Chapter 3

Mechanisms of Spliceosomal Assembly

Ni-ting Chiou and Kristen W. Lynch

Abstract

Pre-mRNA splicing is a key step for generating mature protein-coding mRNA. An RNA–protein complex known as the spliceosome carries out the chemistry of pre-mRNA splicing. However, several pre-spliceosomal intermediates are assembled on the pre-mRNA before the formation of the catalytically activated spliceosome. The progression to the activated spliceosome involves a cascade of the rearrangement events of the RNA–RNA, RNA–protein, and protein–protein interactions within the pre-spliceosomal intermediates. These rearrangements generate multiple combinatorial interactions of the spliceosome with the substrate, which enhances the accuracy of the splice site selection. Each rearrangement also represents a step at which splicing can potentially be subjected to regulation. The aim of this chapter is to provide an overview of the components of the spliceosome and their rearrangements along the spliceosome assembly pathway.

Key words Spliceosome assembly, Ribonucleoproteins, snRNP, Prp19, NTC, Splicing

1 Introduction

In 1977, Phillip Sharp and his colleagues first provided evidence for the presence of the introns in nascent transcripts [1]. In 1985, the spliceosome, the ribonucleoprotein (RNP) machine which catalyzes pre-mRNA splicing, was identified [2]. The components of the spliceosome contain five small nuclear RNAs (snRNAs) and hundreds of proteins. In contrast to other RNPs, the catalytically active spliceosome are not preassembled before they bind to the pre-mRNA substrate. Instead, the components of the spliceosome interact with the substrates in a stepwise way to assemble a series of pre-spliceosomal intermediates, which leads to the formation of active site of the spliceosome. Since the 1990s, these pre-spliceosomal intermediates have been trapped and analyzed in vitro using a variety of approaches [3]. The biochemical characterizations of these intermediate complexes, combined with yeast genetic functional studies of the individual spliceosomal components, have revealed much insight into the spliceosome assembly pathway.

Klemens J. Hertel (ed.), *Spliceosomal Pre-mRNA Splicing: Methods and Protocols*, Methods in Molecular Biology, vol. 1126, DOI 10.1007/978-1-62703-980-2_3, © Springer Science+Business Media, LLC 2014

Recently, with the advance of the mass spectrometry and electric microscopy techniques, the components and structures of the various pre-spliceosomal complexes have been even more precisely mapped [4]. Importantly, this detailed knowledge of spliceosome assembly forms the basis of our understanding of the mechanisms that govern splicing fidelity, alternative splicing, and the regulation of alternative splicing, all of which are the key steps of gene expression and are discussed in Chapters 4–6.

2 Basic Spliceosome Assembly Pathway

2.1 Stepwise Assembly of the snRNPs Across an Intron

Nuclear pre-mRNA splicing involves two transesterification steps to remove the intron from the pre-mRNA to generate mature protein-coding mRNA. The three reactive regions on the pre-mRNA are the 5′ splice site (5′SS), the 3′ splice site (3′SS), and the branch point site (BPS), which are all defined by short consensus sequence. In addition to three regions, metazoan introns contain a conserved polypyrimidine tract (PPT) between the 3′ss and the BPS (Fig. 1a; *see* also Chapter 1).

The spliceosome, the largest RNP machine in nucleus, recognizes and positions these reactive regions to catalyze pre-mRNA splicing (for a recent extensive review, *see* ref. 5). The main building blocks of spliceosome are small nuclear ribonucleoproteins (snRNPs). There are five spliceosomal snRNPs: U1, U2, U4, U5, and U6 snRNP. Each snRNP contains a single snRNA and at least seven protein subunits. The snRNPs and additional non-snRNP-associated proteins (such as SF1, U2AF, and the Prp19 complex (NTC); *see* Subheading 3.2 below) are assembled on the pre-mRNA substrate in a stepwise way to form the pre-spliceosomal E, A, B, and, finally, the catalytic-spliceosomal C complex (Fig. 1b). During the stepwise assembly processes, multiple combinatorial interactions are generated between the spliceosomal components and the reactive regions of the substrate. Although these reactive regions have very limited conservation (*see* Fig. 1a), the multiple interactions provide the spliceosome multiple opportunities to double-check the fidelity of interactions, thereby increasing the accuracy of site selection.

To begin the assembly of the spliceosome, U1 snRNP engages with the 5′SS, while SF1 binds to the BPS, in an ATP-independent manner to form the pre-spliceosomal E complex (Fig. 1b; E). In metazoan systems, the 65 and 35 subunits of the U2AF protein heterodimer also bind to the PPT and 3′ ss respectively during this ATP-independent step to further promote correct identification of the 3′ end of the intron. In the presence of ATP, several rearrangements of the snRNPs then occur to progress assembly from the E to A, B, and C complexes. The first rearrangement is that U2 snRNP displaces SF1 from the BPS to form the A complex (Fig. 1b; A).

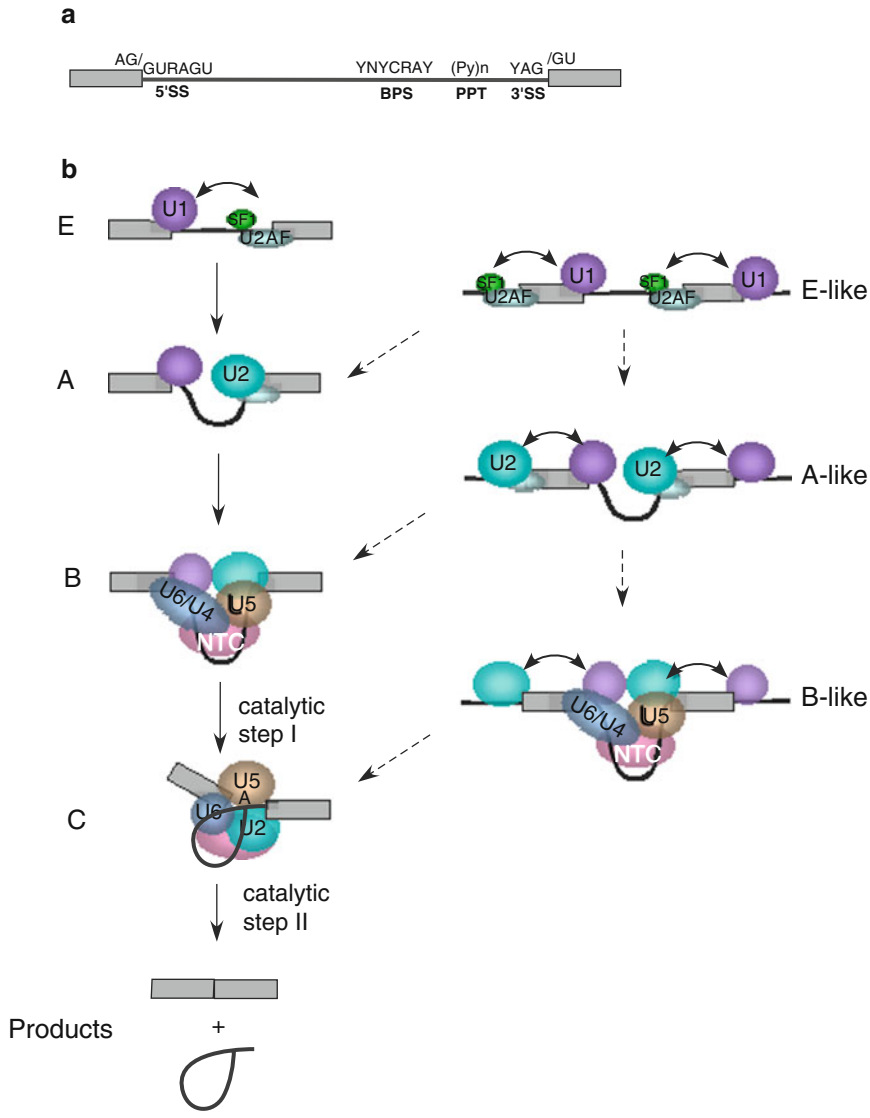

Fig. 1 The pathways of the spliceosome assembly during the pre-mRNA splicing. (**a**) The consensus nucleotide sequences of metazoan pre-mRNAs. Here, two exons (*boxes*) are separated by an intron (*line*). The consensus sequence at the 5′ splice site (5′SS), branch point sequence (BPS), polypyrimidine tract (PPT), and 3′ splice site (3′SS) are indicated above the *line*. In these sequences, R stands for either G or A; Y stands for either U or C. The A within the BPS forms the branch point of the intron lariat produced by splicing. (**b**) The stepwise assembly of U1, U2, U4/U6, U5 snRNP, and NTC on the consensus sequences in the removal of an intron from a pre-mRNA is depicted. *Left*, the cross-intron assembly. *Right*, corresponding exon-defined version of each step

At this point, the preassembled U4/U6.U5 tri-snRNP and NTC are recruited to form the B complex (Fig. 1b; B). The U1 and U4 snRNP are then released followed by the association of the U6 snRNP with the 5′SS and with the U2 snRNA. These rearrangement events promote the first catalytic step to occur, i.e., cleavage of the 5′SS with concurrent formation of a covalent bond between

the first nucleotide of the intron and an A residue at the BPS to results in the C complex formation (Fig. 1b; C). In the C complex, the second catalytic step proceeds to excise the lariat intron and join 5′ and 3′ exons to generate mature mRNA (Fig. 1b; products).

2.2 The Exon-Definition Complex

The combinatorial interactions described above are built across the introns. However, the average lengths for exons and introns of human protein-coding genes are, respectively, 145 and 3,364 nucleotides [6]. Since the exons are significantly shorter than introns, it is expected that initially identifying exons during the spliceosome assembly would help the splicing components to be deposited across the introns more precisely, and hence avoid the use of the cryptic splice sites. Thus, it is envisioned that the cross-exon interactions of the snRNPs occur first or simultaneously with the cross-intron interactions in each stage of the assembly (Fig. 1b, right column). Indeed, the U1 and U2 snRNP-containing exon-defined A complexes have been observed for several exons (Fig. 1b; A-like) [7, 8]. In addition, some exon-definition complexes have been shown to contain the tri-snRNP, and the exon-bound tri-snRNP can directly interact with the upstream 5′SS to assemble the B complex across the intron (Fig. 1b; B-like) [9]. However, despite the characterization of some exon-defined complexes, it is still unknown exactly what interactions of the RNA and protein components are involved in building the exon-defined and the connection or conversion of exon-defined to intron-defined complexes.

3 The Rearrangements of the Spliceosome During the Assembly Processes

3.1 The Dynamics of the RNA–RNA Interactions

Much of the structural rearrangements during spliceosome assembly and establishment of its active site involves the remodeling of the base-pairing interactions among five snRNAs and three reactive regions of the pre-mRNA [10, 11]. In the pre-spliceosomal complexes, such as the A and B complex, the base-pairing interactions with the 5′ SS and BPS involve the 5′ end of U1 snRNA and the internal region of U2 snRNA, respectively (Fig. 2a; purple and cyan box of U1 and U2). For the tri-snRNP in the pre-spliceosomal complex, U4 and U6 snRNA are held firmly by base-pairing interactions, and U5 snRNA is associated through RNA–protein interactions (Fig. 2a; tri-snRNP). The U4 and U6 base-pairing interactions inactivate the catalytically important regions of U6 to prevent from cleaving pre-mRNA prematurely.

During the integration of the tri-snRNP into the spliceosome, U1 is displaced from 5′SS, and U4/U6 base-pairing interactions are taken apart. This rearrangement frees U6 snRNA and allows it to form two new base-pairing interactions. One of the interactions involves the ACAGA motif of U6 snRNA engaging in 5′SS interaction, and the other involves the region downstream of ACAGA

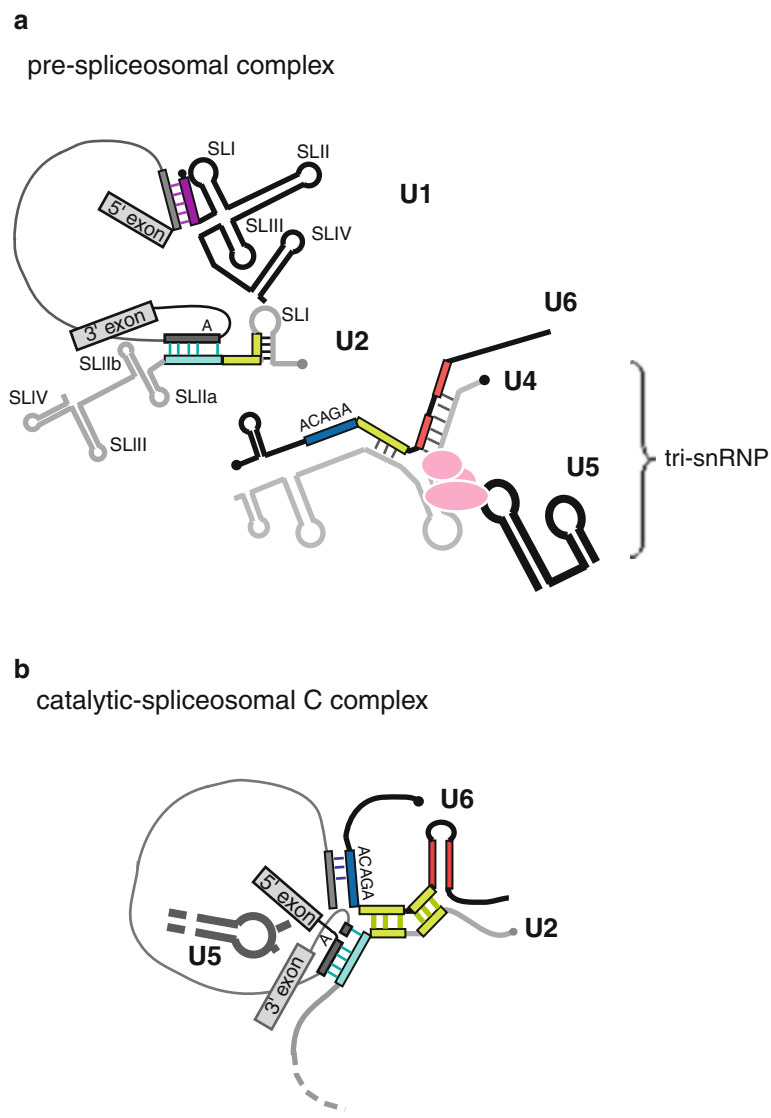

Fig. 2 The rearrangements of RNA–RNA interaction networks during the transition to the catalytic-spliceosomal complex. (**a**) The secondary structures of the five human U snRNAs and the base-pairing interactions of the snRNAs–pre-mRNA within the pre-spliceosomal complex. The stem-loops of the U1 and U2 snRNAs are numbered, and the proteins that associate with these stem-loops are listed in Table 1. The regions of U1 snRNA and U2 snRNA base pair with the 5′SS or BPS are highlighted in *purple* and *cyan*. The tri-snRNP shown here has not integrated into the spliceosome. Within the tri-snRNP, U6 and U4 snRNA are held through base-pairing interactions, while the U5 snRNA and U4/U6 di-snRNA are held by the proteins shown as the *pink circles*. (**b**) During the transition to the catalytic-spliceosomal C complex, U1 and U4 snRNA are released, while U2, U5, and U6 snRNA form the new base-pairing interactions. For simplicity, only the base-pairing interactions of snRNA–snRNA or snRNA–pre-mRNA are shown, but the secondary structures of the snRNAs are not depicted. The regions of the U6 snRNA that are engaged in the base-pairing interactions with the 5′SS and U2 snRNA are highlighted in *purple* and *yellow*, respectively. The stem-loop region of U5 snRNA makes a few contacts with the 5′ and 3′-exon

motif base pairing with the U2 snRNA (Fig. 2b; blue and yellow box of U6). It is interesting that the region of U2 snRNA that base pairs with U6 snRNA is immediately proximal to the region interacting with the pre-mRNA (Fig. 2b; yellow box of U2). Thus, these two newly formed interactions bring the splice sites together to allow the first catalytic step to occur. Following first catalytic step, the stem-loop region of U5 snRNA contacts the nucleotides of the 5′ and 3′exon to bring the two exons into proximity for the second catalytic step.

The RNA base-pairing interactions illustrated above involves from three to eight base pairs that are typically not fully complementary. Thus, proteins certainly play an important role in creating and stabilizing these RNA–RNA interactions. In turn, these RNA–RNA interactions also influence the protein–RNA and protein–protein interactions.

3.2 The Changes of Protein–Protein or Protein–RNA Interactions

The human spliceosomal complexes contain ~45 distinct snRNP-associated proteins and ~170 spliceosome-associated factor [12, 13]. During the splicing processes, proteins enter and leave the spliceosomes from one stage to the next. Thus, the number of the total spliceosomal proteins varies among different pre-spliceosomal intermediates. In general, each of the spliceosomal A, B, and C complex contains ~125 proteins or less (in the case of the A complex) [14–16]. Table 1 lists some of these proteins that have the well-known functions.

During the progression from the A to B complex, ~35 tri-snRNP proteins and ~25 non-snRNP proteins are recruited (*see* Table 1). The major part of these non-snRNP proteins is the Prp19 complex (NTC). The human Prp19 complex is comprised of seven distinct subunits with four copies of prp19 protein. This complex is thus similar to the size of the snRNPs, but unlike the snRNPs, the NTC contains no RNAs [17]. The B complex then transition to the C complex, involving the release not only of the U1 and U4 RNAs but also the protein components of the U1 and U4 snRNP. Although U6 snRNA is not released, most of its protein components also fall off in the C complex. It is possible that the non-snRNP proteins in the C complex, such as the NTC, form new interactions with U6 snRNA to promote its interaction with pre-mRNA in creating the active splice sites of the spliceosome. Moreover, there are ~30 non-snRNP proteins which are recruited during C complex to promote the catalytic site formation (Table 1).

Importantly, many of the proteins that are recruited during the assembly steps are RNA-dependent ATPases/helicases, which are required for the various RNA rearrangements [11]. For example, Brr2, the U5 snRNP component, is involved in unwinding U6/U4 duplex. Subsequently, Prp28, also a U5 snRNP component, mediates the transfer of the 5′SS from the 5′ end of U1 snRNA to the ACAGAG motif of U6 snRNA (Fig. 2). Prp16 and

Table 1
The representative protein components of human U snRNPs, NTC, and spliceosomal factors

U snRNP/NTC/ spliceosomal factors	Representative proteins	Present in the complex			Functions/interactions/modifications
		A	B	C	
A complex factors (~10)	RBM5	+			Block the conversion from a cross-exon to a cross-intron complex
U1 snRNP (~14)	Sm (7)	+	+		Bind to the sm site of U1 snRNA
	U170K	+	+		Bind to the SLI of U1 snRNA
	U1A	+	+		Bind to the SLII of U1 snRNA
	U1C	+	+		Mediate the base-pairing interactions between U1 snRNA and 5′SS
U2 (~17)	Sm (7)	+	+	+	Bind to the sm site of U2 snRNA
	SF3a (3)	+	+	+	Bind to SLI and SLIIb of U2 snRNA
	SF3b (7)	+	+	+	Mediate the base-pairing interactions between U2 and BPS
U2-related (~10)	U2AF35	+	+		Bind to AG nucleotide at 3′SS
	U2AF65	+	+		Bind to PPT
	SPF30	+	+		Bridges an interactions between U2AF35 and Prp3
	Prp5/DDX46	+	+		Bridges a U1 and U2 snRNP interaction network
U5 (~14)	Sm (7)		+	+	Bind to sm site of U5 snRNA
	hSnu114		+	+	GTPase; promote Brr2 helicase activity
	hBrr2		+	+	RNA helicase; unwinding U4/U6 hairpin
	hPrp8		+	+	Bind to both of 5′ and 3′ exon during the catalytic step II
	hPrp6		+	+	Phosphorylated during B complex formation
	hPrp28		+	+	RNA helicase; exchange of U1 for U6 snRNP at 5′SS
U5-related (~11)	hPrp38		+	+	Promote U4/U6 snRNA dissociation
U4 (~12)	Sm (7)		+		Bind to sm site of U4 snRNA
	hPrp31		+		Phosphorylated during B complex formation
	hPrp4		+		Phosphorylate Prp31 and Prp6
	hPrp3		+		Ubiquitinated by NTC
	hPrp24		+		Facilitates the association of U4 and U6 snRNPs
U6 (7)	Lsm2-8 (7)		+		Bind to the Lsm site of U6 snRNA
NTC (7)	Prp19		+	+	Stabilize the association of U5 and U6 with the spliceosome after U4 is dissociated
	CDC5L		+	+	
	SPF27		+	+	
	PRLG1		+	+	
NTC-related (~12)	RBM22		+	+	Promote the catalytic conformation
B complex factors (~8)	UBL5		+		Unknown
C complex factors (~37)	Prp22			+	RNA helicase; required for catalytic step II
	Prp16			+	RNA helicase; required for both catalytic steps
	Slu7			+	Mediate 3′ splice site choice

The functions, interaction, or modifications of the representative proteins are compiled from several sources [5, 11, 23–26]. The association of these representative proteins with the spliceosomal complexes is based on the review paper [5]. Numbers indicate the total number of individual proteins in a particular group

Prp22, the non-snRNP proteins, are recruited during the transition to the C complex and required for the second transesterification step [18, 19].

4 Conclusion and Future Perspectives

The mechanism of spliceosome assembly provides an extraordinary model in illustrating how RNAs and proteins cooperate as they work together to recognize the reactive regions of the pre-mRNA and catalyze its splicing. The other fact that makes the spliceosome an important RNP machine is that 90–95 % of human genome is alternatively spliced and at least 10 % of human genetic disease arises from the mutations either in the splice sites or in the splicing regulatory sequences [20–22]. This indicates that the assembly of the spliceosome is highly regulated in the cell and sensitive to the minor changes of the pre-mRNA sequences. Thus, the mechanism of spliceosome assembly not only is critical for understanding the principles that govern alternative splicing but also brings new opportunities to the possible treatment of human genetic diseases.

References

1. Berget SM, Moore C, Sharp PA (1977) Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci U S A 74:3171–3175
2. 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
3. Jurica MS, Moore MJ (2002) Capturing splicing complexes to study structure and mechanism. Methods 28:336–345
4. Luhrmann R, Stark H (2009) Structural mapping of spliceosomes by electron microscopy. Curr Opin Struct Biol 19:96–102
5. Wahl MC, Will CL, Luhrmann R (2009) The spliceosome: design principles of a dynamic RNP machine. Cell 136:701–718
6. Consortium IHGS (2004) Finishing the euchromatic sequence of the human genome. Nature 431:931–945
7. House AE, Lynch KW (2006) An exonic splicing silencer represses spliceosome assembly after ATP-dependent exon recognition. Nat Struct Mol Biol 13:937–944
8. Sharma S, Kohlstaedt LA, Damianov A, Rio DC, Black DL (2008) Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. Nat Struct Mol Biol 15:183–191
9. Schneider M, Will CL, Anokhina M, Tazi J, Urlaub H et al (2010) Exon definition complexes contain the tri-snRNP and can be directly converted into B-like precatalytic splicing complexes. Mol Cell 38:223–235
10. Nilsen TW (1994) RNA-RNA interactions in the spliceosome: unraveling the ties that bind. Cell 78:1–4
11. Staley JP, Guthrie C (1998) Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell 92:315–326
12. Zhou Z, Licklider LJ, Gygi SP, Reed R (2002) Comprehensive proteomic analysis of the human spliceosome. Nature 419:182–185
13. Rappsilber J, Ryder U, Lamond AI, Mann M (2002) Large-scale proteomic analysis of the human spliceosome. Genome Res 12:1231–1245
14. Behzadnia N, Golas MM, Hartmuth K, Sander B, Kastner B et al (2007) Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. EMBO J 26:1737–1748
15. Deckert J, Hartmuth K, Boehringer D, Behzadnia N, Will CL et al (2006) Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. Mol Cell Biol 26:5528–5543
16. 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

17. Grote M, Wolf E, Will CL, Lemm I, Agafonov DE et al (2010) Molecular architecture of the human Prp19/CDC5L complex. Mol Cell Biol 30:2105–2119
18. Schwer B (2008) A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. Mol Cell 30:743–754
19. Schwer B, Guthrie C (1992) A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. EMBO J 11:5033–5039
20. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet 40: 1413–1415
21. Cooper TA, Wan L, Dreyfuss G (2009) RNA and disease. Cell 136:777–793
22. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L et al (2008) Alternative isoform regulation in human tissue transcriptomes. Nature 456:470–476
23. Schneider M, Hsiao HH, Will CL, Giet R, Urlaub H et al (2010) Human PRP4 kinase is required for stable tri-snRNP association during spliceosomal B complex formation. Nat Struct Mol Biol 17:216–221
24. Pomeranz Krummel DA, Oubridge C, Leung AK, Li J, Nagai K (2009) Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. Nature 458:475–480
25. Dybkov O, Will CL, Deckert J, Behzadnia N, Hartmuth K et al (2006) U2 snRNA-protein contacts in purified human 17S U2 snRNPs and in spliceosomal A and B complexes. Mol Cell Biol 26:2803–2816
26. Song EJ, Werner SL, Neubauer J, Stegmeier F, Aspden J et al (2010) The Prp19 complex and the Usp4Sart3 deubiquitinating enzyme control reversible ubiquitination at the spliceosome. Genes Dev 24:1434–1447