. Except for the U5 snRNP, nearly all other protein and RNA components of the B complex undergo marked changes—either dissociation or translocation and conformational change—in the transition to the  $B^{\text{act}}$  complex. This transition is mediated by Brr2, which is thought to bind the single-stranded U4 snRNA sequences downstream from the U4/U6 duplex (Fig. 5A). Propelled by



**Figure 5.** Spliceosomal transitions are driven by the ATPase/helicases. (A) The B-to- $B^{\text{act}}$  transition is propelled by Brr2. Shown here is a surface view of the *Saccharomyces cerevisiae* B complex (Plaschka et al. 2017). Brr2 is thought to pull on the single-stranded U4 small nuclear RNA (snRNA) sequences, triggering the unwinding of the U4/U6 duplex and subsequent changes. (B) The  $B^{\text{act}}$ -to- $B^*$  transition is mediated by Prp2. Shown here is a surface view of the *S. cerevisiae*  $B^{\text{act}}$  complex (Yan et al. 2016). Prp2 likely pulls on the 3'-end sequences of the pre-mRNA, leading to dissociation of the retention and splicing complex (RES) complex, the SF3b and SF3a complexes, and the splicing factor Cwc24. (C) The C-to- $C^*$  transition is executed by Prp16. Shown here is a surface view of the *S. cerevisiae* C complex (Wan et al. 2016a). Prp16 presumably pulls on the 3'-end sequences of the intron lariat-3'-exon intermediate. (D) The P-to-intron lariat spliceosome (ILS) transition is driven by Prp22. Shown here is a surface view of the *S. cerevisiae* P complex (Bai et al. 2017). Dissociation of the ligated exon is mediated by Prp22 pulling on its 3'-end sequences. (E) The disassembly of the ILS complex is performed by Prp43. Shown here is a surface view of the *S. cerevisiae* ILS complex (Wan et al. 2017). Prp43 may function by binding to and pulling on either the 3' end of U6 snRNA or the intron lariat. The PDB codes for the B,  $B^{\text{act}}$ , C, P, and ILS complexes are 5NRL, 5GM6, 5GMK/5LJ5, 5YLZ, and 5Y88, respectively.

ATP binding and hydrolysis, Brr2 presumably pulls on the single-stranded RNA sequence (Plaschka et al. 2017), leading to the dissociation of the associated protein factors and unwinding of the U4/U6 duplex. The B-to-B<sup>act</sup> transition is speculated to occur in several distinct steps (Plaschka et al. 2017; Sidarovich et al. 2017; Zhang et al. 2018).

At the periphery of the B<sup>act</sup> complex (Yan et al. 2016), Prp2 binds the SF3b protein Hsh155 and is located close to the RES complex (Fig. 5B). The RNA sequences downstream from the BPS are bound by the RES complex, with the disordered 3'-end sequences presumably recognized by Prp2 (Liu and Cheng 2012; Schneider et al. 2015). This structural arrangement is fully consistent with the notion that pulling on the 3'-end intron sequences may result in the dissociation of the RES complex and the SF3b complex during the B<sup>act</sup>-to-B\* transition. In the C complex (Galej et al. 2016), the RNA-binding site of Prp16 is about 70 Å away from the structurally identified nucleotide at the 3' end of the intron (Fig. 5C). In the C-to-C\* transition, Prp16 would presumably pull on the 3'-end sequences of the intron, leading to the dissociation of the step I factors Cwc25 and the amino-terminal fragment of Yju2 (Tseng et al. 2011) and subsequent recruitment of the step II factors Prp18 and Slu7 (Ansari and Schwer 1995; James et al. 2002). In the P complex (Fig. 5D), Prp22 directly interacts with the Linker and the RT Finger/Palm domains of Prp8; but the RNA-binding site of Prp22 is ~80 Å away from the structurally identified 3'-end nucleotide of the 3'-exon. This structural feature explains why the release of the ligated exon from the P complex is inhibited by shortening the 3'-exon sequence (Schwer 2008; Ilagan et al. 2013). In the P-to-ILS transition, Prp22 is thought to pull on the 3'-end RNA sequences to dissociate the ligated exon (Company et al. 1991; Schwer and Gross 1998; Semlow et al. 2016), triggering the dissociation of Cwc21, Cwc22, Prp18, Slu7, and the WD40 domain of Prp17 and subsequent recruitment of Cwc23 and the disassembly Ntr complex (Prp43, Ntr1, and Ntr2).

The disassembly of the spliceosome is mediated by the Ntr complex (Arenas and Abelson 1997; Tsai et al. 2005, 2007; Fourmann et al. 2016). The structure of the ILS complex suggests two potential mechanisms by which the spliceosome is disassembled by the ATPase/helicase Prp43 (Fig. 5E) (Wan et al. 2017). Prp43 binds the middle portion of the superhelical protein Syf1, whereas Ntr1 and Ntr2 are located away from Prp43. The 3'-end nucleotide of U6 snRNA that has been structurally identified is located only ~50 Å away from the RNA-binding site of Prp43. In contrast, the intron lariat is located 85–100 Å away. Therefore, it appears more likely that Prp43 grabs onto and pulls on the 3'-end of U6 snRNA as opposed to the intron lariat (Bohnsack et al. 2009).

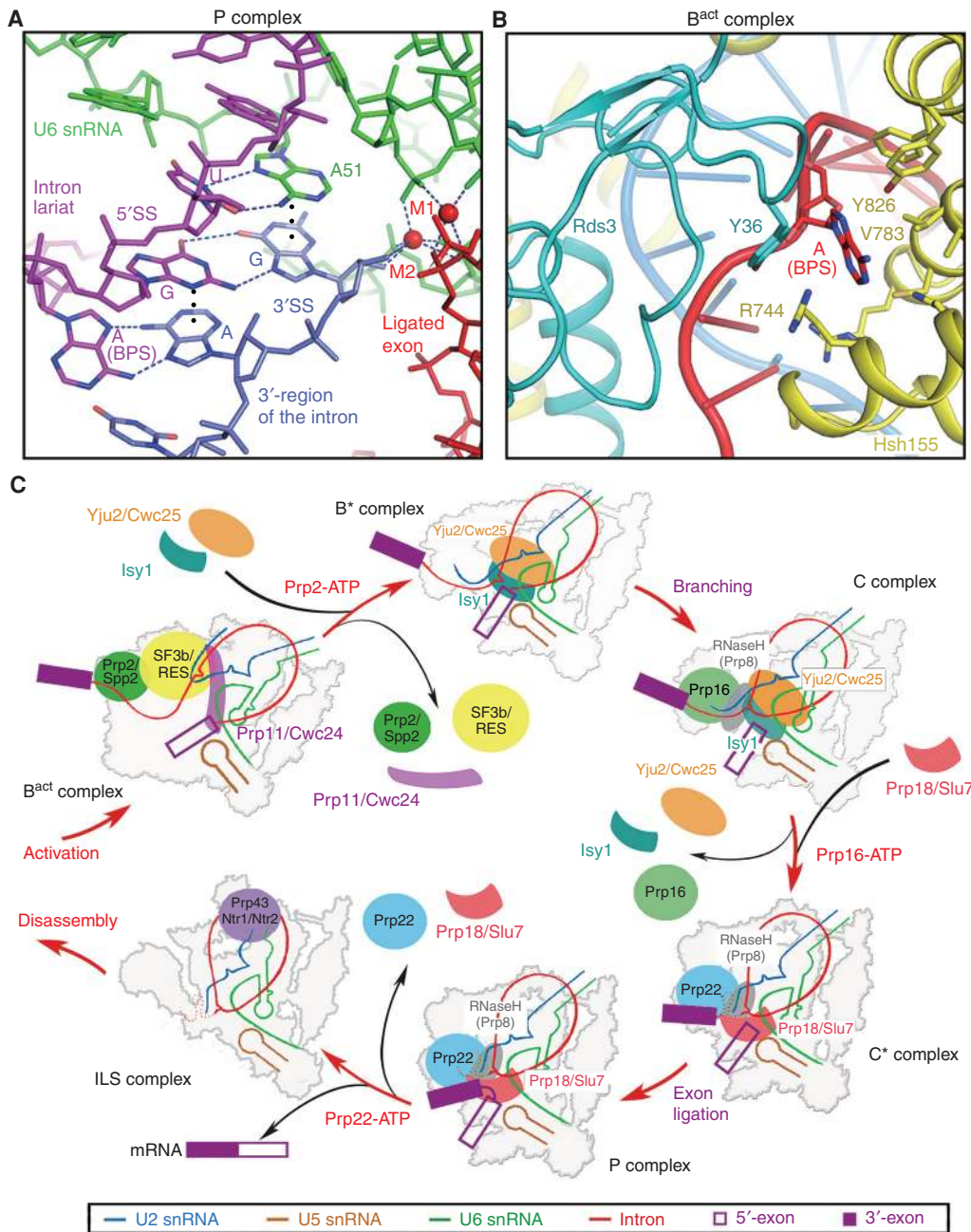
These transitions are, by definition, dynamic and can be inferred only from comparative analysis of the static cryo-EM structures. The steps and mechanisms separating two neighboring spliceosomal complexes could be considerably more dynamic than the qualitative comparison. For example, the human B<sup>act</sup> complex has at least three compositional states—early, mature, and late (Zhang et al. 2018)—and the mature state comprises at least eight distinct conformational states (Haselbach et al. 2018). Importantly, however, the essential features of the various transitions have been structurally observed. These features to a large extent confirm published biochemical and genetic data that have accumulated in the past three decades. Despite the tantalizing interpretation of the structural images, how the ATPase/helicases function mechanically remain to be elucidated, perhaps through single-molecule biophysical investigations. The need for in-depth biophysical studies is reinforced by the fact that these ATPase/helicases are located mostly in the periphery of the spliceosomes and are characterized by medium-to-low resolution, which does not allow accurate assignment of atomic features that are required for faithful mechanistic understanding.

## 6 RECOGNITION OF THE RNA ELEMENTS

The conserved sequence motifs of the pre-mRNA are mainly recognized by conserved RNA elements from both snRNA and pre-mRNA. Three nucleotides at the 3'-end of the 5'-exon are anchored on the loop I of U5 snRNA in the B<sup>act</sup> through P complexes. The BPS and surrounding sequences of the intron form an elongated duplex with complementary sequences of the U2 snRNA in the A through ILS complexes. The dinucleotides GU at the 5' end of the 5'SS are recognized by the Linker domain of Prp8 in the *S. cerevisiae* U4/U6.U5 tri-snRNP (Wan et al. 2016b) and the human B complex (Bertram et al. 2017a) and by Cwc24 in the B<sup>act</sup> complex (Yan et al. 2016); the six consecutive nucleotides downstream from GU of the 5'SS are recognized by U6 snRNA in the B<sup>act</sup> through ILS complexes. In the P (Bai et al. 2017; Liu et al. 2017; Wilkinson et al. 2017) and possibly C\* complexes, the dinucleotides AG of the 3'SS are recognized by the lariat junction that is formed after the branching reaction (Fig. 6A). Notably, AG of the 3'SS pair up with the nucleophile-containing A of the BPS and the invariant G at the 5'-end of the 5'SS, respectively, through noncanonical Watson–Crick H-bonds.

The spliceosome is a protein-directed metalloribozyme. The RNA elements, both the immobile portion (U5, U6, the 5'-end portion of U2, and the 3'-end portion of the 5'-exon) and the mobile segments (U2 downstream from nucleotide 30, the bulk of the intron, and the 3'-exon), are specifically recognized by protein components. This subject has been





**Figure 6.** Choreography of the protein and RNA components during the splicing reaction. (A) Recognition of the AG dinucleotide at the 3'SS by the lariat junction in the P complex (Bai et al. 2017; Liu et al. 2017; Wilkinson et al. 2017). Notably, the AG nucleobases at the 3'SS form noncanonical Watson–Crick base-pairing interactions with two consecutive AG nucleobases of the lariat junction. (B) Recognition of the nucleophile-containing adenine nucleotide of the branch point sequence (BPS) by Hsh155. The adenine nucleobase is already flipped out of the registry of the intron/U2 duplex and is in close contact with a few residues of Hsh155 and Rds3. (C) A cartoon diagram depicting the activation of the B<sup>act</sup> complex, the branching and exon ligation, and the disassembly of the intron lariateosome (ILS) complex.

thoroughly reviewed recently (Shi 2017a) and is thus only briefly discussed here. The immobile portion is mainly recognized by the 17 static protein components. For example, the active site RNA elements are cradled in the positively charged catalytic cavity that is formed between the N-domain and the core of the Prp8 (Fig. 2D). The mobile RNA segments are recognized and directed by conserved protein elements. For example, in the B<sup>act</sup> complex (Rauhut et al. 2016; Yan et al. 2016), the BPS/U2 duplex is surrounded by Hsh155 and the nucleophile-containing adenine nucleobase of the BPS is flipped out of the U2/BPS duplex registry and recognized by specific residues of Hsh155 (Fig. 6B). The 5'-end guanine nucleotide of the 5'SS is recognized by the splicing factor Cwc24 (Fig. 4E) and the SF3a protein Prp11.

## 7 CHOREOGRAPHY OF THE RNA AND PROTEIN COMPONENTS

Three snRNA molecules, U5, U6, and 30 nucleotides at the 5'-end of U2, remain largely static in the spliceosomal B<sup>act</sup> through ILS complexes, with only minor shift (mostly <3 Å) for select local segments. The static segments of U2 and U6 form helix I and helix II. The 5'-exon of the pre-mRNA remains identically anchored to the loop I of U5 snRNA in the spliceosomal B<sup>act</sup> through P complexes. The sequences downstream from nucleotide 30 of U2 snRNA, along with segments of the intron, undergo a series of dramatic movements that are designed to facilitate branching and exon ligation.

In the B<sup>act</sup> complex (Rauhut et al. 2016; Yan et al. 2016), conserved U2 sequences form a duplex with the BPS, which is sequestered away from the splicing active site by components of the SF3b and SF3a complexes and the splicing factor Cwc24. In the B<sup>act</sup>-to-B\* transition, Prp2 presumably pulls on the 3'-end of the pre-mRNA, causing dissociation of the SF3a and SF3b complexes, the RES complex, and Cwc24 (Fig. 6C). The step I factors Cwc25 and Yju2 and the NTC component Isy1 are recruited into the splicing active site to stabilize the conformation required for the branching reaction. Segments of the U2 snRNA and the intron sequences are translocated by 50–100 Å, bringing the nucleophile—the invariant adenine nucleotide of the BPS—into the close proximity of the 5'SS as observed in the C complex (Fig. 6C) (Galej et al. 2016; Wan et al. 2016a).

In the C complex (Galej et al. 2016; Wan et al. 2016a), the lariat junction spatially blocks the entry of the 3'SS-3'-exon sequences into the splicing active site for exon ligation. In the C-to-C\* transition, through the action of Prp16, the lariat junction is translocated by approximately 20 Å to vacate space for the 3'SS-3'-exon. Cwc25, Yju2, and Isy1 dissociate from the C complex, and the step II factors Prp18 and Slu7 are recruited to form the C\* complex (Fig. 6C).

Much of U2 snRNA undergoes drastic movement in the transition, and a large portion of the intron lariat is also flipped by up to 60 Å. In the P complex (Bai et al. 2017; Liu et al. 2017; Wilkinson et al. 2017), the dinucleotides AG of the 3'SS form base-pairing interactions with two consecutive nucleotides AG of the lariat junction, which represent the nucleophile-containing adenine nucleotide of the BPS and the 5'-end guanine nucleotide of the 5'SS.

In the P complex (Bai et al. 2017; Liu et al. 2017; Wilkinson et al. 2017), Prp22 binds the Linker domain of Prp8 at a corner of the spliceosome. In the P-to-ILS transition, Prp22 likely pulls on the 3'-end sequences of the ligated exon, leading to its release from the spliceosome and subsequent dissociation of the step II factors Prp18 and Slu7 (Fig. 6C). Finally, in the ILS complex (Wan et al. 2017), Prp43 may pull on the 3'-end nucleotide of U6 snRNA or the intron lariat, leading to the disassembly of the spliceosome (Fig. 6C).

## 8 THE HUMAN SPliceOSOME

Compared with yeast, the human spliceosome is considerably more dynamic and less stable. Isolation of the intact human spliceosome toward structural studies requires application of chemical cross-linking reagents (Table 1). Cryo-EM investigation of the human spliceosome has yielded structures of the B (Bertram et al. 2017a), B<sup>act</sup> (Haselbach et al. 2018; Zhang et al. 2018), C (Zhan et al. 2018), and C\* (Bertram et al. 2017b; Zhang et al. 2017) complexes at resolutions that allow identification of structural details (Table 1). The overall features as well as the detailed interactions of the human spliceosomal complexes are very similar to those of the corresponding yeast complexes. These shared features include the overall organization, the splicing active site conformation, the location of catalytic and structural metal ions, the static protein and RNA components, and the vast majority of the protein–protein, protein–RNA, and RNA–RNA interactions. Here, we briefly describe only the unique features that have emerged from the cryo-EM structures of the human spliceosome. A detailed structural analysis of the human spliceosome and its comparison with the yeast spliceosome will be more meaningful after structural resolution of the human P and ILS complexes.

The human spliceosome generally contains more protein components than its yeast counterpart. Compared with the *S. cerevisiae* structure (Yan et al. 2016), the structure of the human mature B<sup>act</sup> complex has 16 additional proteins: SF3a120, SF3a60, SF3b14a/p14, U5-40K, nine proteins (Lea1, Msl1, and the Sm complex) of the U2 snRNP core, and three proteins (Aquarius, RBM22, and PPIL1) of the NTR (Haselbach et al. 2018; Zhang et al. 2018). The structure of the human C complex contains 11 more



proteins, which include four PPIs, four proteins of the exon junction complex, U5-40K, RBM22, and Aquarius (Zhan et al. 2018). Compared with the *S. cerevisiae* C\* complex (Yan et al. 2017), the structure of the human complex has six additional proteins: Brr2, Slu7, the functionally mysterious component PRKRIP1, and three proteins (MAGOH, Y14, and MLN51) of the exon junction complex (EJC) (Zhang et al. 2017).

As exemplified by the six additional proteins that may assist exon ligation in the human C\* complex (Zhang et al. 2017), these additional proteins apparently play an important role in the function of the corresponding human spliceosome. The human protein RBM22 appears to be a fusion between two yeast proteins Cwc2 and Ecm2 (Rasche et al. 2012). Notably, RBM22 is present in the mature, but not the early, B<sup>act</sup> complex (Zhang et al. 2018). PRKRIP1 forms an elongated helix and links the U2 snRNP core to the splicing active site (Zhang et al. 2017). Compared with its yeast counterpart, the human spliceosome appears to have more conformational states within the same spliceosomal complex. For example, the human B<sup>act</sup> complex assumes three compositionally different states: early, mature, and late B<sup>act</sup> complexes (Zhang et al. 2018). The mature B<sup>act</sup> complex has at least eight distinct conformational states (Haselbach et al. 2018).

## 9 CONCLUSION

During the three years since 2015, mechanistic understanding of spliceosome function through structural elucidation has taken a qualitative leap, beginning with the 3.6-Å cryo-EM structure of the *S. pombe* ILS complex (Hang et al. 2015; Yan et al. 2015). At present, with 18 independent structures of the spliceosome within the overall resolution range of 3.3–9.9 Å, we have developed a reasonably good understanding of the molecular mechanisms of spliceosome-mediated pre-mRNA splicing. With structures of the B, B<sup>act</sup>, C, C\*, P, and ILS complexes from *S. cerevisiae*, we have recapitulated the processes of spliceosome activation, the branching reaction and exon ligation, and disassembly of the spliceosome. We have the structure of the U4/U6.U5 tri-snRNP, but not the pre-spliceosome (A complex) or the early spliceosome (E complex). Importantly, the structure of the pre-B complex remains elusive. Consequently, we are yet to piece together the assembly process of the spliceosome.

The static structural images of the various spliceosomal complexes provide a physical basis for understanding the functions of the spliceosome. But these structures alone fail to answer some of the important mechanistic questions. For example, the RNA-dependent ATPase/helicases, which are responsible for RNP remodeling, are located at the periph-

ery of the spliceosome and defined by relatively poor EM density maps at medium-to-low resolution, which disallows identification of potentially bound ATP molecules and/or RNA nucleotides. Consequently, how these ATPase/helicases function in the context of the spliceosome remains to be structurally characterized. Other biophysical approaches may provide advantages complementary to cryo-EM. For example, single-molecule fluorescence studies of the spliceosomal transitions by the ATPase/helicases have begun to reveal their underlying molecular mechanisms (Semlow et al. 2016). Adaptation of such investigative approaches to the mammalian PPIs may provide additional insights to regulation of the spliceosome.

The best local resolution achieved for an intact spliceosome is 2.8 Å for the stable core of the *S. cerevisiae* B<sup>act</sup> complex (Yan et al. 2016), which displays clearly distinguishable features for amino acid side chains. At this resolution, water molecules and the metal ions are yet to be conclusively identified. Improved resolution may also allow conclusive identification of posttranslational protein modifications and small-molecule ligands that target the various spliceosomal proteins. This information, in turn, may guide rational design of splicing modulators with therapeutic potential.

For technical reasons, structural advances in 2015 and 2016 have been accomplished almost exclusively for the yeast spliceosome. Major progress was achieved in 2017 for the structures of the human B and C\* complexes, followed by the structures of the B<sup>act</sup> and C complexes in the beginning of 2018. It is anticipated that in the near future major advances for the human spliceosome may parallel those for the yeast spliceosome. We will be able to perform a comprehensive comparison of the yeast and human spliceosomes. Lessons gained from such a comparison will not only allow appreciation of the fine species distinctions but also engender discovery of potential modulators that selectively target the spliceosome from one, but not the other, species.

## ACKNOWLEDGMENTS

We thank Rui Bai, Xiaofeng Zhang, and Xiechao Zhan for help in figure preparation and suggestions on text and references. This work was supported by funds from Beijing Advanced Innovation Center for Structural Biology at Tsinghua University, Beijing, China.

## REFERENCES

Agafonov DE, Kastner B, Dybkov O, Hofele RV, Liu WT, Urlaub H, Luhrmann R, Stark H. 2016. Molecular architecture of the human U4/U6.U5 tri-snRNP. *Science* 351: 1416–1420.



- Andersen CB, Ballut L, Johansen JS, Chamieh H, Nielsen KH, Oliveira CL, Pedersen JS, Seraphin B, Le Hir H, Andersen GR. 2006. Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. *Science* **313**: 1968–1972.
- Ansari A, Schwer B. 1995. SLU7 and a novel activity, SSF1, act during the PRP16-dependent step of yeast pre-mRNA splicing. *EMBO J* **14**: 4001–4009.
- Arenas JE, Abelson JN. 1997. Prp43: An RNA helicase-like factor involved in spliceosome disassembly. *Proc Natl Acad Sci* **94**: 11798–11802.
- Aronova A, Bacikova D, Crotti LB, Horowitz DS, Schwer B. 2007. Functional interactions between Prp8, Prp18, Slu7, and U5 snRNA during the second step of pre-mRNA splicing. *RNA* **13**: 1437–1444.
- Bai R, Yan C, Wan R, Lei J, Shi Y. 2017. Structure of the post-catalytic spliceosome from *Saccharomyces cerevisiae*. *Cell* **171**: 1589–1598.
- Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**: 905–920.
- Behzadnia N, Golas MM, Hartmuth K, Sander B, Kastner B, Deckert J, Dube P, Will CL, Urlaub H, Stark H, et al. 2007. Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. *EMBO J* **26**: 1737–1748.
- Berget SM, Roberson BL. 1986. U1, U2, and U4/U6 small nuclear ribonucleoproteins are required for in vitro splicing but not polyadenylation. *Cell* **46**: 691–696.
- Berget SM, Moore C, Sharp PA. 1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci* **74**: 3171–3175.
- Bertram K, Agafonov DE, Dybkov O, Haselbach D, Leelaram MN, Will CL, Urlaub H, Kastner B, Luhrmann R, Stark H. 2017a. Cryo-EM Structure of a pre-catalytic human spliceosome primed for activation. *Cell* **170**: 701–713.
- Bertram K, Agafonov DE, Liu WT, Dybkov O, Will CL, Hartmuth K, Urlaub H, Kastner B, Stark H, Luhrmann R. 2017b. Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature* **542**: 318–323.
- Bessonov S, Anokhina M, Krasauskas A, Golas MM, Sander B, Will CL, Urlaub H, Stark H, Luhrmann R. 2010. Characterization of purified human B<sup>act</sup> spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. *RNA* **16**: 2384–2403.
- Black DL, Chabot B, Steitz JA. 1985. U2 as well as U1 small nuclear ribonucleoproteins are involved in premessenger RNA splicing. *Cell* **42**: 737–750.
- Boehringer D, Makarov EM, Sander B, Makarova OV, Kastner B, Luhrmann R, Stark H. 2004. Three-dimensional structure of a pre-catalytic human spliceosomal complex B. *Nat Struct Mol Biol* **11**: 463–468.
- Boesler C, Rigo N, Anokhina MM, Tauchert MJ, Agafonov DE, Kastner B, Urlaub H, Ficner R, Will CL, Luhrmann R. 2016. A spliceosome intermediate with loosely associated tri-snRNP accumulates in the absence of Prp28 ATPase activity. *Nat Commun* **7**: 11997.
- Bohnsack MT, Martin R, Granneman S, Ruprecht M, Schleiff E, Tollervy D. 2009. Prp43 bound at different sites on the pre-rRNA performs distinct functions in ribosome synthesis. *Mol Cell* **36**: 583–592.
- Bon E, Casaregola S, Blandin G, Llorente B, Neveuglise C, Munsterkotter M, Guldener U, Mewes HW, Van Helden J, Dujon B, et al. 2003. Molecular evolution of eukaryotic genomes: Hemiascomycetous yeast spliceosomal introns. *Nucleic Acids Res* **31**: 1121–1135.
- Brody E, Abelson J. 1985. The “spliceosome”: Yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. *Science* **228**: 963–967.
- Brys A, Schwer B. 1996. Requirement for SLU7 in yeast pre-mRNA splicing is dictated by the distance between the branchpoint and the 3' splice site. *RNA* **2**: 707–717.
- 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* **280**: 31190–31199.
- Chan SP, Kao DI, Tsai WY, Cheng SC. 2003. The Prp19p-associated complex in spliceosome activation. *Science* **302**: 279–282.
- Chen CH, Yu WC, Tsao TY, Wang LY, Chen HR, Lin JY, Tsai WY, Cheng SC. 2002. Functional and physical interactions between components of the Prp19p-associated complex. *Nucleic Acids Res* **30**: 1029–1037.
- Chen W, Shulha HP, Ashar-Patel A, Yan J, Green KM, Query CC, Rhind N, Weng Z, Moore MJ. 2014. Endogenous U2.U5.U6 snRNA complexes in *S. pombe* are intron lariat spliceosomes. *RNA* **20**: 308–320.
- Chiu YF, Liu YC, Chiang TW, Yeh TC, Tseng CK, Wu NY, Cheng SC. 2009. Cwc25 is a novel splicing factor required after Prp2 and Yju2 to facilitate the first catalytic reaction. *Mol Cell Biol* **29**: 5671–5678.
- Chow LT, Gelinis RE, Broker TR, Roberts RJ. 1977. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* **12**: 1–8.
- Company M, Arenas J, Abelson J. 1991. Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. *Nature* **349**: 487–493.
- Cordin O, Hahn D, Beggs JD. 2012. Structure, function and regulation of spliceosomal RNA helicases. *Curr Opin Cell Biol* **24**: 431–438.
- Cretu C, Schmitzova J, Ponce-Salvatierra A, Dybkov O, De Laurentiis EI, Sharma K, Will CL, Urlaub H, Luhrmann R, Pena V. 2016. Molecular architecture of SF3b and structural consequences of its cancer-related mutations. *Mol Cell* **64**: 307–319.
- Deckert J, Hartmuth M, Boehringer D, Behzadnia N, Will CL, Kastner B, Stark H, Urlaub H, Luhrmann R. 2006. Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol Cell Biol* **26**: 5528–5543.
- DiMaria PR, Kaltwasser G, Goldenberg CJ. 1985. Partial purification and properties of a pre-mRNA splicing activity. *J Biol Chem* **260**: 1096–1102.
- Fabrizio P, Dannenberg J, Dube P, Kastner B, Stark H, Urlaub H, Luhrmann R. 2009. The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol Cell* **36**: 593–608.
- Fica SM, Tuttle N, Novak T, Li NS, Lu J, Koodathingal P, Dai Q, Staley JP, Piccirilli JA. 2013. RNA catalyzes nuclear pre-mRNA splicing. *Nature* **503**: 229–234.
- Fica SM, Oubridge C, Galej WP, Wilkinson ME, Bai XC, Newman AJ, Nagai K. 2017. Structure of a spliceosome remodelled for exon ligation. *Nature* **542**: 377–380.
- Fourmann JB, Dybkov O, Agafonov DE, Tauchert MJ, Urlaub H, Ficner R, Fabrizio P, Luhrmann R. 2016. The target of the DEAH-box NTP triphosphatase Prp43 in *Saccharomyces cerevisiae* spliceosomes is the U2 snRNP-intron interaction. *eLife* **5**: e15564.
- Frendewey D, Keller W. 1985. Stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. *Cell* **42**: 355–367.
- Furman E, Glitz DG. 1995. Purification of the spliceosome A-complex and its visualization by electron microscopy. *J Biol Chem* **270**: 15515–15522.
- Galej WP, Oubridge C, Newman AJ, Nagai K. 2013. Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature* **493**: 638–643.
- Galej WP, Wilkinson ME, Fica SM, Oubridge C, Newman AJ, Nagai K. 2016. Cryo-EM structure of the spliceosome immediately after branching. *Nature* **537**: 197–201.
- Golas MM, Sander B, Bessonov S, Grote M, Wolf E, Kastner B, Stark H, Luhrmann R. 2010. 3D cryo-EM structure of an active step I spliceosome and localization of its catalytic core. *Mol Cell* **40**: 927–938.
- Grabowski PJ, Sharp PA. 1986. Affinity chromatography of splicing complexes: U2, U5, and U4 + U6 small nuclear ribonucleoprotein particles in the spliceosome. *Science* **233**: 1294–1299.
- Grabowski PJ, Padgett RA, Sharp PA. 1984. Messenger RNA splicing in vitro: An excised intervening sequence and a potential intermediate. *Cell* **37**: 415–427.
- Grabowski PJ, Seiler SR, Sharp PA. 1985. A multicomponent complex is involved in the splicing of messenger RNA precursors. *Cell* **42**: 345–353.
- Hang J, Wan R, Yan C, Shi Y. 2015. Structural basis of pre-mRNA splicing. *Science* **349**: 1191–1198.

- Haselbach D, Komarov I, Agafonov DE, Hartmuth K, Graf B, Dybkov O, Urlaub H, Kastner B, Luhrmann R, Stark H. 2018. Structure and conformational dynamics of the human spliceosomal B<sup>act</sup> complex. *Cell* **172**: 454–464.e411.
- Ilgan JO, Chalkley RJ, Burlingame AL, Jurica MS. 2013. Rearrangements within human spliceosomes captured after exon ligation. *RNA* **19**: 400–412.
- James SA, Turner W, Schwer B. 2002. How Slu7 and Prp18 cooperate in the second step of yeast pre-mRNA splicing. *RNA* **8**: 1068–1077.
- Jones MH, Frank DN, Guthrie C. 1995. Characterization and functional ordering of Slu7p and Prp17p during the second step of pre-mRNA splicing in yeast. *Proc Natl Acad Sci* **92**: 9687–9691.
- Jurica MS, Licklider LJ, Gygi SR, Grigorieff N, Moore MJ. 2002. Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA* **8**: 426–439.
- Jurica MS, Sousa D, Moore MJ, Grigorieff N. 2004. Three-dimensional structure of C complex spliceosomes by electron microscopy. *Nat Struct Mol Biol* **11**: 265–269.
- Keating KS, Toor N, Perlman PS, Pyle AM. 2010. A structural analysis of the group II intron active site and implications for the spliceosome. *RNA* **16**: 1–9.
- 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* doi: 10.7554/eLife.04986.
- Krainer AR, Maniatis T. 1985. Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing in vitro. *Cell* **42**: 725–736.
- Kramer A, Keller W, Appel B, Luhrmann R. 1984. The 5' terminus of the RNA moiety of U1 small nuclear ribonucleoprotein particles is required for the splicing of messenger RNA precursors. *Cell* **38**: 299–307.
- Kupfer DM, Drabenstot SD, Buchanan KL, Lai H, Zhu H, Dyer DW, Roe BA, Murphy JW. 2004. Introns and splicing elements of five diverse fungi. *Eukaryot Cell* **3**: 1088–1100.
- Lerner MR, Steitz JA. 1979. Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *Proc Natl Acad Sci* **76**: 5495–5499.
- Lerner MR, Boyle JA, Mount SM, Wolin SL, Steitz JA. 1980. Are snRNPs involved in splicing? *Nature* **283**: 220–224.
- Leung AK, Nagai K, Li J. 2011. Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis. *Nature* **473**: 536–539.
- Li X, Liu S, Jiang J, Zhang L, Espinosa S, Hill RC, Hansen KC, Zhou ZH, Zhao R. 2017. CryoEM structure of *Saccharomyces cerevisiae* U1 snRNP offers insight into alternative splicing. *Nat Commun* **8**: 1035.
- Lin PC, Xu RM. 2012. Structure and assembly of the SF3a splicing factor complex of U2 snRNP. *EMBO J* **31**: 1579–1590.
- Liu HL, Cheng SC. 2012. The interaction of Prp2 with a defined region of the intron is required for the first splicing reaction. *Mol Cell Biol* **32**: 5056–5066.
- Liu S, Li X, Zhang L, Jiang J, Hill RC, Cui Y, Hansen KC, Zhou ZH, Zhao R. 2017. Structure of the yeast spliceosomal postcatalytic P complex. *Science* **358**: 1278–1283.
- Luhrmann R, Stark H. 2009. Structural mapping of spliceosomes by electron microscopy. *Curr Opin Struct Biol* **19**: 96–102.
- Montemayor EJ, Curran EC, Liao HH, Andrews KL, Treba CN, Butcher SE, Brow DA. 2014. Core structure of the U6 small nuclear ribonucleoprotein at 1.7-Å resolution. *Nat Struct Mol Biol* **21**: 544–551.
- 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**: 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**: 80–84.
- Nguyen TH, Li J, Galej WP, Oshikane H, Newman AJ, Nagai K. 2013. Structural basis of Brr2-Prp8 interactions and implications for U5 snRNP biogenesis and the spliceosome active site. *Structure* **21**: 910–919.
- Nguyen TH, Galej WP, Bai XC, Savva CG, Newman AJ, Scheres SH, Nagai K. 2015. The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature* **523**: 47–52.
- Nguyen TH, Galej WP, Bai XC, Oubridge C, Newman AJ, Scheres SH, Nagai K. 2016. Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. *Nature* **530**: 298–302.
- Nissen P, Hansen J, Ban N, Moore PB, Steitz TA. 2000. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**: 920–930.
- Ohi MD, Gould KL. 2002. Characterization of interactions among the Cef1p-Prp19p-associated splicing complex. *RNA* **8**: 798–815.
- Ohi MD, Ren L, Wall JS, Gould KL, Walz T. 2007. Structural characterization of the fission yeast U5.U2/U6 spliceosome complex. *Proc Natl Acad Sci* **104**: 3195–3200.
- Ohr T, Prior M, Dannenberg J, Odenwalder P, Dybkov O, Rasche N, Schmitzova J, Gregor I, Fabrizio P, Enderlein J, et al. 2012. Prp2-mediated protein rearrangements at the catalytic core of the spliceosome as revealed by dcFCCS. *RNA* **18**: 1244–1256.
- Padgett RA, Mount SM, Steitz JA, Sharp PA. 1983. Splicing of messenger RNA precursors is inhibited by antisera to small nuclear ribonucleoprotein. *Cell* **35**: 101–107.
- Padgett RA, Konarska MM, Grabowski PJ, Hardy SF, Sharp PA. 1984. Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. *Science* **225**: 898–903.
- Pikielny CW, Rosbash M. 1986. Specific small nuclear RNAs are associated with yeast spliceosomes. *Cell* **45**: 869–877.
- Plaschka C, Lin PC, Nagai K. 2017. Structure of a pre-catalytic spliceosome. *Nature* **546**: 617–621.
- Pomeranz Krummel DA, Oubridge C, Leung AK, Li J, Nagai K. 2009. Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature* **458**: 475–480.
- Price SR, Evans PR, Nagai K. 1998. Crystal structure of the spliceosomal U2B'-U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature* **394**: 645–650.
- Rasche N, Dybkov O, Schmitzova J, Akyildiz B, Fabrizio P, Luhrmann R. 2012. Cwc2 and its human homologue RBM22 promote an active conformation of the spliceosome catalytic centre. *EMBO J* **31**: 1591–1604.
- Rauhut R, Fabrizio P, Dybkov O, Hartmuth K, Pena V, Chari A, Kumar V, Lee CT, Urlaub H, Kastner B, et al. 2016. Molecular architecture of the *Saccharomyces cerevisiae* activated spliceosome. *Science* **353**: 1399–1405.
- Rogers J, Wall R. 1980. A mechanism for RNA splicing. *Proc Natl Acad Sci* **77**: 1877–1879.
- Ruskin B, Krainer AR, Maniatis T, Green MR. 1984. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. *Cell* **38**: 317–331.
- Schlauenzen F, Tocilj A, Zarivach R, Harms J, Gluehmann M, Janell D, Bashan A, Bartels H, Agmon I, Franceschi F, et al. 2000. Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. *Cell* **102**: 615–623.
- 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**: e1005539.
- Schwer B. 2008. A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. *Mol Cell* **30**: 743–754.
- Schwer B, Gross CH. 1998. Prp22, a DEXH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. *EMBO J* **17**: 2086–2094.
- Semlow DR, Blanco MR, Walter NG, Staley JP. 2016. Spliceosomal DEAH-box ATPases remodel pre-mRNA to activate alternative splice sites. *Cell* **164**: 985–998.
- Shi Y. 2017a. Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat Rev Mol Cell Biol* **18**: 655–670.

- Shi Y. 2017b. The spliceosome: A protein-directed metalloribozyme. *J Mol Biol* **429**: 2640–2653.
- Sickmier EA, Frato KE, Shen H, Paranawithana SR, Green MR, Kielkopf CL. 2006. Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. *Mol Cell* **23**: 49–59.
- Sidarovich A, Will CL, Anokhina MM, Ceballos J, Sievers S, Agafonov DE, Samatov T, Bao P, Kastner B, Urlaub H, et al. 2017. Identification of a small molecule inhibitor that stalls splicing at an early step of spliceosome activation. *eLife* **6**: e23533.
- Sontheimer EJ, Sun S, Piccirilli JA. 1997. Metal ion catalysis during splicing of premessenger RNA. *Nature* **388**: 801–805.
- Spingola M, Grate L, Haussler D, Ares M Jr. 1999. Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. *RNA* **5**: 221–234.
- Steitz TA, Steitz JA. 1993. A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci* **90**: 6498–6502.
- Tarn WY, Hsu CH, Huang KT, Chen HR, Kao HY, Lee KR, Cheng SC. 1994. Functional association of essential splicing factor(s) with PRP19 in a protein complex. *EMBO J* **13**: 2421–2431.
- Toor N, Keating KS, Taylor SD, Pyle AM. 2008. Crystal structure of a self-spliced group II intron. *Science* **320**: 77–82.
- 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**: 2991–3003.
- Tsai RT, Tseng CK, Lee PJ, Chen HC, Fu RH, Chang KJ, Yeh FL, Cheng SC. 2007. Dynamic interactions of Ntr1-Ntr2 with Prp43 and with U5 govern the recruitment of Prp43 to mediate spliceosome disassembly. *Mol Cell Biol* **27**: 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**: 145–154.
- Wan R, Yan C, Bai R, Huang G, Shi Y. 2016a. Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution. *Science* **353**: 895–904.
- Wan R, Yan C, Bai R, Wang L, Huang M, Wong CC, Shi Y. 2016b. The 3.8 Å structure of the U4/U6.U5 tri-snRNP: Insights into spliceosome assembly and catalysis. *Science* **351**: 466–475.
- Wan R, Yan C, Bai R, Lei J, Shi Y. 2017. Structure of an intron lariat spliceosome from *Saccharomyces cerevisiae*. *Cell* **171**: 120–132.
- Warkocki Z, Odenwalder P, Schmitzova J, Platzmann F, Stark H, Urlaub H, Ficner R, Fabrizio P, Luhrmann R. 2009. Reconstitution of both steps of *Saccharomyces cerevisiae* splicing with purified spliceosomal components. *Nat Struct Mol Biol* **16**: 1237–1243.
- Weber G, Trowitzsch S, Kastner B, Luhrmann R, Wahl MC. 2010. Functional organization of the Sm core in the crystal structure of human U1 snRNP. *EMBO J* **29**: 4172–4184.
- 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**: 1283–1288.
- Wimberly BT, Brodersen DE, Clemons WM Jr, Morgan-Warren RJ, Carter AP, Vonnrhein C, Hartsch T, Ramakrishnan V. 2000. Structure of the 30S ribosomal subunit. *Nature* **407**: 327–339.
- Wolf E, Kastner B, Deckert J, Merz C, Stark H, Luhrmann R. 2009. Exon, intron and splice site locations in the spliceosomal B complex. *EMBO J* **28**: 2283–2292.
- Wysoczanski P, Schneider C, Xiang S, Munari F, Trowitzsch S, Wahl MC, Luhrmann R, Becker S, Zweckstetter M. 2014. Cooperative structure of the heterotrimeric pre-mRNA retention and splicing complex. *Nat Struct Mol Biol* **21**: 911–918.
- Yan C, Hang J, Wan R, Huang M, Wong CC, Shi Y. 2015. Structure of a yeast spliceosome at 3.6-Å resolution. *Science* **349**: 1182–1191.
- Yan C, Wan R, Bai R, Huang G, Shi Y. 2016. Structure of a yeast activated spliceosome at 3.5 Å resolution. *Science* **353**: 904–911.
- Yan C, Wan R, Bai R, Huang G, Shi Y. 2017. Structure of a yeast step II catalytically activated spliceosome. *Science* **355**: 149–155.
- Yang VW, Lerner MR, Steitz JA, Flint SJ. 1981. A small nuclear ribonucleoprotein is required for splicing of adenoviral early RNA sequences. *Proc Natl Acad Sci* **78**: 1371–1375.
- Yang K, Zhang L, Xu T, Heroux A, Zhao R. 2008. Crystal structure of the β-finger domain of Prp8 reveals analogy to ribosomal proteins. *Proc Natl Acad Sci* **105**: 13817–13822.
- Yean SL, Wuenschell G, Termini J, Lin RJ. 2000. Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. *Nature* **408**: 881–884.
- Zhan X, Yan C, Zhang X, Lei J, Shi Y. 2018. Structure of a human catalytic step I spliceosome. *Science* **359**: 537–545.
- Zhang X, Schwer B. 1997. Functional and physical interaction between the yeast splicing factors Slu7 and Prp18. *Nucleic Acids Res* **25**: 2146–2152.
- Zhang X, Yan C, Hang J, Finci LI, Lei J, Shi Y. 2017. An atomic structure of the human spliceosome. *Cell* **169**: 918–929.e914.
- Zhang X, Yan C, Zhan X, Lei J, Shi Y. 2018. Structure of the human activated spliceosome in three conformational states. *Cell Res* **28**: 307–322.
- Zhou Z, Sim J, Griffith J, Reed R. 2002. Purification and electron microscopic visualization of functional human spliceosomes. *Proc Natl Acad Sci* **99**: 12203–12207.
- Zhou L, Hang J, Zhou Y, Wan R, Lu G, Yin P, Yan C, Shi Y. 2014. Crystal structures of the Lsm complex bound to the 3' end sequence of U6 small nuclear RNA. *Nature* **506**: 116–120.



## Molecular Mechanisms of pre-mRNA Splicing through Structural Biology of the Spliceosome

Chuangye Yan, Ruixue Wan and Yigong Shi

*Cold Spring Harb Perspect Biol* 2019; doi: 10.1101/cshperspect.a032409

---

Subject Collection [RNA Worlds](#)

---

### Alternate RNA Structures

*Marie Teng-Pei Wu and Victoria D'Souza*

### Approaches for Understanding the Mechanisms of Long Noncoding RNA Regulation of Gene Expression

*Patrick McDonel and Mitchell Guttman*

### Principles and Practices of Hybridization Capture Experiments to Study Long Noncoding RNAs That Act on Chromatin

*Matthew D. Simon and Martin Machyna*

### Linking RNA Sequence, Structure, and Function on Massively Parallel High-Throughput Sequencers

*Sarah K. Denny and William J. Greenleaf*

### Extensions, Extra Factors, and Extreme Complexity: Ribosomal Structures Provide Insights into Eukaryotic Translation

*Melanie Weisser and Nenad Ban*

### Nascent RNA and the Coordination of Splicing with Transcription

*Karla M. Neugebauer*

### Combining Mass Spectrometry (MS) and Nuclear Magnetic Resonance (NMR) Spectroscopy for Integrative Structural Biology of Protein–RNA Complexes

*Alexander Leitner, Georg Dorn and Frédéric H.-T. Allain*

### Structural Biology of Telomerase

*Yaqiang Wang, Lukas Susac and Juli Feigon*

### Structural Insights into Nuclear pre-mRNA Splicing in Higher Eukaryotes

*Berthold Kastner, Cindy L. Will, Holger Stark, et al.*

### What Are 3' UTRs Doing?

*Christine Mayr*

### Single-Molecule Analysis of Reverse Transcriptase Enzymes

*Linnea I. Jansson and Michael D. Stone*

### CRISPR Tools for Systematic Studies of RNA Regulation

*Jesse Engreitz, Omar Abudayyeh, Jonathan Gootenberg, et al.*

### Relating Structure and Dynamics in RNA Biology

*Kevin P. Larsen, Junhong Choi, Arjun Prabhakar, et al.*

### Beyond DNA and RNA: The Expanding Toolbox of Synthetic Genetics

*Alexander I. Taylor, Gillian Houlihan and Philipp Holliger*

For additional articles in this collection, see <http://cshperspectives.cshlp.org/cgi/collection/>



**All Modifications and Oligo Types Synthesized**  
Long Oligos • Fluorescent • Chimeric • DNA • RNA • Antisense

*Oligo Modifications?*  
Your wish is our command.



**Discovering and Mapping the Modified Nucleotides That Comprise the Epitranscriptome of mRNA**

*Bastian Linder and Samie R. Jaffrey*

**Structural Basis of Nuclear pre-mRNA Splicing: Lessons from Yeast**

*Clemens Plaschka, Andrew J. Newman and Kiyoshi Nagai*

For additional articles in this collection, see <http://cshperspectives.cshlp.org/cgi/collection/>



The advertisement banner features the Gene Link logo on the left, which consists of four green cubes arranged in a 2x2 grid. To the right of the logo, the text reads: "All Modifications and Oligo Types Synthesized" in a bold, sans-serif font. Below this, a list of services is provided: "Long Oligos • Fluorescent • Chimeric • DNA • RNA • Antisense". On the right side of the banner, the text "Oligo Modifications?" is written in a cursive font, with the tagline "Your wish is our command." underneath it. The background of the banner is a green gradient with a faint image of a DNA double helix on the right.