ARTICLE

doi:10.1038/nature22799

Structure of a pre-catalytic spliceosome

Clemens Plaschka[1]*, Pei-Chun Lin[1]* & Kiyoshi Nagai[1]

Intron removal requires assembly of the spliceosome on precursor mRNA (pre-mRNA) and extensive remodelling to form the spliceosome's catalytic centre. Here we report the cryo-electron microscopy structure of the yeast *Saccharomyces cerevisiae* pre-catalytic B complex spliceosome at near-atomic resolution. The mobile U2 small nuclear ribonucleoprotein particle (snRNP) associates with U4/U6.U5 tri-snRNP through the U2/U6 helix II and an interface between U4/U6 di-snRNP and the U2 snRNP SF3b-containing domain, which also transiently contacts the helicase Brr2. The 3′ region of the U2 snRNP is flexibly attached to the SF3b-containing domain and protrudes over the concave surface of tri-snRNP, where the U1 snRNP may reside before its release from the pre-mRNA 5′ splice site. The U6 ACAGAGA sequence forms a hairpin that weakly tethers the 5′ splice site. The B complex proteins Prp38, Snu23 and Spp381 bind the Prp8 N-terminal domain and stabilize U6 ACAGAGA stem–pre-mRNA and Brr2–U4 small nuclear RNA interactions. These results provide important insights into the events leading to active site formation.

Splicing of pre-mRNA is an important step in eukaryotic gene expression, in which non-coding introns are removed from newly synthesized pre-mRNA by two step-wise transesterification reactions. Intron excision is catalysed by the spliceosome, a large and dynamic ribonucleoprotein particle (RNP) that assembles on each intron[1]. First, U1 and U2 snRNPs recognize the pre-mRNA 5′ splice site (5′SS) and the branch-point sequence, forming the A complex, which associates with U4/U6.U5 tri-snRNP to form the pre-B complex[2]. The 5′SS–U1 snRNP interaction is then disrupted by the helicase Prp28 (ref. 3), and the 5′SS is transferred to the U6 small nuclear RNA (snRNA) ACAGAGA region[4], forming the pre-catalytic B complex. Subsequent spliceosome activation requires the proteins Spp381, Prp38 and Snu23 (refs 5, 6) and is initiated by ATP-dependent unwinding of the U4/U6 duplex by the helicase Brr2 (refs 7, 8). During this process, U4 snRNA and U4/U6 di-snRNP proteins dissociate[9,10] and nineteen complex (NTC) and nineteen complex-related (NTR)[9,11,12] proteins are recruited. U6 snRNA, freed from U4 snRNA, forms an internal stem–loop and the U2/U6 helix I with U2 snRNA[13], yielding the catalytic RNA centre that harbours the catalytic metal ions[14,15]. This activated spliceosome, B^{act}, is then converted to B*, positioning the branch-point adenosine at the 5′SS for step 1 of splicing (branching). The resultant C complex is remodelled to C*, bringing the 5′ and 3′ exons into proximity for step 2 of splicing (exon ligation)[16,17]. Ligated mRNA is then released and the intron-lariat spliceosome is disassembled to recycle snRNPs for further splicing rounds.

Recent cryo-electron microscopy (cryo-EM) studies of yeast[18–23] and human spliceosomes[24] have provided unprecedented insights into the catalytic stages of splicing. The group II intron-like catalytic RNA centre is fully formed in B^{act} (refs 20, 21), with the pre-mRNA 5′SS positioned near the catalytic metals and the branch-point adenosine located 50 Å away. The C and C* complex structures[18,19,22–24] revealed step I and II conformations and the functions of step-specific factors. By contrast, the molecular details before the catalytic stages are less well understood. Although the structures of isolated tri-snRNP[25–27] and several U2 snRNP parts are available[21,28–30], low-resolution structures of complete yeast[9,31] and human[32,33] pre-catalytic spliceosomes provided only limited insights into their assembly and activation.

Here we report the cryo-EM structure of the pre-catalytic B complex from the yeast *Saccharomyces cerevisiae* at near-atomic resolution, revealing the complete U2 snRNP architecture and its interactions with the U4/U6.U5 tri-snRNP. Comparison with the yeast B^{act} structure[20,21] provides crucial insights into the mechanism of spliceosome activation.

Overall architecture and U2 snRNP

We assembled spliceosomes on *UBC4* pre-mRNA and stalled at the B complex stage by limiting the ATP concentration[9,34] (Methods). Spliceosomes were imaged under cryogenic conditions yielding 496,581 single particle images (Extended Data Fig. 1d). Several rounds of unsupervised 2D and 3D particle sorting, focused refinement, and signal subtraction produced cryo-EM densities for the complete B complex structure (Fig. 1, Extended Data Figs 1e, f, 2, 3, Methods) ranging from 3.6 to 17.2 Å resolution (maps B1–B7). In the B complex, the U2 snRNP resides between the tri-snRNP 'head' and 'arm' regions[26,27], in agreement with low-resolution studies of yeast and human complexes[31,32,35] (Fig. 1, Supplementary Information). The B complex has a concave

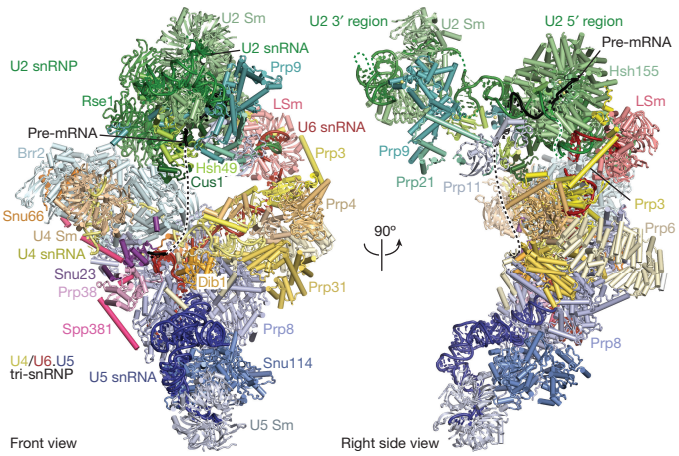


Figure 1 | B complex structure at near-atomic resolution. Two orthogonal views of the B complex structure. Subunits are coloured according to snRNP identity (U2, green; U4, yellow; U5, blue; U6, red; Dib1, Prp6 and Snu66, shades of yellow). B complex proteins are coloured in shades of magenta (Spp381, magenta; Prp38, light magenta; Snu23, violet).

[1]MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK.
*These authors contributed equally to this work.

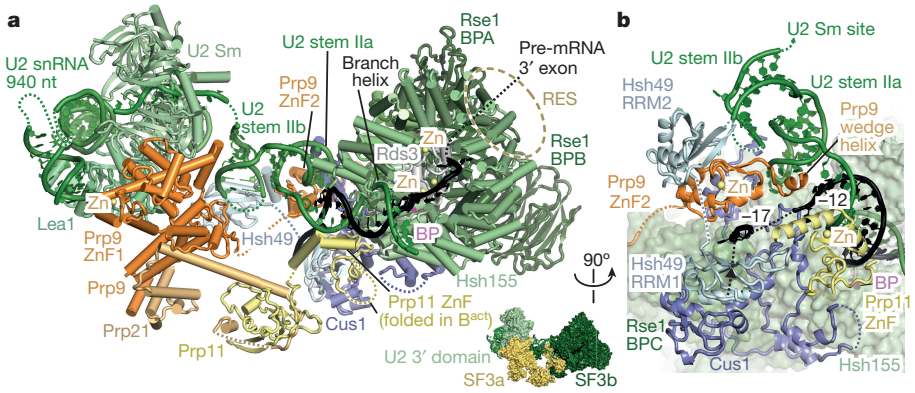


Figure 2 | U2 snRNP architecture and interactions with the intron. **a**, Organization of the U2 snRNP with subunits coloured as in Fig. 1, except for Prp9 (orange), Prp11 (yellow), Prp21 (light orange), Cus1 (dark blue), Hsh49 (light blue), Rds3 (grey) and the branch-point (BP) adenosine (magenta). The thumbnail highlights the U2 snRNP subcomplexes. The RES complex location (dashed brown ellipse) is indicated based on its B^{act} location[20,21]. The non-essential 940-nucleotide (nt) insertion in yeast U2 snRNA[50] is disordered. A loop in the Prp11 Zn-finger (ZnF) domain is disordered in B complex, but folds over the pre-mRNA 5′SS in B^{act} (ref. 21). **b**, SF3a/b proteins chaperone U2 snRNA elements and the pre-mRNA intron. The SF3a Prp9 subunit wedge helix separates the intron from U2 snRNA, guiding it towards Hsh49 RRM1. Pre-mRNA nucleotides at position −12 and −17 from the branch-point adenosine and its 3′ direction (black arrow) are indicated. Colours as in **a**.

shape, formed in part by the U2 snRNP that protrudes over the tri-snRNP region, where the catalytic centre will be created. The yeast U2 snRNP contains 18 proteins, divided into 7 Sm proteins, the Msl1–Lea1 heterodimer, and the SF3a (Prp9, Prp11 and Prp21) and SF3b (Hsh155, Rse1, Cus1, Hsh49, Rds3 and Ysf3) subcomplexes (Fig. 2a, Extended Data Fig. 4). Consistent with negative-stain electron microscopy of its human counterpart[36], the yeast U2 snRNP has a bipartite structure, consisting of 5′ (or SF3b-bound) and 3′ (Sm core domain and the Msl1–Lea1 dimer) domains. The elongated SF3a subcomplex bridges the two domains, and in between U2 snRNA adopts the stem IIa/b configuration[1] (Fig. 2a, Extended Data Fig. 4c). The SF3b subcomplex consists of Rse1, comprising the three β-propeller domains BPA, BPB and BPC, and the HEAT-repeat protein Hsh155, which together envelop Ysf3 and the Zn-containing Rds3 protein[20,21,30]. The pre-mRNA branch-point sequence and U2 snRNA form the branch helix, which is wedged into the split ends of the Hsh155 HEAT repeats, as in the B^{act} structure[20,21]. The nucleotide base of the branch-point adenosine (A70 in *UBC4*) is flipped out of the branch helix and makes contact with Tyr35 of Rds3 and a pocket formed by Hsh155 HEAT repeats H15–H17, as in the B^{act} structure.

The U2 3′ domain and SF3a subcomplexes associate via contacts between the Prp9 Zn-finger 1 (ZnF1) and U2 Sm ring subunits SmD1 and SmD2. SF3a bridges the 5′ and 3′ domains[36] via the conserved Prp9 ZnF2 and Prp11 ZnF, which are anchored largely by SF3b subunit Cus1, U2 snRNA stem IIa/b, and the intron (Fig. 2b). This explains why Prp9 ZnF2 and Prp11 ZnF are required for integration of SF3a into the human U2 snRNP and for A complex assembly[37]. The intron upstream of the branch helix is chaperoned by the Prp11 ZnF and a conserved α-helix in Prp9 (residues 407–415) that we term the 'wedge helix' (Fig. 2b). The wedge helix binds U2 stem IIa, and separates the intron (position −12 from the branch point) from U2 snRNA, guiding the intron towards its binding site on the SF3b Hsh49 RNA recognition motif 1 (RRM1) (around position −17). Hsh49 RRM1 connects to RRM2, which rests on top of the Prp9 ZnF2 domain and binds U2 stem IIb. Thus, both Hsh49 RNA recognition motif domains bind to RNA, and mutation of these interfaces is lethal in yeast[38]. The SF3b protein Cus1 forms an extended polypeptide that folds onto Hsh155 HEAT repeats H16–H20 and the Rse1 BPC domain. Anchored at this site, Cus1 (residues 290–353) forms a globular domain with the Hsh49 RRM1 (ref. 28), and nearby Cus1 regions bind the Prp9 ZnF2, the Prp11 ZnF, and U2 stem IIa (Fig. 2b). Cus1 thereby scaffolds key U2 snRNA and pre-mRNA interactions. The structure thus reveals how SF3a and SF3b form an intricate network of protein–RNA interactions to chaperone U2 snRNA and the intron[38–41].

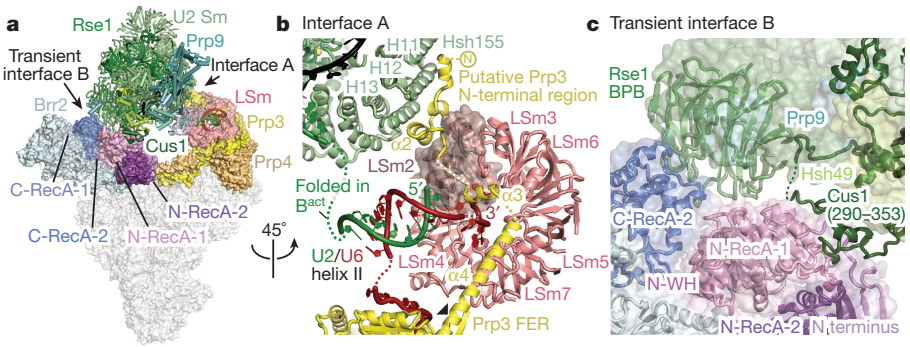


Figure 3 | U2 snRNP and tri-snRNP interfaces. **a**, Overview of the B complex structure, showing interface A and transient interface B. The U2 snRNP (coloured as in Fig. 1) was positioned relative to tri-snRNP using a subset of particles (map B2) (see Methods). Tri-snRNP is shown as a surface (grey), except for Prp3 (yellow), Prp4 (light orange), the LSm ring (salmon) and Brr2 (pale cyan except for N-terminal and C-terminal helicase cassette RecA-1 and RecA-2 lobes (N-RecA-1, C-RecA-1, N-RecA-2 and C-Rec-A-2, respectively), shown in shades of blue and violet). **b**, Interface A. The U2 snRNP binds tri-snRNP via U2/U6 helix II and SF3b subunit Hsh155 HEAT repeats H11–H13, which bind the putative Prp3 N-terminal region. The U2 snRNA nucleotides that form the active site U2/U6 helix I in B^{act} are disordered (dashed green line). The black arrowhead indicates the region of Prp3 helix α4 that bends with different U2 snRNP positions. FER, ferredoxin-like fold. See Extended Data Fig. 5. **c**, Transient interface B. SF3b subunit Cus1 and the Rse1 BPB contact Brr2 in a subset of cryo-EM particles (see Extended Data Fig. 5a; Methods). Colours as in **a**. N-WH, winged helix domain of the Brr2 NHC.

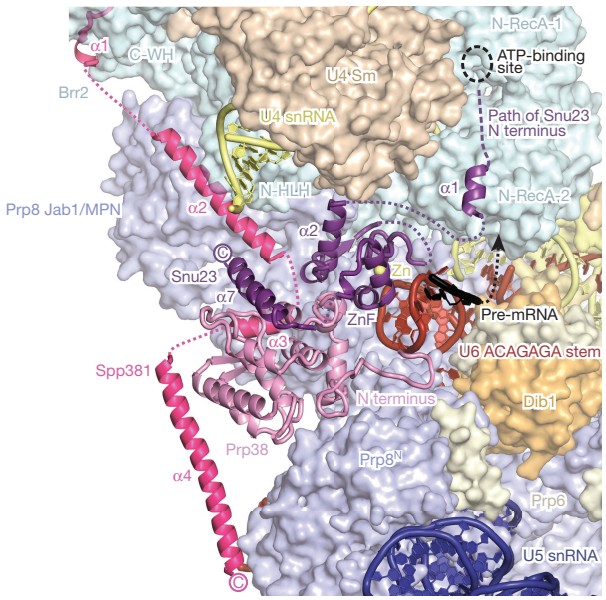


Figure 4 | B complex proteins stabilize Brr2 and U6 snRNA. Protein interactions of the B complex proteins within the pre-catalytic B complex. The extended structures of the B complex proteins Spp381 (magenta), Prp38 (light magenta) and Snu23 (violet) bind the Prp8 N-terminal ($Prp8^N$) domain and contact Prp8 Jab1/MPN, the winged helix domain of the Brr2 CHC (C-WH), the helix–loop–helix of the Brr2 NHC (N-HLH), and the RecA-2 lobe of the Brr2 NHC (N-RecA-2) to position Brr2 on U4 snRNA. The Snu23 ZnF and Prp38 N terminus stabilize the U6 ACAGAGA stem, which may enable tethering of the putative pre-mRNA 5′SS at its tip. The pre-mRNA 3′ direction is indicated (black arrow).

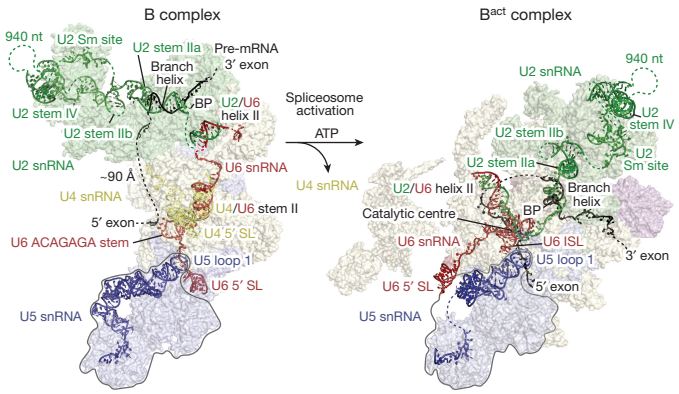


Figure 5 | Changes in the RNA network during spliceosome activation. B (left) and B^{act} (right) complex RNA models are superimposed on transparent surfaces of spliceosome proteins. Proteins are coloured as in Fig. 1, and the two structures are aligned with their U5 snRNP foot domain (black outline). The U2 3′ domain and SF3a proteins are modelled based on a low-pass filtered B^{act} cryo-EM density[20] (EMDB code EMD-4099; see Methods) onto the B^{act} model[21] (PDB accession 5GM6). ISL, internal stem–loop; SL, stem–loop.

B complex stabilization

The U2 snRNP 5′ domain binds to tri-snRNP through two interfaces (A and B; Fig. 3a, Extended Data Fig. 5a). In interface A, the SF3b subunit Hsh155 binds to the U4/U6 di-snRNP protein Prp3 (Fig. 3b). A long α-helix of Prp3 (α4) extends from the Prp4 β-propeller domain towards the U6 LSm ring (LSm5 and LSm7 subunits), reaches across the LSm ring, and inserts its putative N-terminal helices between the LSm ring subunit LSm2 and the Hsh155 HEAT repeats H11–H13 (Fig. 3a, b). Interface A further involves U2/U6 helix II (ref. 13), comprising the 3′ end of U6 snRNA and the 5′ end of U2 snRNA. U6 snRNA thereby tethers U2 and U5 snRNPs during spliceosome activation, when U4/U6 duplex unwinding and additional remodelling occurs. Importantly, the U2 snRNA nucleotides that form the active site U2/U6 helix I in B^{act} remain disordered in B complex and available for pairing with U6 snRNA during activation.

Interface B is transient and found only in a subgroup of cryo-EM particles (map B2) (Fig. 3a, c, Extended Data Figs 2, 5a, Methods). This interface involves weak protein contacts between Cus1 (residues 328–353) and the RecA-1 lobe of the N-terminal helicase cassette (NHC) of Brr2 as well as between the Rse1 BPB and RecA-1 of the NHC and RecA-2 of the C-terminal helicase cassette (CHC) of Brr2. The mobility of interface B may be important to allow SF3b and Brr2 to move independently during spliceosome activation.

B complex protein interactions

Spp381, Prp38 and Snu23 are required for spliceosome activation and are important for viability in yeast[5,42,43]. These proteins are specific to the B complex in humans[44], whereas they associate with isolated U4/U6.U5 tri-snRNP in yeast[6,43,45], and are hereafter called B complex proteins. We obtained a reconstruction of the mobile B complex proteins with a local resolution of 4.0–5.0 Å from a subset of cryo-EM particles by focused classification (map B7) (Fig. 4, Extended Data Figs 1f, g, 2 and 6). Previously, this density was either not interpreted[26] or tentatively attributed to part of Snu66 (ref. 27). Prp38 scaffolds Spp381 and Snu23 (ref. 46), and anchors these proteins at the Prp8 N-terminal domain ($Prp8^N$), consistent with biochemical data[26,31] (Fig. 4). The Spp381 N-terminal region binds the winged helix domain of the Brr2 CHC, and continues between the Prp8 Jab1/MPN domain and the helix–loop–helix domain of the Brr2 NHC with its long helix α2 (Fig. 4). The Spp381 C-terminal region then binds along the globular Prp38 domain as in its human counterpart[46] and extends towards $Prp8^N$. This explains why deletion of the Spp381 N-terminal region impairs splicing, and why Spp381 can rescue the *prp38-1* temperature-sensitive mutation[5]. Snu23 contains a conserved ZnF domain that binds Prp38 and the major groove of the U6 snRNA ACAGAGA stem to stabilize it (Fig. 4). The Snu23 N terminus binds the RecA-2 lobe of the Brr2 NHC and reaches near the NHC RecA-1 ATP-binding pocket. Snu23 may thereby stabilize Brr2 on the U4 snRNA substrate and influence Brr2 helicase activity (Extended Data Fig. 6f). Thus, Spp381 and Snu23 together, held by Prp38, aid in positioning Brr2 to one side of the tri-snRNP. Consistent with the structure, several Prp8 mutations that suppress the U4-cs1 mutant[47], which blocks U4/U6 unwinding, map near Prp8–B complex protein interfaces.

In B complex, the 5′ stem–loop of U6 snRNA binds the $Prp8^N$ domain and the adjacent U6 nucleotides base pair with U5 loop 1. Consistent with this, U5 snRNA loop 1 nucleotides crosslink to a wide region surrounding the pre-mRNA 5′SS in B complex, suggesting that the 5′SS is in the vicinity of U5 loop 1 but does not base pair with it[48]. Pre-mRNA also does not crosslink to U5 loop 1 in the purified human B complex[2]. U6 snRNA continues past U5 loop 1 to form the ACAGAGA stem (map B7 and ref. 27) and is stabilized by the Snu23 ZnF domain, the extended Prp38 N terminus, and tri-snRNP proteins Dib1, Prp6 and Prp8 (Fig. 4, Extended Data Fig. 6g). Destabilization of these interactions by mutations of the Prp38–Snu23 ZnF[5] interface and the Snu23 ZnF domain[43] causes growth defects in yeast. The base of the U6 snRNA ACAGAGA stem is further enclosed by the previously disordered Prp8 residues 2121–2140 (refs 26, 27) that form a β-hairpin, to which Snu23 adds a β-strand. Weak density (maps B2 and B7) connects from the tip of the U6 ACAGAGA stem to the SF3b-bound intron (Extended Data Fig. 6e), approximately 90 Å apart, and RNA crosslinking[2,48] suggests that this is pre-mRNA near the 5′SS. Taken together, the structure indicates that the extensive interactions involving the B complex proteins stabilize Brr2 on U4 snRNA and the weak tethering of the putative 5′SS to U6 snRNA before spliceosome activation.

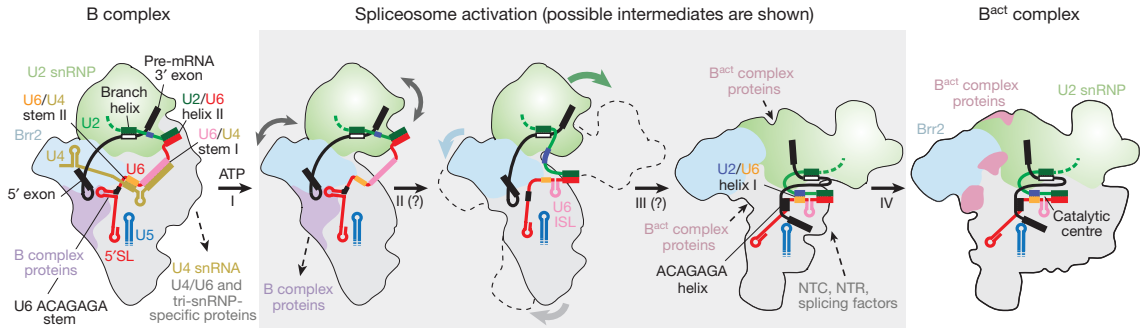


Figure 6 | Model for spliceosome activation. The ATP-dependent activity of Brr2 results in release of U4 snRNA, U4/U6 and tri-snRNP-specific proteins (transition I), followed by U6 internal stem–loop folding and U6 ACAGAGA stem unfolding (II), formation of the U2/U6 helix I, the ACAGAGA helix, the 5′-exon–U5 loop I base pairs (III), and the binding of the NTC, NTR and B^{act} complex proteins (IV). Proposed activation intermediates are shown in a grey box. See also Extended Data Figs 7d and 8.

B complex assembly

Comparisons of yeast B complex and negative-stain EM projections of human B complex[2,32] reveal a conserved architecture, including the location of the Brr2 helicase (referred to as the 'stump' domain in refs 2, 31), suggesting that events before B complex formation may also be conserved. In the human system, an ATPase-deficient mutant of PRP28 accumulates the pre-B complex[2], where U1 snRNA remains paired to the 5′SS[3] (Extended Data Fig. 7). In human U4/U6.U5 tri-snRNP, PRP28 binds the $PRP8^{N}$ and PRP8 endonuclease ($PRP8^{EN}$) domains and SAD1 tethers the BRR2 helicase distant from its U4 snRNA substrate[25]. Importantly, the Prp28-binding site is altered and overlaps with that of the B complex proteins in yeast B complex (Extended Data Fig. 6h). This suggests that Prp28 is released after disrupting the U1 snRNA–pre-mRNA interaction to allow the recruitment of B complex-specific proteins to the $Prp8^{N}$ domain, and binding and repositioning of the helicase Brr2 to its B complex location on U4 snRNA (Extended Data Fig. 6h). The U6 ACAGAGA stem is then stabilized by the B complex proteins to allow tethering of the 5′SS near the tip of the stem before activation, consistent with crosslinking[48]. Prp28 and Sad1 exist in both yeast and humans, suggesting that the conversion from pre-B to B complexes is a conserved feature of splicing. We also note that the surprisingly large space between opposite ends of the concave B complex structure may accommodate the U1 snRNP during early spliceosome assembly (Extended Data Fig. 7c), and allow for structural rearrangements during spliceosome activation.

Model of spliceosome activation

Comparison of the yeast B complex structure presented here with the yeast B^{act} structure[20,21] leads to a more detailed model for spliceosome activation (Figs 5, 6, Extended Data Figs 7d, 8). In the B complex, the helicase Brr2 is positioned by the B complex proteins and engaged with the U4 snRNA substrate. The 3′ end of U6 snRNA is held by pairing with U2 snRNA, while the 5′ stem of U6 snRNA is anchored to the $Prp8^{N}$ domain. Spliceosome activation is initiated by ATP-dependent translocation of Brr2 along the U4 snRNA. Unwinding of the U4/U6 duplex exposes the central region of U6 snRNA, which may fold spontaneously to form the U6 snRNA internal stem–loop. The disordered single-stranded region of U2 snRNA in the B complex (Fig. 3b) then pairs with U6 snRNA to form U2/U6 helix 1a and 1b (ref. 13), followed by formation of the catalytic U2/U6 triplex to complete the catalytic RNA centre[49] (Fig. 6). The loss of U4 snRNA and U4/U6 di-snRNP proteins together with formation of the catalytic centre may induce a large movement of the U2 snRNP towards the Prp8 large ($Prp8^{L}$) domain, and facilitate a new contact between the SF3b Rse1 BPB and Brr2, which also moves from its B complex location (Fig. 6). The B complex proteins may detach from Brr2 in this new position, owing to a steric clash with the $Prp8^{EN}$ domain. This uncovers the binding site for B^{act} proteins Cwc24 and Cwc27 at the $Prp8^{N}$ and $Prp8^{EN}$ domains, which stabilize the new interaction network. Our model suggests that the folding energy of the catalytic RNA core alone could mediate several of the observed remodelling events (Fig. 6) and promote the binding of stabilizing B^{act}, NTC and NTR proteins.

Before activation, U6 snRNA binds $Prp8^{N}$ with its 5′ stem, passes over U5 loop 1, and forms the U6 ACAGAGA stem, where it tethers the pre-mRNA near its 5′SS. U4/U6 unwinding together with the loss of U4/U6 di-snRNP- and tri-snRNP-specific proteins, and the destabilization of B complex proteins allow the U6 5′ stem to reposition to the opposite side of the complex (Fig. 5), where NTR proteins Bud31, Ecm2 and Cwc2 anchor it onto $Prp8^{N}$ in the B^{act} structure. These remodelling events free U5 loop 1 from U6 snRNA and together destabilize the U6 ACAGAGA stem. The tethered 5′ exon and 5′SS can now be recognized by U5 loop 1 and the exposed U6 ACAGAGA box. The 5′ exon is then loaded in the newly created exon channel, formed by rotation of the U5 foot domain and association of Cwc21 and Cwc22 (ref. 18) (Extended Data Fig. 8). A loop of SF3a subunit Prp11 folds in B^{act} and together with Cwc24 shields the 5′SS from the active site, while the SF3b-bound branch helix is immobilized 50 Å away, ready for Prp2-mediated conversion to B* and step 1 of splicing[18].

Activation in yeast, from B to B^{act} complex spliceosomes, involves the release of 24 proteins[9,31] and the recruitment of 22 others (Extended Data Fig. 8a) plus extensive rearrangements of the RNA network. Detailed structural knowledge of both B and B^{act} complexes now provides a framework to dissect the activation mechanism and to determine the precise order of molecular events leading to formation of the spliceosome active site.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 February; accepted 2 May 2017.
Published online 22 May 2017.

1. Will, C. L. & Lührmann, R. Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* **3,** a003707 (2011).
2. Boesler, C. *et al.* A spliceosome intermediate with loosely associated tri-snRNP accumulates in the absence of Prp28 ATPase activity. *Nat. Commun.* **7,** 11997 (2016).
3. Staley, J. P. & Guthrie, C. An RNA switch at the 5′ splice site requires ATP and the DEAD box protein Prp28p. *Mol. Cell* **3,** 55–64 (1999).
4. Lesser, C. F. & Guthrie, C. Mutations in U6 snRNA that alter splice site specificity: implications for the active site. *Science* **262,** 1982–1988 (1993).
5. Lybarger, S. *et al.* Elevated levels of a U4/U6.U5 snRNP-associated protein, Spp381p, rescue a mutant defective in spliceosome maturation. *Mol. Cell. Biol.* **19,** 577–584 (1999).
6. Stevens, S. W. & Abelson, J. Purification of the yeast U4/U6.U5 small nuclear ribonucleoprotein particle and identification of its proteins. *Proc. Natl Acad. Sci. USA* **96,** 7226–7231 (1999).
7. Laggerbauer, B., Achsel, T. & Lührmann, R. The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplices *in vitro*. *Proc. Natl Acad. Sci. USA* **95,** 4188–4192 (1998).
8. Raghunathan, P. L. & Guthrie, C. RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Curr. Biol.* **8,** 847–855 (1998).

9. Fabrizio, P. *et al.* The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol. Cell* **36,** 593–608 (2009).
10. Bessonov, S. *et al.* Characterization of purified human B^{act} spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. *RNA* **16,** 2384–2403 (2010).
11. Tarn, W. Y. *et al.* Functional association of essential splicing factor(s) with PRP19 in a protein complex. *EMBO J.* **13,** 2421–2431 (1994).
12. Hoskins, A. A., Rodgers, M. L., Friedman, L. J., Gelles, J. & Moore, M. J. Single molecule analysis reveals reversible and irreversible steps during spliceosome activation. *eLife* **5,** e14166 (2016).
13. Madhani, H. D. & Guthrie, C. A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell* **71,** 803–817 (1992).
14. Steitz, T. A. & Steitz, J. A. A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl Acad. Sci. USA* **90,** 6498–6502 (1993).
15. Fica, S. M. *et al.* RNA catalyses nuclear pre-mRNA splicing. *Nature* **503,** 229–234 (2013).
16. Newman, A. J. & Norman, C. U5 snRNA interacts with exon sequences at 5′ and 3′ splice sites. *Cell* **68,** 743–754 (1992).
17. Sontheimer, E. J. & Steitz, J. A. The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science* **262,** 1989–1996 (1993).
18. Galej, W. P. *et al.* Cryo-EM structure of the spliceosome immediately after branching. *Nature* **537,** 197–201 (2016).
19. Wan, R., Yan, C., Bai, R., Huang, G. & Shi, Y. Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution. *Science* **353,** 895–904 (2016).
20. Rauhut, R. *et al.* Molecular architecture of the *Saccharomyces cerevisiae* activated spliceosome. *Science* **353,** 1399–1405 (2016).
21. Yan, C., Wan, R., Bai, R., Huang, G. & Shi, Y. Structure of a yeast activated spliceosome at 3.5 Å resolution. *Science* **353,** 904–911 (2016).
22. Yan, C., Wan, R., Bai, R., Huang, G. & Shi, Y. Structure of a yeast step II catalytically activated spliceosome. *Science* **355,** 149–155 (2017).
23. Fica, S. M. *et al.* Structure of a spliceosome remodelled for exon ligation. *Nature* **542,** 377–380 (2017).
24. Bertram, K. *et al.* Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature* **542,** 318–323 (2017).
25. Agafonov, D. E. *et al.* Molecular architecture of the human U4/U6.U5 tri-snRNP. *Science* **351,** 1416–1420 (2016).
26. Wan, R. *et al.* The 3.8 Å structure of the U4/U6.U5 tri-snRNP: insights into spliceosome assembly and catalysis. *Science* **351,** 466–475 (2016).
27. Nguyen, T. H. D. *et al.* Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. *Nature* **530,** 298–302 (2016).
28. van Roon, A.-M. M. *et al.* Crystal structure of U2 snRNP SF3b components: Hsh49p in complex with Cus1p-binding domain. *RNA* **23,** 968–981 (2017).
29. Lin, P.-C. & Xu, R.-M. Structure and assembly of the SF3a splicing factor complex of U2 snRNP. *EMBO J.* **31,** 1579–1590 (2012).
30. Cretu, C. *et al.* Molecular architecture of SF3b and structural consequences of its cancer-related mutations. *Mol. Cell* **64,** 307–319 (2016).
31. Rigo, N., Sun, C., Fabrizio, P., Kastner, B. & Lührmann, R. Protein localisation by electron microscopy reveals the architecture of the yeast spliceosomal B complex. *EMBO J.* **34,** 3059–3073 (2015).
32. Boehringer, D. *et al.* Three-dimensional structure of a pre-catalytic human spliceosomal complex B. *Nat. Struct. Mol. Biol.* **11,** 463–468 (2004).
33. Deckert, J. *et al.* Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol. Cell. Biol.* **26,** 5528–5543 (2006).
34. Tarn, W.-Y., Lee, K.-R. & Cheng, S. C. Yeast precursor mRNA processing protein PRP19 associates with the spliceosome concomitant with or just after dissociation of U4 small nuclear RNA. *Proc. Natl Acad. Sci. USA* **90,** 10821–10825 (1993).
35. Wolf, E. *et al.* Exon, intron and splice site locations in the spliceosomal B complex. *EMBO J.* **28,** 2283–2292 (2009).
36. Krämer, A., Grüter, P., Gröning, K. & Kastner, B. Combined biochemical and electron microscopic analyses reveal the architecture of the mammalian U2 snRNP. *J. Cell Biol.* **145,** 1355–1368 (1999).
37. Nesic, D. & Krämer, A. Domains in human splicing factors SF3a60 and SF3a66 required for binding to SF3a120, assembly of the 17S U2 snRNP, and prespliceosome formation. *Mol. Cell. Biol.* **21,** 6406–6417 (2001).
38. Igel, H., Wells, S., Perriman, R. & Ares, M. Jr. Conservation of structure and subunit interactions in yeast homologues of splicing factor 3b (SF3b) subunits. *RNA* **4,** 1–10 (1998).
39. Pauling, M. H., McPheeters, D. S. & Ares, M. Jr. Functional Cus1p is found with Hsh155p in a multiprotein splicing factor associated with U2 snRNA. *Mol. Cell. Biol.* **20,** 2176–2185 (2000).
40. Dybkov, O. *et al.* U2 snRNA-protein contacts in purified human 17S U2 snRNPs and in spliceosomal A and B complexes. *Mol. Cell. Biol.* **26,** 2803–2816 (2006).
41. Schneider, C. *et al.* Dynamic contacts of U2, RES, Cwc25, Prp8 and Prp45 proteins with the pre-mRNA branch-site and 3′ splice site during catalytic activation and step 1 catalysis in yeast spliceosomes. *PLoS Genet.* **11,** e1005539 (2015).
42. Xie, J., Beickman, K., Otte, E. & Rymond, B. C. Progression through the spliceosome cycle requires Prp38p function for U4/U6 snRNA dissociation. *EMBO J.* **17,** 2938–2946 (1998).
43. Stevens, S. W. *et al.* Biochemical and genetic analyses of the U5, U6, and U4/U6 x U5 small nuclear ribonucleoproteins from *Saccharomyces cerevisiae*. *RNA* **7,** 1543–1553 (2001).
44. Agafonov, D. E. *et al.* Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. *Mol. Cell. Biol.* **31,** 2667–2682 (2011).
45. Gottschalk, A. *et al.* Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6·U5] tri-snRNP. *EMBO J.* **18,** 4535–4548 (1999).
46. Ulrich, A. K. C., Seeger, M., Schütze, T., Bartlick, N. & Wahl, M. C. Scaffolding in the spliceosome via single α helices. *Structure* **24,** 1972–1983 (2016).
47. Kuhn, A. N., Li, Z. & Brow, D. A. Splicing factor Prp8 governs U4/U6 RNA unwinding during activation of the spliceosome. *Mol. Cell* **3,** 65–75 (1999).
48. Chan, S.-P. & Cheng, S.-C. The Prp19-associated complex is required for specifying interactions of U5 and U6 with pre-mRNA during spliceosome activation. *J. Biol. Chem.* **280,** 31190–31199 (2005).
49. Fica, S. M., Mefford, M. A., Piccirilli, J. A. & Staley, J. P. Evidence for a group II intron-like catalytic triplex in the spliceosome. *Nat. Struct. Mol. Biol.* **21,** 464–471 (2014).
50. Shuster, E. O. & Guthrie, C. Two conserved domains of yeast U2 snRNA are separated by 945 nonessential nucleotides. *Cell* **55,** 41–48 (1988).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. Savva, S. Chen, G. Cannone, K. R. Vinothkumar, G. McMullan, J. Grimmett and T. Darling for maintaining electron microscopy and computing facilities; the mass spectrometry facility for protein identification, and A. Newman, L. Strittmatter, S. Fica, M. Wilkinson and C. Norman for help and critical reading of the manuscript. We thank J. Löwe, V. Ramakrishnan, D. Barford and R. Henderson for their continuing support. The project was supported by the Medical Research Council (MC_U105184330) and European Research Council Advanced Grant (693087 - SPLICE3D). C.P. was supported by an EMBO Long-Term Fellowship.

Author Contributions C.P. established complex preparation. C.P. and P.-C.L. performed cryo-EM structure determination and model building. C.P. refined the model. C.P., P.-C.L. and K.N. analysed the structure and wrote the manuscript. K.N. initiated and supervised the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to C.P. (cplasch@mrc-lmb.cam.ac.uk), P.-C.L. (pclin@mrc-lmb.cam.ac.uk) or K.N. (kn@mrc-lmb.cam.ac.uk).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

Preparation and purification of B complex. To prepare B complex spliceosomes for structural study, we grew yeast *Saccharomyces cerevisiae* containing a TAPS affinity tag on Brr2 (ref. 51) in a 120 l fermenter. Splicing extract was prepared using the liquid nitrogen method essentially as described[52], with an additional 18 h dialysis against buffer A (20 mM HEPES, pH 7.9, 50 mM KCl, 20% glycerol, 0.2 mM EDTA, 0.5 mM DTT), to deplete ATP. *UBC4* pre-mRNA[53] containing two MS2 stem–loops at the 5′-end was made by *in vitro* transcription. The RNA product was subsequently labelled with fluorescein[54] at its 3′ end to monitor complex purification. *In vitro* splicing reactions were assembled from pre-mRNA substrate pre-bound to MS2–MBP fusion protein as described[55]. Splicing reactions contained 50 μM ATP to stall splicing at the B complex stage as described[9,34], and proceeded for 50 min at 23 °C. The reaction mixture was centrifuged through a 40% glycerol cushion in buffer B (20 mM HEPES, pH 7.9, 50 mM KCl, 1% glycerol, 0.2 mM EDTA, 1 mM DTT, 0.04% NP-40). The cushion was collected, diluted with buffer B, and applied to amylose resin (NEB) pre-washed with buffer C (20 mM HEPES, pH 7.9, 75 mM KCl, 5% glycerol, 0.2 mM EDTA, 1 mM DTT, 0.04% NP-40). After 12 h incubation at 4 °C, the resin was washed with buffer C and eluted in buffer C containing 50 mM KCl and 12 mM maltose. Fractions containing spliceosomes were pooled and applied to Strep-Tactin resin (GE Healthcare), pre-washed with buffer C, and incubated for 4 h at 4 °C. The resin was washed with buffer C, and the sample was eluted with buffer D (20 mM HEPES, pH 7.9, 50 mM KCl, 0.2 mM EDTA, 1 mM DTT, 2.5 mM desthiobiotin). Fractions containing spliceosomes were pooled and crosslinked using 1 mM BS3 (Sigma) on ice for 1 h, and subsequently quenched with 50 mM ammonium bicarbonate. The sample was concentrated to around 1.5 mg ml^{-1} and immediately used for EM sample preparation. The isolated complexes contained U2, U4, U5 and U6 snRNA and pre-mRNA, and all proteins previously assigned to pre-catalytic spliceosomes[9,31]. The U1 snRNP was identified in sub-stoichiometric amounts, consistent with its destabilization in B complex[2,31] (Extended Data Fig. 1a, b).

To confirm that purified B complex spliceosomes were functional, we carried out an *in vitro* splicing assay. B complex was prepared as for EM studies using labelled *UBC4* pre-mRNA, but was not crosslinked, and was flash-frozen in liquid nitrogen and stored at −80 °C until further use. *In vitro* splicing was carried out in the presence of 2 mM ATP and unlabelled 60 nM pre-mRNA competitor, inhibiting *de novo* spliceosome assembly. Purified B complex was incubated with nuclear extract for 10 min before adding ATP, and product formation was visualized after 50 min of splicing at 23 °C on a denaturing 18% polyacrylamide TBE gel with a Typhoon scanner (GE Healthcare) (Extended Data Fig. 1c).

Electron microscopy. B complex spliceosomes were first studied by negative-stain EM. Copper grids (Quantifoil) were coated with a ~10-nm homemade carbon film and glow-discharged 20 s before deposition of 3.5 μl sample (0.15 mg ml^{-1}). The sample was incubated for 1 min and grids were blotted with three 3.5 μl drops of distilled water. The sample was stained in a 3.5 μl drop of saturated uranyl formate solution for 1 min, and blotted until dry. 409 images were recorded using a FEI Spirit microscope operated at 120 kV and a 2k × 2k CCD camera with a defocus of approximately 1.5 μm and a nominal magnification of 26,000 (3.84 Å pixel^{-1}).

For cryo-EM analysis the sample was applied to R1.2/1.3 holey carbon grids (Quantifoil) and coated with a 5–7 nm homemade carbon film. Grids were glow-discharged for 15 s before deposition of 3 μl sample (approximately 1.5 mg ml^{-1}), and subsequently incubated for 2–3.5 s before blotting and vitrification by plunging into liquid ethane with a Vitrobot Mark III (FEI) operated at 4 °C and 100% humidity. Cryo-EM data was acquired on a FEI Titan Krios operated in EFTEM mode at 300 keV, equipped with a K2 Summit direct detector (Gatan) and a GIF Quantum energy filter (slit width of 20 eV, Gatan). Data acquisition was carried out with Digital Micrograph (Gatan) to record 5,115 movies with a defocus range of −0.35 μm to −5.3 μm at a nominal magnification of 81,000 (1.43 Å pixel^{-1}). The camera was operated in 'super-resolution' mode (0.715 Å pixel^{-1}) with a total exposure time of 16 s fractionated into 20 frames, a dose rate of approximately 1.25 e$^-$ pixel^{-1} s^{-1}, and a total dose of 56 e$^-$ Å^{-2} per movie. Movies were binned once and aligned using MOTIONCORR[56].

Image processing. For single-particle analysis, CTF parameters were estimated using CTFFIND4 (ref. 57) and CTF correction and subsequent image processing were carried out using RELION 2.0 beta[58], unless otherwise noted. Resolution is reported based on the gold-standard Fourier shell correlation (FSC) (0.143 criterion) as described[59] and temperature factors were determined and applied automatically in RELION[58]. For negative-stain analysis, 12,229 particles were selected semi-automatically using e2boxer.py from EMAN2 (ref. 60), extracted using a 176^2 pixel box and pre-processed to normalize the images. The negative stain map of human BΔU1 (ref. 32) (EMDB code EMD-1066) was used as an initial reference for 3D refinement of the 12,229 particles set. A single round of 3D classification revealed a B complex density from 3,707 particles (Extended Data Fig. 3a) that was refined to an estimated resolution of 50 Å (Extended Data Fig. 3b). This density was low-pass filtered to 60 Å and used for processing of cryo-EM data.

For cryo-EM analysis, an initial set of 7,208 particles were selected semi-automatically using e2boxer.py[60], and extracted with a 480^2 pixel box and pre-processed to normalize the images. Reference-free 2D class averages were calculated and filtered to 30 Å resolution and used for automated picking in RELION[61] of all micrographs. The resulting particles were screened manually and by reference-free 2D classification, yielding 496,581 particles for subsequent processing. A 3D reconstruction of all particles was calculated to an overall resolution of 4 Å, and subjected to particle polishing in RELION[58], improving the resolution to 3.7 Å.

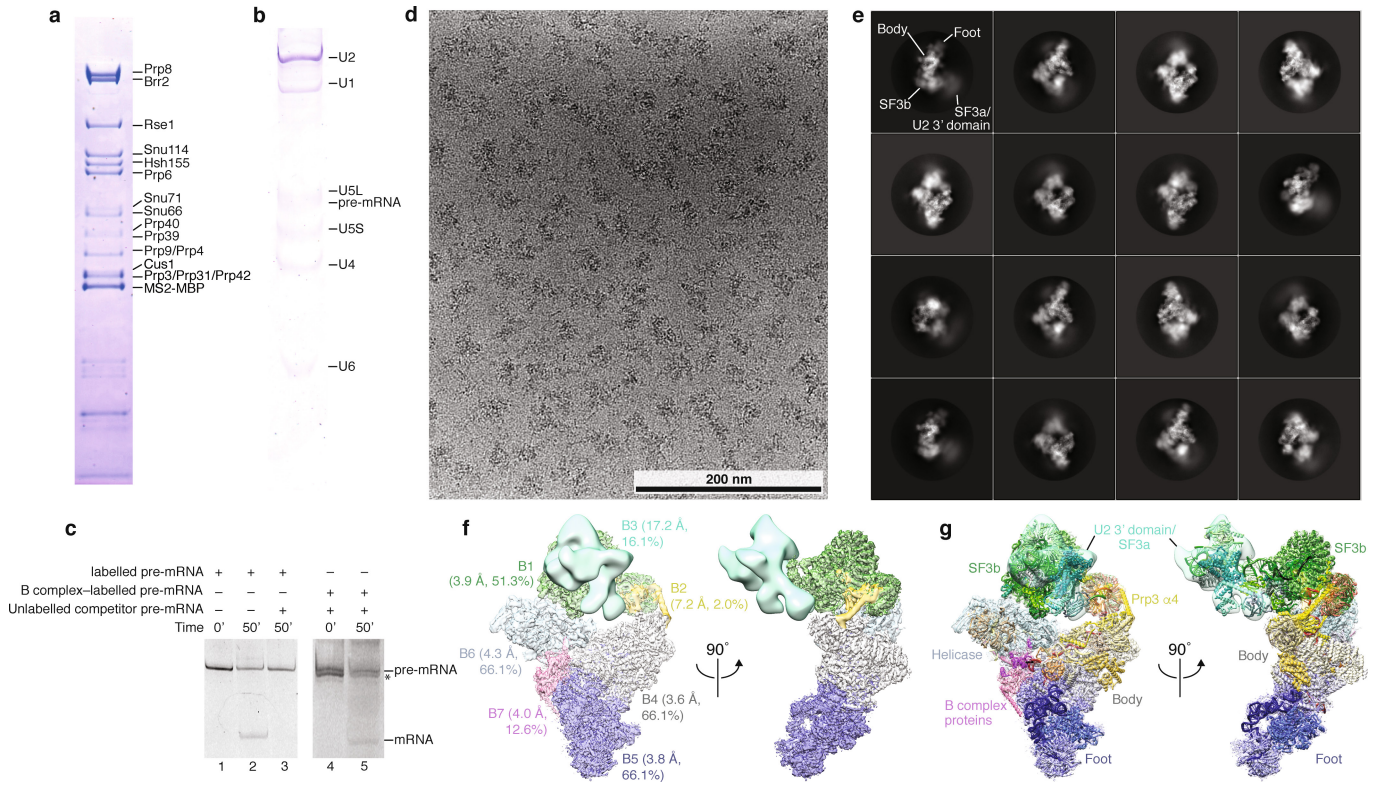
Hierarchical 3D classification was carried out without image alignment to reduce computational costs and identify homogenous particle groups (Extended Data Fig. 2). Soft masks enveloping complete B complex or smaller regions were generated using the volume eraser in UCSF Chimera[62] and RELION[58]. This included masks for tri-snRNP (excluding the helicase domain), B complex proteins, U2 snRNP SF3b and U2 3′ domain–SF3a subcomplexes. Each resultant class was refined using the 3D auto-refine procedure against the respective particles within that class with a soft reference mask in the shape of the feature of interest, yielding B1–B7 reconstructions (Extended Data Figs 1f, 2 and 3c, d). To improve the resolution of regions with weak density, we used focused refinements (B4, B5 and B7) together with signal subtraction (B1, B3 and B6) as described[63]. The B1 reconstruction (improved SF3b density) was calculated in a focused refinement from 254,095 particles from which the tri-snRNP signal was subtracted[63], and was determined to a resolution of 3.9 Å with a temperature factor of 125 Å^{-2}. The B2 reconstruction was determined from 9,559 particles to a resolution of 7.2 Å with a temperature factor of 170 Å^{-2}, and was used to determine the relative positions of U2 snRNP and tri-snRNP for the complete B complex model (see 'Structural modelling'). The B3 reconstruction (improved U2 snRNP 3′ domain–SF3a densities) was calculated in a focused refinement from 79,166 particles from which the U2 snRNP SF3b and tri-snRNP signal was subtracted[63], and was determined to a resolution of 17.2 Å with a temperature factor of 200 Å^{-2}. The low resolution of the B3 density is probably due to flexibility of the U2 3′ domain relative to SF3a. The B4 reconstruction (improved tri-snRNP body density) was calculated in a focused refinement from 327,490 particles to a resolution of 3.6 Å with a temperature factor of 105 Å^{-2}. The B5 reconstruction (improved tri-snRNP foot domain density) was calculated in a focused refinement from 327,490 particles to a resolution of 3.8 Å with a temperature factor of 127 Å^{-2}. The B6 reconstruction (improved helicase density) was calculated in a focused refinement from 327,490 particles from which the tri-snRNP signal, excluding the helicase domain, was subtracted[63], and was determined to a resolution of 4.3 Å with a temperature factor of 140 Å^{-2}. The B7 reconstruction (improved B complex protein density) was calculated in a focused refinement from 62,406 particles to a resolution of 4.0 Å with a temperature factor of 80 Å^{-2}. Local resolution estimates were determined using ResMap[64] (Extended Data Fig. 3e, f).

Structural modelling. A composite model of B complex was obtained using B1–B7 densities (Extended Data Figs 1f, g, 9, Extended Data Table 1). Model building was carried out in COOT[65]. Coordinates were refined in real space using phenix.real_space_refine in PHENIX[66] into the respective sharpened density, applying secondary structure, rotamer, nucleic acid, and metal coordination restraints, unless otherwise noted (Extended Data Fig. 9). The B^{act} SF3b and Prp11 ZnF cryo-EM model[21] (PDB accession 5GM6) was fitted into the B1 density and adjusted. We additionally modelled the conserved Prp9 ZnF2, Prp9 C terminus (residues 503–528), Cus1 N- and C-terminal regions, U2 snRNA stem IIb, and re-assigned Hsh49 RRM2. The Cus1 (residues 290–368)–Hsh49 RRM1 crystal structure[28] was fitted into the B1 density and adjusted. The pre-mRNA sequence was changed to that of *UBC4* and an altered pre-mRNA path downstream from the branch-point adenosine could be determined compared to the B^{act} model[21] (Extended Data Fig. 8e). Furthermore, the crystal structure of the LSm ring[67] (PDB code 4M7A) could be fitted in a unique orientation in the B1 density and the putative Prp3 N-terminal region (residues 35–71) and Prp3 helix α3, bridging SF3b to the LSm ring, were modelled as poly-alanine. The amino acid register for Prp3 residues 35–138 is uncertain. A homology model of the yeast Sm ring was generated based on the human Sm ring[68] (PDB code 4WZJ) using MODELLER[69] and was fitted into B5 and B6 maps for U5 and U4 snRNPs, respectively, and adjusted. The cryo-EM structure of the U2 3′ domain from C complex[18], where the Sm ring was replaced with the adjusted MODELLER[69] model, and the SF3a crystal structure[29], where selenomethionine residues were changed to the original protein sequence in COOT[65], were fitted automatically into the B3 density using a global 6D search in Situs[70]. The U2 3′ domain–SF3a complex was positioned relative to SF3b (B1 density) by fitting into the reconstruction from round 5 class 6 (Extended Data Fig. 2). This yielded a model for the complete U2 snRNP as well

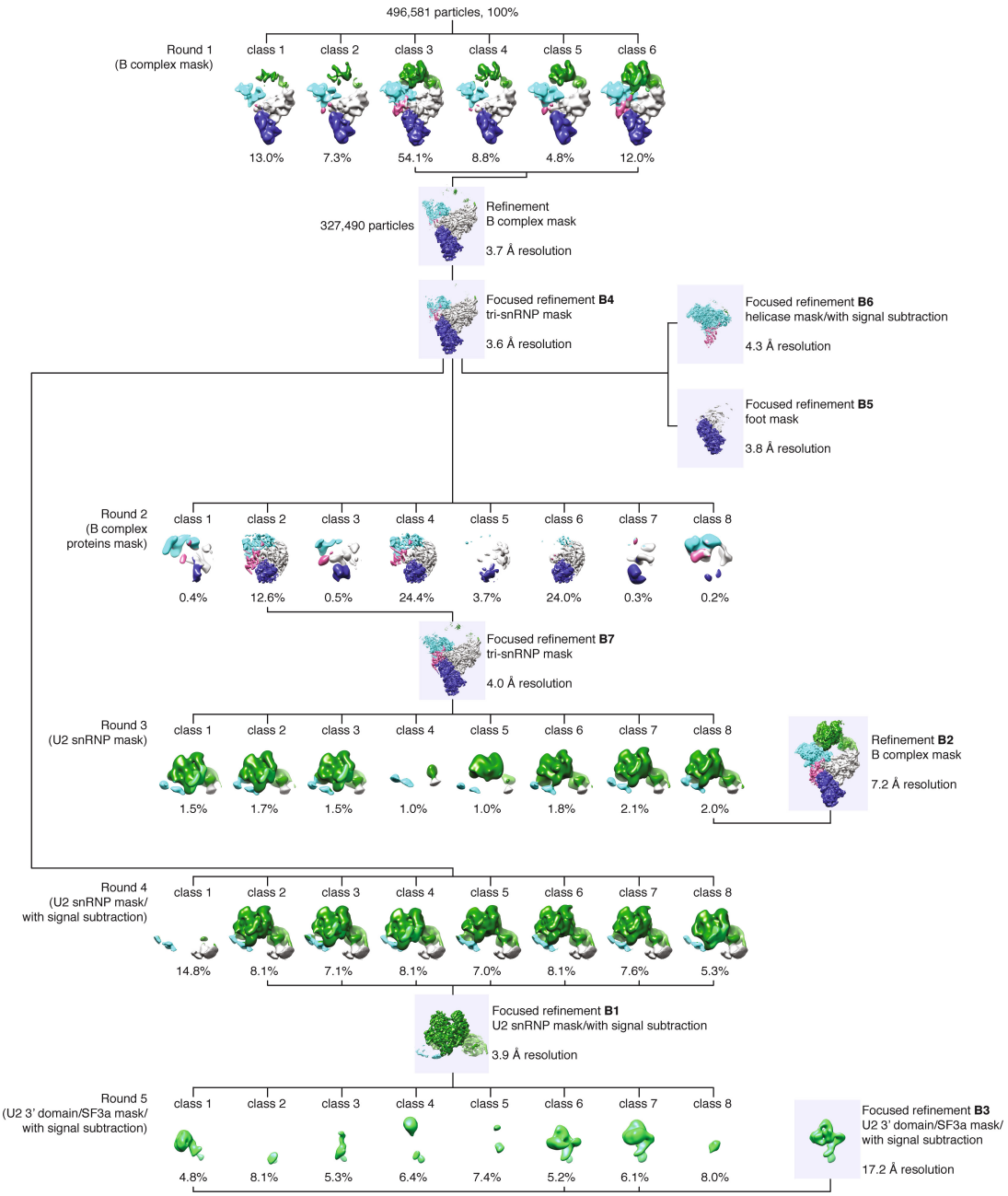
as its interaction with Prp3 and the nearby LSm ring. Previous cryo-EM models of the yeast tri-snRNP[26,27] (PDB codes 3JCM and 5GAN) were fitted into B4, B5 and B6 cryo-EM densities, and were manually adjusted and extended in COOT. The B4 density was used in particular to improve models for Prp3, Prp4, Prp6 and Prp31. A Prp6 helix (residues 64–80) was assigned based on weak density connectivity. Snu66 (residues 148–236) and two peptides of unknown polarity, which bind Prp6 and the Prp8 switch loop, were built as well. The U6 5′ stem and the U6 nucleotides that bind U5 snRNA loop 1 were located in the B4 map, as in an earlier study[27]. The U5 snRNA variable stem–loop II, stem III and stem IV (Extended Data Fig. 7a) were predicted by RNAfold[71] and RNAComposer[72], and fitted into the B5 density and adjusted. Using the B6 density models for Brr2, U4 snRNA, and the Sm ring were adjusted and the register of Snu66 bound to Brr2 was revised (residues 297–323 and 343–364), consistent with protein crosslinking[26]. Additionally, the Spp381 N-terminal region (residues 98–151) was identified based on density connectivity to the Spp381 C terminus (Extended Data Fig. 6a) and protein crosslinking[26] and was modelled as poly-alanine. The Snu23 helix α1 was placed as a poly-alanine model into the B6 density, based on connectivity to the Snu23 ZnF. To model the globular domain of the B complex proteins Spp381, Snu23 and Prp38, we first generated a homology model based on the *Chaetomium thermophilum* Prp38–Snu23 (residues 137–157)–MFAP1 (yeast Spp381; residues 217–296) heterotrimer[46] (PDB code 5F5V) using MODELLER[69]. The homology model was fitted into the B7 density and adjusted and extended in COOT[65]. Spp381 helices α3 and α4 were modelled as poly-alanine into weak density (B7) based on homology to the *C. thermophilum* structure[46]. Snu23 ZnF and Prp38 N-terminal regions and the previously disordered Prp8 residues 2121–2148 (refs 26, 27), connecting the RNase H and Jab1/MPN domains, were built into the B7 density. The complete U6 ACAGAGA stem could be modelled in the B7 map, together with a short stretch of density that we assigned to pre-mRNA. This density connects at low threshold to the intron bound by the U2 snRNP SF3b subcomplex, suggesting it is pre-mRNA (Extended Data Fig. 6e). This density could not be assigned to another part of B complex as surrounding protein and RNA parts are accounted for. The register of pre-mRNA at the U6 ACAGAGA stem is unknown, and was modelled here based on complementarity with *UBC4* pre-mRNA upstream and nearest to the 5′-exon, consistent with RNA crosslinking[48]. To generate a model for tri-snRNP, we combined models refined into B4, B5, B6 and B7 densities, and placed them together into the B7 map. To generate the complete B complex model, we used a reconstruction of B complex with a defined position of the U2 snRNP SF3b complex relative to tri-snRNP (B2 map, round 3 class 8, Extended Data Fig. 2), to reveal their transient contact at interface B (Fig. 3a, c) and to model the complete U2 snRNP on tri-snRNP. The U2 snRNP is flexibly linked to tri-snRNP, and adopts a continuum of conformations (round 4, Extended Data Fig. 2). The Prp3 helix α4, modelled as an ideal α-helix as poly-alanine, and U2/U6 helix II, modelled as an idealized double-stranded RNA helix, were rigid-body fitted into the B2 density. To visualize the transition from B to B^{act} complex spliceosomes, we revised the SF3b model in the B^{act} structure[21] (PDB code 5GM6) in the corresponding density (EMDB code EMD-9524) and modelled the U2 3′ domain–SF3a subcomplexes based on another B^{act} cryo-EM map[20] (EMDB code EMD-4099) that was low-pass filtered to 30 Å resolution. The final B complex structure comprises 52 proteins, 4 snRNAs, and the pre-mRNA substrate (Fig. 1, Extended Data Table 1). Figures were generated with PyMol (http://www.pymol.org) and UCSF Chimera[62].

Data availability. Three-dimensional cryo-EM density maps B1, B2, B3, B4, B5, B6 and B7 have been deposited in the Electron Microscopy Data Bank (EMDB) under the accession numbers EMD-3682, EMD-3683, EMD-3684, EMD-3685, EMD-3686, EMD-3687 and EMD-3688, respectively. The coordinate file of B complex has been deposited in the Protein Data Bank (PDB) under the accession number 5NRL. All other data are available from the corresponding authors upon reasonable request.

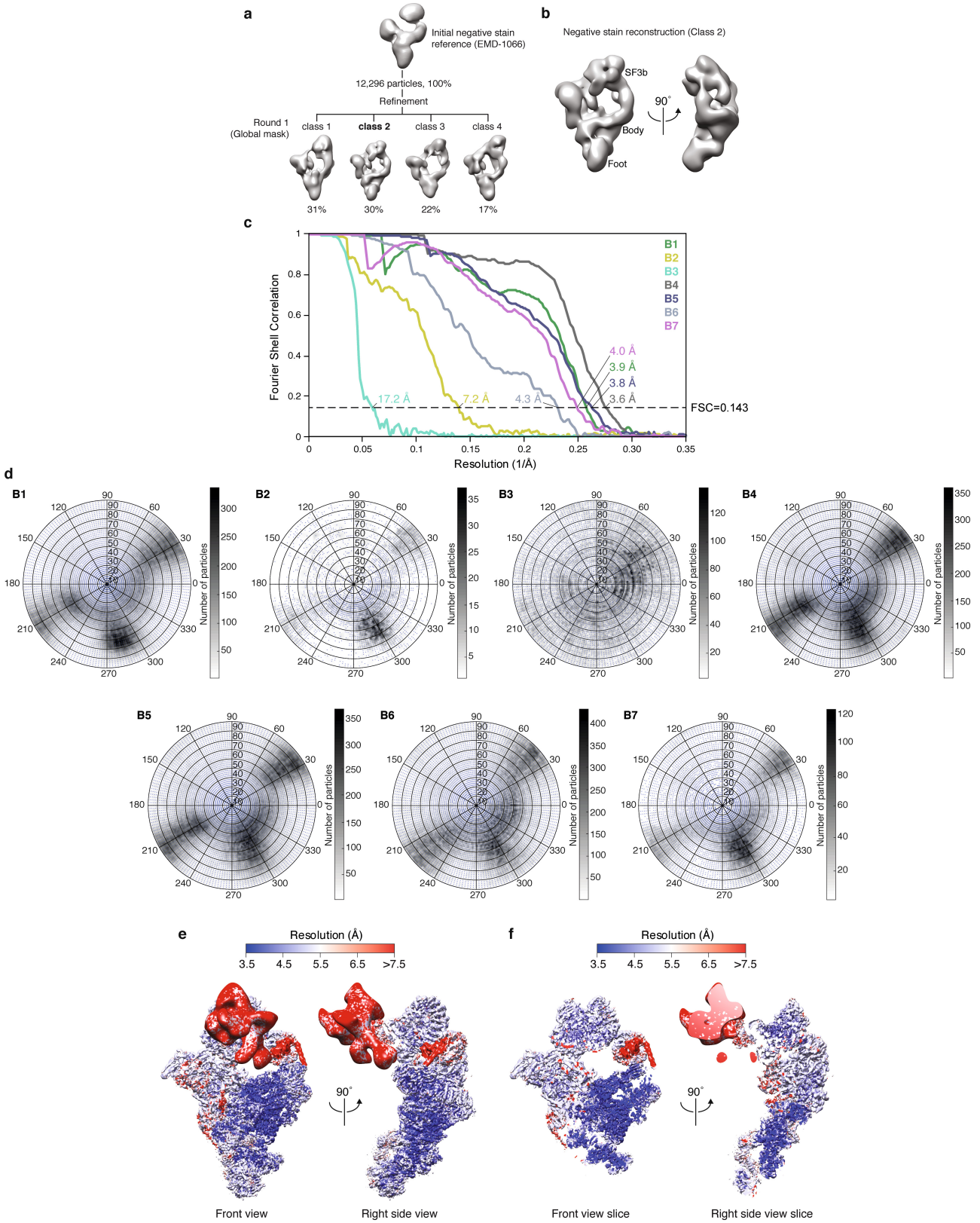
51. Nguyen, T. H. D. *et al.* The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature* **523,** 47–52 (2015).
52. Umen, J. G. & Guthrie, C. A novel role for a U5 snRNP protein in 3′ splice site selection. *Genes Dev.* **9,** 855–868 (1995).
53. Abelson, J., Hadjivassiliou, H. & Guthrie, C. Preparation of fluorescent pre-mRNA substrates for an smFRET study of pre-mRNA splicing in yeast. *Methods Enzymol.* **472,** 31–40 (2010).
54. Wu, T. P., Ruan, K. C. & Liu, W. Y. A fluorescence-labeling method for sequencing small RNA on polyacrylamide gel. *Nucleic Acids Res.* **24,** 3472–3473 (1996).
55. Zhou, Z. & Reed, R. Purification of functional RNA-protein complexes using MS2-MBP. *Curr. Protoc. Mol. Biol.* **63,** 27.3.1—27.3.7 (2001).
56. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10,** 584–590 (2013).
57. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192,** 216–221 (2015).
58. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180,** 519–530 (2012).
59. Scheres, S. H. W. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* **9,** 853–854 (2012).
60. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157,** 38–46 (2007).
61. Scheres, S. H. Semi-automated selection of cryo-EM particles in RELION-1.3. *J. Struct. Biol.* **189,** 114–122 (2015).
62. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).
63. Bai, X.-C., Rajendra, E., Yang, G., Shi, Y. & Scheres, S. H. W. Sampling the conformational space of the catalytic subunit of human γ-secretase. *eLife* **4,** e11182 (2015).
64. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11,** 63–65 (2014).
65. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
66. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66,** 213–221 (2010).
67. Zhou, L. *et al.* Crystal structures of the Lsm complex bound to the 3′ end sequence of U6 small nuclear RNA. *Nature* **506,** 116–120 (2014).
68. Leung, A. K. W., Nagai, K. & Li, J. Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis. *Nature* **473,** 536–539 (2011).
69. Eswar, N. *et al.* Comparative protein structure modeling using Modeller. *Curr. Protoc. Bioinformatics* **Chapter 5,** 5.6.1–5.6.3 (2014).
70. Wriggers, W. Conventions and workflows for using Situs. *Acta Crystallogr. D* **68,** 344–351 (2012).
71. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6,** 26 (2011).
72. Popenda, M. *et al.* Automated 3D structure composition for large RNAs. *Nucleic Acids Res.* **40,** e112 (2012).
73. Mozaffari-Jovin, S. *et al.* Inhibition of RNA helicase Brr2 by the C-terminal tail of the spliceosomal protein Prp8. *Science* **341,** 80–84 (2013).
74. McGrail, J. C. & O'Keefe, R. T. The U1, U2 and U5 snRNAs crosslink to the 5′ exon during yeast pre-mRNA splicing. *Nucleic Acids Res.* **36,** 814–825 (2008).
75. Sharma, S., Wongpalee, S. P., Vashisht, A., Wohlschlegel, J. A. & Black, D. L. Stem-loop 4 of U1 snRNA is essential for splicing and interacts with the U2 snRNP-specific SF3A1 protein during spliceosome assembly. *Genes Dev.* **28,** 2518–2531 (2014).
76. van Nues, R. W. & Beggs, J. D. Functional contacts with a range of splicing proteins suggest a central role for Brr2p in the dynamic control of the order of events in spliceosomes of *Saccharomyces cerevisiae*. *Genetics* **157,** 1451–1467 (2001).



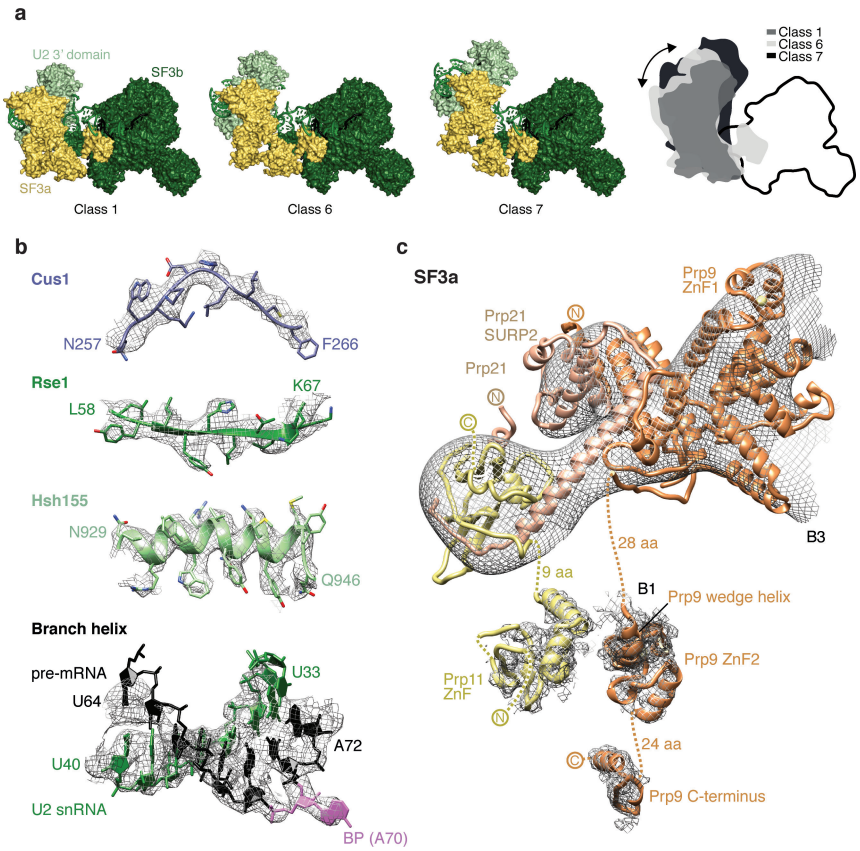
Extended Data Figure 1 | Biochemical characterization and cryo-EM of the B complex spliceosome. a, Protein analysis of purified B complex (SDS–PAGE stained with Coomassie blue). U1 snRNP components and U1 snRNA are sub-stoichiometric (see **b**), consistent with U1 snRNP destabilization in B complex[2,31]. For gel source data see Supplementary Fig. 1a. **b**, RNA analysis of purified B complex (denaturing 9% polyacrylamide TBE gel stained with Toluidine blue). For gel source data see Supplementary Fig. 1b. **c**, Purified B complex is active in an *in vitro* splicing assay. Splicing reactions were carried out in yeast extract in the absence (lanes 1 and 2) or presence of 60 nM unlabelled competitor pre-mRNA (lane 3), prohibiting the assembly of new spliceosomes. B complex was assembled on labelled pre-mRNA and purified (see Methods), and added to yeast extract for 10 min together with 60 nM unlabelled competitor pre-mRNA, before the addition of ATP to initiate the reaction (lanes 4 and 5). Splicing reactions contained 2 mM ATP (lanes 1–5). The asterisk marks a degradation product. For gel source data see Supplementary Fig. 1c. **d**, Cryo-EM micrograph of B complex. Scale bar, 200 nm. **e**, Representative 2D class averages of B complex reveal flexibility of peripheral regions relative to the tri-snRNP body. **f**, Composite cryo-EM density of B complex shown in two orthogonal views. Colours indicate the respective cryo-EM densities used for modelling (B1, green; B2, yellow; B3, light green; B4, grey; B5, blue; B6, light blue; B7, magenta). The sharpened densities are shown and are aligned using overlapping regions (see Extended Data Fig. 2). The percentage of particles from the full set of 496,581 that contribute to the respective density is indicated together with the overall resolution (see Extended Data Fig. 9). **g**, Composite cryo-EM density of B complex superimposed on a ribbon model of the B complex structure, coloured as in Fig. 1. The B complex, excluding the U1 snRNP, has a molecular mass of 2.5 MDa of which we modelled 1.8 MDa.



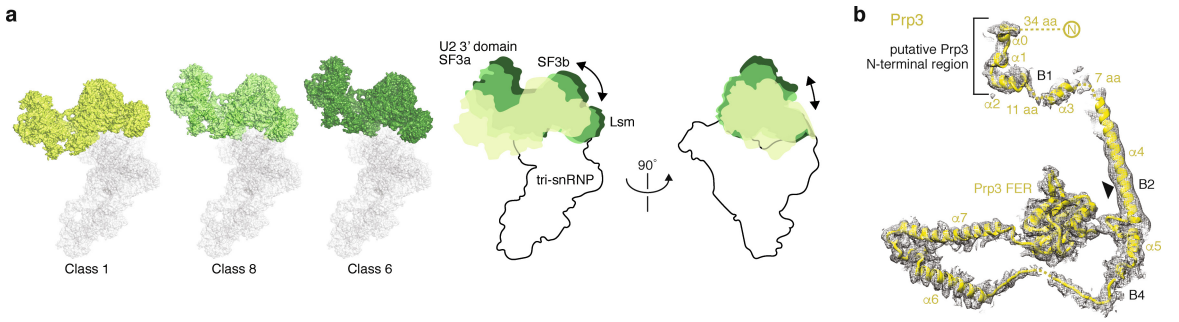
Extended Data Figure 2 | Three-dimensional classification of cryo-EM data. Three-dimensional image classification of the cryo-EM dataset using the B complex negative-stain reconstruction (Methods; Extended Data Fig. 3b) as the initial reference model. The percentage of single particles contributing to each class is provided. To help visualize structural differences, 3D reconstructions of B complex are coloured according to mobile regions: SF3b (green), U2 3′ domain–SF3a (light green), helicase (cyan), body (grey), foot (navy blue), and B complex proteins (magenta). For each classification round, the type of mask and use of signal subtraction is indicated. Additionally, the type of mask, overall resolution, and use of signal subtraction is also indicated for each 3D refinement after classification. For additional details, see the Methods and Extended Data Fig. 9.



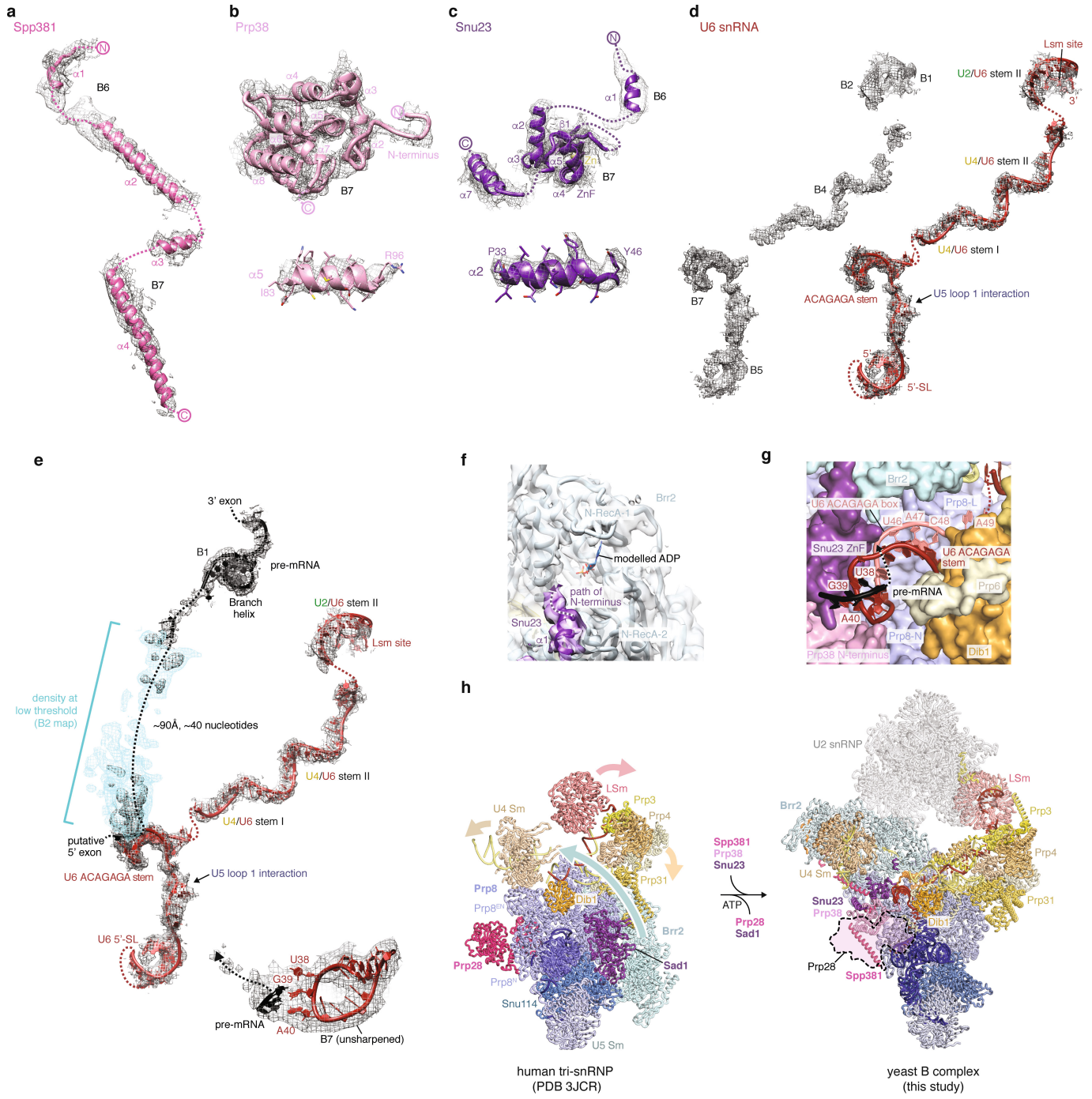
Extended Data Figure 3 | Negative-stain and cryo-EM reconstructions of B complex. a, Three-dimensional image classification of negative-stain EM data. A total of 12,296 particles were refined using the negative-stain reconstruction of the human BΔU1-complex[32] (EMDB code EMD-1066) as the initial reference, and were subsequently classified. Class 2 contained most features and was used for 3D refinement. The percentage of single particles contributing to each class is provided. **b**, Two orthogonal views of the yeast B complex negative-stain reconstruction used as the initial reference for processing of the cryo-EM dataset. **c**, Gold-standard Fourier shell correlation (FSC = 0.143) of the respective B1, B2, B3, B4, B5, B6 and B7 cryo-EM single-particle reconstructions. **d**, Orientation distribution plot of all particles that contribute to the respective B1, B2, B3, B4, B5, B6 and B7 cryo-EM single-particle reconstructions. **e**, The composite B complex cryo-EM density (maps B1–B7) is shown in two orthogonal views and coloured by local resolution, as determined by ResMap[64]. Compare with Extended Data Fig. 1f. **f**, A central slice through the composite B complex cryo-EM density (maps B1–B7) is shown in two orthogonal views and coloured by local resolution, as determined by ResMap[64].



Extended Data Figure 4 | Details of the U2 snRNP. a, Multiple conformations of the U2 snRNP 3′ domain–SF3a subcomplexes relative to SF3b indicate flexibility. This apparent mobility may be important for formation of the A complex, consistent with dynamic contacts of the U2 snRNA 5′ end with SF3a60 (human PRP9) and SF3b49 (human HSH49) in the isolated U2 snRNP[40] that differ from U2 snRNP protein–snRNA interactions observed in the yeast B complex structure. Surface representations of the U2 3′ domain (light green), SF3a (light yellow), and SF3b (dark green) are shown. The U2 3′ domain–SF3a complex is positioned according to 3D classifications 1, 6 and 7 from round 5 (compare with Extended Data Fig. 2). A cartoon summarizes the movements. **b**, Representative regions of the sharpened SF3b-containing density (B1) at 3.9 Å resolution are superimposed on the refined coordinate model. The density shows side-chain features for a loop in Cus1, a β-strand in Rse1, an α-helix in Hsh155, and separation of RNA nucleotides in the U2–pre-mRNA branch helix. Colours are as in Fig. 2. **c**, Cryo-EM densities for SF3a are superimposed on the B complex coordinate model. The crystal structure of the Y-shaped core of SF3a (ref. 29) is superimposed on the B3 density. Prp11 ZnF, Prp9 ZnF2 and the Prp9 C terminus are superimposed on the B1 density. Structural elements of SF3a, including the Prp9 wedge helix, and disordered regions are indicated. For cryo-EM density nomenclature, see Extended Data Fig. 1f. Colours are as in Fig. 2.

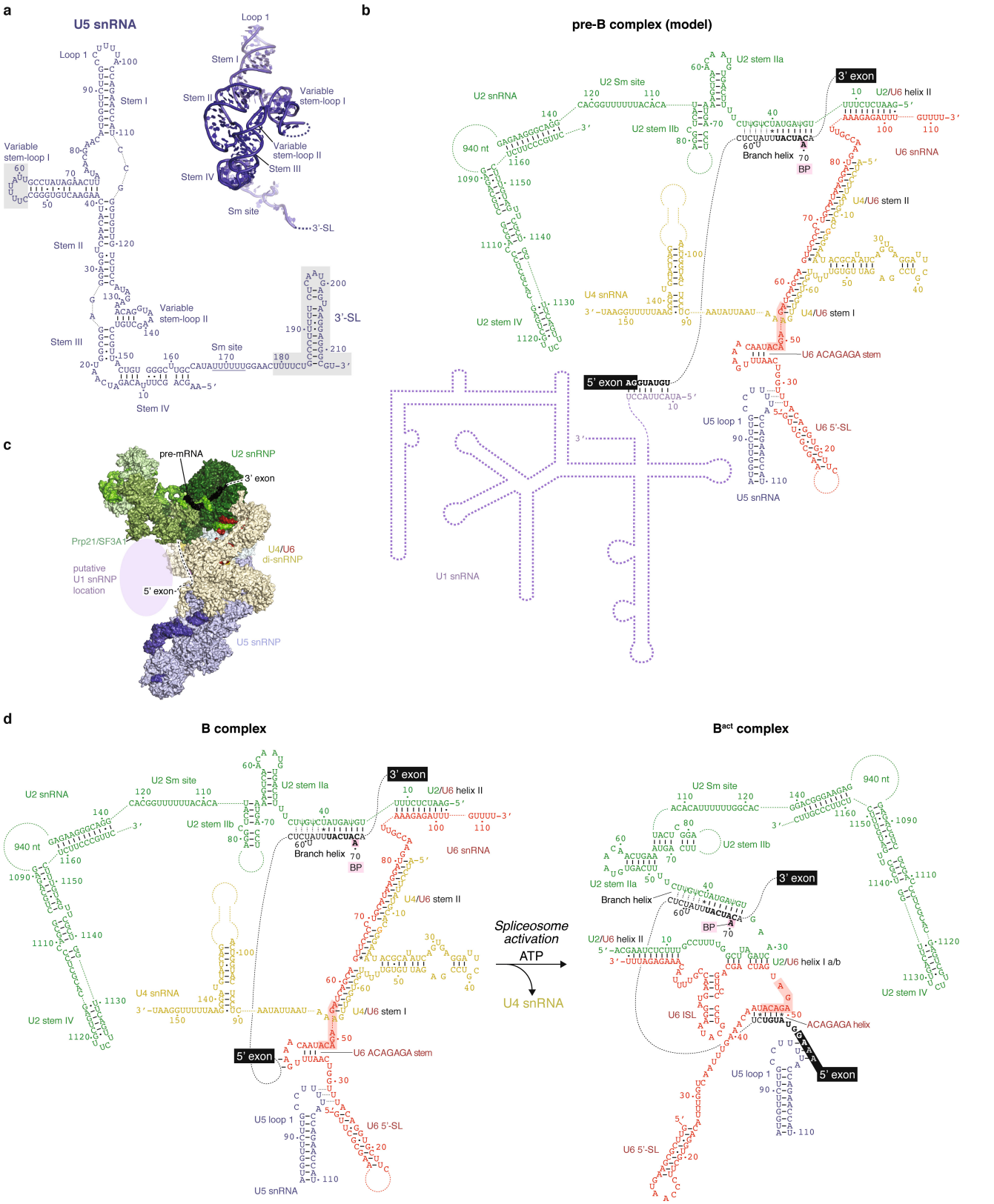


Extended Data Figure 5 | Flexibility of the U2 snRNP relative to tri-snRNP. a, Multiple positions of the U2 snRNP relative to tri-snRNP. Representative classes are shown (classes 1, 6 and 8 from round 3, see Extended Data Fig. 2) that reside along a continuum of conformations (compare with Extended Data Fig. 1e). The U2 snRNP moves together with the U6 LSm ring, which is anchored via the putative Prp3 N terminus. A cartoon summarizes the movements in two orthogonal views. The location of the U2 snRNP has an apparent effect on the strength of the putative pre-mRNA 5′ exon cryo-EM density (compare Extended Data Fig. 6e). When the U2 snRNP is positioned away from the U6 ACAGAGA stem, the pre-mRNA density is weaker than when the U2 snRNP is positioned closer. This suggests how Brr2 helicase activity may perform a kinetic proofreading of the pre-mRNA 5′-exon–U6 snRNA interaction: When the U2 snRNP is close to the tri-snRNP, activation may occur normally (productive activation). However, when the U2 snRNP is positioned further away and the 5′ exon is not tethered, Brr2 activity may instead lead to dissociation of the U2 snRNP from the tri-snRNP (non-productive activation)[12]. **b**, The Prp3 model is superimposed on B1 (putative N-terminal region), B2 (helix α4) and B4 (C-terminal region) cryo-EM densities. The Prp3 ferredoxin-like fold (FER) and secondary structure elements are labelled. The black arrowhead indicates the region of Prp3 helix α4 that bends with different U2 snRNP positions (see **a**). For cryo-EM density nomenclature, see Extended Data Fig. 1f.



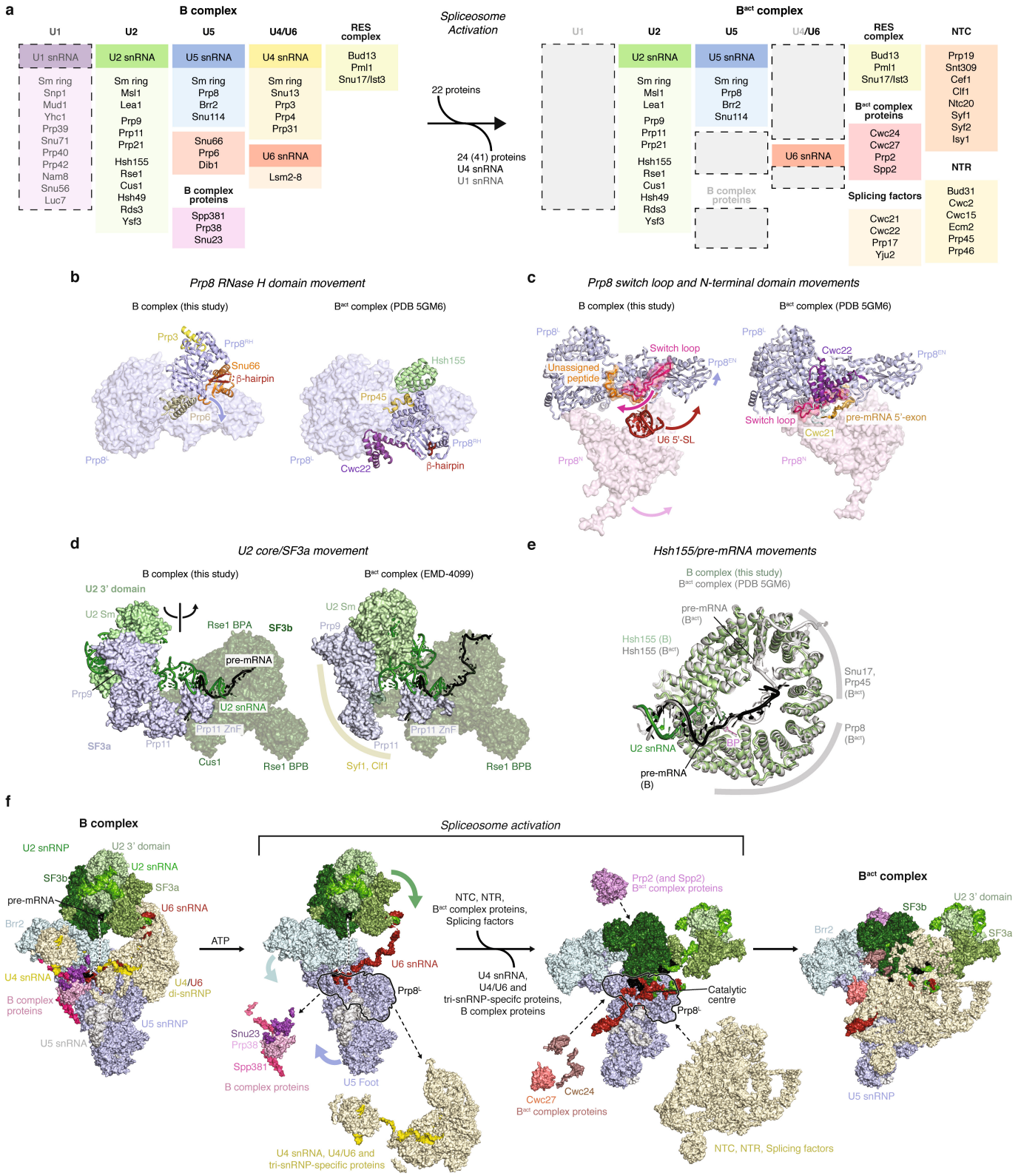
Extended Data Figure 6 | See next page for caption.

Extended Data Figure 6 | Details of B complex proteins and U6 snRNA. **a**, Fit of the Spp381 model to B6 (helices α1–α2) and B7 (helices α3–α4) cryo-EM densities. Helix α4 was modelled into weak density (B7) based on homology to the human and *Chaetomium thermophilum* crystal structures[46]. For cryo-EM density nomenclature, see Extended Data Fig. 1f. **b**, Fit of the Prp38 model to the B7 cryo-EM density. Helix α5 is shown below, revealing side-chain features in the density. **c**, Fit of the Snu23 model to B6 (helix α1) and B7 (remainder of Snu23) cryo-EM densities. Helix α2 is shown below, revealing side-chain features in the density. **d**, Composite cryo-EM density and fit of the U6 snRNA model. The B1, B2, B4, B5 and B7 densities are shown without (left) and with (right) the U6 snRNA model superimposed. U6 elements and the site of interaction with the U5 snRNA loop 1 are indicated. **e**, A weak density for pre-mRNA connects from the U6 snRNA ACAGAGA stem to the U2 SF3b-bound intron. The connecting density (map B2) is shown at intermediate (grey, threshold of 0.023) and low (light blue, threshold of 0.0173) thresholds. The U6 snRNA densities are shown as in **d**. The register of pre-mRNA near the U6 ACAGAGA stem is uncertain and was tentatively modelled based on complementarity with *UBC4* pre-mRNA upstream and nearest to the 5′ exon, consistent with RNA crosslinking[48]. According to this register, approximately 40 pre-mRNA nucleotides separate U6- and SF3b-bound regions. The bottom right panel shows the fit of the pre-mRNA–U6 snRNA ACAGAGA stem–loop model to the unsharpened B7 density. For cryo-EM density nomenclature, see Extended Data Fig. 1f. **f**, Snu23 binds near the nucleotide-binding pocket of the Brr2 NHC, where it may influence Brr2 activity. Brr2 (pale cyan) and Snu23 (violet) models are superimposed on the B6 cryo-EM density, coloured as the underlying proteins. The RecA-1 and RecA-2 lobes of the Brr2 NHC are labelled, and an ADP nucleotide was modelled by aligning the NHC of human BRR2 bound to ADP[73] (PDB code 4KIT) on the equivalent yeast Brr2 residues. The path of the Snu23 N terminus cryo-EM density is indicated, and is near to the Brr2 nucleotide-binding pocket. **g**, The U6 ACAGAGA stem is chaperoned by Dib1, Prp6, Prp8 and B complex proteins. Stabilization of the U6 ACAGAGA stem may facilitate tethering of pre-mRNA at its tip, whereas the U6 ACAGAGA box is buried in the stem. Surface models of Snu23 ZnF, Prp38 N terminus, Prp6 N terminus, Dib1, $Prp8^{L}$ and $Prp8^{N}$ domains, and Brr2 are shown and reveal a network of protein–RNA contacts to maintain the U6 ACAGAGA stem. **h**, Comparison of human tri-snRNP[25] (PDB code 3JCR) and yeast B complex (this study) reveals that the $Prp8^{N}$ and $Prp8^{EN}$ domains serve as a platform for mutually exclusive binding of Prp28 and B complex proteins. Their $Prp8^{N}$ binding sites overlap and the altered location of the $Prp8^{EN}$ domain between the two complexes provides additional interfaces for either Prp28 or B complex proteins to bind. Movements compared to B complex in Brr2 and U4/U6 di-snRNP are indicated with arrows, and may occur after A complex association. The U2 snRNP is shown in grey to highlight tri-snRNP components, which are coloured as in Fig. 1. Comparison with B^{act} and C* structures[20–24] further suggests that the $Prp8^{N}$ and $Prp8^{EN}$ domains serve as a general platform for step-specific splicing factors during the splicing cycle.



Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | U5 snRNA model, location of the U1 snRNP, and RNA secondary structure diagrams. a, An improved model for U5 snRNA. A secondary structure diagram (left) and refined coordinate model (right) are shown, and the newly determined variable stem–loop II, stem III and stem IV are labelled together with known U5 snRNA elements. The long form of U5 snRNA is shown, where the short form ends with nucleotide 179. The grey boxes indicate regions not included in the model. Lines indicate Watson–Crick base pairs, and dots indicate G–U wobble base pairs. The U5 snRNA model was prepared by M. Wilkinson. **b**, Putative model of the RNA interaction network in the pre-B complex. The RNA network is unchanged compared to B complex, except for the interaction of the pre-mRNA 5′-exon with U1 snRNA[74]. Only loop 1 of U5 snRNA is shown. Lines indicate Watson–Crick base pairs, dots G–U wobble base pairs, stars denote non-canonical base pairs, and dotted lines putative nucleotide interactions. **c**, Putative location of the U1 snRNP. To gain insights into U1 snRNP location relative to U2 snRNP and tri-snRNP, we combined genetic, biochemical and structural observations. The U1 snRNP may bind between the human SF3a subunit SF3A1 (yeast Prp21)[75], the Prp28-binding site[25], Brr2 (ref. 76), and the U6 ACAGAGA stem. In B complex, the U1 snRNP may be destabilized owing to a steric clash with Brr2, which is likely to be repositioned compared to the pre-B complex, as in the human tri-snRNP structure[25]. In humans, the U1 snRNP may be further destabilized by a loss of A complex-specific proteins[2]. Brr2 repositioning may therefore serve as a checkpoint to signal the release of U1 snRNA from pre-mRNA. **d**, RNA secondary structure diagrams of regions modelled in B and B^{act} complex spliceosomes, using *UBC4* pre-mRNA. The pre-mRNA substrate in B^{act} (ref. 21) is a mixture of cellular pre-mRNAs and its sequence is replaced by that of *UBC4*. The consensus nucleotides at the 5′SS and branch point sequence in yeast are shown in bold. Only loop 1 of U5 snRNA is shown. Lines indicate Watson–Crick base pairs, dots denote G–U wobble base pairs, stars denote non-canonical base pairs, and dotted lines putative nucleotide interactions. Highlighted are the branch-point adenosine, pre-mRNA 5′ and 3′ exons, and the U6 ACAGAGA box.



Extended Data Figure 8 | See next page for caption.

Extended Data Figure 8 | Compositional and conformational changes during spliceosome activation. a, List of RNA and protein components in B and B^{act} complex spliceosomes. During spliceosome activation, 22 proteins join the spliceosome, whereas 24 proteins (or 41 including the U1 snRNP) and U1 and U4 snRNAs are released. The U1 snRNP is indicated by a dashed line owing to its substoichiometric binding in B complex. RES complex proteins were not modelled in B complex, however, weak density is visible at the same binding site as in B^{act} (refs 20, 21), consistent with mass spectrometry data (ref. 9 and data not shown). **b**, Movement of the Prp8 RNase H domain ($Prp8^{RH}$) between B and B^{act} complex spliceosomes. In B complex (left) regions of Prp3, Prp6 and Snu66 (residues 148–236) contact the $Prp8^{RH}$ and its β-hairpin (red). The Prp8 large domain ($Prp8^{L}$) is shown as a surface and subunits are coloured as in Fig. 1. An arrow indicates the movement of $Prp8^{RH}$ to its B^{act} position[21] (right; PDB code 5GM6), where it is stabilized by Hsh155, Prp45 (yellow) and Cwc22 (dark violet). B and B^{act} models were aligned on the $Prp8^{L}$. **c**, Movements of the Prp8 switch loop and N-terminal domain between B and B^{act} complex spliceosomes. In B complex (left), an unassigned peptide (orange) binds the Prp8 switch loop, stabilizing it on the $Prp8^{L}$ domain. The Prp8 N-terminal domain ($Prp8^{N}$) is shown as a surface (magenta) and binds the U6 snRNA 5′ stem. The Prp8 endonuclease ($Prp8^{EN}$) domain is labelled. Arrows indicate the movements required to transition to B^{act} (PDB code 5GM6). The unassigned peptide is released in B^{act} (right) and Cwc21 (yellow) and Cwc22 (dark violet) bind in its stead to stabilize the new position of the Prp8 switch loop and the loaded pre-mRNA 5′ exon (light orange) in the exon channel. The repositioned $Prp8^{N}$ completes this channel. B and B^{act} models were aligned on the $Prp8^{L}$ domain. **d**, Movements of U2 3′ domain–SF3a subcomplexes between B and B^{act} complex spliceosomes. In B complex (left) the U2 3′ domain–SF3a subcomplexes are flexibly linked to SF3b and assume several positions relative to SF3b (compare Extended Data Fig. 4a). For comparison to B^{act}, the model for U2 3′ domain–SF3a (light green/grey) was rigid-body fitted into a low-pass filtered cryo-EM density of yeast B^{act} (EMDB code EMD-4099)[20]. This indicated that in B^{act} (right), the U2 3′ domain–SF3a complex is repositioned due to binding of NTC proteins Syf1 and Clf1 (yellow arch), to avoid a steric clash. The NTC subunits Syf1 and Clf1 may thus be positioned in B^{act} to bind the U2 3′ domain after the release of SF3a, during conversion to B* (Extended Data Fig. 8d). This repositioning is distinct from U2 3′ domain–SF3a conformations that are sampled in B complex. B and B^{act} models of the U2 snRNP were aligned on SF3b subunit Rse1. **e**, Movements of SF3b subunit Hsh155 and pre-mRNA between B and B^{act} complex spliceosomes. B and B^{act} (PDB code 5GM6) models were aligned on Hsh155, revealing small conformational changes in Hsh155 HEAT repeats, possibly owing to extended interaction of Hsh155 with Prp8, Snu17 and Prp45 (grey arches) in B^{act}. The U2–pre-mRNA branch helix is bound in the same manner in B and B^{act} complex spliceosomes. However, nucleotides downstream of the branch point (magenta) are bound differently, possibly due to movements in Hsh155. **f**, Structural differences between B and B^{act} complex spliceosomes suggest a model for activation. The ATP-dependent helicase Brr2 is positioned in B complex (left) and unwinds the U4/U6 duplex to release U4 snRNA and U4/U6 and B complex proteins (middle left). Their removal enables movements in Brr2, U2 and U5 snRNPs, and U6 snRNA to facilitate the loading of pre-mRNA, and formation of the U2/U6 catalytic centre (middle right). This intermediate is subsequently stabilized by NTC, NTR, splicing factors and B^{act} complex proteins to form B^{act} (ref. 21) (right; PDB code 5GM6). U5 snRNP (blue; U5 snRNA, light grey), U4/U6 di-snRNP (light yellow; U4 snRNA, yellow; U6 snRNA, red), U2 snRNP (U2 3′ domain, light green; SF3a, olive; SF3b, dark green), B complex proteins (shades of magenta), NTC, NTR and splicing factors (light yellow), and B^{act} complex proteins (shades of salmon) are indicated. The B^{act} position of U2 snRNP 3′ domain–SF3a was modelled as in Fig. 5. B and B^{act} spliceosomes were aligned on their $Prp8^{L}$ domain. Spliceosome activation intermediates are modelled.

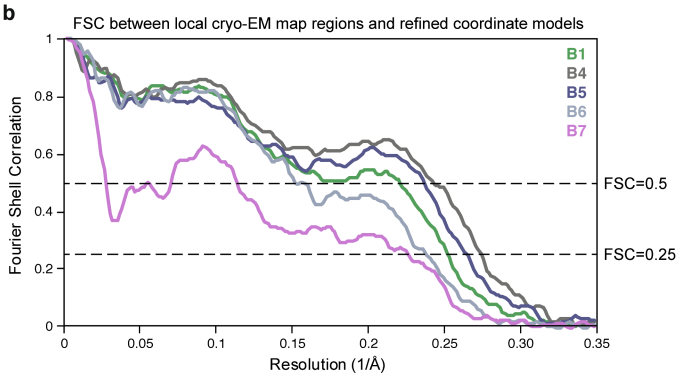
a

Cryo-EM data collection and refinement statistics of the B-complex structure

	B1 (SF3b)	B2 (B)[a]	B3 (U2 3'domain/SF3a)[b]	B4 (Body)	B5 (Foot)	B6 (Helicase)	B7 (B-specific proteins)
Data collection							
Particles	254,095	9,559	79,166	327,490	327,490	327,490	62,406
Pixel Size (Å)	1.43	1.43	1.43	1.43	1.43	1.43	1.43
Defocus range (μm)	–0.35 to –5.3	–0.35 to –5.3	–0.35 to –5.3	–0.35 to –5.3	–0.35 to –5.3	–0.35 to –5.3	–0.35 to –5.3
Voltage (kV)	300	300	300	300	300	300	300
Electron dose ($e^- A^{-2}$)	56	56	56	56	56	56	56
Reconstruction (RELION)							
Accuracy of rotations (°)	1.06	0.90	3.0	0.65	0.83	1.53	0.66
Accuracy of translations (pixel)	0.83	0.71	2.9	0.51	0.76	0.97	0.51
Resolution (Å)	3.9	7.2	17.2	3.6	3.8	4.3	4.0
Map sharpening B-factor ($Å^2$)	-125	-170	-200	-105	-127	-140	-80
Model composition							
Non-hydrogen atoms	28,833		13,410	33,940	20,644	21,609	8,239
Protein residues	3,358		1,414	3,786	1,992	2,624	850
RNA bases	105		92	162	215	48	69
Refinement (PHENIX)							
Map CC (around atoms)	0.680			0.536	0.662	0.376	0.496
Rms deviations							
Bond lengths (Å)	0.005			0.004	0.005	0.007	0.003
Bond angles (°)	1.00			0.94	0.96	1.24	0.84
Validation							
Molprobity score	2.22			2.19	2.12	2.09	2.04
All-atom clashscore	19.07			19.86	18.50	20.54	13.00
Rotamer outliers (%)	0.03			0.00	0.00	0.09	0.00
C-beta deviations	0			0	0	0	0
Ramachandran plot							
Outliers (%)	0.3			0.2	0.1	0.3	0.1
Allowed (%)	6.5			5.8	5.0	3.8	6.2
Favoured (%)	93.2			94.0	94.9	95.9	93.7
RNA validation							
Correct sugar puckers (%)	95.2			92.6	94.0	89.6	92.8
Good backbone conformations (%)	81.9			68.5	63.3	62.5	62.3
Data Deposition							
EMDB ID	EMD-3682	EMD-3683	EMD-3684	EMD-3685	EMD-3686	EMD-3687	EMD-3688
PDB ID for the complete model		5NRL					

[a]The B2 density was used to model the position of U2 snRNP relative to tri-snRNP (see Methods).
[b]The U2 3' domain/SF3a model was generated by rigid-body fitting of previously determined U2 3' domain (PDB ID 5LJ5) and SF3a (PDB ID 4DGW) structures (see Methods).

b



Extended Data Figure 9 | Data collection, refinement statistics and structure validation. a, Cryo-EM data collection and refinement statistics of the B complex structure. Different regions of the composite B complex structure were refined into B1, B4, B5, B6 and B7 maps as described (see Methods). **b**, FSC between local cryo-EM map regions and corresponding regions of the refined coordinate models. Note that the FSC curve for the B-specific proteins correlates with the local resolution in this sub-region of the B7 density (4.0–5.0 Å; Extended Data Fig. 3e), below the nominal resolution (4.0 Å).

Extended Data Table 1 | Summary of components modelled into B complex cryo-EM densities

Proteins and RNA included in the model

Sub-complexes	Protein/RNA	Domains	Total residues	M.W. (kDa)	Modelled residues	Modelling template (PDB ID)	Modelling	Chain ID	Human/*S. pombe* names
U5 snRNP	Prp8	N-terminal	1-870	101.7	128-870	5GAN, 3JCM	Docked and rebuilt	A	U5-220K/*Spp42*
		Large	871-1827	111.5	871-1827	5GAN, 3JCM	Docked and rebuilt		
		RNaseH	1828-2085	29.5	1828-2085	5GAN, 3JCM	Docked and rebuilt		
		Jab1/MPN	2086-2413	36.8	2086-2109; 2121-2401	5GAN	Docked		
	Brr2		2,163	246.2	447-1712; 1716-1825; 1841-2163	5GAN	Docked and rebuilt	B	U5-200K/*Brr2*
	Snu114		1008	114.0	105-517; 531-693; 706-725; 742-770; 774-979; 986-1005	5GAN, 3JCM	Docked and rebuilt	C	U5-116K/*Cwf10*
	Dib1		143	16.8	2-143	5GAN	Docked and rebuilt	D	U5-15K/*Brr2*
	SmB		196	22.4	12-54; 76-102	4WZJ	Homology modelled	b	SmB/*SmB*
	SmD3		101	11.2	4-86	4WZJ	Homology modelled	d	SmD3/*SmD3*
	SmD1		146	16.3	1-48; 78-109	4WZJ	Homology modelled	h	SmD1/*SmD1*
	SmD2		110	12.9	17-108	4WZJ	Homology modelled	i	SmD2/*SmD2*
	SmE		96	9.7	10-63; 71-93	4WZJ	Homology modelled	e	SmE/*SmE*
	SmF		86	10.4	13-84	4WZJ	Homology modelled	f	SmF/*SmF*
	SmG		77	8.5	3-73	4WZJ	Homology modelled	g	SmG/*SmG*
	U5 snRNA-S/-L		179/214	68.8	1-53; 62-178	5GAN	Docked and de novo	5	
U2 snRNP	Msl1		111	12.8	28-111	1A9N	Homology modelled	Y	U2-B"
	Lea1		238	27.2	1-170	1A9N	Homology modelled	W	U2-A'
	SmB		196	22.4	12-54; 76-102	4WZJ	Homology modelled	s	SmB/*SmB*
	SmD3		101	11.2	4-85	4WZJ	Homology modelled	v	SmD3/*SmD3*
	SmD1		146	16.3	1-48; 78-101	4WZJ	Homology modelled	t	SmD1/*SmD1*
	SmD2		110	12.9	17-108	4WZJ	Homology modelled	u	SmD2/*SmD2*
	SmE		96	9.7	10-63; 71-93	4WZJ	Homology modelled	w	SmE/*SmE*
	SmF		86	10.4	12-84	4WZJ	Homology modelled	x	SmF/*SmF*
	SmG		77	8.5	2-76	4WZJ	Homology modelled	y	SmG/*SmG*
	Hsh155		971	110.0	132-149; 157-971	5GM6	Docked and rebuilt	O	SF3B1/*Sap155*
	Rse1		1361	153.8	53-305; 323-571; 581-784; 814-890; 918-1265; 1292-1361	5GM6	Docked and rebuilt	P	SF3B3/*Sap130*
	Cus1		436	50.3	125-213; 239-353; 361-376	5GM6	Docked and rebuilt	Q	SF3B2/*Sap145*
	Hsh49		213	24.5	9-86; 106-144; 147-185; 189-203	5GM6	Docked and rebuilt	R	SF3B4/*Sap49*
	Rds3		107	12.3	2-104	5GM6	Docked and rebuilt	S	SF3B14b/*Ini1*
	Ysf3		85	10.0	2-84	5GM6	Docked and rebuilt	Z	SF3B5/*Sab10*
	Prp9		530	63.0	1-97; 112-378; 407-478; 503-528	4DGW	Docked and de novo	T	SF3A3/*Sap61*
	Prp11		266	29.9	34-47; 51-105; 115-136; 149-253	4DGW, 5GM6	Docked and de novo	U	SF3A2/*Sap62*
	Prp21		280	33.1	89-206; 220-228	4DGW	Docked	V	SF3A1/*Sap114*
	U2 snRNA		1175	363.8	3-13; 30-73; 79-86; 108-122; 139-150; 1089-1109; 1115-1130; 1138-1154; 1159-1169	5LJ5, 5GM6	Docked and de novo	2	
U4/U6 di-snRNP	Snu13		126	13.6	3-126	5GAN, 3JCM	Docked and rebuilt	K	U4/U6.U5-15.5K/*Snu13*
	Prp31		494	56.3	42-458	5GAN, 3JCM	Docked and rebuilt	F	U4/U6.U5-61K/*Prp31*
	Prp3		469	55.9	35-71; 83-92; 100-138; 140-173; 211-224; 228-410; 413-467	5GAN, 3JCM	Docked and rebuilt	G	U4/U6.U5-90K/*Prp3*
	Prp4		465	52.4	23-91; 108-465	5GAN, 3JCM	Docked and rebuilt	H	60K/*Rna4*
	SmB		196	22.4	12-54; 76-102	4WZJ	Homology modelled	k	SmB/*SmB*
	SmD3		101	11.2	4-85	4WZJ	Homology modelled	n	SmD3/*SmD3*
	SmD1		146	16.3	1-48; 78-119; 138-146	4WZJ	Homology modelled	l	SmD1/*SmD1*
	SmD2		110	12.9	17-108	4WZJ	Homology modelled	m	SmD2/*SmD2*
	SmE		96	9.7	10-63; 71-93	4WZJ	Homology modelled	p	SmE/*SmE*
	SmF		86	10.4	12-84	4WZJ	Homology modelled	q	SmF/*SmF*
	SmG		77	8.5	2-76	4WZJ	Homology modelled	r	SmG/*SmG*
	LSm2		95	11.2	1-90	4M7A	Docked and rebuilt	a	LSm2/*LSm2*
	LSm3		89	10.0	3-79	4M7A	Docked	3	LSm3/*LSm3*
	LSm4		172	20.3	1-47; 64-90	4M7A	Docked	j	LSm4/*LSm4*
	LSm5		93	10.4	4-54; 61-84	4M7A	Docked	o	LSm5/*LSm5*
	LSm6		86	9.4	11-84	4M7A	Docked	z	LSm6/*LSm6*
	LSm7		115	13.0	26-70; 85-105	4M7A	Docked	7	LSm7/*LSm7*
	LSm8		109	12.4	4-67	4M7A	Docked	8	LSm8/*LSm8*
	U4 snRNA		160	51.4	1-68; 71-79; 90-102; 131-154	5GAN	Docked and rebuilt	4	
	U6 snRNA		112	36.1	1-9; 16-51; 55-88; 92-102; 108-112	5GAN	Docked and de novo	6	
tri-snRNP	Prp6		899	104.2	4-26; 64-80; 98-178; 212-526; 530-896	5GAN, 3JCM	Docked and rebuilt	J	PRPF6/*Prp1*
	Snu66		587	66.4	148-152; 159-223; 232-236; 297-323; 343-364		De novo	E	U5-102k/*Snu66*
B complex proteins	Snu23		194	22.7	11-19; 33-62; 71-118; 125-146	5F5V	Homology modelled and de novo	L	ZMAT2/*Snu23*
	Prp38		242	28.0	2-64; 74-79; 83-133; 144-169; 182-209	5F5V	Homology modelled and de novo	M	hPrp38/*Prp38*
	Spp381		291	33.8	98-108; 124-151; 164-173; 181-216	5F5V	Homology modelled and de novo	N	MFAP1/Saf3
Substrate	UBC4 pre-mRNA		135	40.6	6-8; 51-53; 57-79		Docked and de novo	I	
Unknown					1-18; 101-111		de novo	X	