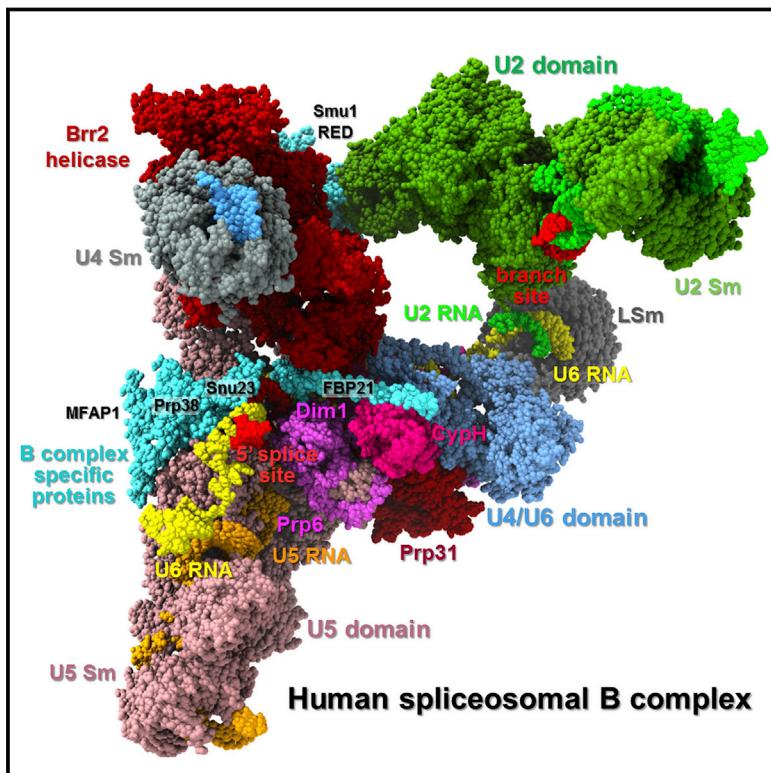


Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation

Graphical Abstract



Authors

Karl Bertram, Dmitry E. Agafonov,
Olexandr Dybkov, ..., Berthold Kastner,
Reinhard Lührmann, Holger Stark

Correspondence

b.kastner@mpibpc.mpg.de (B.K.),
reinhard.luehrmann@mpibpc.mpg.de
(R.L.),
hstark1@gwdg.de (H.S.)

In Brief

Structure of the human spliceosome getting ready for action.

Highlights

- Molecular architecture of the human spliceosomal B complex elucidated by cryo-EM
- Dramatic restructuring of the human tri-snRNP following its integration into complex B
- Elucidation of the location and likely functions of B-specific proteins in complex B
- Marked differences in the early activation pathways of human and yeast spliceosomes

Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation

Karl Bertram,^{1,5} Dmitry E. Agafonov,^{2,5} Alexandr Dubkov,² David Haselbach,¹ Majety N. Leelaram,² Cindy L. Will,² Henning Urlaub,^{3,4} Berthold Kastner,^{2,*} Reinhard Lührmann,^{2,6,*} and Holger Stark^{1,*}

¹Department of Structural Dynamics, MPI for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

²Department of Cellular Biochemistry, MPI for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

³Bioanalytical Mass Spectrometry, MPI for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

⁴Bioanalytics Group, Institute for Clinical Chemistry, University Medical Center Göttingen, Robert-Koch-Straße 40, 37075 Göttingen, Germany

⁵These authors contributed equally

⁶Lead Contact

*Correspondence: b.kastner@mpibpc.mpg.de (B.K.), reinhard.luehrmann@mpibpc.mpg.de (R.L.), hstark1@gwdg.de (H.S.)

<http://dx.doi.org/10.1016/j.cell.2017.07.011>

SUMMARY

Little is known about the spliceosome's structure before its extensive remodeling into a catalytically active complex. Here, we report a 3D cryo-EM structure of a pre-catalytic human spliceosomal B complex. The U2 snRNP-containing head domain is connected to the B complex main body via three main bridges. U4/U6.U5 tri-snRNP proteins, which are located in the main body, undergo significant rearrangements during tri-snRNP integration into the B complex. These include formation of a partially closed Prp8 conformation that creates, together with Dim1, a 5' splice site (ss) binding pocket, displacement of Sad1, and rearrangement of Brr2 such that it contacts its U4/U6 substrate and is poised for the subsequent spliceosome activation step. The molecular organization of several B-specific proteins suggests that they are involved in negatively regulating Brr2, positioning the U6/5'ss helix, and stabilizing the B complex structure. Our results indicate significant differences between the early activation phase of human and yeast spliceosomes.

INTRODUCTION

The spliceosome forms stepwise on its pre-mRNA substrate by sequential recruitment of snRNPs (small nuclear ribonucleoproteins) and numerous other proteins (Papasaikas and Valcárcel, 2016; Wahl et al., 2009). During early spliceosome assembly, U1 and U2 snRNPs interact with the 5' splice site (ss) and branch site (BS), respectively, of the intron, forming the spliceosomal A complex. The pre-formed U4/U6.U5 tri-snRNP then joins the A complex, generating the pre-B complex in which the tri-snRNP is not yet stably bound (Boesler et al., 2016). After conformational changes, including Prp28 RNA helicase-mediated exchange of U1 with U6 at the 5'ss, the pre-catalytic B complex with stably associated tri-snRNP is formed (Boesler et al., 2016; Staley

and Guthrie, 1999). Complex B undergoes extensive compositional and conformational rearrangements, including dissociation of U1 and U4, yielding the activated B^{act} complex. The latter is converted into a catalytically active spliceosome (designated B*) that catalyzes step I of splicing, yielding the cleaved 5' exon and intron-3' exon lariat intermediates. At this stage, the spliceosomal C complex is generated and after additional rearrangements, the C* complex catalyzes step II, resulting in ligation of the 5' and 3' exons and release of the spliced-out intron.

During spliceosome assembly and activation a dynamic RNA-RNA network involving snRNAs and the pre-mRNA is formed (Staley and Guthrie, 1998). U4 and U6 snRNA are extensively base paired in the tri-snRNP and B complex. During activation, the U4/U6 helices are unwound and a highly structured RNA network forms between the pre-mRNA and U2, U6, and U5 snRNAs, generating the spliceosome's catalytic RNA core (Fica et al., 2013; Staley and Guthrie, 1998). The U5 snRNP proteins Prp8 and RNA helicase Brr2 play central roles during catalytic activation. Prp8 is a major scaffolding protein that interacts with Brr2 and Snu114 (Wahl et al., 2009) and also forms a pocket that encompasses the catalytic RNA network of activated spliceosomes (Galej et al., 2014; Hang et al., 2015). Brr2 initiates spliceosome activation by unwinding the U4/U6 snRNA helices (Laggerbauer et al., 1998; Raghunathan and Guthrie, 1998). As Brr2 and its RNA substrate are present in the tri-snRNP and B complex, its activity must be regulated to ensure that U4/U6 unwinding first occurs during activation. In contrast to the situation in the *S. cerevisiae* tri-snRNP, Brr2 is located in human tri-snRNPs at a large distance from the U4/U6 helices, thereby preventing their premature unwinding (Agafonov et al., 2016). Whether Brr2 is repositioned and engages its substrate in the human B complex, and if so, how Brr2 is negatively regulated at this stage of spliceosome assembly, is unclear.

Yeast and humans share a common set of core spliceosomal proteins that are evolutionarily conserved, but human spliceosomes contain many additional proteins not present in *S. cerevisiae* (Fabrizio et al., 2009). For example, human B complexes contain a set of B-specific proteins, which include hSnu23, RED, Smu1, hMFAP1, FBP21, hPrp38, NPW38, and

NPW38BP (Agafonov et al., 2011). These proteins are conserved in higher eukaryotes, but only hSnu23, hPrp38, and hMFAP1 have homologs in *S. cerevisiae* (Agafonov et al., 2011; Ulrich and Wahl, 2017). B-specific proteins first bind during B complex formation and are missing or much less abundant in B^{act} complexes (Agafonov et al., 2011). The function of most of these proteins remains unclear, but they are not required for tri-snRNP recruitment to the B complex (Boesler et al., 2016). Instead, they likely contribute to the activation process, as indicated for Prp38 (Schütze et al., 2016; Xie et al., 1998), via currently unknown mechanisms. Several B-specific proteins are implicated in alternative splicing decisions, but the mechanisms whereby they achieve this is unclear (Papasaikas et al., 2015; Spartz et al., 2004).

3D electron cryo-microscopy (cryo-EM) structures of the *S. cerevisiae* tri-snRNP, B^{act}, C, and C* complexes and *S. pombe* post-splicing ILS spliceosomes (Galej et al., 2016; Nguyen et al., 2016; Rauhut et al., 2016; Wan et al., 2016a, 2016b; Yan et al., 2016, 2017; Fica et al., 2017), and of the human tri-snRNP and C* complex were recently reported (Agafonov et al., 2016; Bertram et al., 2017; Zhang et al., 2017). These studies revealed the spliceosome's molecular architecture during its activation and catalytic activity. However, relatively little information is available about the molecular organization of the spliceosome at early assembly stages before its activation. The structural organization of the human U4/U6.U5 tri-snRNP revealed by cryo-EM indicates that this major spliceosomal subunit is initially maintained in an inactive state. Thus major rearrangements in tri-snRNP components must occur to generate the functional centers required to prime the B complex for the subsequent activation step. To elucidate these structural changes, we determined the 3D cryo-EM structure of the human spliceosomal B complex at a core resolution of 4.5 Å and determined its spatial organization with the aid of protein crosslinking.

RESULTS AND DISCUSSION

Structure Determination and Model Building

By lowering the Mg²⁺ concentration of the in vitro splicing reaction, we could affinity-purify human spliceosomes containing stoichiometric amounts of U2, U4, U5, and U6 snRNAs, but which were essentially devoid of U1 snRNA, indicating that they were stalled prior to Brr2-mediated activation and thus are spliceosomal B complexes (Figure S1). Consistent with this, highly abundant proteins in these complexes included U2 and tri-snRNP proteins (except RBM42, Prp28, and Sad1, which were absent or present in low amounts), U2AF, B-specific proteins, and pre-mRNA binding proteins, such as the CBPs. Except for moderate amounts of Skip (hPrp45) and RBM22, Prp19 complex proteins, or other proteins typically present in activated spliceosomes were absent or present in very low amounts (Figure S1D). Chase experiments with micrococcal nuclease treated extract showed that our purified B complexes are functional (Figure S1G). However, incubation of the latter with ATP, did not induce significant displacement of U4 snRNP (Figure S1F), indicating that Brr2 is still inhibited at this stage.

The 3D structure of the human B complex was determined by cryo-EM (Figure S2), revealing a globular head domain and a

triangular body with a central domain with adjacent "foot," "stump," and "neck" domains (Figure 1; Movie S1). Three major densities (B1-B3) bridge the head to the neck (B1) or stump (B2), or run as an extended density element almost parallel to the central axis of the main body (B3). Most of the triangular domain is well-defined and its structure was determined at an overall resolution of 4.5 Å (Figures 1 and S2). However, the head and connecting bridges and some areas of the triangular body are more dynamic or contain components with substoichiometric occupancies and are thus less well resolved (Figures 1 and S2). By fitting known X-ray structures or homology models of structured regions of B complex components into the EM density map (summarized in Table S1), and aided by chemical protein cross-linking coupled with mass spectrometry (Table S2), we generated a pseudo-atomic model for the more stable parts of the B complex (Figure 1).

U2 snRNP Is Located in the B Complex Head Domain

Several U2-SF3a and SF3b proteins contact the pre-mRNA at or near the BS in A, B and B^{act} complexes, stabilizing the U2/BS helix (Gozani et al., 1996). In the *S. cerevisiae* B^{act} complex the U2/BS helix is sequestered between the terminal HEAT repeats of the C-terminal HEAT domain of Hsh155 (SF3B1/SF3b155 in human) (Rauhut et al., 2016; Yan et al., 2016). Consistent with earlier immuno-EM studies (Wolf et al., 2009), we could fit the core of the SF3b complex together with the U2/BS helix, in a closed conformation as found in the B^{act} complex (Rauhut et al., 2016; Yan et al., 2016), into the B complex head domain (Figure 1). The SF3B3 WD40 domains are oriented toward the stump and the circular SF3B1 HEAT domain is oriented toward the neck of the body, while the U2 Sm core structure is located close to bridge B3 (Figure 1). One end of the U2/BS helix is located at the interface between the head and B3 (Figure 1). Based on protein crosslinks, U2 SF3a proteins are also likely located in this region, while the largely intrinsically unfolded C-terminal region of U2 SF3A1 appears to extend from the head through B3 into the lower part of the main body, where it is crosslinked to Prp8 and U5-40K (Figure S1H).

Organization of U4, U6, and U5 snRNA and the 5'ss Region of the Pre-mRNA

We were able to trace most nucleotides (nts) of U4, U6, and U5 snRNA in the B complex body (Figure 1). The three-way helical junction of the U4/U6 snRNA is located in the upper part close to the neck. The U4 Sm and U6 LSm protein rings are located in the stump or at the B1 bridge between U2 snRNP and the neck, respectively. B1 contains, among others, U2/U6 helix II, in which the 5' end of U2 and 3' end of U6 are base paired (Figures 1B and 3C). The major U5 stem-loop (SL) has a similar conformation in B and the tri-snRNP, with the U5 Sm core located in the foot of both (Figure 1B). However, in B, U5 loop 1 adopts a different conformation and loop 1 nucleotides U41 and U43 are base paired with 3' terminal nucleotides of the 5' exon, i.e., A-3 and G-1 (Figures 1C and S3A). An additional RNA helical element comprising the U6 ACAGA box base paired to several nucleotides near the 5' end of the intron is located ca 3–4 nm from the U5 loop 1/5' exon helix (Figures 1B and 1C), confirming that the U6 ACAGA/5'ss helix is formed in our

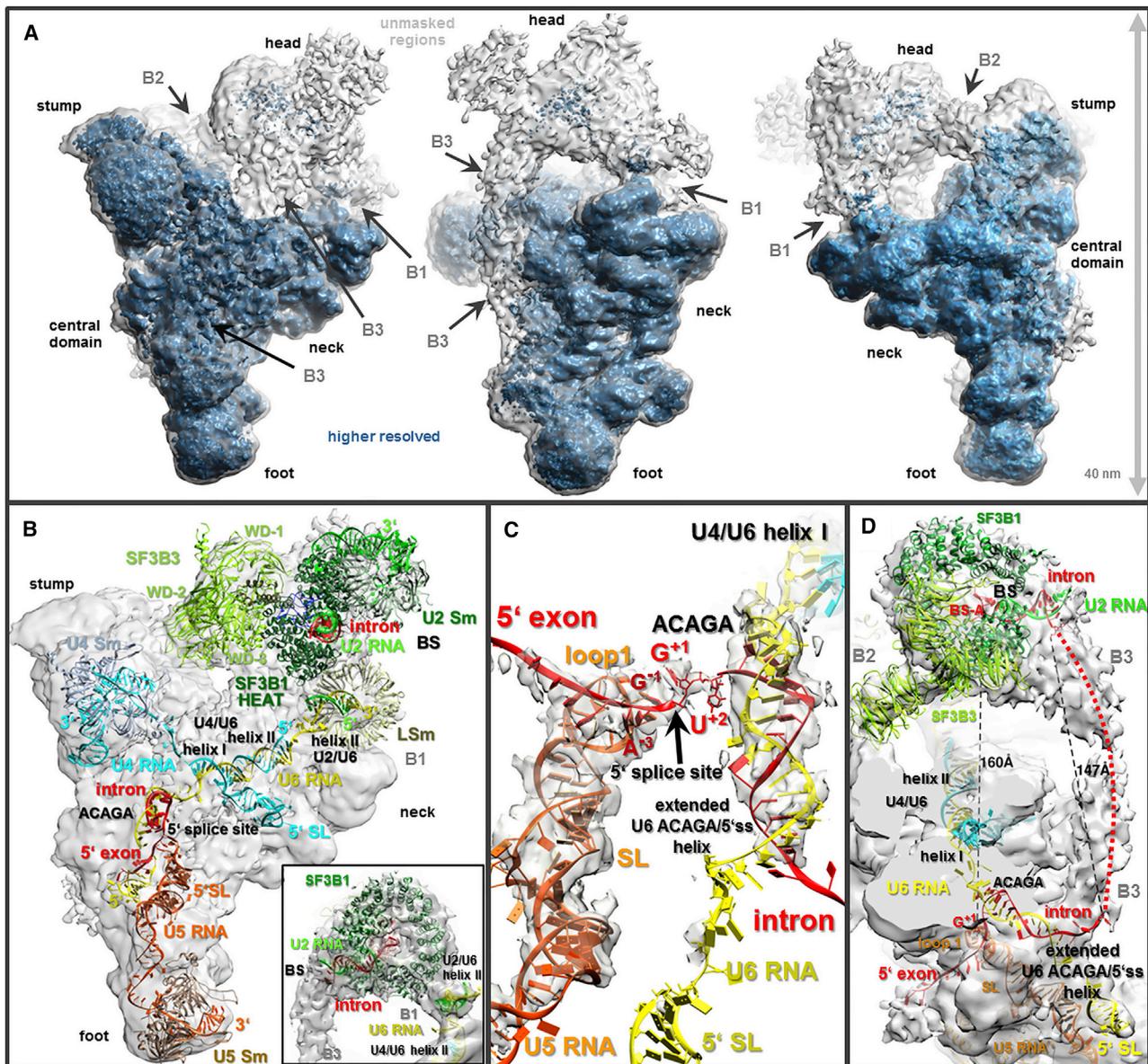


Figure 1. Cryo-EM Reconstruction of the Human B Complex

(A) Different views of the B complex EM density map (rotated around the vertical axis). Blue, better-resolved densities. Gray, masked regions not seen in the 4.5 Å structure. B1, B2, B3, and density bridges connecting the head with the triangular body.

(B) Location in the unmasked density of the U2 SF3b protein core and U2 Sm core in the head, and the U4 and U5 Sm cores, U6 LSm core, and major RNA elements in the main body. Inset, fit of the SF3B1 HEAT domain in complex with the U2/BS duplex.

(C) Fit of the U5 SL, 3' end of the 5' exon, 5'ss nucleotides, and extended U6 ACAGA/5'ss helix into the 4.5 Å EM density map.

(D) Path of B3 (in the unmasked density) from the U2/BS helix to the end of the extended U6 ACAGA/5'ss helix. Stippled red line, possible path of the intron. Distances between the branch adenosine and 5'ss, and between the U2/BS helix and extended U6/5'ss helix, are indicated.

See also Figures S1, S2, and S3, Tables S1 and S2, and Movie S1.

B complexes. The 5'ss GU nucleotides at the intron's 5' end (G+1 and U+2) are positioned between both helical elements in an extended conformation with the two bases pointing away from each other (Figure 1C), which differs significantly from their conformation in the activated spliceosome (Figure S3A).

Interestingly, U6 nucleotides 30 to 40, immediately upstream of the ACAGA box, also form base pairs (including non-canonical

ones) with additional intron nucleotides downstream of the 5'ss (Figure 1C). As an extended U6 ACAGA/5'ss helix is also present in the human C* complex (Bertram et al., 2017; Zhang et al., 2017), but is not observed in yeast spliceosomes (Figure S3B–3D), it may be a structural feature of the spliceosome solely in higher eukaryotes. Such extended helical elements may help to stabilize short RNA helices, such as the human U6 ACAGA

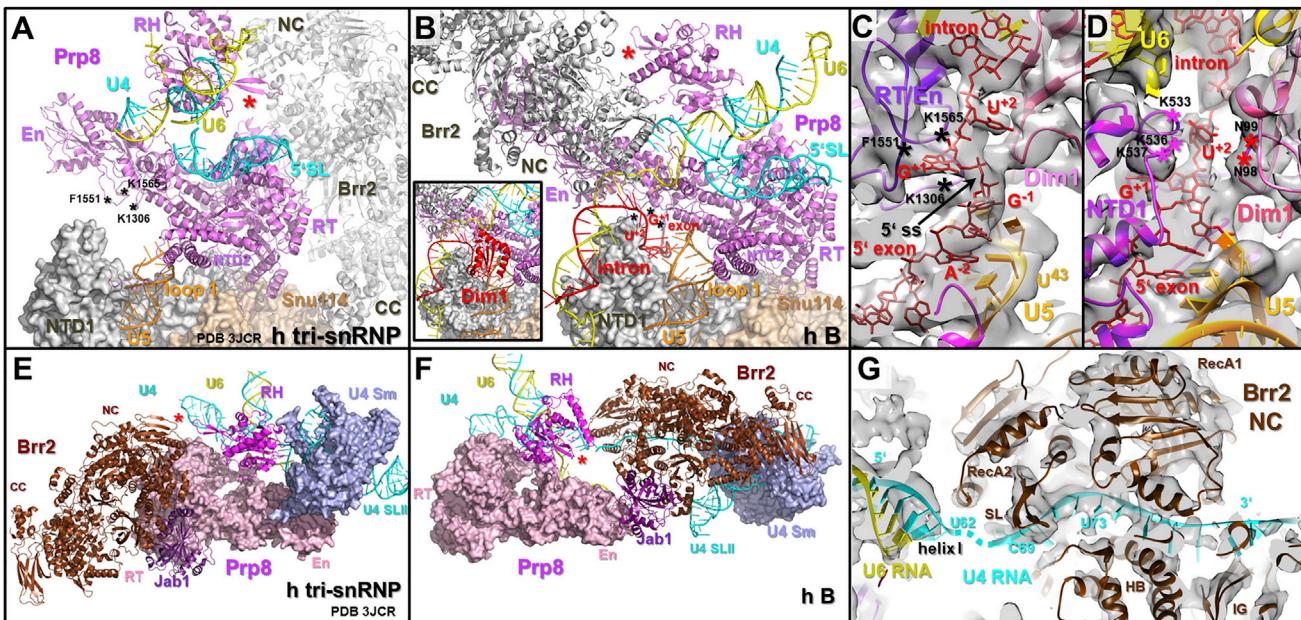


Figure 2. Substantial Prp8 and Brr2 Rearrangements and Formation of a 5'ss Binding Pocket in the B Complex

(A and B) Open and partially closed conformation of Prp8 in the human tri-snRNP (A), and the B complex (B), respectively, aligned relative to the Prp8 NTD1 domain and Snu114. Black asterisks, amino acids in Prp8 linker loops close to G+1 of the 5'ss in the B complex. For clarity, the Dim1 protein is only shown in the inset of (B). Red asterisk, the β hairpin of the Prp8 RH domain. For a general overview of the position of U5 proteins in the B complex EM density map, see Figure S4A.

(C) Fit of the 5'ss nucleotides into the 4.5 Å EM density map of the B complex. G+1 is bound in a pocket comprised of amino acids of Prp8 linker loops (indicated by black asterisks). Base pairing of 5' exon nucleotides with U5 loop 1 is also shown.

(D) U+2 of the 5'ss is located in a protein pocket comprised of loops of both Prp8 NTD1 and of Dim1. Amino acids of Prp8 NTD1 and Dim1 close to U+2 in the B complex are indicated.

(E and F) Large-scale rotational movement of Brr2's helicase domain from the Prp8 RT domain in the tri-snRNP (E) to the Prp8 En domain in the B complex (F), aligned relative to the RT/En domain.

(G) Docking of Brr2's NC onto the central single-stranded region of U4 snRNA close to U4/U6 helix I. SL, separator loop. U4 nucleotides between U62 and C69 are flexible and are thus indicated by a stippled line.

See also Figures S3 and S4 and Tables S1 and S2.

box/5'ss helix. The extended bridge B3 (spanning ca 15 nm) connects the region where the U2/BS helix is located with the end of the extended U6 ACAGA/5'ss helix, suggesting B3 contains, in addition to protein, the middle part of the intron that connects these functionally important RNA helices (Figure 1D). The 5'ss and BS, which later must be juxtaposed for catalytic step 1 to occur, are physically separated by ca 16 nm in the B complex (Figure 1D). The 5'SL of U6 is located in a position similar to that observed in the yeast B^{act} or human C^{*} complexes, close to the U5 Sm core (Figures 1B and S3). As the U6 5'SL and ACAGA box are likely located in the upper part of the isolated human tri-snRNP, they must be substantially repositioned upon tri-snRNP integration into the spliceosome.

Prp8 Adopts a Partially Closed Conformation and, Together with Dim1, Forms a 5'ss Binding Pocket

The B complex central domain contains the U5 Prp8, Snu114, U5-40K and Dim1 proteins (Figures 1A and S4A). Whereas the positions and structures of the latter three, and of the Prp8 N-terminal domain 1 (NTD1), are very similar in the human tri-snRNP and B complex, the position of the Prp8 RT/En domain is clearly different. In the tri-snRNP, Prp8 has an open conformation,

whereby the RT/En domain points upward and the tip of the En domain is separated from NTD1 by ca 5 nm (Figure 2A). In the B complex, the RT/En domain has rotated by ca 20° around its long axis compared to its position in the tri-snRNP, and moved downward, such that several loops emanating from the RT/En linker are juxtaposed with Prp8 NTD1 and Dim1 (Figures 2B and S4). While Prp8 adopts a partially closed conformation in the B complex, Prp8's active site pocket (formed by the RT/En, NTD1, and NTD2 domains), into which the catalytic U2/U6 RNA network docks during activation, is not completely closed, as found in catalytically activated spliceosomes (Figure S4B).

The partial rearrangement of Prp8 after tri-snRNP integration into the spliceosome, generates a protein pocket, comprised of residues of the Prp8 RT/En linker region, Prp8 NTD1 and Dim1, that binds the 5'ss GU dinucleotide in an extended conformation. That is, G+1 of the intron is sandwiched between loop regions (containing amino acids [aa] K1306 and F1551 to K1565, respectively) emanating from the Prp8 RT/En linker, while U+2 is contacted by a loop (aa 93–101) of Dim1, and by a helical region of Prp8 NTD1 (aa 532–537) (Figures 2B to 2D). Our structure thus suggests that the evolutionarily conserved Dim1 protein plays a direct, previously unknown role in 5'ss recognition in

the B complex. Recognition of the extended 5'ss by an intricate network of RNA-RNA and RNA-protein interactions, as shown here, provides the structural basis for sequestering the 5'ss at this pre-catalytic stage of splicing.

The Prp8 RH domain is located above the RT/En linker, in both the B complex and human tri-snRNP, indicating that it has undergone a similar downward shift as the RT/En domain. However, the RH domain has rotated by ca 180° about its central axis in B and thus its spatial orientation differs dramatically between the two complexes (Figures 2A and 2B). The 5'ss GU can be crosslinked to Prp8's RH domain at an early stage of spliceosome assembly (Reyes et al., 1999). As the RH domain is separated from the 5'ss by ca 6 nm in our B complex (Figure 2B), the proposed RH-5'ss interaction appears to take place at an earlier assembly step, presumably before disruption of the U1/5'ss interaction.

Brr2 Is Dramatically Rearranged and Binds to Its RNA Substrate in the B Complex

In human tri-snRNPs, Brr2 and the Prp8 Jab1 domain that tightly binds to it, are located close to the RT end of the Prp8 RT/En domain, and the active N-terminal helicase cassette (NC) of Brr2 is located ca 10 nm away from the U4/U6 duplex and the U4 Sm core structure (Figures 2E and S4A). In contrast, in the B complex, Brr2 is positioned near Prp8's En domain, ca 20 nm away from its position in the tri-snRNP, and the Prp8 Jab1 domain now contacts the tip of Prp8's En domain (Figures 2F and S4A). This large-scale movement of Brr2 would require its rotation by ca 180° around the long axis of the tri-snRNP part of the B complex. Moreover, in the latter, the U4 Sm core is now located at the interface between Brr2's helicase cassettes, such that Brr2's NC is positioned between the U4 Sm core and U4/U6 helix I, which are connected via the central single-stranded region of U4 snRNA (Figures 2F, 2G, and S1). This region of U4 is required for U4/U6 duplex unwinding by Brr2 in vitro, indicating that it functions as a docking site for Brr2's NC (Mozaffari-Jovin et al., 2012). Consistent with this, the central single-stranded region of U4 runs across the two RecA domains of the NC, and RecA2 binds U4 nucleotides 69 to 73 downstream of U4/U6 helix I and is thus positioned close to helix I (Figure 2G). Thus in our B complex, Brr2 is bound to its RNA substrate and is poised to unwind the U4/U6 duplex and initiate the spliceosome activation process, but its helicase activity is still negatively regulated (Figure S1F). Interestingly, Brr2 does not appear to contact the double-stranded region of U4/U6 helix I (Figure 2G), which would be required for its ultimate unwinding. As described below, this contact is potentially prevented by the B-specific protein FBP21.

Remodeling of the U4/U6 Di-snRNP and Prp6 during B Complex Formation

The major domains of the U4/U6 proteins interacting with the U4/U6 three-way junction are also located close to the B complex neck (Figure 3). Snu13, which directly interacts with the U4 k-turn motif, is located between the U4 5'SL and U4/U6 stem II, and also interacts with the C-terminal WD40 domain of Prp4 (Figure 3B). The Prp31 Nop domain interacts with Snu13 and the U4 k-turn motif in a manner similar to that observed in

the co-crystal structure of Snu13, Prp31 and the U4 5'SL (Liu et al., 2007). The Prp31 coiled-coil domain, however, has rotated ca 50° and is now closer to the Prp31 Nop domain in the B complex, and thus adopts a more compact structure (Figure 3B, see also below). The C-terminal part of Prp31 has an extended conformation and runs between the phosphodiester backbone of the U4 5'SL and Dim1 (Figure 3B).

Prp3 interacts closely with various parts of the U4/U6 duplex. Its C-terminal ferredoxin-like domain (FLD) binds to the 3' terminal single-stranded region of U6 (Liu et al., 2015) and likely stabilizes the neighboring U2/U6 helix II in bridge B1 (Figure 3C). Three α helices (H1-H3), located N-terminal to Prp3's FLD, interact directly with U4/U6 helices I (H1 and H2) and II (H3) (Figure 3D). As both U4/U6 helices are subsequently unwound by Brr2, Prp3 may contribute to Brr2 regulation at this stage. Based on numerous protein-protein crosslinks, the N-terminal part of Prp3 runs across the Prp4 WD40 domain, and then passes along bridge B1, where it crosslinks with several proteins of the U6 LSm core and finally contacts the U2 SF3A1 and SF3B1 proteins in the head (Figure S5A). Thus, Prp3 may also play an important role in stabilizing this major connection between U2 and the tri-snRNP during early spliceosome assembly. Prp4 bridges several U4/U6 proteins with each other. For example, its C-terminal WD40 domain interacts with Snu13, Prp3's FLD and Prp6, while its N-terminal-most helical bundle (NHB) bridges cyclophilin H (CypH) to α helix H1 of Prp3 (Figures 3B and 3D). Moreover, the N-terminal region of Prp4 additionally interacts with U2 SF3A proteins and Brr2's helicase domain, as indicated by multiple crosslinks (Figure S5B).

A comparison of the cryo-EM structures of the isolated tri-snRNP and B complex reveals that in addition to Prp8 and Brr2, several U4/U6 components and Prp6 also undergo significant rearrangements upon/after tri-snRNP integration into the spliceosome. For example, in the B complex, the U4 Sm and U6 LSm core structures have undergone large-scale translocations (Figures 2E, 2F, 4A, and 4B). Moreover, the entire complex of U4/U6 proteins, which includes Snu13, Prp4's WD40, Prp31's Nop, and Prp3's FLD domains, and the C-terminal TPR repeat domain of Prp6, has rotated counter-clockwise by ca 20–30° with respect to the Prp8 RT domain. Thus they, together with the U4/U6 three-way junction, are shifted closer to the Prp8 RT/En domain in the B complex (Figure 4B). The Prp6 C-terminal TPRs interact with Prp4, Snu13 and Prp3 in a very similar manner in both complexes, but they additionally contact Prp31's rearranged coiled-coil domain only in B (Figure 3E, and see below). After tri-snRNP integration into the spliceosome, a major rearrangement also occurs within the N-terminal-most TPR repeats of Prp6, which are more extended in the B complex and whose interaction with Prp8's RT/En and RH domains has changed substantially (Figures 4A and 4B). The rearranged Prp6-Prp8 interaction appears to be stabilized in the B complex by α -helical regions of Snu66, which form a bridge stretching from Prp6's N-terminal TPRs to the Prp8 RT/En linker region, close to the switch loop (Figure 4D). The remaining part of Snu66, which is thought to be largely intrinsically unfolded, appears to be located close to Brr2's NC and CC, Prp8's linker, En, RH, and Jab1 domains, and Smu1, as indicated by protein-protein crosslinks (Tables S1 and S2).

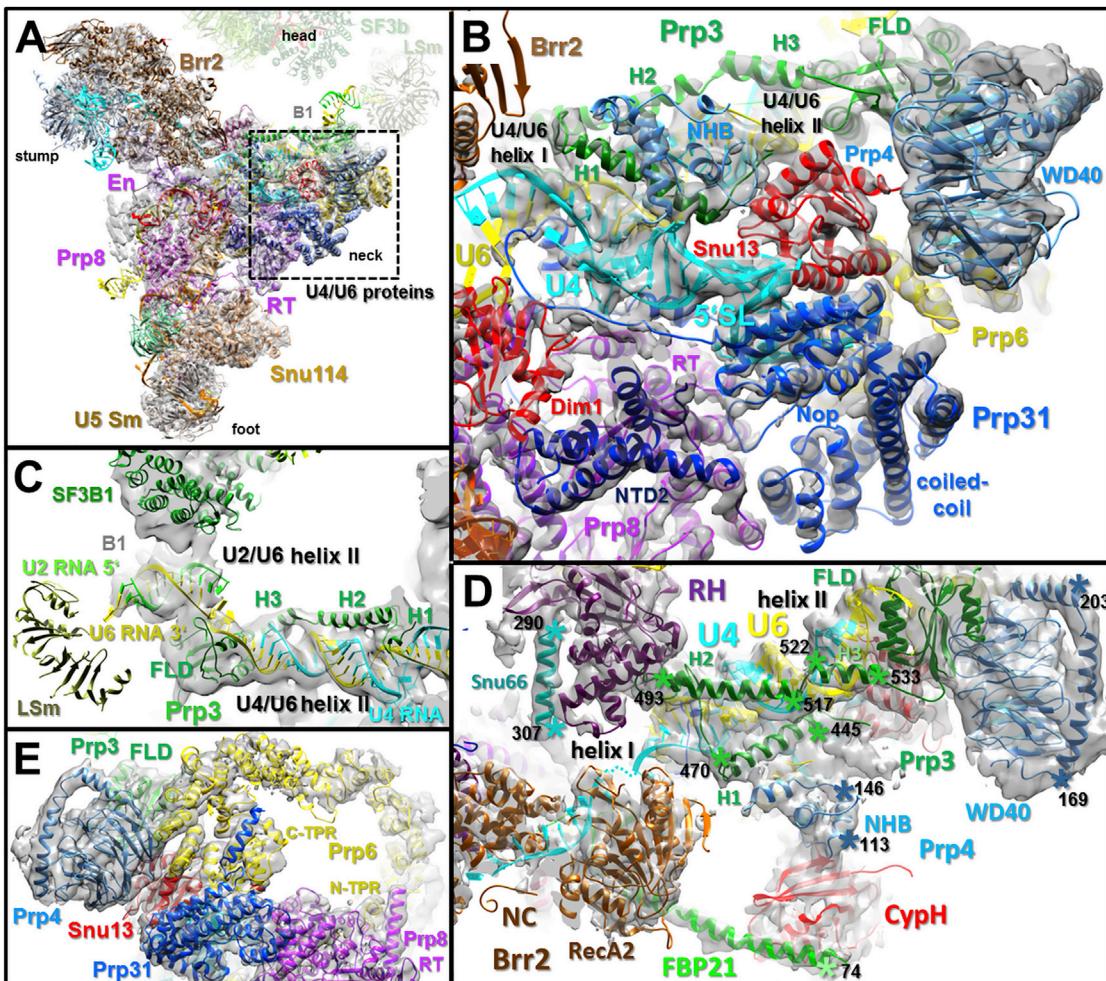


Figure 3. Location of U4/U6 Proteins and Their Interaction with the U4/U6 Three-Way Junction

(A) Location of the U4/U6 proteins in the B complex neck region.

(B) Fit of the indicated U4/U6 protein domains/regions and U4/U6 three-way junction into the 4.5 Å EM density map of the B complex.

(C) Location of U2/U6 helix II, Prp3's FLD, and the U6 LSm domain in or close to the B1 bridge at the interface between the head and neck of complex B.

(D) A helical domain (NHB) of Prp4, located N-terminal of its WD40 domain, links CypH via Prp3 helix H1 to the U4/U6 three-way junction. This network is connected via FBP21 to Brr2's NC and the U6 ACAGA/5'ss helix (see also Figure 5). Selected amino acids (black numbers) are indicated by asterisks.

(E) Fit of Prp6's C-terminal TPR repeats into the 4.5 Å EM density of the B complex and their interactions with the major domains of U4/U6 proteins.

See also Figure S5 and Tables S1 and S2.

Location and Potential Functions of the B-Specific Proteins

A number of B-specific proteins, whose domain organization is shown in Figure 5A, could be located in the B complex EM density map (Figure 5B), providing first insights into their potential functions. The B-specific protein Smu1 possesses an N-terminal, ca 180-aa-long, structured helical region (NTR), that contains LisH and CTLH domains, which is connected via a short linker region to a C-terminal WD40 domain (Ulrich et al., 2016a). In the B complex, Smu1's C-terminal WD40 domain is located at the interface between Brr2's CC and the WD40 domain 2 of SF3B3. Moreover, based on crosslinks, we could place Smu1's NTR in a density element that bridges the WD40 domains of Smu1 and SF3B3 (Figure 5C). Thus, Smu1 and SF3B3's WD40 domains are the major constituents of the B2

bridge, and Smu1 may help to stabilize the position of Brr2 after its rearrangement. Smu1's WD40 domain is positioned at the interface of Brr2's NC and CC (Figure 5D), and thus might influence the relative orientation of both helicase cassettes and thereby potentially modulate the helicase activity of the NC (Santos et al., 2012). The strategically important position of Smu1 in the B complex could explain why its mutation or deletion has profound effects on pre-mRNA splicing (Papasaikas et al., 2015; Spartz et al., 2004).

Smu1 forms a heterodimer with the B-specific RED protein, stably binding a short region of the latter (i.e., aa 209–222) via its NTR (Ulrich et al., 2016a). RED is largely intrinsically unfolded and we could not locate its precise position in the B complex EM density map. However, in our purified B complex, crosslinks were detected between the Smu1 NTR and the short region of

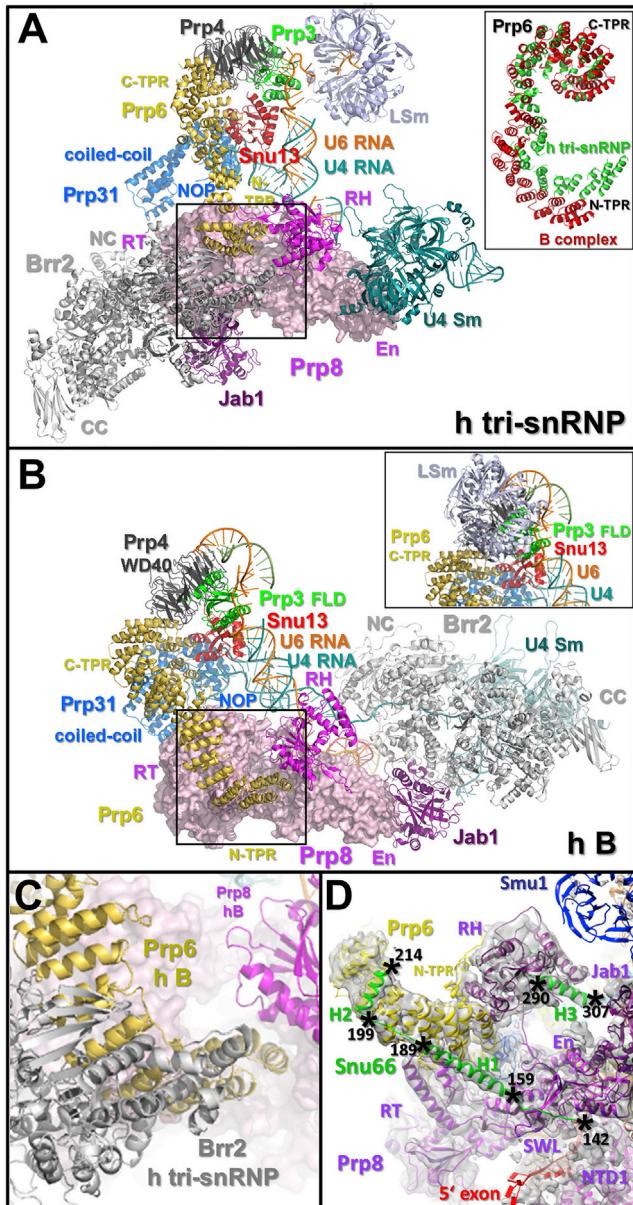


Figure 4. U4/U6 snRNP and Prp6 Remodeling during B Complex Formation

(A and B) Organization of the indicated U5 and U4/U6 protein domains, as well as C-terminal TPR repeats of Prp6 in the human tri-snRNP (A) and the human B complex (B). Both complexes are aligned relative to the Prp8 RT/En domain. As they obscure the U4/U6 proteins, the LSm proteins were omitted from (B), but instead are shown in the inset of (B). Inset in (A), Superposition of Prp6's TPR domain in the tri-snRNP (green) and B complex (red), aligned relative to the Prp6 C-terminal TPR repeats.

(C) Superposition of the Brr2 helicase domain from the isolated human tri-snRNP onto the B complex structure, aligned relative to the RT/En domain, indicating a clash with the rearranged Prp6 TPR domain in the B complex in the absence of Brr2 rearrangement. Superimposed regions of the tri-snRNP and B complex are indicated by boxes in (A) and (B).

(D) N-terminal α -helices of Snu66 connect the Prp6 N-terminal TPR repeats with the Prp8 RT/En, RH, and NTD1 domains. SWL, switch loop within the Prp8 RT/En domain. Asterisks, selected amino acid positions of Snu66.

See also Tables S1 and S2.

RED that binds to it, and between the latter and Prp8's RH domain. Moreover, we observed numerous crosslinks between the N-terminal part of RED and several U2 proteins in the head, while its C-terminal region was crosslinked to the Prp8 RT/En domain and NTD1 (Figure S6A). Thus, like Smu1, RED appears to play a role in bridging U2 with U5 proteins in the B complex.

Several B-specific proteins contact the U6/5'ss helix and may aid in its repositioning within the spliceosome. The N-terminal domain (NTD) of Prp38 interacts simultaneously with several B complex proteins, and was crystallized in complex with short helices of MFAP1 and Snu23 (Ulrich et al., 2016b). The Prp38 NTD is located close to the U6/5'ss helix, and appears to interact with the U6 ACAGA sequence (Figures 5B and 5E). The α helices of Snu23 and MFAP1 that bind Prp38 fit into neighboring EM densities, and the N-terminal Snu23 zinc finger (ZnF) also interacts with the upper part of the extended U6 ACAGA/5'ss helix (Figures 5E and 5F). Prp38 is required to convert B into a B^{act} complex (Schütze et al., 2016; Xie et al., 1998), but little is known about the molecular mechanism whereby it contributes to the activation process. A comparison of the position of the U6 ACAGA/5'ss helix in B and B^{act} complexes indicates that it must be repositioned during activation to ultimately juxtapose the 5'ss and U2/U6 catalytic center (Figure S3). As Prp38 and Snu23 contact the U6 ACAGA/5'ss helix, they may facilitate the repositioning of the latter during activation. In addition, or alternatively, they may help recruit proteins such as Cwc24 to the 5'ss during activation (Yan et al., 2016). MFAP1 appears to play a role in connecting the head and main body of the B complex, as indicated by protein crosslinks (Figure S6B).

Finally, in addition to Smu1, the B-specific protein FBP21 also contacts Brr2 and thus potentially aids in maintaining Brr2 in an inactive state in the B complex. The N-terminal region of FBP21 contains tandem WW domains (Huang et al., 2009), and is predicted to contain at its very N terminus a ZnF that appears to be followed by a long α -helical element (Table S1). In the B complex, FBP21's N-terminal region indeed adopts a matrin-like ZnF structure that binds across the minor groove of the U6 ACAGA/5'ss helix (Figure 5F). The adjacent α helix fits into a density element that runs from the U6/5'ss helix toward the RecA2 domain of Brr2's NC and the Cyph protein (Figures 3D and 5E). Cyph interacts with Prp4, which in turn communicates via Prp3 with U4/U6 helix I (Figure 3D). The position of FBP21's N-terminal α helix between Brr2's NC (bound just upstream of U4/U6 helix I) and Cyph, which indirectly binds U4/U6 helix I via Prp4 and then Prp3, suggests that it may act to maintain U4/U6 helix I and its associated proteins at a sufficient distance from Brr2's NC (i.e., hold it at bay), thereby preventing Brr2 contact with the double-stranded region of its substrate and potentially maintaining it in an inactive state. Our crosslinking results also suggest that the C-terminal region of FBP21 contacts Brr2's helicase domains and Prp4 (Figure S6C). Thus, a complex interaction network involving FBP21, Brr2, and the U4/U6 snRNP proteins potentially enables the tight regulation of Brr2 helicase activity at this stage. Release of FBP21 during the B to B^{act} transition could play a key role in triggering Brr2 activity during spliceosome activation. Consistent with this idea, FBP21 is the only B-specific protein lost from spliceosomes that are blocked at an

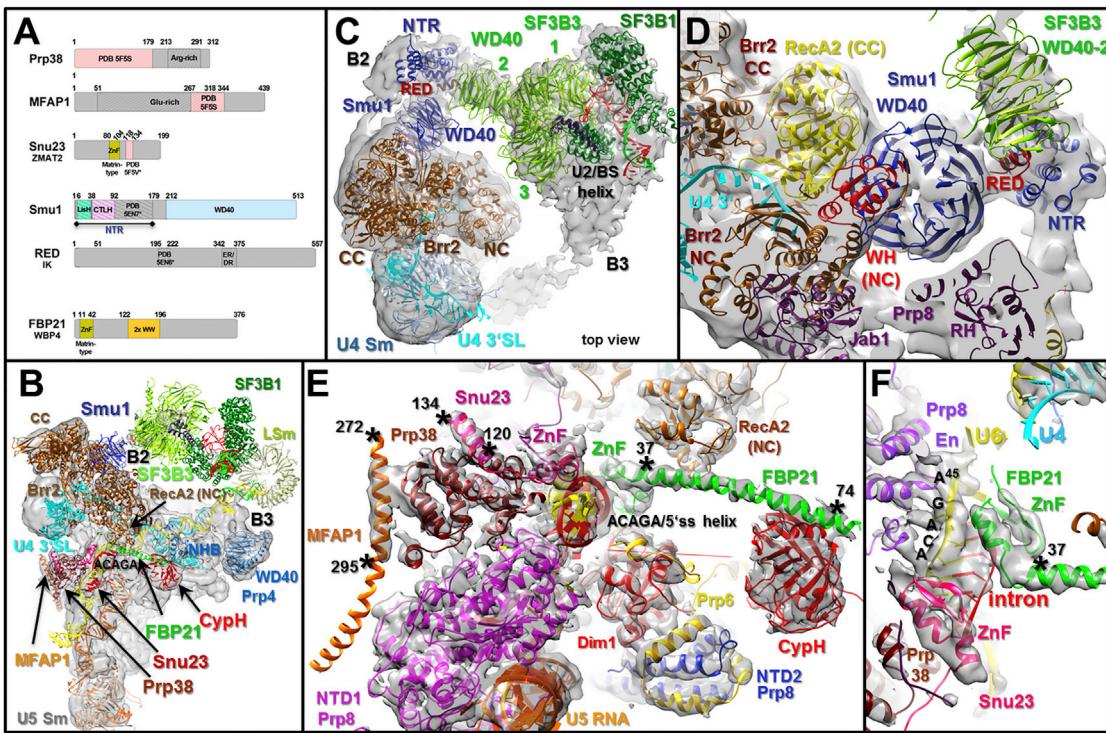


Figure 5. Organization of B-Specific Proteins in the Human B Complex

(A) Domain organization of the indicated human B-specific proteins. Domain boundaries are indicated. Protein regions for which high resolution structures are available are indicated by PDB numbers. Structures derived from non-human species are marked with an asterisk.

(B) Overview of the location of major parts of various B-specific proteins.

(C) Location of the N-terminal region (NTR) and the C-terminal WD40 domain of Smu1 (in the unmasked density) at the interface between U2 SF3B3 and Brr2. The short α helix of RED (aa 209–222) that binds Smu1's NTR is shown.

(D) Smu1's WD40 domain is located at the interface between Brr2's NC and CC domains and contacts the WD40-2 domain of U2 SF3B3. WH, wing helix.

(E) Close up of the fit of the N-terminal helical domain of Prp38 and associated α helices of Snu23 and MFAP1, and the N-terminal α helix of FBP21, near the extended U6 ACAGA/5' ss helix. Black asterisks: amino acid positions of the indicated B-specific proteins.

(F) Interaction of the FBP21 N-terminal zinc finger (ZnF) with the U6 ACAGA/5' ss helix and of Snu23's ZnF with the extended region of the U6 ACAGA/5' ss helix.

See also Figure S6 and Tables S1 and S2.

early stage of the B to B^{act} transition, but in which the U4/U6 duplex has been unwound (Sidarovich et al., 2017).

Tri-snRNP Rearrangements during Spliceosome Assembly Likely Involve Coordinated Remodeling Events

Comparison of the 3D structures of the B complex and human U4/U6.U5 tri-snRNP reveals that most tri-snRNP proteins undergo extensive repositioning and/or structural rearrangements during B complex formation (Figure 6G). These rearrangements likely occur in a highly coordinated and ordered manner, as the interaction sites of several components in the tri-snRNP and B complex are mutually exclusive. An intriguing question is how these coordinated remodeling events are triggered. Initial docking of the tri-snRNP to the A complex first generates a 37S pre-B complex. In the latter, U1 snRNA is still base paired to the 5' ss, and U2/U6 helix II has formed, but the tri-snRNP is not yet stably bound (Boesler et al., 2016) (Figures 6G and S1). Displacement of U1 from the 5' ss and establishment of the U6 ACAGA/5' ss interaction, which is mediated by the helicase Prp28, leads to a substantial structural change that converts

the 37S pre-B complex into a 50S B complex with stably associated tri-snRNP (Boesler et al., 2015, 2016). Thus many of the rearrangements in tri-snRNP proteins described here likely occur during the pre-B to B transition, and formation of the U6 ACAGA/5' ss helix likely plays a key role in tri-snRNP remodeling. This suggests that Prp28 action may trigger or contribute to tri-snRNP rearrangements, and would likely create new conformational space and enable sampling of different RNP conformations within the spliceosome. The handover of U1 for U6 at the 5' ss likely influences the movement of the RT/En domain toward the Prp8 NTD1. In human tri-snRNPs, Prp28 is located between Prp8's RT/En and NTD1 (Figure 6A). The position of Prp28 in the tri-snRNP, and both the partially closed conformation of Prp8 and the formation/position of the U6 ACAGA/5' ss helix in the B complex, are mutually exclusive (Figures 6A and 6B). Prp28-mediated rearrangements must thus be coordinated with Prp28 displacement from the spliceosome. Interestingly, displacement of Prp28 is likewise a prerequisite for subsequent binding of the Prp38/Snu23/MFAP1 protein complex, whose binding site close to the U6/5' ss helix is also mutually exclusive with that of Prp28 (Figures 5B, 6A, and 6B).

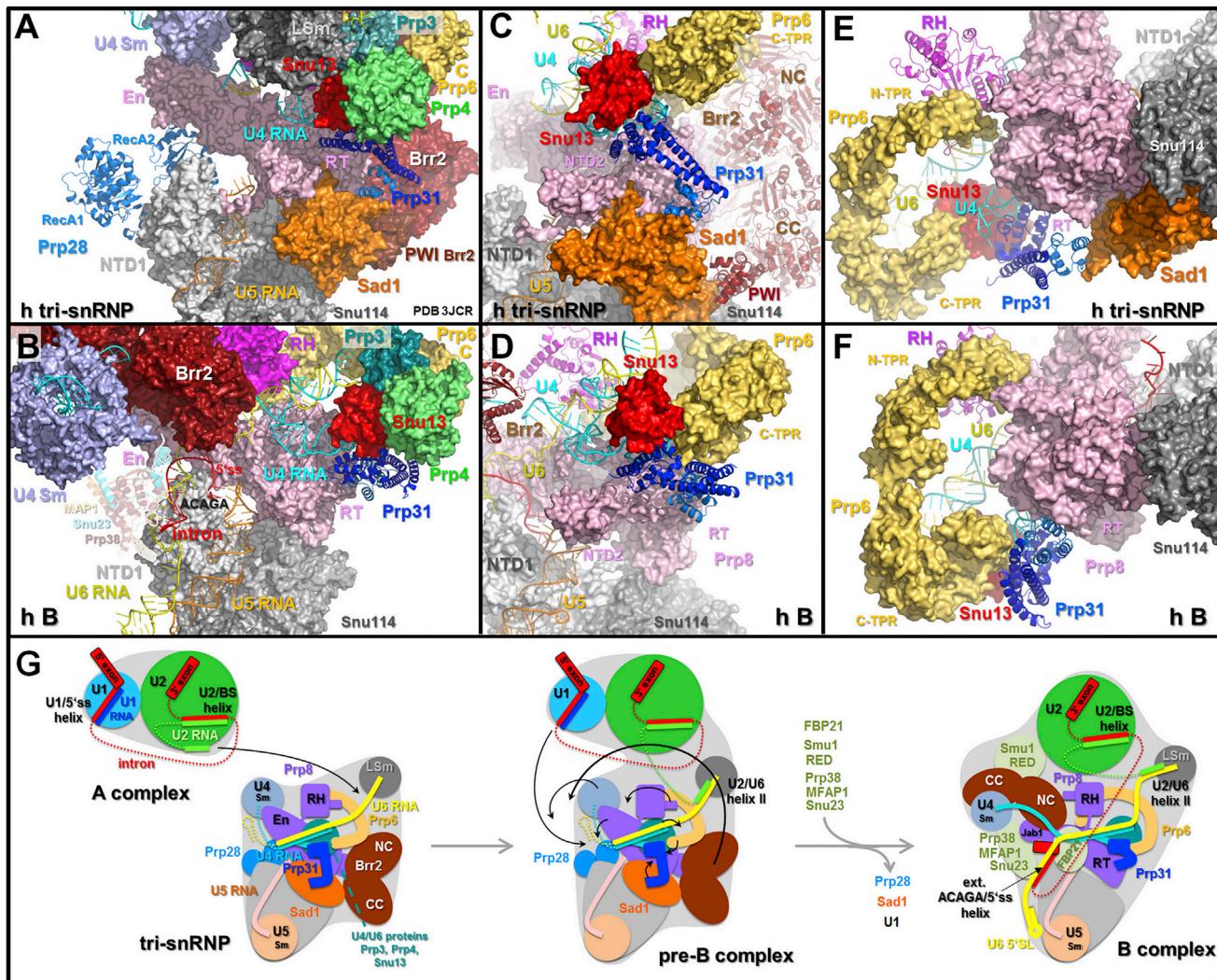


Figure 6. Coordinated Remodeling of Tri-snRNP Proteins and Displacement of Sad1 during B Complex Formation

(A and B) Overview of the location of various tri-snRNP proteins and snRNAs in the isolated human tri-snRNP (A) and B complex (B), aligned relative to Prp8 NTD1 and Snu114. Whereas Sad1 bridges Prp8, Brr2 (via Brr2's PWI domain), and Prp31 in the tri-snRNP, it is missing in the B complex.

(C–F) Side (C and D) and top views (E and F) of the tri-snRNP and B complex, showing the remodeling of Prp31's coiled-coil domain upon/after Sad1 displacement (C and E) and creation of new binding sites for Prp6's C-terminal TPRs with the rearranged Prp31 domain in the B complex (D and F). In the tri-snRNP (C and E), Prp31's coiled-coil tip α -helices (light blue) bind to Sad1, whereas in the B complex (D and F) they interact with Prp8's RT domain in a mutually exclusive manner.

(G) Schematic overview of tri-snRNP rearrangements upon its integration into the spliceosome. Selected proteins and RNAs are shown schematically in the human tri-snRNP (left panel), after initial docking of the tri-snRNP in the pre-B complex (middle panel) and after stable tri-snRNP integration in the pre-catalytic B complex (right panel). Major protein domain movements occurring prior/to/during B complex formation are indicated by black arrows in the middle panel. See also Tables S1 and S2.

The Sad1 protein likely also plays a key role in triggering tri-snRNP rearrangements. In the tri-snRNP (Figure 6A), Sad1 contacts the Prp31 coiled-coil domain and the Prp8 NTD2 and RT domains, as well as Snu114 and the Brr2 N-terminal PWI domain, which in turn is associated with Brr2's CC cassette (Agafonov et al., 2016). This suggests that Sad1 not only plays a role in stabilizing the interaction of the U4/U6 and U5 snRNPs, but that it may also help to tether Brr2 in a pre-activation position, i.e., away from the U4/U6 duplex. Strikingly, in the B complex cryo-EM structure we cannot localize any density accommoda-

ting Sad1, consistent with its underrepresentation in our B complex preparation. Thus Sad1 is clearly displaced from its original position in the tri-snRNP, which would allow most of the rearrangements in tri-snRNP proteins observed upon B complex formation. For example, in the B complex, the Prp31 coiled-coil domain is rearranged and is now juxtaposed with Prp31's Nop domain, whereas the α helices at the coiled-coil tip now directly interact with the Prp8 RT domain (Figures 6C and 6D). This creates new binding sites for Prp6's C-terminal TPRs on Prp31, which would then facilitate the observed shift of the U4/U6

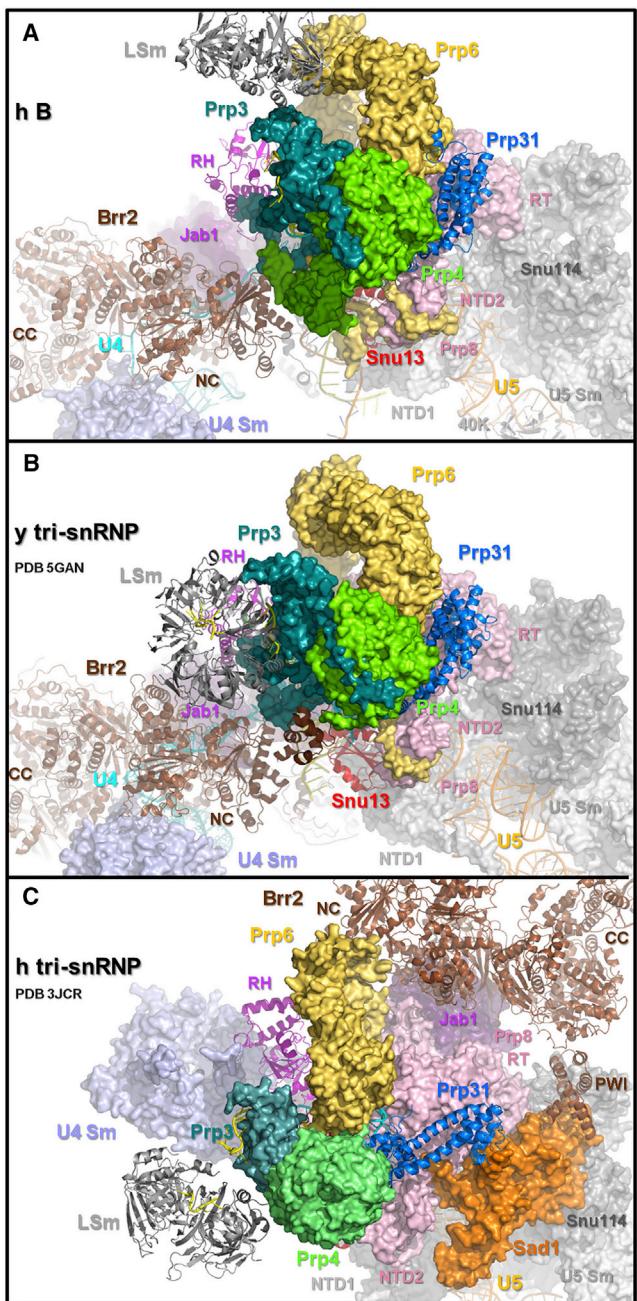


Figure 7. U4/U6/U5 Tri-snRNP Proteins Are Arranged in the *S. cerevisiae* Tri-snRNP and Human B Complex in a Similar Manner
 (A–C) Positions of various tri-snRNP proteins in the purified human B complex (A), *S. cerevisiae* tri-snRNP (Nguyen et al., 2016) (B), and human tri-snRNP (C), as revealed by cryo-EM. All complexes are aligned relative to Snu114 and Prp8 NTD1, whose structures are very similar in all of the cryo-EM models. Similar results were obtained with the *S. cerevisiae* tri-snRNP cryo-EM structure from Wan et al. (2016b). The organization of the tri-snRNP proteins is highly similar in the yeast tri-snRNP and recently published yeast B complex (Plaschka et al., 2017).

See also Figure S7 and Tables S1 and S2.

protein complex with respect to the Prp8 RT domain (Figures 6E and 6F). Moreover, Sad1 displacement and the major rearrangements occurring within Prp6's N-terminal most TPRs (Figure 4) also disrupt interactions of these proteins with Brr2, which would allow the large-scale movement of Brr2 to its RNA substrate. Indeed, Prp6's position in the B complex and the position of Brr2's helicase domain, as found in the tri-snRNP, would be mutually exclusive in complex B (Figure 4C). As Sad1 is highly abundant in the human pre-B complex (Boesler et al., 2016), its displacement likely occurs during the pre-B to B transition, which is consistent with the idea that most tri-snRNP rearrangements occur after its initial docking with the A complex. These tri-snRNP structural changes would thus appear to be only possible within the context of the spliceosome where the interaction of U6 with the 5'ss is first possible. Prp4 kinase, which is required for human B complex formation, phosphorylates Prp6 and Prp31 (and possibly additional proteins) specifically during B complex formation (Boesler et al., 2016; Schneider et al., 2010). Thus, phosphorylation of Prp31 and Prp6 could also potentially play an important role in tri-snRNP remodeling, by triggering conformational changes (e.g., in the Prp31 coiled-coil domain) and/or by stabilizing newly generated protein-protein and protein-RNA interactions.

Differences in the Activation Pathways of Human and *S. cerevisiae* Spliceosomes

A comparison of the cryo-EM structures of human and yeast U4/U6/U5 tri-snRNPs surprisingly revealed extensive differences, foremost in the position of Brr2 relative to its substrate, which already interacts with the U4/U6 duplex in the yeast tri-snRNP (Agafonov et al., 2016; Nguyen et al., 2015, 2016; Wan et al., 2016b). While our B complex structure shows that in humans the large-scale rearrangements that lead to loading of Brr2 onto its RNA substrate occur first during B complex formation, cryo-EM structures of purified *S. cerevisiae* tri-snRNPs suggest that these rearrangements already occur in isolated yeast tri-snRNPs. This suggests that either there is a fundamental difference in the structural organization of the yeast and human tri-snRNP, which is unlikely given the high conservation of their protein components, or that the human and yeast tri-snRNP structures represent different conformational states. Intriguingly, the structural organization of Prp8, Brr2, Prp6, and the U4/U6 proteins is highly similar in the yeast tri-snRNP and human B complex (Figures 7 and S7). This suggests that yeast tri-snRNPs analyzed by cryo-EM potentially could be derived from endogenous B complexes and thus represent a spliceosome dissociation product. Consistent with this, one of the yeast tri-snRNP preparations analyzed by cryo-EM was reported to contain not only pre-mRNA but also U2 proteins (Wan et al., 2016b). Furthermore, purified yeast tri-snRNPs appear to be activated pre-maturely, as they dissociate in the presence of ATP in a Brr2-dependent manner into U5, U4, and U6 snRNPs (Nguyen et al., 2015, 2016; Wan et al., 2016b), while purified human tri-snRNPs and B complexes remain stable (Agafonov et al., 2016) (Figure S1F).

Alternatively, in yeast the tri-snRNP may readily be rearranged prior to its docking with the spliceosomal A complex, and thus not require additional protein or RNA contacts to trigger a

structural rearrangement. Consistent with this, Sad1 is no longer present in the purified yeast tri-snRNP, which could potentially shift the equilibrium between different conformational states toward the B complex structural organization (Huang et al., 2014). This would imply that the tri-snRNP rearrangement that leads to the loading of Brr2 onto its U4/U6 RNA substrate is more highly regulated in higher eukaryotes. Consistent with this idea, Prp4 kinase, which is essential for B complex formation in higher eukaryotes, is absent from *S. cerevisiae*.

Our results indicate that there are significant differences between the activation pathways of human and yeast spliceosomes. The presence of a number of additional proteins in the human B complex, which have no counterparts in *S. cerevisiae* and appear to function first during activation, is a first indication that the transformation of B into an activated B^{act} complex is likely to be more complex in higher eukaryotes. Our cryo-EM structure of the human B complex further reveals that at least two of these proteins, namely, FBP21 and Smu1, contact Brr2 and potentially play important roles in regulating its helicase activity. Thus, the initial steps toward the formation of an activated spliceosome, namely, triggering U4/U6 duplex unwinding, appear to be different in the *S. cerevisiae* versus human spliceosome.

After submitting this paper, a cryo-EM structure of the *S. cerevisiae* spliceosomal B complex was reported (Plaschka et al., 2017). A comparison of the molecular architecture of the human and yeast B complexes, reveals that the structure and organization of most proteins common to the yeast and human B complexes are conserved. The location of most human B complex proteins in our model that have homologs in *S. cerevisiae*, for which only short pieces of structural information were available, e.g., MFAP1 (Spp381 in *S. cerevisiae*) or whose position was supported mainly by protein-protein crosslinking, is consistent with the position of their homologs in the yeast B complex structure. A notable difference between the yeast and human B complex structures is the lack of a U6 ACAGA/5' ss interaction in the yeast B complex despite that U1 snRNP is largely absent, which may indicate that the yeast complex is stalled at an earlier stage, but after release of U1.

Conclusions

Our cryo-EM structure of the human B complex provides important insights into the organization of the spliceosome prior to its activation. The B complex possesses a U2-containing head domain attached via three main bridges (B1-B3) to the main body where the U4/U6.U6 tri-snRNP and B-specific proteins are located. B1, which contains U2/U6 helix II, is likely involved in the initial docking of the tri-snRNP to the spliceosomal A complex. B2 may be important for stabilizing the association of the tri-snRNP within the B complex, whereas B3 allows communication between the U2/BS helix in the head and the extended U6 ACAGA/5' ss helix in the main body of the B complex. One of the most striking observations in this work is how tri-snRNP integration into the B complex leads to extensive rearrangements and/or repositioning of the majority of its proteins. These rearrangements lead, among others, to a partially closed conformation of Prp8 such that its RT/En and NTD1 domains, together with Dim1, form a protein pocket that binds the 5' ss. This is likely fol-

lowed by the movement of the U4/U6 snRNP and Brr2 helicase, such that the latter now contacts its U4/U6 substrate and is poised for the subsequent spliceosome activation step. Thus, these rearrangements generate the functional centers required to prime the B complex for the subsequent activation step. In addition, they create binding sites for the B-specific proteins, ensuring that they are first recruited at this specific stage of spliceosome assembly. Our structure also reveals how several of the human B-specific proteins are organized in the B complex, and suggests they are involved in negatively regulating Brr2, positioning the U6/5' ss helix and/or stabilizing the B complex structure. Finally, our work solidifies the intriguing idea that there is a significant difference between the early events of the activation pathways of human and yeast spliceosome. Consistent with the large exchange of spliceosomal proteins and major rearrangements in the RNA network that occur during the B to B^{act} transition, the general architecture of the B and B^{act} complex are dramatically different (Fabrizio et al., 2009). The exchange of more than 40 proteins in the human system will likely occur in multiple steps, with multiple intermediate complexes formed during the transformation of B to an activated B^{act} complex, hindering conclusions about the direction or trajectory of most protein rearrangements, as well as about the sequence of remodeling events at the atomic level, during the B to B^{act} transition. Future elucidation of the molecular architecture of the human B^{act} complex, as well as complexes at intermediate stages of activation, will thus greatly aid in elucidating the RNP remodeling events during spliceosome activation in higher eukaryotes.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
 - In vitro splicing
 - MS2 affinity selection of splicing complexes
 - 2D gel electrophoresis and mass spectrometry
 - ATP sensitivity of purified B complexes
 - Chase of purified B complexes with nuclear extract
 - Crosslinking of the B complex and crosslink identification
 - EM and image processing
 - Model fitting and building
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, two tables, and one movie and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2017.07.011>.

AUTHOR CONTRIBUTIONS

D.E.A. and M.N.L. developed the purification strategy and characterized the B complex. O.D. and H.U. analyzed protein-protein crosslinking data. K.B,

D.H., and H.S. prepared grids and collected the EM data. B.K., K.B., H.S., and R.L. analyzed the structure. All authors were involved in data interpretation. The manuscript was written by R.L., B.K., and C.L.W. with input from all authors. R.L., D.E.A., B.K., and H.S. initiated and orchestrated the project.

ACKNOWLEDGMENTS

We thank T. Conrad for HeLa cell production in a bioreactor; H. Kohansal for preparing HeLa cell nuclear extract; U. Steuerwald, W. Lendeckel, M. Raabe, and U. Pleßmann for excellent technical assistance; and K. Hartmuth for advice and discussions. Majety N. Leelaram was supported by a fellowship from the Alexander von Humboldt Stiftung. This work was supported by the Max Planck Society and the Deutsche Forschungsgemeinschaft (grants SFB 860 to R.L., H.S., and H.U. and LU294/15-1 to R.L.)

Received: May 14, 2017

Revised: June 22, 2017

Accepted: July 11, 2017

Published: August 3, 2017

REFERENCES

- Adams, P.D., Afonine, P.V., Bunkóczki, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 66, 213–221.
- Agafonov, D.E., Deckert, J., Wolf, E., Odenwälder, P., Bessonov, S., Will, C.L., Urlaub, H., and Lührmann, R. (2011). Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. *Mol. Cell. Biol.* 31, 2667–2682.
- Agafonov, D.E., Kastner, B., Dybkov, O., Hofele, R.V., Liu, W.-T., Urlaub, H., Lührmann, R., and Stark, H. (2016). Molecular architecture of the human U4/U6.U5 tri-snRNP. *Science* 351, 1416–1420.
- Bertram, K., Agafonov, D.E., Liu, W.-T., Dybkov, O., Will, C.L., Hartmuth, K., Urlaub, H., Kastner, B., Stark, H., and Lührmann, R. (2017). Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature* 542, 318–323.
- Boesler, C., Rigo, N., Agafonov, D.E., Kastner, B., Urlaub, H., Will, C.L., and Lührmann, R. (2015). Stable tri-snRNP integration is accompanied by a major structural rearrangement of the spliceosome that is dependent on Prp8 interaction with the 5' splice site. *RNA* 21, 1993–2005.
- Boesler, C., Rigo, N., Anokhina, M.M., Tauchert, M.J., Agafonov, D.E., Kastner, B., Urlaub, H., Ficner, R., Will, C.L., and Lührmann, R. (2016). A spliceosome intermediate with loosely associated tri-snRNP accumulates in the absence of Prp28 ATPase activity. *Nat. Commun.* 7, 11997.
- Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., 3rd, Shoeyink, J., Richardson, J.S., and Richardson, D.C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 35, W375–W383.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2126–2132.
- Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., and Lührmann, R. (2009). The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol. Cell* 36, 593–608.
- Fica, S.M., Tuttle, N., Novak, T., Li, N.-S., Lu, J., Koodathingal, P., Dai, Q., Staley, J.P., and Piccirilli, J.A. (2013). RNA catalyses nuclear pre-mRNA splicing. *Nature* 503, 229–234.
- Fica, S.M., Oubridge, C., Galej, W.P., Wilkinson, M.E., Bai, X.-C., Newman, A.J., and Nagai, K. (2017). Structure of a spliceosome remodelled for exon ligation. *Nature* 542, 377–380.
- Galej, W.P., Oubridge, C., Newman, A.J., and Nagai, K. (2013). Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature* 493, 638–643.
- Galej, W.P., Nguyen, T.H.D., Newman, A.J., and Nagai, K. (2014). Structural studies of the spliceosome: zooming into the heart of the machine. *Curr. Opin. Struct. Biol.* 25, 57–66.
- Galej, W.P., Wilkinson, M.E., Fica, S.M., Oubridge, C., Newman, A.J., and Nagai, K. (2016). Cryo-EM structure of the spliceosome immediately after branching. *Nature* 537, 197–201.
- Gozani, O., Feld, R., and Reed, R. (1996). Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes Dev.* 10, 233–243.
- Guex, N., and Peitsch, M.C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18, 2714–2723.
- Hang, J., Wan, R., Yan, C., and Shi, Y. (2015). Structural basis of pre-mRNA splicing. *Science* 349, 1191–1198.
- Huang, X., Beullens, M., Zhang, J., Zhou, Y., Nicolaescu, E., Lesage, B., Hu, Q., Wu, J., Bollen, M., and Shi, Y. (2009). Structure and function of the two tandem WW domains of the pre-mRNA splicing factor FBP21 (formin-binding protein 21). *J. Biol. Chem.* 284, 25375–25387.
- Huang, Y.-H., Chung, C.-S., Kao, D.-I., Kao, T.-C., and Cheng, S.-C. (2014). Sad1 counteracts Brp2-mediated dissociation of U4/U6.U5 in tri-snRNP homeostasis. *Mol. Cell. Biol.* 34, 210–220.
- Kastner, B., Fischer, N., Golas, M.M., Sander, B., Dube, P., Boehringer, D., Hartmuth, K., Deckert, J., Hauer, F., Wolf, E., et al. (2008). GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat. Methods* 5, 53–55.
- Korneta, I., Magnus, M., and Bujnicki, J.M. (2012). Structural bioinformatics of the human spliceosomal proteome. *Nucleic Acids Res.* 40, 7046–7065.
- Laggerbauer, B., Achsel, T., and Lührmann, R. (1998). The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplexes in vitro. *Proc. Natl. Acad. Sci. USA* 95, 4188–4192.
- Leung, A.K.W., Nagai, K., and Li, J. (2011). Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis. *Nature* 473, 536–539.
- Liu, S., Li, P., Dybkov, O., Nottrott, S., Hartmuth, K., Lührmann, R., Carologno, T., and Wahl, M.C. (2007). Binding of the human Prp31 Nop domain to a composite RNA-protein platform in U4 snRNP. *Science* 316, 115–120.
- Liu, S., Ghalei, H., Lührmann, R., and Wahl, M.C. (2011). Structural basis for the dual U4 and U4atac snRNA-binding specificity of spliceosomal protein hPrp31. *RNA* 17, 1655–1663.
- Liu, S., Mozaffari-Jovin, S., Wollenhaupt, J., Santos, K.F., Theuser, M., Dunin-Horkowicz, S., Fabrizio, P., Bujnicki, J.M., Lührmann, R., and Wahl, M.C. (2015). A composite double-/single-stranded RNA-binding region in protein Prp3 supports tri-snRNP stability and splicing. *eLife* 4, e07320.
- Macías, S., Bragulat, M., Tardiff, D.F., and Vilardell, J. (2008). L30 binds the nascent RPL30 transcript to repress U2 snRNP recruitment. *Mol. Cell* 30, 732–742.
- Mozaffari-Jovin, S., Santos, K.F., Hsiao, H.-H., Will, C.L., Urlaub, H., Wahl, M.C., and Lührmann, R. (2012). The Prp8 RNase H-like domain inhibits Brp2-mediated U4/U6 snRNA unwinding by blocking Brp2 loading onto the U4 snRNA. *Genes Dev.* 26, 2422–2434.
- Mozaffari-Jovin, S., Wandersleben, T., Santos, K.F., Will, C.L., Lührmann, R., and Wahl, M.C. (2013). Inhibition of RNA helicase Brp2 by the C-terminal tail of the spliceosomal protein Prp8. *Science* 341, 80–84.
- Nguyen, T.H.D., Galej, W.P., Bai, X.C., Savva, C.G., Newman, A.J., Scheres, S.H.W., and Nagai, K. (2015). The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature* 523, 47–52.
- Nguyen, T.H.D., Galej, W.P., Bai, X.C., Oubridge, C., Newman, A.J., Scheres, S.H.W., and Nagai, K. (2016). Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. *Nature* 530, 298–302.
- Papasaikas, P., and Valcárcel, J. (2016). The spliceosome: the ultimate RNA chaperone and sculptor. *Trends Biochem. Sci.* 41, 33–45.
- Papasaikas, P., Tejedor, J.R., Vigevani, L., and Valcárcel, J. (2015). Functional splicing network reveals extensive regulatory potential of the core spliceosomal machinery. *Mol. Cell* 57, 7–22.

- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612.
- Plaschka, C., Lin, P.-C., and Nagai, K. (2017). Structure of a pre-catalytic spliceosome. *Nature* 546, 617–621.
- Raghunathan, P.L., and Guthrie, C. (1998). A spliceosomal recycling factor that reanneals U4 and U6 small nuclear ribonucleoprotein particles. *Science* 279, 857–860.
- Rauhut, R., Fabrizio, P., Dybkov, O., Hartmuth, K., Pena, V., Chari, A., Kumar, V., Lee, C.-T., Urlaub, H., Kastner, B., et al. (2016). Molecular architecture of the *Saccharomyces cerevisiae* activated spliceosome. *Science* 353, 1399–1405.
- Reidt, U., Wahl, M.C., Fasshauer, D., Horowitz, D.S., Lührmann, R., and Ficner, R. (2003). Crystal structure of a complex between human spliceosomal cyclophilin H and a U4/U6 snRNP-60K peptide. *J. Mol. Biol.* 331, 45–56.
- Reuter, K., Nottrott, S., Fabrizio, P., Lührmann, R., and Ficner, R. (1999). Identification, characterization and crystal structure analysis of the human spliceosomal U5 snRNP-specific 15 kD protein. *J. Mol. Biol.* 294, 515–525.
- Reyes, J.L., Gustafson, E.H., Luo, H.R., Moore, M.J., and Konarska, M.M. (1999). The C-terminal region of hPrp8 interacts with the conserved GU dinucleotide at the 5' splice site. *RNA* 5, 167–179.
- Santos, K.F., Jovin, S.M., Weber, G., Pena, V., Lührmann, R., and Wahl, M.C. (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.
- Schellenberg, M.J., Wu, T., Ritchie, D.B., Fica, S., Staley, J.P., Atta, K.A., LaPointe, P., and MacMillan, A.M. (2013). A conformational switch in PRP8 mediates metal ion coordination that promotes pre-mRNA exon ligation. *Nat. Struct. Mol. Biol.* 20, 728–734.
- Scheres, S.H.W. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* 180, 519–530.
- Schneider, M., Hsiao, H.-H., Will, C.L., Giet, R., Urlaub, H., and Lührmann, R. (2010). Human PRP4 kinase is required for stable tri-snRNP association during spliceosomal B complex formation. *Nat. Struct. Mol. Biol.* 17, 216–221.
- Schütze, T., Ulrich, A.K.C., Apelt, L., Will, C.L., Bartlick, N., Seeger, M., Weber, G., Lührmann, R., Stelzl, U., and Wahl, M.C. (2016). Multiple protein-protein interactions converging on the Prp38 protein during activation of the human spliceosome. *RNA* 22, 265–277.
- Sidarovich, A., Will, C.L., Anokhina, M.M., Ceballos, J., Sievers, S., Agafonov, D.E., Samatov, T., Bao, P., Kastner, B., Urlaub, H., et al. (2017). Identification of a small molecule inhibitor that stalls splicing at an early step of spliceosome activation. *eLife* 6, e23533.
- Singer, A., Coifman, R.R., Sigworth, F.J., Chester, D.W., and Shkolnisky, Y. (2010). Detecting consistent common lines in cryo-EM by voting. *J. Struct. Biol.* 169, 312–322.
- Spartz, A.K., Herman, R.K., and Shaw, J.E. (2004). SMU-2 and SMU-1, *Caenorhabditis elegans* homologs of mammalian spliceosome-associated proteins RED and fSAP57, work together to affect splice site choice. *Mol. Cell. Biol.* 24, 6811–6823.
- Staley, J.P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* 92, 315–326.
- Staley, J.P., and Guthrie, C. (1999). An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. *Mol. Cell* 3, 55–64.
- Ulrich, A.K.C., and Wahl, M.C. (2017). Human MFAP1 is a cryptic ortholog of the *Saccharomyces cerevisiae* Spp381 splicing factor. *BMC Evol. Biol.* 17, 91.
- Ulrich, A.K.C., Schulz, J.F., Kamprad, A., Schütze, T., and Wahl, M.C. (2016a). Structural basis for the functional coupling of the alternative splicing factors Smu1 and RED. *Structure* 24, 762–773.
- Ulrich, A.K.C., Seeger, M., Schütze, T., Bartlick, N., and Wahl, M.C. (2016b). Scaffolding in the spliceosome via single α helices. *Structure* 24, 1972–1983.
- Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* 136, 701–718.
- Wan, R., Yan, C., Bai, R., Huang, G., and Shi, Y. (2016a). Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution. *Science* 353, 895–904.
- Wan, R., Yan, C., Bai, R., Wang, L., Huang, M., Wong, C.C.L., and Shi, Y. (2016b). The 3.8 Å structure of the U4/U6.U5 tri-snRNP: Insights into spliceosome assembly and catalysis. *Science* 351, 466–475.
- Wolf, E., Kastner, B., Deckert, J., Merz, C., Stark, H., and Lührmann, R. (2009). Exon, intron and splice site locations in the spliceosomal B complex. *EMBO J.* 28, 2283–2292.
- Xie, J., Beickman, K., Otte, E., and Rymond, B.C. (1998). Progression through the spliceosome cycle requires Prp38p function for U4/U6 snRNA dissociation. *EMBO J.* 17, 2938–2946.
- Yan, C., Wan, R., Bai, R., Huang, G., and Shi, Y. (2016). Structure of a yeast activated spliceosome at 3.5 Å resolution. *Science* 353, 904–911.
- Yan, C., Wan, R., Bai, R., Huang, G., and Shi, Y. (2017). Structure of a yeast step II catalytically activated spliceosome. *Science* 355, 149–155.
- Yang, B., Wu, Y.J., Zhu, M., Fan, S.B., Lin, J., Zhang, K., Li, S., Chi, H., Li, Y.X., Chen, H.-F., et al. (2012). Identification of cross-linked peptides from complex samples. *Nat. Methods* 9, 904–906.
- Zhang, K. (2016). Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* 193, 1–12.
- Zhang, X., Yan, C., Hang, J., Finci, L.I., Lei, J., and Shi, Y. (2017). An atomic structure of the human spliceosome. *Cell* 169, 918–929.e14.
- Zheng, S.Q., Palovcak, E., Armache, J.-P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* 14, 331–332.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
Rosetta 2(DE3) Competent Cells	EMD Millipore	Cat# 71397
Deposited Data		
EM map of the human B complex, soft mask 2 (4.5 Å)	This paper	EMDB: 3766
EM map of the human B complex, soft mask 1 (5.4 Å)	This paper	EMDB: 3767
EM map of the human B complex, unmasked sharpened	This paper	EMDB: 3768
EM map of the human B complex, unmasked (9.9 Å)	This paper	EMDB: 3769
PDB coordinates of the human B complex	This paper	PDB: 5O9Z
Cryo-EM model of the U4/U6.U5 tri-snRNP of <i>Saccharomyces cerevisiae</i>	Wan et al., 2016b	PDB: 3JCM
Cryo-EM model of the human U4/U6.U5 tri-snRNP	Agafonov et al., 2016	PDB: 3JCR EMDB: 6581
Cryo-EM model of the U4/U6.U5 tri-snRNP of <i>Saccharomyces cerevisiae</i>	Nguyen et al., 2016	PDB: 5GAN
Cryo-EM model of the Bact complex of <i>Saccharomyces cerevisiae</i>	Rauhut et al., 2016	PDB: 5LQW
Cryo-EM model of the human C* complex	Bertram et al., 2017	PDB: 5MQF, EMDB: 3545
Crystal structure of the human Prp4-CypH complex	Reidt et al., 2003	PDB: 1MZV
Crystal structure of the human Dim1 protein	Reuter et al., 1999	PDB: 1QGV
Crystal structure of the human Prp31-Snu13-U4 snRNA complex	Liu et al., 2007	PDB: 2OZB
Crystal structure of the human Prp31-Snu13-U4atac 5' stem loop complex	Liu et al., 2011	PDB: 3SIU
Crystal structure of Prp8-Aar2 complex from <i>S. cerevisiae</i>	Galej et al., 2013	PDB: 3ZEF
Crystal structure of the human PRP8 RNase H-like domain	Schellenberg et al., 2013	PDB: 4JK7
Crystal structure of the human Brr2 helicase with the Prp8 Jab1/MPN domain	Mozaffari-Jovin et al., 2013	PDB: 4KIT
Crystal structure of the human U4 snRNP core domain	Leung et al., 2011	PDB: 4WZJ
Crystal structure of the Prp3 ferredoxin-like domain bound to a fragment of U4/U6 di-snRNA of <i>S. cerevisiae</i>	Liu et al., 2015	PDB: 4YHW
Crystal structure of the Smu1-RED complex of <i>Caenorhabditis elegans</i>	Ulrich et al., 2016a	PDB: 5EN6, 5EN7
Crystal structure of the human Prp38-MFAP1 complex	Ulrich et al., 2016b	PDB: 5F5S
Crystal structure of the Snu23-Prp38-MFAP1 complex of <i>Chaetomium thermophilum</i>	Ulrich et al., 2016b	PDB: 5F5U, 5F5V
Experimental Models: Cell Lines		
Human: HeLa S3 cells	Helmholtz Center for Infection Research, Brunswick	N/A
Recombinant DNA		
Plasmid: pMBP-MS2	Macías et al., 2008	Addgene plasmid # 65104
Plasmid: pT7-MINX-M3	Boesler et al., 2016	N/A
Software and Algorithms		
Coot v0.8.3	Emsley and Cowtan, 2004	https://www2.mrc-lmb.cam.ac.uk/ personal/pemsley/coot/
Gautomatch	Dr. Kai Zhang	http://www.mrc-lmb.cam.ac.uk/kzhang/
Gctf	Zhang, 2016	http://www.mrc-lmb.cam.ac.uk/kzhang/

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
MotionCor2	Zheng et al., 2017	http://www.msg.ucsf.edu/em/software/index.html
PHENIX suite	Adams et al., 2010	https://www.phenix-online.org
pLink v.1.23	Yang et al., 2012	http://pfind.ict.ac.cn/software/pLink/
PyMOL	Schrödinger LCC	http://www.pymol.org
RELION v2.0.2	Scheres, 2012	http://www2.mrc-lmb.cam.ac.uk/relion/index.php/Main_Page
SpliProt3D	Korneta et al., 2012	http://iimcb.genesisilico.pl/SpliProt3D/home/
SWISS-MODEL suite	Guex and Peitsch, 1997	http://spdbv.vital-it.ch/disclaim.html
UCSF Chimera v.1.11.2	Pettersen et al., 2004	http://www.cgl.ucsf.edu/chimera/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Reinhard Lührmann (reinhard.luehrmann@mpibpc.mpg.de).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

For in vitro splicing and purification of spliceosomes, nuclear extract from HeLa S3 cells was used. HeLa cells were grown in DMEM/F12 (1:1) medium supplemented with 5% NCS serum in a 30l fermenter (Applikon Bioteck) at a density of 6.5×10^6 cells/ml and harvested using a Sorvall BIOS 16 centrifuge.

MBP-MS2 fusion protein was expressed in the *Escherichia coli* strain Rosetta 2 (DE3) (Novagen) which was grown in 2YT medium at 37°C.

METHOD DETAILS**In vitro splicing**

Uniformly [³²P]-labeled, m⁷G(5')ppp(5')G-capped MINX pre-mRNA was synthesized in vitro by T7 runoff transcription. HeLa S3 cells were obtained from GBF, Braunschweig (currently Helmholtz Zentrum für Infektionsforschung, Braunschweig) and tested negative for mycoplasma.

To prepare splicing active nuclear extracts, HeLa cells were grown to a density of 6.5×10^6 cells/ml and harvested by centrifuging for 10 min at 2000 rpm in a 8 × 2000 mL BIOS rotor (Thermo Scientific). Cells were washed twice with ice cold PBS (pH 7.4) and resuspended in 1.25 volumes of MC buffer [10 mM HEPES-KOH, pH 7.6, 10 mM KOAc, 0.5 mM Mg(OAc)₂, 0.5 mM DTT] supplemented with 2 protease inhibitor cocktail tablets (Roche) per 50 mL of the buffer. They were then incubated for 5 min on ice and lysed with 18 strokes of a Dounce homogenizer at 4°C. Nuclei were pelleted for 5 min at 10000 rpm in a F14-14x50cy rotor (Thermo Scientific). After removing the supernatant, 1.3 volumes of Roeder C buffer [25% (v/v) glycerol, 20 mM HEPES-KOH, pH 7.6, 0.2 mM EDTA pH 8.0, 420 mM NaCl] supplemented with 0.5 mM DTT and 0.5 mM PMSF were added per gram of nuclei. The latter were then lysed with 20 strokes of a Dounce homogenizer. The mixture was stirred slowly for 40 min at 4°C, followed by centrifugation for 30 min at 12300 rpm in a F14-14x50 rotor (Thermo Scientific). The supernatant was recovered and was immediately used for B complex assembly as described below without dialysis or freezing.

To isolate B complexes, splicing was performed with 5 nM of ³²P-labeled pre-mRNA and 20% (v/v) HeLa nuclear extract, in buffer containing 0.3 mM MgCl₂, 0.2 mM EDTA, 50 mM NaCl, 20 mM HEPES-KOH pH 7.9, 2 mM ATP and 20 mM creatine phosphate, and was incubated at 30°C for 2 hr.

MS2 affinity selection of splicing complexes

Spliceosomal complexes were isolated by MS2 affinity selection. MINX pre-mRNA containing three MS2 aptamers at its 3' end RNA was incubated with a tenfold molar excess of MBP-MS2 fusion protein and then added to a splicing reaction. After incubating at 30°C for 2 h, centrifuging to remove aggregates, and adding 100 mM NaCl, the reaction was loaded onto a MBP Trap HP column (GE Healthcare). The column was washed with G-150 buffer (20 mM HEPES-KOH pH 7.9, 1.5 mM MgCl₂, 150 mM NaCl) and complexes were eluted with G-150 buffer containing 1 mM maltose. Eluted complexes were loaded onto a 36 ml linear 5%–20% (w/v) sucrose gradient containing G-150 buffer, centrifuged at 25,000 rpm for 10 h at 4°C in a Surespin 630 (Thermo Scientific) rotor, and gradient fractions were harvested from the bottom. The distribution of ³²P-labeled MINX pre-mRNA across the gradient was

determined by Cherenkov counting. Fractions were analyzed by denaturing 4%–12% NuPAGE (Life Technologies) followed by autoradiography and SYBR Gold staining to detect RNA. Peak fractions containing B complexes were pooled, concentrated by centrifugation with an Amicon 50 kD cut-off unit, diluted to decrease the sucrose concentration and reloaded on the same gradients with glutaraldehyde as fixative (Kastner et al., 2008). The GraFix gradient contained 0%–0.15% of glutaraldehyde and fractions were quenched with 100 mM aspartate immediately after harvesting. For biochemical sample validation, the same procedure was performed but without fixation in the second gradient. The RNA and protein compositions of purified complexes were determined by denaturing PAGE and two-dimensional (2D) gel electrophoresis.

2D gel electrophoresis and mass spectrometry

2D gel electrophoresis of affinity-purified spliceosomal complexes was performed as described previously (Agafonov et al., 2011) using a 7.5% acrylamide mono gel in the second dimension for analysis of proteins larger than 50 kDa, or 15% acrylamide for proteins smaller than 50 kDa. For mass spectrometry, Coomassie-stained protein-spots were cut out of the 1D or 2D gels, and proteins were digested in-gel with trypsin and extracted. The extracted peptides were analyzed in a liquid-chromatography coupled electrospray ionization mass spectrometer (LTQ Orbitrap XL) under standard conditions. Proteins were identified by searching fragment spectra against the NCBI non-redundant (nr) database using Mascot as a search engine.

ATP sensitivity of purified B complexes

Affinity-purified B complexes formed on ^{32}P -labeled MINX-MS2 pre-mRNA were incubated with or without 2 mM ATP at 30°C for 30 min. The integrity of the complexes was checked by sedimentation in a 5%–20% (w/v) sucrose gradient containing G-150 buffer followed by Bradford assay of gradient fractions to determine the peak. Aliquots of peak fractions were separated by denaturing 4%–12% NuPAGE (Life Technologies) and RNA was visualized by SYBR Gold staining.

Chase of purified B complexes with nuclear extract

Affinity-purified B complexes formed on ^{32}P -labeled MINX-MS2 pre-mRNA were incubated with splicing buffer alone (20 mM HEPES-KOH pH 7.9, 50 mM NaCl, 3 mM MgCl₂, 2 mM ATP, 20 mM creatine phosphate) or additionally in the presence of 20% untreated HeLa nuclear extract or extract pre-treated with micrococcal nuclease (NEB) as described previously (Bertram et al., 2017). The splicing reaction was initiated by addition of ^{32}P -labeled MINX-MS2 pre-mRNA or purified B complexes, and then incubated at 30°C for 0–90 min. Time point aliquots were analyzed by SDS-PAGE, and the ^{32}P -labeled RNA was visualized with a Typhoon phosphorimager (GE Healthcare).

Crosslinking of the B complex and crosslink identification

After gradient centrifugation, MS2 affinity-purified spliceosomal complexes were crosslinked with 150 μM BS3 for 30 min at 20°C and further purified by a second gradient centrifugation step. Approximately 25 pmol of B complexes were pelleted by ultracentrifugation and analyzed essentially as described before (Bertram et al., 2017). After digestion with trypsin, peptides were reverse-phase extracted and fractionated on a Superdex Peptide PC3.2/30 column (GE Healthcare). 50 μL fractions corresponding to an elution volume of 1.2–1.8 ml were analyzed on Thermo Scientific Q Exactive HF, Orbitrap Fusion Tribrid or Orbitrap Fusion Lumos Tribrid mass spectrometers. Protein–protein crosslinks were identified by pLink 1.23 search engine (<http://pfind.ict.ac.cn/software/pLink>) and filtered at FDR 1% as recommended by the developer (Yang et al., 2012). For simplicity, the crosslink score is reported as a negative value of the common logarithm of the original pLink score, i.e., Score = $-\log_{10}(\text{pLink Score})$. For model building, a maximum distance of 30 Å between the C α atoms of the crosslinked lysines was allowed.

EM and image processing

A negative-stain starting model was built and refined essentially as described previously (Bertram et al., 2017; Singer et al., 2010). 8,168 cryo-images were recorded at -193°C in a Titan Krios electron microscope (FEI Company, the Netherlands) on a Falcon III direct electron detector at 120,700x magnification resulting in a pixel size of 1.16 Å at the specimen level. 20 frames were recorded for each micrograph with an average dose of 1.5 e⁻ per frame per Å². Motion correction and spatial frequency weighted frame summation was achieved using the MotionCor2 software (Zheng et al., 2017) (<http://www.msg.ucsf.edu/em/software/index.html>). Summed micrograph images were then evaluated based on real space appearance and CTF parameters and only ca 6,000 with good contrast and isotropic Thon rings were used for particle picking and extraction. Using the particle picking software Gautomatch (<http://www.mrc-lmb.cam.ac.uk/kzhang/>) and 40° projections of the negative stain model filtered to 40 Å as a reference, we extracted \sim 550,000 particle images from the pre-sorted, dose-weighted cryo-micrographs and applied several particle sorting steps at the 2D and 3D level. 2D multivariate statistics and classification were first applied to the non-aligned particle images and subsequently to the aligned particles. In each round, only particles comprising better resolved classes were included in further processing. The remaining \sim 407,000 particles were then re-extracted using RELION 2.0 and coarsened 2x in the process. 3D classification of these particles in RELION, featuring ten classes, resulted in \sim 49,000 particles in the overall best defined class. Subsequent rounds of 3D refinement followed by 2D classifications, where particles from the most poorly defined class-sums in RELION were discarded, yielded a 2x coarsened model at 9.9 Å. For the highest resolution structure, the \sim 44,600 remaining particles were re-extracted at their native pixel size using coordinates refined by previous rounds of classification and consequently utilized for an additional round of

refinement yielding a 9.3 Å resolution unmasked structure. A soft mask (soft mask 1, [Figure S2](#)) that included the density harboring Brr2 helicase and the U4 Sm core, with a cut-off of 7 voxel was then used for the refinement and for the determination of resolution. A map with a resolution of 5.4 Å as determined by Fourier shell correlation calculated from two independent datasets with a threshold of 0.143, was obtained. A second soft mask, that excluded the more flexible density region of Brr2 helicase and the U4 Sm core (soft mask 2, [Figure S2](#)) was applied during refinement in a similar way as described above and yielded an additional map of this area with a resolution of 4.5 Å as determined by Fourier shell correlation calculated from two independent datasets with a threshold of 0.143. A local resolution plot revealed that there are indeed areas of higher resolution at the RNP core of the B complex that approach 3.5 Å. Some peripheral regions have somewhat lower resolution ([Figures 1](#) and [S2](#)). To obtain the most complete B complex structure at the overall highest possible resolution, sorting schemes were adjusted accordingly. Using a starting model that did not include the head of the B complex during 3D classification, the ~407,000 2x coarsened particles (see above) were split into six equally sized groups and 3D classified accordingly, with three classes each. Particles from all classes yielding a (partially) defined structure, regardless of the appearance of the head, were pooled and split again into two subsets of ~120,000 particles each. Both particle sets were then again 3D classified into 6 classes each and inspected visually. Particles from the two best defined classes were then pooled and 3D refined without masking in RELION, yielding a model revealing the best achievable head definition at 9.9 Å resolution according to the FSC 0.143 criterion (which we refer to as the unmasked EM density map).

Model fitting and building

Available X-ray or homology models of proteins were fit into the EM density using Chimera ([Pettersen et al., 2004](#)). Individual models of substructures (e.g., domains or structural motifs) were further fitted as rigid bodies using Coot ([Emsley and Cowtan, 2004](#)). The models were adjusted manually to fit into the EM density after visual inspection. Disordered regions were removed and regions that were reorganized or were not present in the initial models (e.g., loops and secondary structure elements) were built using Coot. Homology models of proteins were either obtained using the SWISS-MODEL suite ([Guex and Peitsch, 1997](#)), or were directly adapted from the SpIProt3D database ([Korneta et al., 2012](#)). Details of the processing of protein models incorporated into the B structure are described in [Table S1](#). An initial model of the human U5 snRNA was obtained from the C* complex (PDB: 5MQF, [Bertram et al., 2017](#)) and fitted into the 4.5 Å EM density using Coot. MINX pre-mRNA intron bases A56 and G58 could subsequently be modeled as base-pairing with U5 loop 1 bases U41 and U43. This then allowed the placement of all other modeled pre-mRNA bases by tracing in the EM density. U4 snRNA A1-C16 were initially modeled as an idealized double helix base paired with U6 snRNA G59-U74 (helix II) and subsequently refined in Coot. U4 snRNA nucleotides A20-U52 were modeled in a similar manner, based on its crystal structure (PDB: 2OZB, [Liu et al., 2007](#)). Missing U4 bases up to U62 were then modeled in Coot by tracing the EM density. The remaining U4 snRNA bases up to A68 were modeled from its 3' terminal helix (which was modeled as an idealized helix) by tracing their path in the EM density. U6 snRNA bases preceding U4/U6 stem II and comprising the helical ACAGA and extended ACAGA box elements, were modeled tracing their path in the EM density using Coot beginning at A30. The U6 5' terminal loop and adjacent bases were adapted from the human C* reference structure (PDB: 5MQF, [Bertram et al., 2017](#)). Once the entire coordinate model was built, all proteins were truncated to poly-alanine and a global minimization real space refinement was conducted against the 4.5 Å or 5.4 Å cryo-EM density, respectively, using the real space refine program from the PHENIX suite ([Adams et al., 2010](#)) (https://www.phenix-online.org/documentation/reference/real_space_refine.html). The RNA model was validated using the MolProbity server ([Davis et al., 2007](#)) and exhibited an all atom clash score of less than 12 and no bad bond-lengths or -angles. The scores for individual RNA nucleotides shown in the final B complex model are provided in a separate HTML file. Final visualization was performed with Chimera and PyMOL (<http://www.pymol.org>).

DATA AND SOFTWARE AVAILABILITY

The cryo-EM maps have been deposited in the Electron Microscopy Data Bank with accession codes EMDB: 3766 (4.5 Å map), EMDB: 3767 (5.4 Å map), and EMDB: 3769 (unmasked 9.9 Å map). A sharpened version of the unmasked map is deposited under EMDB: 3768. The atomic model has been deposited in the Protein Data Bank under accession code PDB: 5O9Z.

Supplemental Figures

Cell

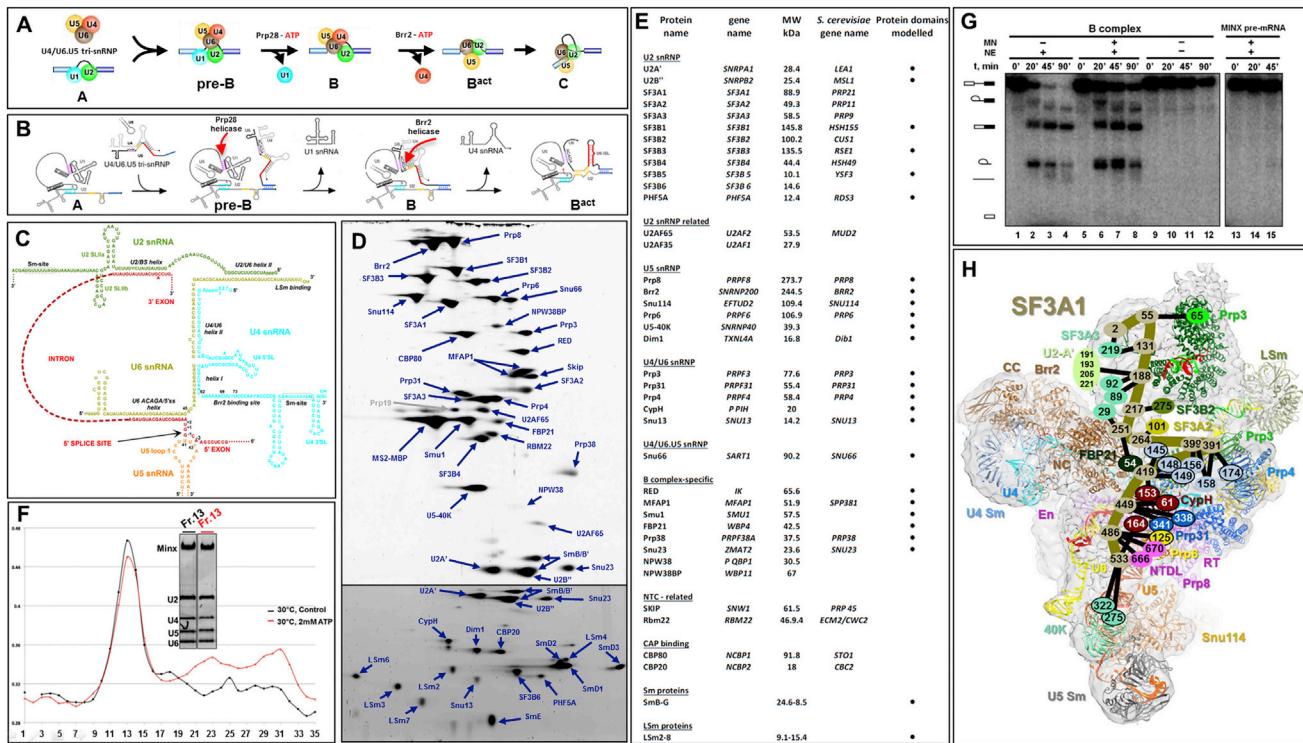


Figure S1. Compositional and Functional Characterization of Affinity-Purified Human B Complexes, Related to Figure 1

- (A) Spliceosome assembly pathway, starting with complex A.
- (B) RNA-RNA rearrangements occurring during spliceosome activation and catalysis of splicing.
- (C) Schematic representation of the secondary structure of RNA in the spliceosomal B complex. The complete secondary structure of the U4/U6 duplex is shown. Only selected regions of the pre-mRNA, U2 and U5 snRNAs are shown.
- (D) Identification of abundant B complex proteins by 2D gel electrophoresis followed by mass spectrometry. Upper and lower panel: proteins > 30 kDa or < 30 kDa, respectively.
- (E) Summary of proteins in purified, human B complexes. Proteins, or domains thereof, that were modeled into the B complex EM density map are indicated by a bullet point.
- (F) Purified B complexes are stable in the presence of ATP. Affinity-purified B complexes formed on ³²P-labeled MINX-MS2 pre-mRNA were incubated with (red line) or without 2 mM ATP (black line) at 30°C for 30 min and fractionated by sedimentation on a 5%-20% (w/v) sucrose gradient. RNA in peak fraction #13 was separated by denaturing 4%-12% NuPAGE and visualized by SYBR Gold staining. B complexes incubated with ATP not only exhibit the same sedimentation behavior as the control B complexes, but also the same RNA composition, indicating that Brr2 helicase is negatively regulated in the purified B complexes and does not displace U4 from U6 snRNA in the presence of ATP.
- (G) Chase of purified B complexes with nuclear extract. Affinity-purified B complexes formed on ³²P-labeled MINX-MS2 pre-mRNA were incubated with untreated HeLa nuclear extract (lanes 1-4) or extract pre-treated with micrococcal nuclease (MN) (lanes 5-8), or with splicing buffer alone (lanes 9-12). The pre-mRNA was efficiently chased into mRNA in the presence of both types of nuclear extract, but not buffer alone, indicating that our purified B complexes are functional and not dead-end complexes. As a control, ³²P-labeled MINX-MS2 pre-mRNA was incubated in the presence of extract pre-treated with MN (lanes 13-15). The pre-mRNA was not spliced with nuclear extract pre-treated with MN, confirming that the vast majority of snRNAs in the MN treated nuclear extract were destroyed. All 2D analyses and in vitro splicing experiments were performed at least twice in two independent experiments.
- (H) Network of crosslinks between U2 SF3A1 and proteins in the head and the central part of the B complex body. Overview of the positions of selected proteins and RNA in the B complex EM density map (shown is the front view of the B complex; see Figure 1A, left panel), with a schematic diagram showing intermolecular crosslinks between U2 SF3A1 and other indicated proteins, and the likely path of SF3A1. Numbers indicate the positions of crosslinked lysine residues (connected by black lines) in each protein. Ovals share the same color as that of the crosslinked protein's name. The crosslinking pattern suggests a path for U2 SF3A1 from the head region along the central axis of the main body close to its foot region, via bridge B3.

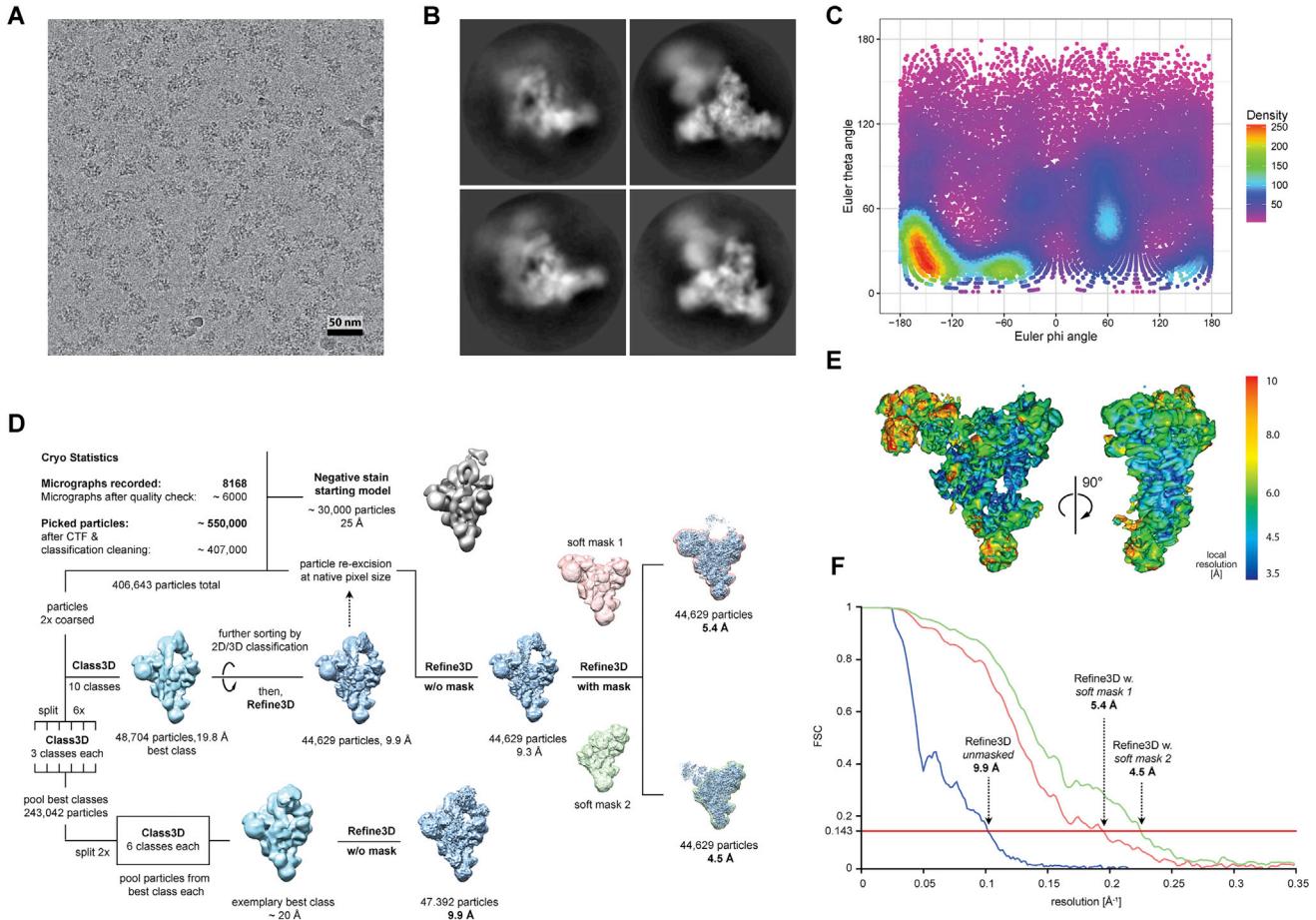


Figure S2. Cryo-EM and Image Processing of the Human B Complex, Related to Figure 1

(A) Typical cryo-EM raw image of *H. sapiens* B spliceosomes recorded with a Titan Krios (FEI Company) electron microscope at a nominal magnification of 120,700 \times with a Falcon III direct electron detector resulting in a pixel size of 1.16 Å.

(B) Several representative class averages showing different views of the B complex after reference free 2D classification.

(C) Euler angle distribution of all particle images that contributed to the final high resolution 3D map. The coordinates describe the phi and theta angles. The color and size of the dots reflect the number of particles at any given Euler angle.

(D) Computational sorting scheme. Imaged micrographs were first evaluated according to their real space appearance and the Thon ring quality of local power spectra. Roughly 550,000 particle images were then selected from the remaining micrographs. In a second sorting step, particle images were again discarded based on the quality of Thon rings in classified, local power spectra. After evaluations in Fourier space, particles were subsequently excluded according to multiple rounds of 2D classifications. The remaining ca 407,000 particles were then coarsened 2x to undergo further classification in RELION 2.0. A 3D classification featuring 10 classes was then performed and 48,704 particles were extracted from the dataset. In subsequent rounds of RELION 2D and 3D refinement, only those particles comprising well resolved classes were selected from the dataset, resulting in a 3D volume reconstructed from 44,629 particles at 9.9 Å. To achieve a reconstruction at maximum resolution, only those particles contributing to this final 3D model were re-extracted from their micrographs at the native pixel size of 1.16 Å, using coordinates refined by previous rounds of classification. A further round of 3D refinement in RELION revealed a final, unmasked structure with 9.3 Å overall resolution. To improve details in the more stable areas of the complex, two soft masks (1 & 2) were applied separately, each during one subsequent round of 3D refinement, yielding a final model at 4.5 Å resolution according to the FSC 0.143 criterion. To find the overall best defined structure of the entire complex (including its head region), a modified sorting and classification scheme was applied. Using the 2x coarsened, pre-processed 407,000 single particle images and a 3D starting model lacking the head, the particle dataset was split into 6 equally sized groups and 3D classified into three classes each in RELION. After pooling all particles from the best defined classes, the particle dataset was split again into two equally sized datasets. Subsequent 3D classification of each individual particle dataset, using six classes each, followed by 3D refinement finally yielded the overall best defined structure at 9.9 Å.

(E) Local resolution plot of the complex refined with soft mask 1 reveals a resolution distribution from approximately 3.5 to 10 Å with some less well-defined areas at the periphery of the complex. Higher resolution regions (in blue, up to 3.5 Å resolution) were obtained for the centrally-located core of the spliceosome.

(F) Fourier-shell correlation function of two independently refined half datasets calculated during the 3D Refinement or PostProcessing procedure in RELION indicates a global resolution of 9.9 Å for the unmasked B complex. The same function calculated for the volumes refined with soft mask 1 and 2 reveals a resolution of 5.4 Å and 4.5 Å, respectively.

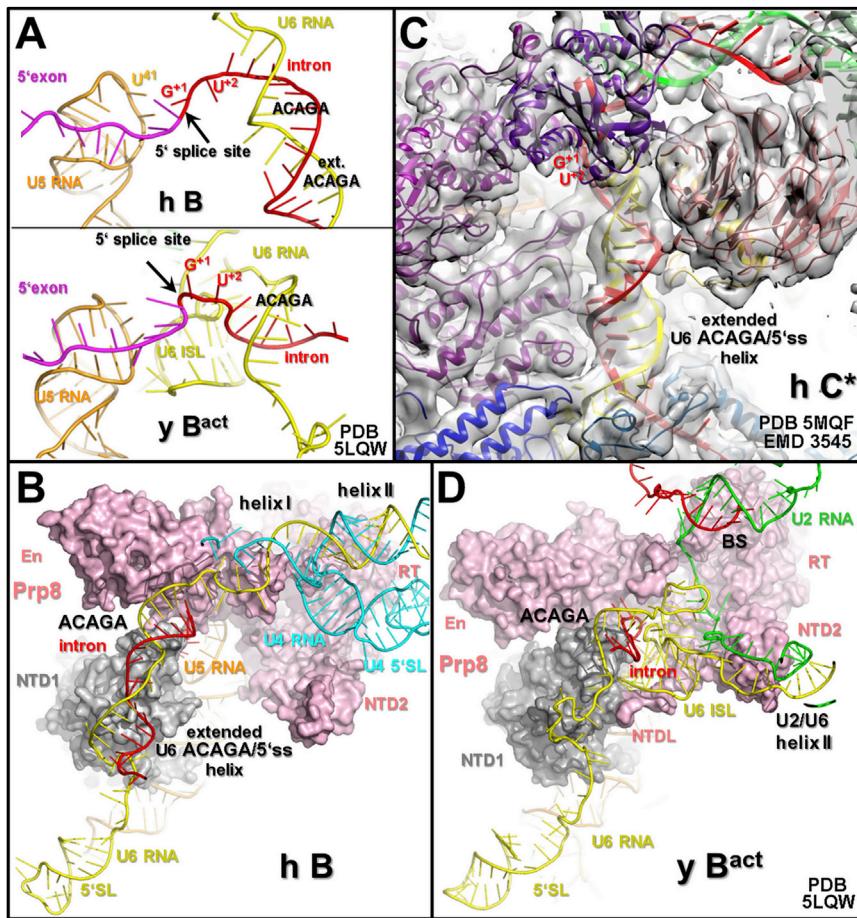


Figure S3. Arrangements of U6 snRNA and the U6 ACAGA/5' ss Helix in Various Spliceosomal Complexes, Related to Figures 1 and 2

(A) Substantial differences between the conformations of the first two nucleotides of the 5'ss and the positions of the U6 ACAGA/5' ss helices in the human B complex (upper panel) and the yeast B^{act} complex (lower panel).

(B-D) An extended U6 ACAGA/5' ss helix is present in the human B (B) and C* (C) complexes, but not in the yeast B^{act} complex (D). Complexes in (B) and (D) are aligned relative to Snu114 and Prp8's NTD1. In panel C, the extended U6 ACAGA/5' ss helix from the human B complex was fit into the corresponding density of the C* cryo-EM map. Panels B and D also show that the 5' stem-loop of U6 snRNA adopts already in the B complex a position very similar to its position at later functional stages of the spliceosome.

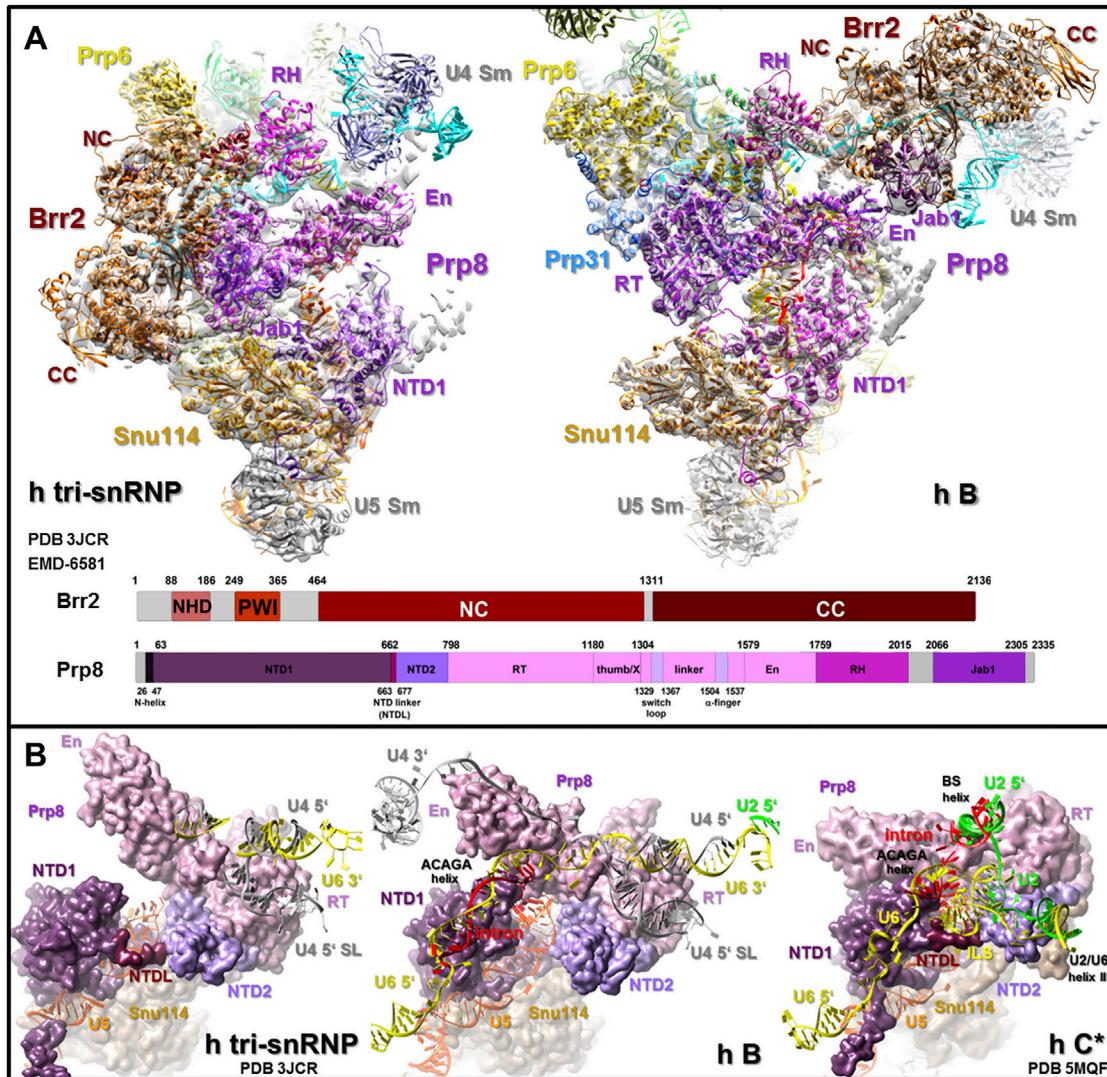


Figure S4. Differential Arrangements of Selected U5 and U4/U6 Proteins in the Human Tri-snRNP, B and C* Complex, Related to Figure 2

(A) Location of major U5 proteins in the isolated human tri-snRNP and the human B complex. While the U5 protein Snu114 and Prp8's NTD1 domain are structurally organized in a very similar manner in the tri-snRNP (left) and B complex (right), the helicase domain of Brr2 is located at radically different positions and is found at opposite ends of Prp8's RT/En domain in the two complexes. The domain organization of human Prp8 and Brr2 is shown at the bottom of panel A. Amino acids at the domain boundaries are indicated by numbers.

(B) Prp8 adopts significantly different conformations in the human U4/U6.U5 tri-snRNP, B and C* complexes. (Left panel) Open conformation of Prp8 in the human U4/U6.U5 tri-snRNP, where the En end of the elongated RT/En domain is well-separated from the upper region of the NTD1 domain (see also Agafonov et al., 2016). (Middle panel) Partially closed conformation of Prp8 in the human B complex. To achieve the B complex conformation, the RT/En domain must move toward the NTD1 domain, such that the En domain just touches the upper region of the NTD1 domain. (Right panel) Closed conformation of Prp8 in the human, catalytically-active C* complex. To achieve this conformation, the RT/En domain must move even closer to the NTD1 domain, generating a large interface between the Prp8 NTD1 and RT/En domains, which now clamps the 5' exon instead of the 5' ss. In the closed conformation of Prp8, the pocket that accommodates the rearranged catalytic U2/U6 RNA network is generated (see also Bertram et al., 2017). All complexes are aligned relative to Snu114 and the Prp8 NTD1 domain. Prp8 also adopts a very similar closed conformation in the *S. cerevisiae* B^{act}, C, and C* complexes, as well as in the *S. pombe* intron-lariat spliceosome (Fica et al., 2017; Galej et al., 2016; Hang et al., 2015; Rauhut et al., 2016; Wan et al., 2016a; Yan et al., 2016, 2017).

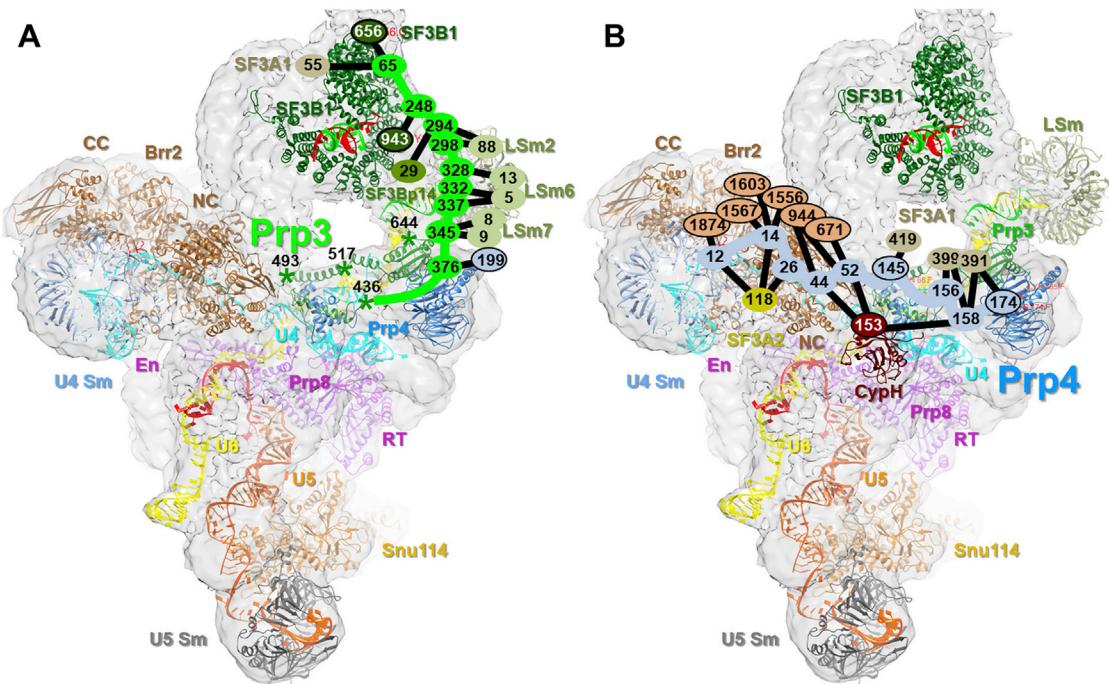


Figure S5. Crosslinks of Prp3 and Prp4 with Other Proteins in the Human B Complex, Related to Figure 3

(A and B) Overview of the position of selected proteins and RNA in the B complex EM density map (shown is the front view of the B complex in both panels; see Figure 1A, left panel), with a schematic diagram showing intermolecular crosslinks between Prp3 (A) or Prp4 (B) and other proteins, as indicated. The likely paths of both proteins in the B complex are indicated by a thick line. Numbers indicate the positions of crosslinked lysine residues (connected by black lines) in each protein. Numbers in ovals with black borders indicate the residues in the modeled regions of the proteins, whereas those in ovals without borders are residues within non-modeled regions. The latter are arbitrarily placed close (less than 30 Å) to their crosslinking partners observed in our model. The ovals share the same color as that of the crosslinked protein's name. Black numbers with asterisks (*) in (A) represent residues of Prp3 α helices that are modeled in the high resolution B complex structure.

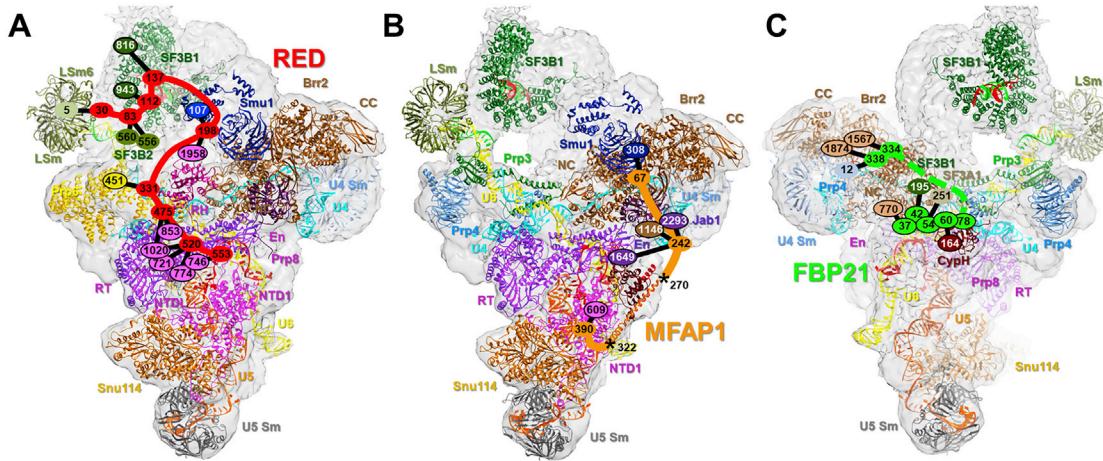


Figure S6. Crosslinks of RED, MFAP1, and FBP21 with Other Proteins in the Human B Complex, Related to Figure 5

(A–C) Overview of the position of selected proteins and RNA in the B complex EM density map, with a schematic diagram showing intermolecular crosslinks between RED (panel A), MFAP1 (panel B), FBP21 (panel C) and other indicated proteins, and the likely paths of RED, MFAP1 and FBP21 (thick lines), respectively, in the B complex. The back view of the B complex is shown in panels A and B (see Figure 1A, right panel), whereas the front view is shown in panel C (see Figure 1, left panel). Numbers indicate the positions of crosslinked lysine residues (connected by black lines) in each protein. Numbers in ovals with black borders indicate the residues in the modeled regions of the proteins, whereas those in ovals without borders are residues within non-modeled regions. The latter are arbitrarily placed close (less than 30 Å) to their crosslinking partners observed in our model. The ovals share the same color as that of the crosslinked protein's name.

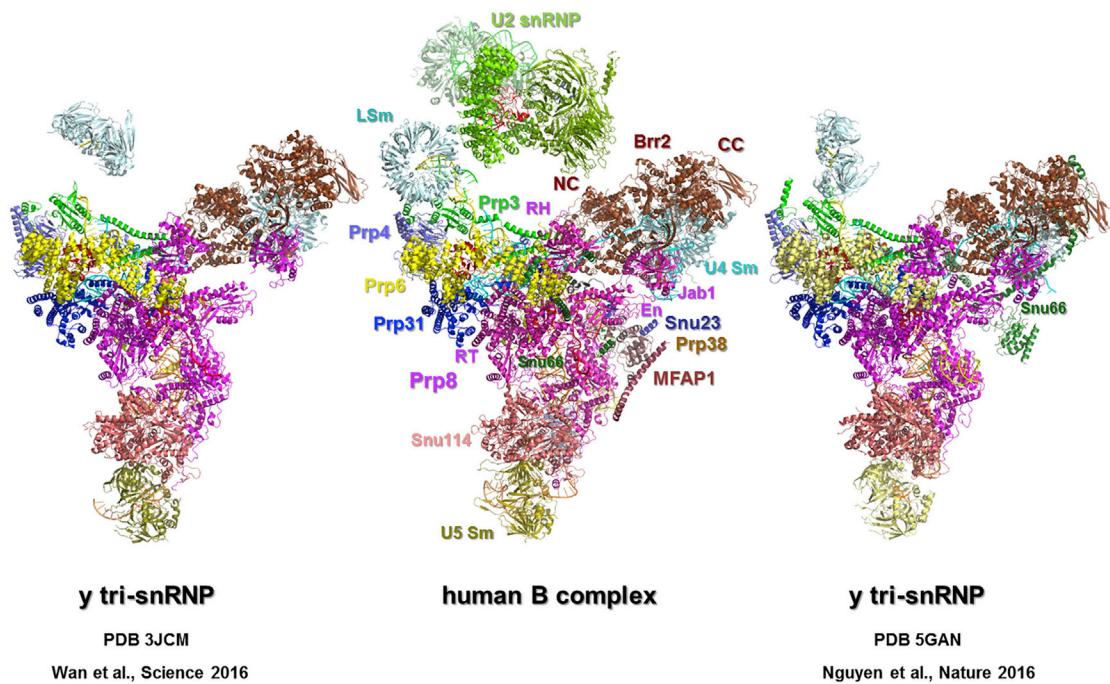


Figure S7. Comparison of the Molecular Architecture of the *S. cerevisiae* U4/U6.U5 Tri-snRNP and Human B Complex, Related to Figure 7

Overview of the structure and organization of proteins in the purified *S. cerevisiae* tri-snRNP as reported by Wan et al. (2016b) (left) or Nguyen et al. (2016) (right), and in the human spliceosomal B complex (middle) as determined by cryo-EM. All complexes are aligned relative to Snu114 and the Prp8 NTD1 domain. The coloring of proteins (labeled in the middle panel) is conserved in all three panels. The structural organization of Prp8, Brr2, Prp6 and the U4/U6 proteins in the yeast tri-snRNP is highly similar to their organization in the human B complex, consistent with the idea that the purified yeast tri-snRNPs are either activated at an early stage, i.e., before they join the spliceosome to form the B complex, or that they might potentially be derived by dissociation of spliceosomal B complexes. In addition, unlike human tri-snRNPs, purified yeast tri-snRNPs contain Prp38 and Snu23, but lack Prp28. Interestingly, one of the cryo-EM structures of a yeast tri-snRNP exhibits density elements, into which the yeast Prp38 NTD in complex with α helices of Snu23 and Spp381 (the yeast likely homolog of MFAP1) would fit nicely. While this density was tentatively assigned to helical regions of yeast Snu66 by Nguyen et al. (2016), they more likely contain the yeast Prp38 complex, as they are located at the equivalent position as the hPrp38 complex in the human B complex and can be even superimposed. Indeed, in the recently published yeast B complex structure (Plaschka et al., 2017), this density was shown to comprise the yeast Prp38/Snu23/Spp381 protein complex.