THE PROTEIN SOCIETY

Structural analyses of the pre-mRNA splicing machinery

Lingdi Zhang, Xueni Li, and Rui Zhao*

Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, Colorado 80045

Received 28 February 2013; Revised 3 April 2013; Accepted 8 April 2013
DOI: 10.1002/pro.2266
Published online 17 April 2013 proteinscience.org

Abstract: Pre-mRNA splicing is a critical event in the gene expression pathway of all eukaryotes. The splicing reaction is catalyzed by the spliceosome, a huge protein-RNA complex that contains five snRNAs and hundreds of different protein factors. Understanding the structure of this large molecular machinery is critical for understanding its function. Although the highly dynamic nature of the spliceosome, in both composition and conformation, posed daunting challenges to structural studies, there has been significant recent progress on structural analyses of the splicing machinery, using electron microscopy, crystallography, and nuclear magnetic resonance. This review discusses key recent findings in the structural analyses of the spliceosome and its components and how these findings advance our understanding of the function of the splicing machinery.

Keywords: pre-mRNA splicing; spliceosome; structure; electron microscopy; crystallography; NMR

Introduction

In eukaryotes, DNA is first transcribed into pre-mRNAs that often contain introns. The removal of introns (pre-mRNA splicing) is an essential step in the gene-expression pathway of all eukaryotes. In higher eukaryotes such as mammals, an average of 95% of the nucleotides in the primary transcript of a protein-encoding gene are introns.[1] These introns need to be removed precisely by splicing before the mRNA can be transported out of the nucleus and translated. Alternative splicing greatly expands the gene coding capacity, and over 95% of human genes are alternatively spliced.[2] It is also becoming increasingly clear that alternative splicing is a fundamental component of eukaryotic gene regulation, influencing cell differentiation,[3] development,[4] and many processes in the nervous system.[5] Errors in splicing contribute to at least 15% of human genetic disorders, such as retinitis pigmentosa, spinal muscular atrophy, and myotonic dystrophy.[6–8] Aberrant splicing also plays a significant role in the onset and development of many other diseases.[6,7] For example, alterations in the alternative splicing of several pre-mRNAs (such as WT1 and CD44) correlate with neoplastic conversion and metastatic potential in cancer.[7]

A typical intron contains a conserved 5′ splice site (5′ ss), a branch point sequence (BPS) followed by a polypyrimidine tract (PYT), and a 3′ splice site (3′ ss).[9,10] Introns are removed through two transesterification reactions. In the first transesterification reaction, the 2′ hydroxyl group of the critical adenosine residue in the BPS attacks the phosphate group at the 5′ ss, generating a lariat intermediate. In the second transesterification reaction, the newly freed 3′ hydroxyl group from the cleaved 5′ exon attacks the phosphate group at the 3′ ss, releasing the lariat and ligating the two exons.

Pre-mRNA splicing is catalyzed by the spliceosome, a huge RNA/protein complex. The spliceosome contains five small nuclear RNAs (U1, U2, U4, U5, and U6 snRNAs), which form five small nuclear ribonucleic particles (snRNPs) with their associated proteins, and numerous non-snRNP associated

Additional Supporting Information may be found in the online version of this article.

Lingdi Zhang and Xueni Li contributed equally to this work

Grant sponsor: NIH; Grant number: R01GM080334 (to R.Z.).

*Correspondence to: R. Zhao, Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, CO 80045. E-mail: rui.zhao@ucdenver.edu

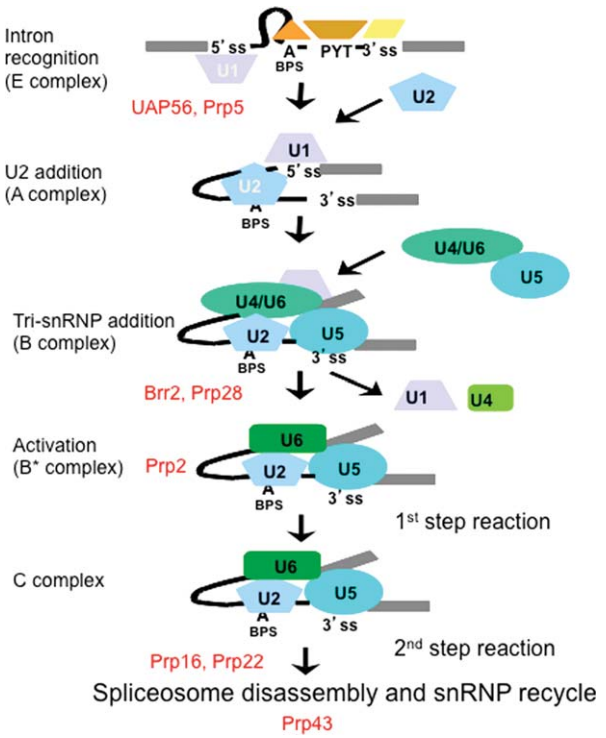

Figure 1. A schematic representation of the spliceosomal assembly, activation, and disassembly pathway. The eight DExD/H-box RNA helicases are indicated in red at the major stages of the splicing pathway that they are involved in.

protein factors.[11,12] Proteome studies indicate that the *Saccharomyces cerevisiae* spliceosome contains ~80 and the human spliceosome contains ~170 different proteins.[13] Spliceosome components assemble on pre-mRNA in a step-wise manner[13] (Fig. 1). The first spliceosome assembly step involves the initial recognition of an intron by the spliceosome and the formation of the E complex. The 5′ ss is recognized by U1 snRNP, the BPS by SF1 (BBP in yeast), and the PYT by U2AF^{65} (MUD2 in yeast). Subsequently, the U2 snRNP joins the spliceosome, replaces SF1, and forms the A complex. The subsequent joining of the U4/U6.U5 tri-snRNP signifies the B complex. Extensive structural rearrangements occur at this stage to activate the spliceosome.[14] During this activation process, the base-pairing between the 5′ ss and U1 snRNA is disrupted, and the 5′ ss interacts with U6 snRNA instead, using largely the same nucleotides that base-paired with U1 snRNA. The base-pairing between U4 snRNA and U6 snRNA is also disrupted, and new interactions between U2 snRNA and U6 snRNA are formed, which are mutually exclusive with those in the original U4/U6 complex. These rearrangements help convert the spliceosome to the catalytically active B* complex that is ready for the first catalytic reaction. After the first transesterification reaction, the spliceosome repositions the substrate to form the C complex that is ready for the second catalytic reaction. The second reaction is followed by postcatalytic rearrangements to liberate the mature mRNA for export and release the lariat intron to be degraded and snRNPs to be recycled. The spliceosomal assembly, activation, and disassembly process involves many ATP-dependent conformational rearrangements facilitated by DExD/H-box RNA helicases,[9,10,15] which are potentially important for the fidelity control of splicing.

The large size and highly dynamic nature (both in composition and conformation) of the spliceosome presented daunting challenges to structural studies. However, there has been significant progress in structural analyses of the splicing machinery using a variety of approaches. This review will focus on recent progress on understanding the structural basis of pre-mRNA splicing and how it relates to the function of the splicing machinery. We will highlight a number of electron microscopy (EM) structures that generated the overall (but less detailed) view of the shape and structural organization of the spliceosome, as well as crystallographic and nuclear magnetic resonance (NMR) studies that provided high resolution structures of the individual components, fragments, or subcomplexes.

Overall Structural Organization of the Spliceosome Revealed by EM

Single particle EM reconstruction remains the method of choice for determining the overall morphology and architecture of the spliceosome. Although often limited to low or medium resolutions, EM has the advantage of requiring 100- to 1000-fold less samples compared to crystallography and NMR. Spliceosomes used for EM studies are typically assembled *in vitro* by incubating a well-defined pre-mRNA substrate with HeLa nuclear extract or yeast whole cell extract, then purified by affinity chromatography using antibodies specific for spliceosomal components, affinity tag on protein splicing factors, biotinylated pre-mRNA, or pre-mRNA containing MS2-binding aptamers. Double affinity purification or gradient centrifugation is often used to further improve homogeneity, but sample heterogeneity is still the major challenge for improving resolutions of EM structures. The resolution of most EM structures of spliceosomes is too low to directly identify secondary structures and individual components. Labeling techniques using antibody or genetic tags (often large fusion proteins) are often needed to decipher spatial arrangements of components in the spliceosome.

Three-dimensional (3D) structures of several spliceosomal complexes are determined, including the human A complex at 40–50 Å resolution,[16] BΔU1 complex at 40 Å resolution,[17] the core C complex at 30 Å resolution from the Moore laboratory,[18] and a larger C complex at 20–29 Å resolution from the Stark and Luhrmann laboratories[19] (Fig. 2). The

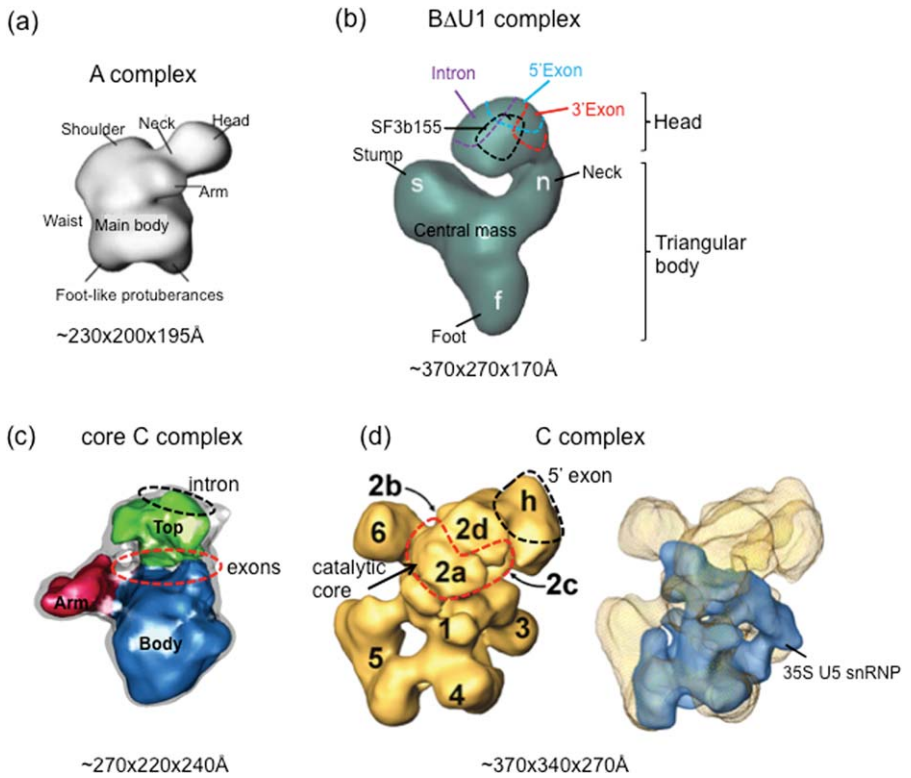

Figure 2. Representative EM structures of the human spliceosome. (a) A complex (figure modified from Ref. 16). Major structural features are labeled and the rough size of the complex is also indicated. (b) The BΔU1 complex with its head, neck, stump, and foot domains (figure modified from Ref. 17). The rough positions of SF3b155, exon, and intron obtained from labeling studies are indicated. (c) The core C complex with its top, arm, and body domains (figure modified from Ref. 18). The rough positions of the exons and the intron region upstream of BPS are indicated. (d) A larger C complex with its various domains (left) and with the 35S U5 snRNP fitted in (right) (figure modified from Ref. 19). The estimated positions of the 5′ exon and the catalytic core are indicated.

A complex (overall size $\sim 230 \times 200 \times 195$ Å) contains a globular main body with several smaller protruding elements, including a prominent head domain and foot-like protrusions[16] [Fig. 2(a)]. The BΔU1 complexes (overall size $\sim 370 \times 270 \times 170$ Å) consist of a flexible head domain attached to a rigid triangular body domain that resembles the tri-snRNP in size and shape[17] [Fig. 2(b)]. The head domain can accommodate the A complex structure. Labeling studies show that the 5′ exon, 3′ exon, intron, and SF3b155 (a protein in U2 snRNP that is in close proximity to the BPS and 3′ ss) are all located in the head domain at several distinct areas.[20] The C complex structure determined by Jurica *et al.*[18] is $\sim 270 \times 220 \times 240$ Å, which is similar in size and shape to the triangular domain in the B complex and may represent a stable core C complex [Fig. 2(c)]. The core C complex contains a cylindrical top domain and a larger bottom domain with an extended arm domain. These three domains are arranged in a relatively open conformation surrounding a central cavity. U5 snRNP, U2 snRNP, and Prp19 (a core component of the NineTeen Complex involved in spliceosomal activation) potentially fit into these different domains. By introducing a RNA hairpin that binds to the bacterial phage coat protein PP7, the two exons in pre-mRNA are shown to be close to each other and located in the general area between the top and bottom domain.[21] Using actin filament labeling, the intron region upstream of the branchpoint is located to the top domain.[22] The Luhrmann laboratory purified and determined the 3D EM structure of a larger C complex ($\sim 360 \times 340 \times 270$ Å), whose larger size is potentially due to a milder purification condition compared to the core C complex[19] [Fig. 2(d)]. Labeling studies indicate that the 5′ end of the 5′ exon is located in domain h or the upper portion of domain 2. Fitting of a postcatalytic 35S U5 snRNP and a salt stable C complex core (containing all components of the 35S U5 snRNP plus U2 and U6 snRNAs) indicate that the 35S U5 snRNP occupies the central and lower region of the C complex, and the catalytic core containing the U2 and U6 snRNAs is likely located in domain 2a.

In addition to the 3D EM structures discussed above, 2D EM images of the following spliceosomal complexes have been obtained: human spliceosomal B[23] and B^{act} (a complex that has lost U1 and U4 snRNAs compared to the B complex but has not

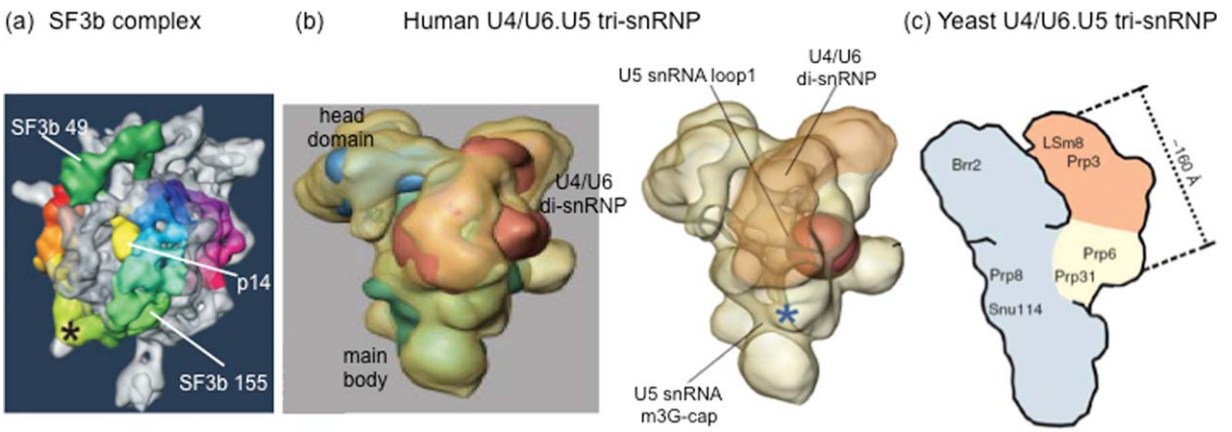

Figure 3. Representative EM structures of the snRNPs. (a) The SF3b complex in U2 snRNP (figure modified from Ref. 34). The HEAT repeats of the SF3b155 protein are in rainbow colors, the RRM domain of SF3b49 is in green, and the p14 RRM is in yellow. Asterisk indicate a potential hinge region in SF3b155. (b) Left: The human tri-snRNP with U5 snRNP (green and blue) and U4/U6 snRNP (red) fitted in (figure modified from Ref. 35). Right: Positions of the 5′ end and Loop 1 of U5 snRNA identified using labeling studies are indicated as an asterisk and a red sphere. (c) The yeast tri-snRNP and positions of U5 snRNP proteins (Brr2, Prp8, and Snu114) and U4/U6 snRNP proteins (LSm8, Prp3, Prp6, and Prp31) (figure modified from Ref. 36).

undergone the conformational change catalyzed by helicase Prp2 to become the B* complex) complexes,[24] yeast spliceosomal B, B^{act}, B*, and C complexes,[25,26] and fly spliceosomal B and C complexes.[27] In general, the same complex in different species is well conserved in terms of shape and size. For example, B complex from human, yeast, and fly all have the triangular or rhombic shape with some subtle differences at the head region. On the other hand, although all complexes are ~400 Å in their maximum dimensions, the different spliceosomal complexes look significantly different, likely reflecting the compositional and conformational change that occurred during the splicing process. For example, B complex has a triangular or rhombic shape, but the B^{act} and C complex are progressively more compact. The difference between B^{act} and B* before and after Prp2-mediated conformational arrangements is also obvious.[26]

Improving sample homogeneity will be the key for improving the resolution of EM structures of spliceosomal complexes. The yeast spliceosome is less complex in its components, and there are many genetic approaches that can arrest the spliceosomal complexes at particular stages. We may see more EM structures of yeast spliceosome in the future.

Structures of snRNPs Revealed by EM and Crystallography

EM structures of snRNPs

snRNPs are much more stable in terms of composition and conformation than the entire spliceosomal complex. There have been medium resolution EM studies of several snRNPs. Human U1 snRNP is the best-studied snRNP biochemically and structurally since it is the smallest snRNP in the cell. The cryo-EM structure of U1 snRNP is determined at 10–14Å resolution.[28] The EM structure is highly consistent with the 5.5Å crystallographic structure of U1 snRNP[29] and the structural details will be discussed later in the crystallographic section of U1 snRNP.

Human U2 snRNP contains U2 snRNA, seven Sm proteins, and 15 U2-specific proteins, most of which are components of the SF3a and SF3b subcomplexes.[30] SF3b plays an important role in BPS recognition and its p14 protein component directly crosslinks to the branch point adenosine.[31–33] The 3D structure of the entire U2 snRNP is unknown but the EM structure of SF3b has been determined to less than 10 Å resolution[34] [Fig. 3(a)]. SF3b has a shell like body roughly $125 \times 150 \times 160$ Å in size, which encloses a large cavity. The intermediate resolution of the EM structure allows the tentative assignment of several protein components with known structural folds, including SF3b49 and p14 [both containing RNA recognition motif (RRM) domain] as well as Sf3b155 (containing 22 tandem HEAT repeats) into the complex without labeling. Protein p14 is likely located in the central cavity of the SF3b complex and the 22 tandem HEAT repeats of SF3b155 are located in the outer shell of the complex enclosing p14. Interestingly, since p14 seems to be located in the interior of the closed SF3b structure, conformational changes of SF3b may be needed for p14 to access the BPS.

The U4/U6.U5 tri-snRNP is the largest snRNP in the spliceosome. The EM structures of the human tri-snRNP (21 Å resolution, $\sim 305 \times 200 \times 175$ Å) and its U5 (26–32 Å resolution, $\sim 265 \times 150 \times 120$ Å) and U4/U6 (~40 Å resolution) snRNP components,[35] as well as the yeast tri-snRNP[36] and U6 snRNP,[37] have been determined [Fig. 3(b)]. The human and yeast tri-snRNP have a similar triangular shape with a head domain and the main body connected at the center.

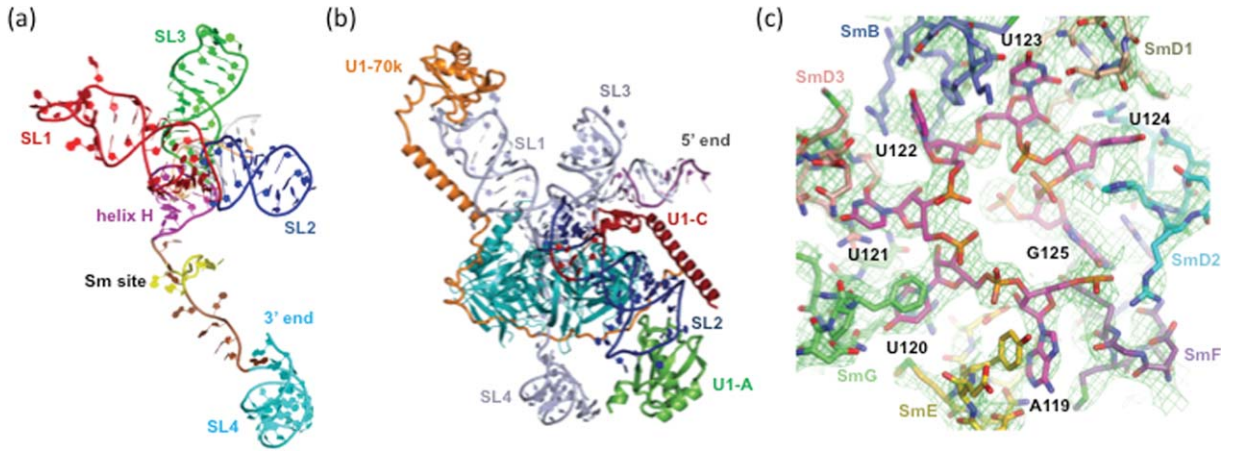

Figure 4. Crystal structures of U1 snRNP (figures modified from Ref. 29) and the Sm core of U4 snRNP (figure modified from Ref. 52). (a) Structure of the U1 snRNA demonstrating the four way junction formed by coaxial stacking of SL1 and 2, as well as SL3 and helix H. (b) Structure of U1 snRNP demonstrating the organization of U1-A, U1-C, U1-70K, and Sm proteins on U1 snRNA. (c) Structure of the Sm core of U4 snRNP demonstrates that each nucleotide of the heptad Sm site (AUUUUUG) interacts with one of the Sm proteins.

Labeling studies show that U4/U6 snRNP is located in the arm domain and the head domain contains U5 protein Brr2, while Prp8 and Snu114 are located around the center of tri-snRNP where the head and body domain meet. Loop 1 of U5 snRNA that was thought to align the two exons during the splicing reaction[38–40] is located at the center of the tri-snRNP. The head and arm domains are flexible, which may have functional implications during spliceosomal activation.

Crystal structure of snRNPs

Human U1 snRNP is the largest subcomplex in the spliceosome whose crystallographic structure has been determined, either by reconstituting with all recombinant components[29] or by direct purification from HeLa cells.[41] Human U1 snRNP contains U1 snRNA, seven Sm proteins common to all snRNPs, and three U1-specific proteins (U1-A, U1-70K, and U1-C).[42] The base-pairing between the 5′ end of U1 snRNA and the 5′ ss is critical for 5′ ss recognition.[43] There have been extensive biochemical studies and structures of U1 snRNP components before the entire U1 snRNP structure was determined. U1 snRNA was thought to form four stem loops (SL1–4) through previous biochemical studies and computer modeling.[44] U1-70K and U1-A bind to SL1 and 2, respectively, through their RRM domains.[45,46] While U1-C does not bind to U1 snRNA by itself, the addition of the first 97 residues of U1-70K enables the association of U1-C with U1 snRNP.[47] Crystal structures of several Sm proteins were determined, leading to a model of seven Sm proteins arranged in a ring-like structure.[48] The crystal structures of the RRM domain of U1-A in complex with SL2,[49] and the NMR structure of the Zn-finger domain of U1-C,[50] were also determined. Building on these successes, Nagai and coworkers[29] reconstituted the entire human U1 snRNP using recombinant proteins (omitting the nonessential U1-A) and *in vitro* transcribed U1 snRNA and determined the crystal structure of U1 snRNP at 5.5 Å resolution[29] (Fig. 4). The Sm proteins, U1C Zn-finger, and a homology model of U1-70K RRM domain were fitted into the electron density map using anomalous peaks from zinc and Se-Met substituted proteins as landmarks. Path of the long *N*-terminal arm of U1-70K was established utilizing seven Met mutations engineered in the *N*-terminus.

U1 snRNA upstream of the Sm binding site forms a four-way junction where SL1 and 2 as well as SL3 and helix H coaxially stack [Fig. 4(a)]. The seven Sm proteins form a heptameric ring on the Sm binding site, which interacts and stabilizes the four-way junction [Fig. 4(a)]. The RRM domain of U1-70K interacts with SL1 [Fig. 4(b)]. The *N*-terminal arm of U1-70K extends over 180Å from its RRM domain, wraps around the core Sm domain, and finally contacts U1-C [Fig. 4(b)], which explains why the *N*-terminus of U1-70K is necessary and sufficient for U1-C binding to the U1 core domain. In the crystal, the 5′ end of U1 snRNA basepairs with its counterpart from a neighboring molecule and mimics the interaction between U1 snRNA and the 5′ ss [Fig. 4(b)]. Helix A of U1-C binds in the minor groove of this duplex. The loop between the two His residues in the Zn-finger of U1-C is in close proximity to the proposed pre-mRNA strand. Arg21/Lys22 on Helix A and Arg28/Lys29 in this loop can potentially interact with the phosphate backbone of RNA. These structural features are consistent with the essential role of U1-C and R28/K29 in mediating the binding between U1 snRNA and the 5′ ss to form the E complex.[51]

Weber *et al.*[41] also determined the crystal structure of human U1 snRNP purified from HeLa cells to

Table I. *Major Crystallographic or NMR Structures (Complete or Partial) of Unique Spliceosomal Protein or RNA Components Determined to Date*

snRNP	Unique protein or RNA structures (complete or partial) determined
U1	U1 snRNP, U1-70K, U1A, U1C, and Prp40
U2	SF3a components (Prp9, Prp11, Prp21), SF3b components (SF3b155, SF3b14, SF3b2, SF3b4), U2AF[35], U2AF[65], PUF60, SPF45, U2 A', U2 B", and BPS/U2 snRNA
U4	U4 snRNP Sm core, and U4 snRNA
U5	Prp8, Brr2, Aar2, 15K, and 52K
U6	Prp24, U6 snRNA, and Lsm3–7
U4/U6.U5	Snu66, 15.5K, Prp31, and Snu13p
U4/U6	Prp3, Cyclophilin H, and Prp4
Components common to all snRNPs	Sm D1, D2, D3, B, E, F, and G
Non-snRNP associated	UAP56, SF1, Pml1, Dim2, Cwc2, Prp19, Rbm22, Prp43, Prp18, Prp22, and U2/U6 snRNA

When more than one homologous structure of a protein is available, only the protein from one species is listed. More details of the structure are provided in Supporting Information Table I.

4.4 Å resolution, after *in situ* limited proteolysis in the crystallization drop.[41] The overall structure is similar to the structure of the reconstituted human U1 snRNP. The structure revealed more details on the organization of Sm proteins on RNA at the Sm core, although the interpretation cannot yet reach atomic level at this resolution.

Leung *et al*.[52] assembled Sm proteins on a fragment of U4 snRNA and determined a 3.6 Å resolution structure of the U4 snRNP core, providing significant more details on the structural organization of the Sm core.[52] The structure demonstrates that the Sm site heptad (AUUUUUG) resides inside the central channel of the heptameric ring of Sm proteins, interacting one-to-one with SmE-SmG-SmD3-SmB-SmD1-SmD2-SmF [Fig. 4(c)]. Each base of the Sm binding site interacts in a distinct manner with four key residues at equivalent positions in the L3 and L5 loops of the Sm fold. These interactions explain the specific requirement for an adenine at the first position of the heptad, the three uredines at Positions 2–4, and the tolerance of positions five and six for other bases. The same Sm proteins demonstrate some structural differences (particularly in the terminal extensions) between U1 and U4 snRNP, potentially providing selectivity for snRNP specific proteins during snRNP assembly.

High Resolution Structures of Spliceosomal Components Revealed by Crystallography and NMR

Although the composition and conformation dynamics made crystallographic studies of the entire spliceosome difficult, crystallography and NMR studies of protein or RNA components of the spliceosome continue to make important contributions to our understanding of the structure and function of the splicing machinery (Table I; Supporting Information Table 1). We will review below high resolution structural studies of representative classes of spliceosomal proteins and snRNAs.

Prp8

Prp8 is arguably the most intriguing protein in the spliceosome. Prp8 is a component of the U5 snRNP and is also present in the tri-snRNP and the spliceosome (reviewed in Ref. 53). Prp8 is one of the largest proteins in the nucleus and Prp8 from all species are over 2000 amino acids in length. Prp8 is also one of the most conserved proteins in the nucleus. There is over 60% sequence identity between human and yeast Prp8. However, Prp8 has remarkably low sequence similarities with other proteins, making it difficult to deduce function from sequence analyses. On the other hand, functional studies clearly point out Prp8 as a key player at the center of the splicing reaction. UV crosslinking experiments demonstrate that Prp8 extensively crosslinks with the 5′ ss, 3′ ss, BPS, U5, and U6 in *in vitro*-assembled snRNPs and spliceosomes. Analyses of *in vivo* RNA-binding sites of Prp8 in the entire yeast cell using crosslinking experiments with intact cells followed by next generation sequencing revealed extensive Prp8 footprints on U5 snRNA, followed by U6, U1, U2, and pre-mRNAs.[54] The Prp8 footprints and crosslinking sites on U6, U2, and U5 snRNAs and pre-mRNA led to the hypothesis that Prp8 may help form/stabilize the catalytic core or even contribute functional groups to the splicing reaction (reviewed in Refs. 53 and 54). Genetic analyses identified numerous Prp8 mutants that either suppress or exacerbate other splicing mutations, many of which are on splicing factors important in spliceosomal activation, leading to the hypothesis that Prp8 may also be a master regulator of spliceosomal activation.

Crystal structures of the *C*-terminal domain (CTD) of Prp8 from *Caenorhabditis elegans* and yeast (Residues 2147–2413) reveal a central MPN domain with *N*- and *C*-terminal extensions.[55,56] This domain will be referred to as the MPN domain of Prp8. A subset of the MPN domains containing a JAMM motif was thought to coordinate a Zn^{++} ion and functions as metalloproteases, although the general function of most MPN domains remains unknown. Prp8 CTD contains a partial JAMM motif, which does not coordinate a Zn^{++} ion and is unlikely to be a Zn-dependent metalloprotease. On the other hand, GST pull down and yeast two-hybrid experiments demonstrate that the CTD of Prp8 interacts with Brr2 and Snu114, which may have implications for the role of Prp8 in the regulation of

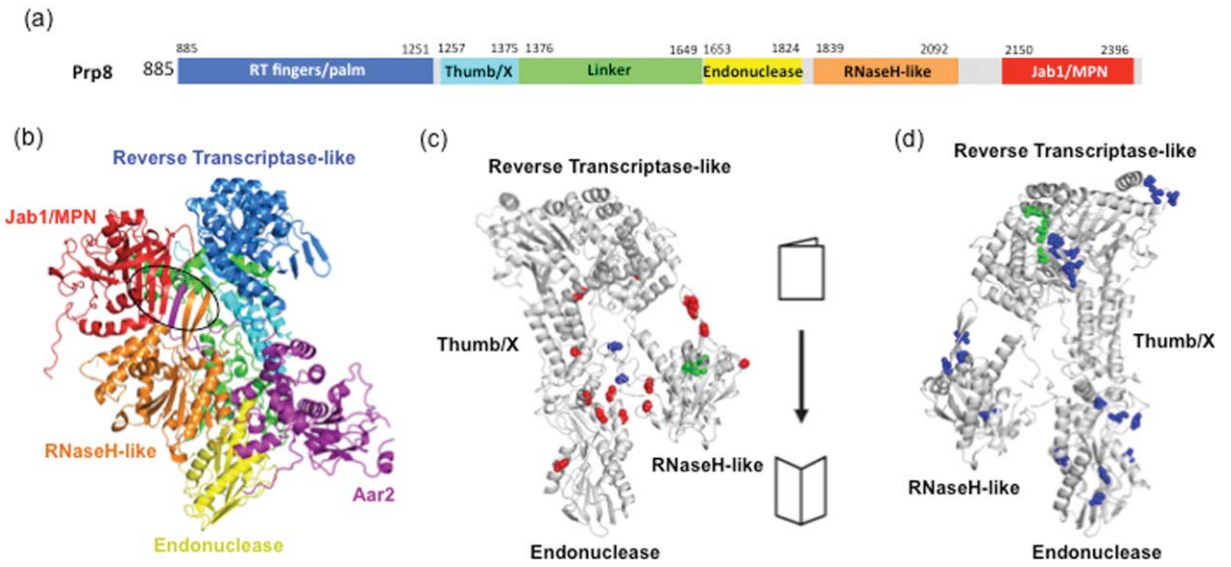

Figure 5. Crystal structure of a large fragment of yeast Prp8 (Residues 885–2413) (figures modified from Ref. 60). (a) Domain organization in this large fragment of Prp8. Gray represents disordered regions not observed in the crystal structure. (b) Structural representation of the reverse transcriptase, endonuclease, MPN, and RNase H domains of Prp8 and Aar2. Circle highlights the β-sheet formed by MPN (red), the *C*-terminus of Aar2 (purple), and RNase H-like (orange) domain. (c) The RNase H-like domain is flipped open to show suppressors of splice site mutations (red spheres) in an open book view. Green is the peptide in human Prp8 that crosslinks to the 5′ss. Blue is the region that crosslinks to the BPS. (d) U4-cs1 (blue spheres) and *brr2-1* (green spheres) suppressor mutants map onto one face of the RT/En domain.

spliceosomal activation. Prp8 contains a number of mutations implicated in retinitis pigmentosa type 13, a human genetic disorder that leads to photoreceptor degeneration and eventual blindness. All retinitis pigmentosa type 13 mutations are located on the *C*-terminal extension of the Prp8 CTD structure. Deletion of the *C*-terminal extension reduces the binding between Prp8 CTD and Brr2 or Snu114, suggesting a potential mechanism for the disease phenotype of these mutations.

The structure of another domain immediately upstream of the CTD was determined (Residues 1822–2095 in yeast),[57–59] which contains an RNase H core and a prominent β-hairpin finger protruding out of the middle of the protein. This domain will be referred to as the RNase H-like domain of Prp8. Canonical RNase H domains contain a conserved DDE active site that coordinates a Mg^{++} ion critical for catalysis. The RNase H domain in Prp8 does not contain a complete DDE triad and does not coordinate Mg^{++} ion. However, the RNase H domain in Prp8 seems to have maintained the RNA-binding capacity of canonical RNase H domains. This Prp8 domain binds a U2/U6 mimic and U4/U6 much better than arbitrary RNAs. Modeling RNA on the RNase H domain of Prp8 places the 5′ ss close to the previously observed 5′ ss cross-linking site on Prp8 in a cleft close to the β-finger. The protruding β-finger is strongly reminiscent of many ribosomal proteins with extensions (in the form of extended loops, α-helix, or β-hairpin fingers) protruding from globular protein bodies. These protrusions in ribosomal proteins insert into folded 16S or 23S rRNAs and stabilize the RNA structure. Mutations at the β-finger of Prp8 affect the conformational equilibrium between the first and second catalytic step and suppress U4-cs1 cold sensitivity, suggesting that Prp8 and its β-finger may interact extensively with RNAs in the spliceosomal complex or tri-snRNP, stabilizing these complexes. Mutations on the β-finger alter these interactions, resulting in the observed phenotypes.

The crystal structure of a large fragment of yeast Prp8 (Residues 885–2413, 176 kD) in complex with Aar2, a U5 snRNP assembly factor, was recently determined[60] (Fig. 5). The structure of this fragment contains a large ~1000-residue domain (Residues 885–1824) spanning the entire length of the complex, the RNaseH-like domain, and the MPN domain [Fig. 5(a)]. The latter two domains are connected by flexible linkers, fold back, and interact with the large domain through Aar2 (see the next paragraph for details) [Fig. 5(b)]. The large domain can be further divided into a large polymerase-like domain (Residues 885–1375), a linker domain, and a small type II restriction endonuclease-like domain (Residues 1650–1810). The polymerase-like domain is composed of three canonical subdomains: palm, fingers, and thumb. In canonical polymerase, the palm subdomain contains four conserved motifs harboring three Asp residues that coordinate an Mg^{++} ion required for catalysis. Consistent with the same theme observed in the Prp8 MPN and RNase H-like domains, the polymerase-like domain in Prp8 only

contains one of the three conserved Asp residues and is unlikely to bind divalent metal ions. The palm domain of Prp8 has been noticed recently to have sequence similarity with the reverse transcriptase (which is a RNA-dependent DNA polymerase and a class of polymerase) palm domain of bacterial group II intron-encoded protein.[61] Therefore, the polymerase-like domain will be more appropriately referred to as the reverse transcriptase domain. The type II restriction endonuclease domain (Residues 1650–1810) of Prp8 is structurally most similar to the influenza virus polymerase acidic subunit, even though there is no detectable sequence similarity. In the influenza polymerase acidic endonuclease domain, two Glu, one Asp, and one His residues are involved in coordinating two metal ions critical for catalysis. Although these residues are all present in Prp8, mutations of these residues do not have any effect on yeast viability. This large domain in Prp8 will be referred to as the reverse transcriptase-endonuclease (RT/En) domain.

The RT/En, RNase H-like, and MPN domains are connected by flexible linkers but form a large assembly through a network of interactions involving Aar2. For example, Aar2 interacts extensively with the En and linker subdomains. In addition, the *C*-terminal helical domain of Aar2 interacts with the RNase H-like domain. The very *C*-terminal end of Aar2 forms an incredible intermolecular β-sheet zipping together the β-finger of RNase H-like domain and the β-sheet in the MPN domain [Fig. 5(b)]. While the MPN and the RNase H-like domains on their own have limited to no interaction with the RT/En domain, Aar2 holds these domains together to form a large assembly. The relative orientations of these domains may be different when Prp8 is bound to Brr2 and/or Snu114 (the binding partners of Prp8 in mature U5 snRNP).

The crystal structure of this large Prp8 fragment reveals the potential active site cavity for the spliceosome. Prp8 extensively crosslinks to key components of the splicing reactions, including U2, U5, U6 snRNAs, and pre-mRNA.[53,54] These crosslinkings almost all map to the RT/En domain.[62] In particular, UV crosslinking in the spliceosome captured right before the second catalytic step maps to the region between Residues 1585 and 1598.[60] This region is disordered in the crystal structure but is located in the thumb subdomain of RT [Fig. 5(c)]. The thumb/En domains and the RNase H domain face each other and form a large cavity. Numerous Prp8 mutants that suppress mutations in the splice sites and branch point map to the surface of this cavity, on both the thumb/En domains and the RNase H domain, suggesting that this cavity holds the spliceosome active site. As discussed in the RNase H domain structure of Prp8, multiple Prp8 suppressor mutations map onto the β-finger of RNase H domain, which may sense and regulate the spliceosomal conformational switch between the first- and second-step reactions, supported by the flexibility of the RNase H domain relative to the RT/En domain in different crystal forms.

The crystal structure of this large Prp8 fragment also revealed potential mechanisms for Prp8's role in regulating spliceosomal activation. Over 40 Prp8 mutants (clustered into Regions a–e on the primary sequence) were identified that suppress the U4-cs1 mutant, a mutant on U4 snRNA that hyperstabilizes U4/U6 by extending Stem 1 in U4/U6.[63] Regions d and e of the U4-cs1 suppressor mutants are located on the crystal structure, most of which map to the same face of the RT/En domain, and the remaining five on the RNase H-like domain. Mutants on Prp8 were also identified that suppress the *brr2-1* mutant that has defective Brr2 helicase activity.[60,64] These *brr2-1* suppressors map on the same face of RT/En as the U4-cs1 suppressors. More specifically, these *brr2-1* suppressors are located in the same region in the palm domain as some of the U4-cs1 suppressors. This face of the RT/En may be a RNA or protein-binding surface that is critical for U4/U6 unwinding and spliceosomal activation. Brr2 is a candidate that can bind to this face of the RT/En, especially the palm domain. This binding will place Brr2 in a perfect position to feed U6 into the spliceosomal active site after U4/U6 unwinding.

The crystal structure of this large fragment of Prp8 revealed intriguing insight into the evolutionary origin of the spliceosome. The palm and thumb domain have considerable sequence and structural similarity with the bacterial and fungal Group II intron-encoded proteins (IEPs). Group II introns are mobile genetic elements found in organelles and bacteria genomes that self-splice through two transesterification steps identical to the eukaryotic splicing reaction. The self-splicing of group II introns is facilitated by the maturase activity of their IEPs, which usually contain a reverse transcriptase and a endonuclease domain. The IEP binds to the spliced intron, targeting it to a homing site on the genomic DNA where integration is achieved through reverse splicing. The opposite DNA strand is then cleaved by the endonuclease domain and used as a primer for reverse transcription of the intron by the reverse transcriptase domain of IEP. The ancestral nuclear pre-mRNA splicing could have evolved from the IEP (ancestral Prp8) and Group II intron RNAs on separate transcripts. When Group II intron ceases to be mobile elements, the selective pressure to maintain an active reverse transcriptase is lost. But the reverse transcriptase domain continues to serve as an assembly platform and maturase that facilitate splicing. The crystal structure of Group II intron reveals a tightly packed active site core organized by

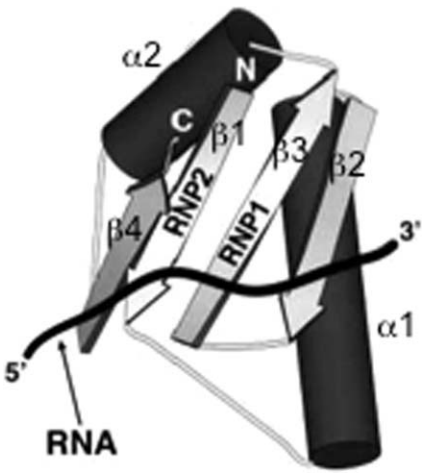

Figure 6. Structural representation of a canonical RRM domain based on the U1-A/RNA structure (figure modified from Ref. 71).

surrounding RNA scaffold. Biochemical studies indicate that the IEP binds to the Group II intron catalytic core to promote splicing. The Prp8 active site cavity is roughly the right size to accommodate the functional core of Group II intron RNA. This notion is supported by the observation that the spliceosomal RNA catalytic core crosslinks to the RT/En and RNase H domains, and the surface in these regions of Prp8 are remarkably conserved and highly positive. Prp8, maybe like an IEP, has replaced the RNA scaffold surrounding the functional core of Group II intron RNA, providing an intriguing evolutionary link between Group II self-splicing introns and the spliceosome.

RNA-binding proteins

There are many RNA-binding proteins in the spliceosome. We will discuss in this section some major RNA-binding domains and their representative structures.

RRM, sometimes referred to as the RBD, is the most common RNA-binding motif in eukaryotes (reviewed in Ref. 65) and is highly abundant in spliceosomal proteins. The RRM is approximately 90 amino acids long containing an RNP1 motif ([RK]-G-[FY]-[GA]-[FY]-[ILV]-X-[FY]) and an RNP2 motif ([ILV]-[FY]-[ILV]-X-N-L). The first 3D structure of the RRM domain determined was the U1-A RRM domain.[49] The structure is made of two α-helices packed against four antiparallel β-strands with a topology of β1α1β2β3α2β4 (Fig. 6). RNP1 is located on β3 and RNP2 on β1. Typically, the three aromatic rings from the first two conserved [FY] residues of RNP1 and the [FY] in RNP2 interact with single-stranded RNA through base stacking or interaction with sugar rings.

A number of spliceosomal proteins contain multiple RRM domains. For example, splicing factor $U2AF^{65}$ contains 3 RRM domains. RRM1 and 2 of $U2AF^{65}$ are responsible for recognizing the polypyrimidine tract near the 3′ ss. The crystal structure of an engineered $U2AF^{65}$ RRM1–2 domain (linker shortened from 30 to 10 residues) in complex with a 7nt poly U revealed specific hydrogen bonds between the protein and uracil bases, providing a structural basis for the preference for polyuridine.[66] A recent study using NMR demonstrates that the two RRM domains adopt a closed conformation with RRM1 and 2 interacting with each other when there is short polypyrimidine or no RNA present. The two RRM domains adopt an open conformation in the presence of long and high-affinity polypyrimidine tracks, with both RRMs contacting the RNA. The two RRM domains undergo a population shift between the open and the closed conformation when bound to polypyrimidine tract with different length and/or sequence. The equilibrium between the open and close conformation serves as a molecular rheostat that quantitatively correlates the natural variation in polypyrimidine tract (sequence and lengths) to the efficiency of recruiting U2 snRNP to the premRNA. Mutations that affect the equilibrium but not RNA binding also affect the splicing activity. This multidomain conformational selection represents an interesting mechanism in the recognition of degenerate nucleotide or amino acid motifs by multidomain proteins and may be applicable to other biological systems.

Prp24, an essential component of the U6 snRNP, is another example of a splicing factor with multiple RRM domains (four in Prp24) performing both RNA binding and other unique functions. Prp24 functions to anneal U4/U6 during spliceosomal assembly.[67] The crystal structure of RRM1–3 and the solution structure of RRM1–2,[68] RRM2 in complex with a hexa-ribonucleotide in U6 snRNA,[69] and RRM4[70] were determined. Each of RRM1–3 adopts a canonical fold. RRM1 and 2 are tightly packed against each other, forming a single RNA-binding surface. RRM2 binds U6 snRNA using its canonical RNA-binding β-sheet face and RRM1 binds noncanonically through a positively charged surface. RRM3 makes no stable contact with RRM2 and its RNA binding property is unclear. The structure of RRM4 assumes a noncanonical RRM fold with two additional flanking α-helices that occlude the β-sheet face (designated as the occluded RRM or oRRM). The flanking α-helices form a large positive-charged surface. The oRRM binds to and unwinds the U6 ISL, the internal stem loop in U6 snRNA that has to be unwound before U6 and U4 annealing.

Some RRM domains also evolved to perform protein-binding functions. The crystal structure of the $U2AF^{35}$/$U2AF^{65}$ hetereodimer[71] and the NMR structure of $U2AF^{65}$/SF1[72] reveal that the RRM domain in $U2AF^{35}$ and RRM3 of $U2AF^{65}$ both interact with

protein instead of RNA. These RRM-like domains are termed U2AF[65] homology motifs (UHM).[73] UHMs often contain aliphatic instead of basic or aromatic residues in RNP1 and RNP2. A Trp residue in the protein ligand inserts into a hydrophobic pocket formed between the α-helix and the RNP1 and RNP2-like motifs in UHM. A conserved R-X-F motif on the α2/β4 loop contributes to the Trp-binding pocket. In addition, a series of acidic residues in α1 of UHM interact with basic residues at the *N*-terminus of the protein ligand. More proteins were subsequently found to contain UHM that interact with proteins.[73]

In addition to RRM domains, other RBDs exist in spliceosomal proteins. For example, U1-C, a component of the U1 snRNP, contains a C2H2 type of Zn-finger domain.[29,50] The C2H2 type of Zn-finger is very common in transcription factors.[74] It forms a simple ββα fold with the α-helix typically binding to the major groove, specifically interacting with the DNA bases. The α-helix in the U1-C Zn-fingers, as well as the loop between the two His residues in U1-C, are in close proximity to the proposed 5′ss and U1 snRNA duplex,[29] potentially stabilizing the 5′ss and U1 snRNA interaction. Other Zn-finger containing spliceosomal proteins include Prp9, a component of the SF3a complex, which is a part of U2 snRNP.[75]

Another RBD present in spliceosomal proteins is the K homology (KH) domain, which is a ~70-amino acid fold that typically binds RNAs or single-stranded DNAs. It has a βααββα topology and is present in splicing factor 1 (SF1, which recognizes the BPS and helps the formation of the E complex). NMR studies demonstrated that the minimal BPS-binding region of SF1 contains a KH domain followed by a QUA2 (quaking homology 2) region.[76] The KH domain recognizes the 3′ end (UAAC) of the BPS (a preferred BPS sequence UACUAAC was used for this study) through a hydrophobic cleft formed by the G-P-R-G motif and the variable loop of the KH domain, a recognition mechanism common to other KH domains.[77] The QUA2 region of SF1 augments the KH domain and recognizes the 5′ end of the BPS (ACU).

RNA helicases

At least eight DExD/H-box proteins (UAP56, Prp5, Prp28, Brr2, Prp2, Prp16, Prp22, and Prp43) are involved in various steps of the spliceosomal assembly and activation process (Fig. 1).[15] DExD/H-box proteins belong to superfamily 2 (SF2) of helicase superfamilies, a class of enzymes that utilize the energy of NTP hydrolysis to unwind double-stranded (ds) DNAs or RNAs.[78] All superfamily 1 and 2 helicases contain the minimal helicase core, some with additional domains.[79–81] The minimal helicase core is composed of two RecA domains encompassing at least eight conserved helicase motifs. Key amino acids in Motifs I and II are highly conserved, with residues "DExD/H" in Motif II.[78,82] Other motifs are conserved within each family but not across the entire SF1 or SF2 superfamily, so these motifs are used to further classify SF1 or SF2 helicases into subfamilies. Multiple DExD/H-box proteins in the spliceosome have been demonstrated to have weak helicase activity *in vitro*.[83–88] Two major unwinding mechanisms have been proposed for DExD/H-box RNA helicases.[89] The first mechanism is used by DEAD-box RNA helicases that bind dsRNA and unwind by local strand separation without translocation.[90] The second mechanism is used by viral RNA helicases NPH-II and NS3 and potentially other eukaryotic RNA helicases, which bind to a single-stranded region adjacent to the duplex and translocate along the loading strand with specific polarity/directionality (3′ to 5′ or *vice versa*), displacing the complementary strand.[91,92] The spliceosomal DExD/H-box proteins have been proposed to be critical for the rearrangement of many mutually exclusive RNA/RNA and RNA/protein interactions in the splicing process. Kinetic proofreading mediated by DExD/H-box proteins through timely product progression and substrate discard is thought to be responsible for the maintenance of splicing fidelity at every ATP-dependent transition throughout the splicing cycle.[93–97] The regulation of these DExD/H-box proteins' activities are likely important for the fidelity of splicing.

Structural information on domains or full-length of spliceosomal helicases are available, including DECD protein UAP56,[98,99] the DEAH/RNA helicase A (RHA) proteins Prp43 and Prp22 (only the CTD), as well as ski2 type helicase Brr2. UAP56 is required for the formation of A complex.[100] The structure of UAP56 contains essentially the minimal helicase core of two RecA domains connected by an interdomain linker. It likely unwinds its RNA substrates through RNA bending and local strand separation, similar to other DEAD-box RNA helicases.[90,101]

Prp2, 16, 22, and 43 belong to the DEAH/RHA helicase family whose members all contain two conserved CTDs (a helicase-associated domain HA2 and another domain with unknown function) following the helicase core. Prp43 is involved in the release of intron lariat after the splicing reaction[102] as well as ribosomal RNA biogenesis.[103] Crystal structure of yeast Prp43 in complex with ADP and Mg^{++} revealed unexpected structural similarity with ski2 type DNA helicase Hel308[104,105] [Fig. 7(a,b)]. Prp43 contains 6 domains, with a Prp43 specific *N*-terminal domain (Domain 1), two RecA domains (Domain 2 and 3), two domains structurally homologous to the corresponding domains in Hel308, and a CTD with an oligonucleotide binding (OB)-fold that is different

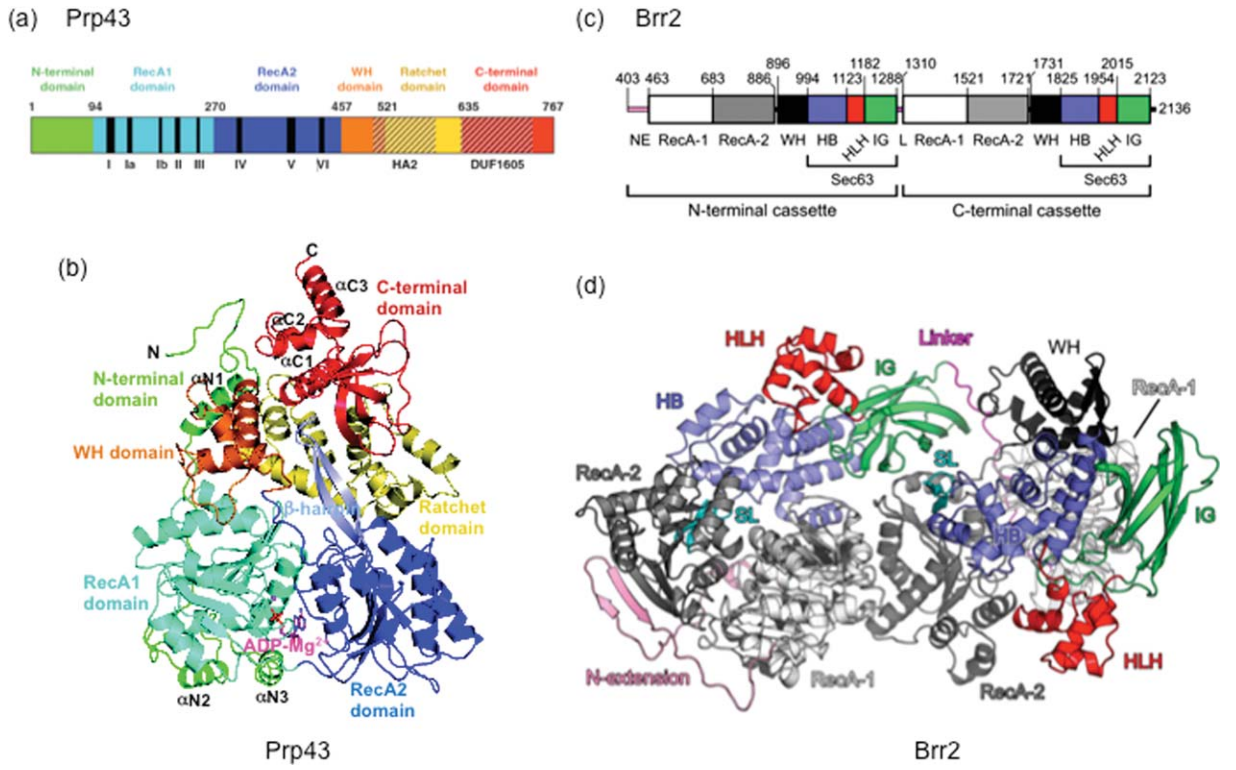

Figure 7. Crystal structure of spliceosomal helicases Prp43 (figure modified from Ref. 104) and Brr2 (figure modified from Ref. 113). (a) Schematic representation of the domain organizations of yeast Prp43. Structural domains and boundaries are indicated on the top. Conserved helicase motifs (roman numerals) and domains (hatched areas) on the bottom. (b) The overall structure of Prp43 in complex with ADP-Mg^{++}. The six structural domains are colored as in (a). (c) Schematic representation of the domain organization of the two consecutive helicase cassettes in Brr2. NE, *N*-terminal extension; RecA-1 and RecA-2, the first and second RecA domain; WH, HB, HLH, IG: winged helix, helical bundle, helix-loop-helix, and immunoglobulin domains. The same colors are used to represent the same domain in both the primary sequence and the structural representation. (d) The overall structure of the two helicase cassettes in Brr2. The domains are colored as in (c).

from the CTD of Hel308. Two key features in Hel308 that were thought to be crucial for the processive unwinding of DNA by Hel308 are both present in Prp43: a prominent β-hairpin in the second RecA that was thought to be responsible for separating the two nucleotide strands, and the ratchet helix in domain 4 that was thought to be responsible for single strand translocation. The OB-fold is generally associated with nucleic acid-binding activity. The nucleic acids-binding groove in the OB-fold of Prp43 is solvent exposed and highly positively charged and is in a perfect position to interact with the RNA substrate entering the unwinding cavity. Indeed, truncation of the *C*-terminal OB-fold in Prp43 retains basal ATPase activity but has significantly reduced RNA-binding affinity and RNA-stimulated ATPase activity. In addition, the *C*-terminal OB-fold serves as a binding site for G-patch proteins (such as Paf1) and is responsible for the G-ptach protein-mediated stimulation of the Prp43 ATPase and helicase activity.

The Prp43 structure revealed a general structural model for DEAH/RHA family of RNA helicases, including splicing helicases Prp2, Prp16, Prp22, and Prp43. They all likely contain two RecA domains, a winged helix (WH) and a ratchet domain similar to Hel308, and a *C*-terminal OB-fold domain. The crystal structure of the *C*-terminal region of human Prp22 downstream of the two RecA domains indeed confirmed that this region of Prp22 also contains the WH and ratchet domain of Hel308 and an OB-fold domain.[106] The *N*-terminal domain upstream of the first RecA domain is varied in lengths (~100 residues for Prp2 and Prp43 but over 500 residues for Prp16 and Prp22) and seems to be unique for each DEAH/RHA helicase. The structural similarity between DEAH/RHA family of RNA helicases and Hel308 suggests that these RNA helicases use a similar unwinding mechanism as Hel308, which involves the threading and translocation of single-stranded RNA in a 3′ to 5′ direction inside an enclosure formed by four domains in these RNA helicases and the separation of the dsRNA by the β-hairpin in the second RecA domain. This is consistent with the proposed function for Prp22, which potentially translocate on the mRNA and disrupt the mRNA and U5 interaction. Prp43 can potentially also translocate on the lariat intron to remove proteins or RNAs associated with the intron. The OB-fold facilitates RNA binding and is responsible

for the binding and stimulation by G-patch proteins, at least in the case of Prp43 (which interacts with G-patch protein Paf1)[104] and Prp2 (interacts with G-patch protein Spp2).[107]

Brr2, a U5 snRNP protein, is a large (250 kD) and unique helicase that contains two tandem helicase cassettes. It is responsible for U4/U6 unwinding during spliceosomal activation and U2/U6 unwinding during spliceosomal disassembly.[85,108–110] Helicase motifs in the first but not the second helicase cassette are critical for ATPase activity, U4/U6 unwinding, and cell viability.[109] The crystal structure of the *C*-terminal region (Sec63 domain) of the second helicase cassette revealed unexpected resemblance to Domains 4 and 5 of DNA helicase Hel308 with an additional fibronectin 3 CTD,[111,112] leading to the hypothesis that the full-length Brr2 is composed of an *N*-terminal domain and two consecutive Hel308-like cassettes (Hel308-I and -II). Crystal structure of a large fragment of human Brr2 encompassing the two helicase cassettes (Residues 403–2136) were also recently determined,[113] revealing that the two Hel308-like cassettes are tightly packed against each other [Fig. 7(c,d)]. The N-terminal helicase cassette can unwind U4/U6 on its own, but with a much lower activity. Based on the Hel308-DNA structure, the RNA substrate was thought to thread through the enclosure formed by Domains 1–5, translocate in a 3′ to 5′ direction with the help of the ratchet helix and unwind with the help of the β-hairpin in the second RecA domain. Since both 3′ and 5′ end of U4 and U6 are occupied by Sm proteins, the authors propose that the *N*-terminal helicase cassette can open up its enclosure between the second RecA domain and the ratchet domain and bind the ss U4 snRNA in the middle. The *C*-terminal helicase cassette retained ATP binding ability but not ATP hydrolysis. The tight packing between the two cassettes may help the *N*-terminal cassette to achieve the best conformation for catalysis. The large size and expanded surface of the *C*-terminal cassette present an opportunity for long distance regulation of the activity of the *N*-terminal cassette, potentially by multiple protein factors. Indeed, the second Hel308 cassette interacts with Prp8 and Snu114 *in vitro* and *in vivo*, potentially serving as a mediator for the regulation of Brr2's activity by Prp8.[111] The *C*-terminal region of Prp8 (Prp8-CTR, including the MPN domain and the RNase H-like domain) facilitates the binding of the Brr2/Prp8-CTR complex to U4/U6, suggesting a potential role of Prp8-CTR as an auxiliary substrate binding and specificity domain for Brr2.

Sequence and structural analyses reveal that most spliceosomal helicases contain a minimal helicase core with additional domains. These additional domains function to stabilize the helicase core and interact with substrate RNA or other proteins that modulate the helicase function. The Hel308 fold seems to be present predominantly in splicing helicases. In spite of the low sequence similarity between DEAH/RHA or Brr2 and Hel308, both the DEAD/RHA and Brr2 helicases have the Hel308 fold. This indicates that Brr2 and DEAH/RHA family helicases (Prp2, 16, 22, and 43) are all more processive than the typical RNA helicases that are thought to act through local strand separation, although confirming this will require detailed biochemical analyses. There are a few helicases whose structures remain unknown (for example, Prp5 and Prp28). Structural studies of these helicases will likely shed new light into their function just as the study of the other splicing helicases. In addition, although we know roughly which step each splicing helicase is involved, the precise targets are unknown. Identifying the physiological targets of spliceosomal helicases and structural analyses of these helicases with their endogenous RNA targets will be a significant step toward understanding the unwinding mechanism, function, and regulation of these helicases.

snRNAs

Since the chemistry of the splicing reaction is identical to the Group II self-splicing introns and both reactions require Mg^{++} ions, the spliceosome was hypothesized to be mainly a ribozyme and snRNAs play important roles in the splicing reaction. U6 snRNA is most likely to play a catalytic role in the splicing reaction based on phylogenetic, genetic, and biochemical analyses.[114] The structure of the highly conserved U6 intramolecular stem loop (ISL) from yeast was determined using NMR.[115–117] The Watson-Crick paired region of the stem has standard A-form helical geometry. U80 in the 3nt internal loop of ISL binds an Mg^{++} ion, which is possibly the Mg^{++} ion required for catalysis. The GCAUA pentaloop in ISL makes a GNRA-type fold that often mediates tertiary interaction with RNA and was hypothesized to bring the metal ion in ISL close to the 5′ ss. There are intriguing structural and functional similarities between U6 ISL and domain V of Group II intron[118] [Fig. 8(a)]. U6 ISL and Domain V both form stem-loops with similar length and secondary structures. Each stem-loop has two helices separated by an internal Mg^{++}-binding bulge and both stem-loops have a GNRA-type loop.

Butcher and coworkers[119,120] also analyzed the structures of various yeast U2/U6 constructs (including the ISL region) using NMR and small angle x-ray scattering (SAXS).[119,120] The shorter U2/U6 constructs truncated at Helix I or II (or both) form a four-helix junction structure,[119] which is potentially an artifact of the truncated constructs. The longer U2/U6 construct (110nt) containing full length Helix I and II [Fig. 8(b)] form a three-helix

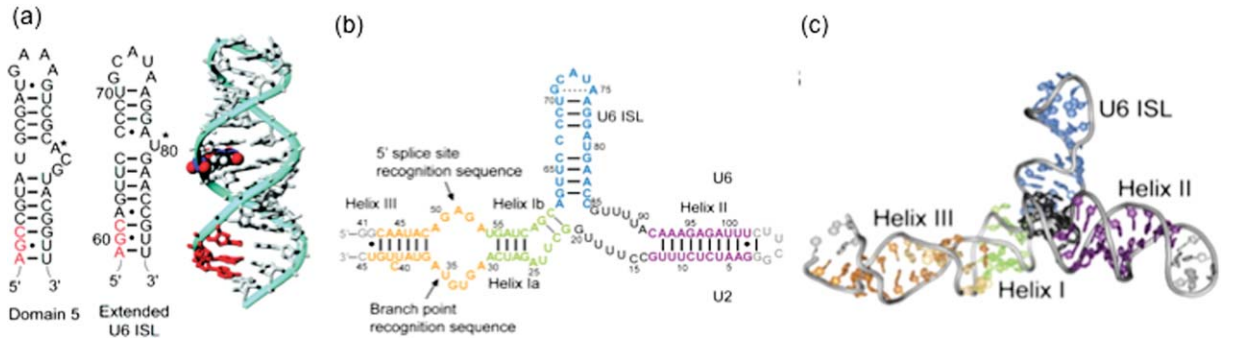

Figure 8. Structural analyses of snRNAs. (a) U6 ISL determined using NMR (right) and domain 5 of group II self-splicing introns demonstrate striking similarities (figure modified from Ref. 118). Metal binding sites are indicated with an asterisk. The AGC triad is shown in red. The U80 nucleotide is represented as a space-filling model. (b) Secondary structure of the 110nt U2/U6 construct used for NMR and SAXS studies (figure modified from Ref. 120). Structural features are Helix I (green), Helix II (purple), Helix III, and an internal loop that binds to 5′ splice site and branch point (orange), U6 ISL (blue), U-rich loop (dark gray), and nonnative sequences (light gray). Black lines or circles denote base pairs determined by NMR experiments. (c) A structural model of the 110nt U2/U6 construct derived from SAXS (figure modified from Ref. 120). Colors in the structural model are the same as that in the secondary structure in (b).

junction consistent with extensive genetic studies.[120] SAXS studies suggest that the 110nt U2/U6 complex assumes a Y shape, with the U6 ISL, Helix Ia, Ib, and III forming a continuous stacking[120] [Fig. 8(c)]. Components in this U2/U6 complex that are critical for catalysis, including the U80 metal-binding site, AGC triad, and the pre-mRNA recognition site, all localize to one face of the structure, potentially interacting with the pre-mRNA substrate or other cofactors.

Over the past several decades, we have made significant progress in understanding the structure and function of the spliceosome. However, high resolution structures of large spliceosome or subcomplexes are still lacking. In the future, improving sample homogeneity will be the key for improving resolution of the EM structure and enabling crystallographic studies of large spliceosome or subcomplexes. Purifying the spliceosome or snRNPs under more stringent conditions or from different species, using small molecule inhibitors or temperature sensitive mutants to trap the spliceosome, or reconstituting spliceosomal subcomplexes from recombinant proteins, are all approaches that can potentially improve sample homogeneity. The hybrid approach that combines multiple structural biology approaches (EM, crystallography, and NMR) as well as biochemical and genetic data have made significant contributions to our understanding of other large cellular machinery (such as the nuclear pore complex[121]). The same method is likely the best approach in the foreseeable future to obtain a detailed and comprehensive view of this complicated molecular machinery.

ACKNOWLEDGMENT

We apologize to our colleagues whose work is not cited due to space limit.

REFERENCES

1. Mattick JS, Gagen MJ (2001) The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. Mol Biol Evol 18:1611–1630.
2. Nilsen TW, Graveley BR (2010) Expansion of the eukaryotic proteome by alternative splicing. Nature 463: 457–463.
3. Venables JP (2002) Alternative splicing in the testes. Curr Opin Genet Dev 12:615–619.
4. Lopez AJ (1998) Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. Annu Rev Genet 32:279–305.
5. Grabowski PJ, Black DL (2001) Alternative RNA splicing in the nervous system. Prog Neurobiol 65: 289–308.
6. Faustino NA, Cooper TA (2003) Pre-mRNA splicing and human disease. Genes Dev 17:419–437.
7. Philips AV, Cooper TA (2000) RNA processing and human disease. Cell Mol Life Sci 57:235–249.
8. Krawczak M, Reiss J, Cooper DN (1992) The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. Hum Genet 90:41–54.
9. Moore MJ, Query CC, Sharp PA, Splicing of precursors to mRNA by the spliceosome. In: Gesteland R, and Atkins J, Ed.(1993) The RNA World. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp 303–357.
10. Hastings ML, Krainer AR (2001) Pre-mRNA splicing in the new millennium. Curr Opin Cell Biol 13: 302–309.
11. Jurica MS, Moore MJ (2003) Pre-mRNA splicing: awash in a sea of proteins. Mol Cell 12:5–14.
12. Nilsen TW (2003) The spliceosome: the most complex macromolecular machine in the cell? Bioessays 25: 1147–1149.
13. Wahl MC, Will CL, Luhrmann R (2009) The spliceosome: design principles of a dynamic RNP machine. Cell 136:701–718.
14. Brow DA (2002) Allosteric cascade of spliceosome activation. Annu Rev Genet 36:333–360.
15. Staley JP, Guthrie C (1998) Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell 92:315–326.

16. Behzadnia N, Golas MM, Hartmuth K, Sander B, Kastner B, Deckert J, Dube P, Will CL, Urlaub H, Stark H, Lührmann R (2007) Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. EMBO J 26: 1737–1748.
17. Boehringer D, Makarov EM, Sander B, Makarova OV, Kastner B, Lührmann R, Stark H (2004) Three-dimensional structure of a pre-catalytic human spliceosomal complex B. Nat Struct Mol Biol 11:463–468.
18. 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.
19. Golas MM, Sander B, Bessonov S, Grote M, Wolf E, Kastner B, Stark H, Lührmann R (2010) 3D cryo-EM structure of an active step I spliceosome and localization of its catalytic core. Mol Cell 40:927–938.
20. Wolf E, Kastner B, Deckert J, Merz C, Stark H, Lührmann R (2009) Exon, intron and splice site locations in the spliceosomal B complex. EMBO J 28: 2283–2292.
21. Alcid EA, Jurica MS (2008) A protein-based EM label for RNA identifies the location of exons in spliceosomes. Nat Struct Mol Biol 15:213–215.
22. Stroupe ME, Xu C, Goode BL, Grigorieff N (2009) Actin filament labels for localizing protein components in large complexes viewed by electron microscopy. RNA 15:244–248.
23. Deckert J, Hartmuth K, Boehringer D, Behzadnia N, Will CL, Kastner B, Stark H, Urlaub H, Lührmann 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.
24. Bessonov S, Anokhina M, Krasauskas A, Golas MM, Sander B, Will CL, Urlaub H, Stark H, Lührmann R (2010) Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. RNA 16:2384–2403.
25. Fabrizio P, Dannenberg J, Dube P, Kastner B, Stark H, Urlaub H, Lührmann R (2009) The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. Mol Cell 36:593–608.
26. Warkocki Z, Odenwälder P, Schmitzová J, Platzmann F, Stark H, Urlaub H, Ficner R, Fabrizio P, Lührmann R (2009) Reconstitution of both steps of Saccharomyces cerevisiae splicing with purified spliceosomal components. Nat Struct Mol Biol 16: 1237–1243.
27. Herold N, Will CL, Wolf E, Kastner B, Urlaub H, Lührmann R (2009) Conservation of the protein composition and electron microscopy structure of Drosophila melanogaster and human spliceosomal complexes. Mol Cell Biol 29:281–301.
28. Stark H, Dube P, Luhrmann R, Kastner B (2001) Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. Nature 409:539–542.
29. 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.
30. Behrens SE, Tyc K, Kastner B, Reichelt J, Luhrmann R (1993) Small nuclear ribonucleoprotein (RNP) U2 contains numerous additional proteins and has a bipartite RNP structure under splicing conditions. Mol Cell Biol 13:307–319.
31. MacMillan AM, Query CC, Allerson CR, Chen S, Verdine GL, Sharp PA (1994) Dynamic association of proteins with the pre-mRNA branch region. Genes Dev 8: 3008–3020.
32. Query CC, Strobel SA, Sharp PA (1996) Three recognition events at the branch-site adenine. EMBO J 15: 1392–1402.
33. Will CL, Urlaub H, Achsel T, Gentzel M, Wilm M, Lührmann R (2002) Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. EMBO J 21:4978–4988.
34. Golas MM, Sander B, Will CL, Luhrmann R, Stark H (2003) Molecular architecture of the multiprotein splicing factor SF3b. Science 300:980–984.
35. Sander B, Golas MM, Makarov EM, Brahms H, Kastner B, Lührmann R, Stark H (2006) Organization of core spliceosomal components U5 snRNA loop I and U4/U6 Di-snRNP within U4/U6.U5 Tri-snRNP as revealed by electron cryomicroscopy. Mol Cell 24: 267–278.
36. Hacker I, Sander B, Golas MM, Wolf E, Karagöz E, Kastner B, Stark H, Fabrizio P, Lührmann R (2008) Localization of Prp8, Brr2, Snu114 and U4/U6 proteins in the yeast tri-snRNP by electron microscopy. Nat Struct Mol Biol 15:1206–1212.
37. Karaduman R, Dube P, Stark H, Fabrizio P, Kastner B, Lührmann R (2008) Structure of yeast U6 snRNPs: arrangement of Prp24p and the LSm complex as revealed by electron microscopy. RNA 14:2528–2537.
38. Wyatt JR, Sontheimer EJ, Steitz JA (1992) Site-specific cross-linking of mammalian U5 snRNP to the 5' splice site before the first step of pre-mRNA splicing. Genes Dev 6:2542–2553.
39. Sontheimer EJ, Steitz JA (1993) The U5 and U6 small nuclear RNAs as active site components of the spliceosome. Science 262:1989–1996.
40. McConnell TS, Steitz JA (2001) Proximity of the invariant loop of U5 snRNA to the second intron residue during pre-mRNA splicing. EMBO J 20: 3577–3586.
41. 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.
42. Bringmann P, Luhrmann R (1986) Purification of the individual snRNPs U1, U2, U5 and U4/U6 from HeLa cells and characterization of their protein constituents. EMBO J 5:3509–3516.
43. Ruby SW, Abelson J (1988) An early hierarchic role of U1 small nuclear ribonucleoprotein in spliceosome assembly. Science 242:1028–1035.
44. Krol A, Westhof E, Bach M, Lührmann R, Ebel JP, Carbon P (1990) Solution structure of human U1 snRNA. Derivation of a possible three-dimensional model. Nucleic Acids Res 18:3803–3811 (1990).
45. Patton JR, Pederson T (1988) The Mr 70,000 protein of the U1 small nuclear ribonucleoprotein particle binds to the 5' stem-loop of U1 RNA and interacts with Sm domain proteins. Proc Natl Acad Sci USA 85: 747–751.
46. Bach M, Krol A, Luhrmann R (1990) Structure-probing of U1 snRNPs gradually depleted of the U1-specific proteins A, C and 70k. Evidence that A interacts differentially with developmentally regulated mouse U1 snRNA variants. Nucleic Acids Res 18:449–457.
47. Nelissen RL, Will CL, van Venrooij WJ, Luhrmann R (1994) The association of the U1-specific 70K and C

proteins with U1 snRNPs is mediated in part by common U snRNP proteins. EMBO J 13:4113–4125.
48. Kambach C, Walke S, Young R, Avis JM, de la Fortelle E, Raker VA, Lührmann R, Li J, Nagai K (1999) Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. Cell 96:375–387.
49. Oubridge C, Ito N, Evans PR, Teo CH, Nagai K (1994) Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. Nature 372:432–438.
50. Muto Y, Pomeranz Krummel D, Oubridge C, Hernandez H, Robinson CV, Neuhaus D, Nagai K (2004) The structure and biochemical properties of the human spliceosomal protein U1C. J Mol Biol 341:185–198.
51. Will CL, Rumpler S, Klein Gunnewiek J, van Venrooij WJ, Luhrmann R (1996) In vitro reconstitution of mammalian U1 snRNPs active in splicing: the U1-C protein enhances the formation of early (E) spliceosomal complexes. Nucleic Acids Res 24:4614–4623.
52. 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.
53. Grainger RJ, Beggs JD (2005) Prp8 protein: at the heart of the spliceosome. RNA 11:533–557.
54. Li X, Zhang W, Xu T, Ramsey J, Zhang L, Hill R, Hansen KC, Hesselberth JR, Zhao R (2013) Comprehensive in vivo RNA-binding site analyses reveal a role of Prp8 in spliceosomal assembly. Nucleic Acids Res 41:3805–3518.
55. Pena V, Liu S, Bujnicki JM, Luhrmann R, Wahl MC (2007) Structure of a multipartite protein-protein interaction domain in splicing factor prp8 and its link to retinitis pigmentosa. Mol Cell 25:615–624.
56. Zhang L, Shen J, Guarnieri MT, Heroux A, Yang K, Zhao R (2007) Crystal structure of the C-terminal domain of splicing factor Prp8 carrying retinitis pigmentosa mutants. Protein Sci 16:1024–1031.
57. Yang K, Zhang L, Xu T, Heroux A, Zhao R (2008) Crystal structure of the beta-finger domain of Prp8 reveals analogy to ribosomal proteins. Proc Natl Acad Sci USA 105:13817–13822.
58. Ritchie DB, Schellenberg MJ, Gesner EM, Raithatha SA, Stuart DT, Macmillan AM (2008) Structural elucidation of a PRP8 core domain from the heart of the spliceosome. Nat Struct Mol Biol 15:1199–1205.
59. Pena V, Rozov A, Fabrizio P, Luhrmann R, Wahl MC (2008) Structure and function of an RNase H domain at the heart of the spliceosome. EMBO J 27: 2929–2940.
60. Galej WP, Oubridge C, Newman AJ, Nagai K (2013) Crystal structure of Prp8 reveals active site cavity of the spliceosome. Nature 493:638–643.
61. Dlakic M, Mushegian A (2011) Prp8, the pivotal protein of the spliceosomal catalytic center, evolved from a retroelement-encoded reverse transcriptase. RNA 17:799–808.
62. Turner IA, Norman CM, Churcher MJ, Newman AJ (2006) Dissection of Prp8 protein defines multiple interactions with crucial RNA sequences in the catalytic core of the spliceosome. RNA 12:375–386.
63. Kuhn AN, Brow DA (2000) Suppressors of a cold-sensitive mutation in yeast U4 RNA define five domains in the splicing factor Prp8 that influence spliceosome activation. Genetics 155:1667–1682.
64. Kuhn AN, Reichl EM, Brow DA (2002) Distinct domains of splicing factor Prp8 mediate different aspects of spliceosome activation. Proc Natl Acad Sci USA 99:9145–9149.
65. Clery A, Blatter M, Allain FH (2008) RNA recognition motifs: boring? Not quite. Curr Opin Struct Biol 18: 290–298.
66. 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.
67. Raghunathan PL, Guthrie C (1998) A spliceosomal recycling factor that reanneals U4 and U6 small nuclear ribonucleoprotein particles. Science 279:857–860.
68. Bae E, Reiter NJ, Bingman CA, Kwan SS, Lee D, Phillips GN Jr, Butcher SE, Brow DA (2007) Structure and interactions of the first three RNA recognition motifs of splicing factor prp24. J Mol Biol 367: 1447–1458.
69. Martin-Tumasz S, Reiter NJ, Brow DA, Butcher SE (2010) Structure and functional implications of a complex containing a segment of U6 RNA bound by a domain of Prp24. RNA 16:792–804.
70. Martin-Tumasz S, Richie AC, Clos LJ 2nd, Brow DA, Butcher SE (2011) A novel occluded RNA recognition motif in Prp24 unwinds the U6 RNA internal stem loop. Nucleic Acids Res 39:7837–7847.
71. Kielkopf CL, Rodionova NA, Green MR, Burley SK (2001) A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. Cell 106:595–605.
72. Selenko P, Gregorovic G, Sprangers R, Stier G, Rhani Z, Krämer A, Sattler M (2003) Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. Mol Cell 11:965–976.
73. Kielkopf CL, Lucke S, Green MR (2004) U2AF homology motifs: protein recognition in the RRM world. Genes Dev 18:1513–1526.
74. Klug A, Rhodes D (1987) Zinc fingers: a novel protein fold for nucleic acid recognition. Cold Spring Harbor Symp Quant Biol 52:473–482.
75. Lin PC, Xu RM (2012) Structure and assembly of the SF3a splicing factor complex of U2 snRNP. EMBO J 31:1579–1590.
76. Liu Z, Luyten I, Bottomley MJ, Messias AC, Houngninou-Molango S, Sprangers R, Zanier K, Krämer A, Sattler M (2001) Structural basis for recognition of the intron branch site RNA by splicing factor 1. Science 294:1098–1102.
77. Lewis HA, Musunuru K, Jensen KB, Edo C, Chen H, Darnell RB, Burley SK (2000) Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome. Cell 100:323–332.
78. Gorbalenya AE, Koonin EV (1993) Helicases: amino acid sequence comparisons and structure-function relationships. Curr Opin Struct Biol 3:419–429.
79. Pyle AM (2011) RNA helicases and remodeling proteins. Curr Opin Chem Biol 15:636–642.
80. Jankowsky E, Fairman ME (2007) RNA helicases—one fold for many functions. Curr Opin Struct Biol 17: 316–324.
81. Cordin O, Banroques J, Tanner NK, Linder P (2006) The DEAD-box protein family of RNA helicases. Gene 367:17–37.
82. Tanner NK, Linder P (2001) DExD/H box RNA helicases: from generic motors to specific dissociation functions. Mol Cell 8:251–262 (2001).
83. Wagner JD, Jankowsky E, Company M, Pyle AM, Abelson JN (1998) The DEAH-box protein PRP22 is an ATPase that mediates ATP-dependent mRNA release from the spliceosome and unwinds RNA duplexes. EMBO J 17:2926–2937.

84. 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.
85. Laggerbauer B, Achsel T, Luhrmann R (1998) The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplices in vitro. Proc Natl Acad Sci USA 95: 4188–4192.
86. Wang Y, Wagner JD, Guthrie C (1998) The DEAH-box splicing factor Prp16 unwinds RNA duplexes in vitro. Curr Biol 8:441–451.
87. Tanaka N, Schwer B (2006) Mutations in PRP43 that uncouple RNA-dependent NTPase activity and pre-mRNA splicing function. Biochemistry 45:6510–6521.
88. Shen J, Zhang L, Zhao R (2007) Biochemical characterization of the ATPase and helicase activity of UAP56, an essential pre-mRNA splicing and mRNA export factor. J Biol Chem 282:22544–22550.
89. Jankowsky E (2011) RNA helicases at work: binding and rearranging. Trends Biochem Sci 36:19–29.
90. Yang Q, Del Campo M, Lambowitz AM, Jankowsky E (2007) DEAD-box proteins unwind duplexes by local strand separation. Mol Cell 28:253–263.
91. Jankowsky E, Gross CH, Shuman S, Pyle AM (2000) The DExH protein NPH-II is a processive and directional motor for unwinding RNA. Nature 403: 447–451.
92. Pang PS, Jankowsky E, Planet PJ, Pyle AM (2002) The hepatitis C viral NS3 protein is a processive DNA helicase with cofactor enhanced RNA unwinding. EMBO J 21:1168–1176.
93. Burgess S, Couto JR, Guthrie C (1990) A putative ATP binding protein influences the fidelity of branchpoint recognition in yeast splicing. Cell 60:705–717.
94. Burgess SM, Guthrie C (1993) A mechanism to enhance mRNA splicing fidelity: the RNA-dependent ATPase Prp16 governs usage of a discard pathway for aberrant lariat intermediates. Cell 73:1377–1391.
95. Burgess SM, Guthrie C (1993) Beat the clock: paradigms for NTPases in the maintenance of biological fidelity. Trends Biochem Sci 18:381–384.
96. Mayas RM, Maita H, Staley JP (2006) Exon ligation is proofread by the DExD/H-box ATPase Prp22p. Nat Struct Mol Biol 13:482–490.
97. Xu YZ, Query CC (2007) Competition between the ATPase Prp5 and branch region-U2 snRNA pairing modulates the fidelity of spliceosome assembly. Mol Cell 28:838–849.
98. Zhao R, Shen J, Green MR, MacMorris M, Blumenthal T (2004) Crystal structure of UAP56, a DExD/H-box protein involved in pre-mRNA splicing and mRNA export. Structure 12:1373–1381.
99. Shi H, Cordin O, Minder CM, Linder P, Xu RM (2004) Crystal structure of the human ATP-dependent splicing and export factor UAP56. Proc Natl Acad Sci USA 101:17628–17633.
100. Fleckner J, Zhang M, Valcarcel J, Green MR (1997) U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. Genes Dev 11:1864–1872.
101. Sengoku T, Nureki O, Nakamura A, Kobayashi S, Yokoyama S (2006) Structural basis for RNA unwinding by the DEAD-box protein Drosophila vasa. Cell 125:287–300.
102. Arenas JE, Abelson JN (1997) Prp43: an RNA helicase-like factor involved in spliceosome disassembly. Proc Natl Acad Sci USA 94:11798–11802.
103. Lebaron S, Papin C, Capeyrou R, Chen YL, Froment C, Monsarrat B, Caizergues-Ferrer M, Grigoriev M, Henry Y (2009) The ATPase and helicase activities of Prp43p are stimulated by the G-patch protein Pfa1p during yeast ribosome biogenesis. EMBO J 28:3808–3819.
104. Walbott H, Mouffok S, Capeyrou R, Lebaron S, Humbert O, van Tilbeurgh H, Henry Y, Leulliot N (2010) Prp43p contains a processive helicase structural architecture with a specific regulatory domain. EMBO J 29:2194–2204.
105. He Y, Andersen GR, Nielsen KH (2010) Structural basis for the function of DEAH helicases. EMBO Rep 11: 180–186.
106. Kudlinzki D, Schmitt A, Christian H, Ficner R (2012) Structural analysis of the C-terminal domain of the spliceosomal helicase Prp22. Biol Chem 393: 1131–1140.
107. Silverman EJ, Maeda A, Wei J, Smith P, Beggs JD, Lin RJ (2004) Interaction between a G-patch protein and a spliceosomal DEXD/H-box ATPase that is critical for splicing. Mol Cell Biol 24:10101–10110.
108. Raghunathan PL, Guthrie C (1998) RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. Curr Biol 8:847–855.
109. Kim DH, Rossi JJ (1999) The first ATPase domain of the yeast 246-kDa protein is required for in vivo unwinding of the U4/U6 duplex. RNA 5:959–971.
110. Small EC, Leggett SR, Winans AA, Staley JP (2006) The EF-G-like GTPase Snu114p regulates spliceosome dynamics mediated by Brr2p, a DExD/H box ATPase. Mol Cell 23:389–399.
111. Zhang L, Xu T, Maeder C, Bud LO, Shanks J, Nix J, Guthrie C, Pleiss JA, Zhao R (2009) Structural evidence for consecutive Hel308-like modules in the spliceosomal ATPase Brr2. Nat Struct Mol Biol 16: 731–739.
112. Pena V, Jovin SM, Fabrizio P, Orlowski J, Bujnicki JM, Lührmann R, Wahl MC (2009) Common design principles in the spliceosomal RNA helicase Brr2 and in the Hel308 DNA helicase. Mol Cell 35:454–466.
113. Santos KF, Jovin SM, Weber G, Pena V, Lührmann R, Wahl MC (2012) Structural basis for functional cooperation between tandem helicase cassettes in Brr2-mediated remodeling of the spliceosome. Proc Natl Acad Sci USA 109:17418–17423.
114. Valadkhan S (2010) Role of the snRNAs in spliceosomal active site. RNA Biol 7:345–353.
115. Huppler A, Nikstad LJ, Allmann AM, Brow DA, Butcher SE (2002) Metal binding and base ionization in the U6 RNA intramolecular stem-loop structure. Nat Struct Biol 9:431–435.
116. Blad H, Reiter NJ, Abildgaard F, Markley JL, Butcher SE (2005) Dynamics and metal ion binding in the U6 RNA intramolecular stem-loop as analyzed by NMR. J Mol Biol 353:540–555.
117. Venditti V, Clos L 2nd, Niccolai N, Butcher SE (2009) Minimum-energy path for a u6 RNA conformational change involving protonation, base-pair rearrangement and base flipping. J Mol Biol 391:894–905.
118. Butcher SE, Brow DA (2005) Towards understanding the catalytic core structure of the spliceosome. Biochem Soc Trans 33:447–449.
119. Sashital DG, Cornilescu G, McManus CJ, Brow DA, Butcher SE (2004) U2-U6 RNA folding reveals a group II intron-like domain and a four-helix junction. Nat Struct Mol Biol 11:1237–1242.
120. Burke JE, Sashital DG, Zuo X, Wang YX, Butcher SE (2012) Structure of the yeast U2/U6 snRNA complex. RNA 18:673–683.
121. Hoelz A, Debler EW, Blobel G (2011) The structure of the nuclear pore complex. Annu Rev Biochem 80: 613–643.