Cryo-EM structure of the spliceosome immediately after branching

Wojciech P. Galej1, Max E. Wilkinson1, Sebastian M. Fica2, Chris Oubridge1, Andrew J. Newman1 & Kiyoshi Nagai1

Precursor mRNA (pre-mRNA) splicing proceeds by two consecutive transesterification reactions via a lariat–intron intermediate. Here we present the 3.8 Å cryo-electron microscopy structure of the spliceosome immediately after lariat formation. The 5′-splice site is cleaved but remains close to the catalytic Mg2+ site in the U2/U6 small nuclear RNA (snRNA) triplex, and the 5′-phosphate of the intron nucleotide G(+)1 is linked to the branch adenosine 2′OH. The 5′-exon is held between the Prp8 amino-terminal and linker domains, and base-pairs with U5 snRNA loop 1. Non-Watson–Crick interactions between the branch helix and 5′-splice site dock the branch adenosine into the active site, while intron nucleotides +3 to +6 base-pair with the U6 snRNA ACAGAGA sequence. Isy1 and the step-one factors Yju2 and Cwc25 stabilize docking of the branch helix. The intron downstream of the branch site emerges between the Prp8 reverse transcriptase and linker domains and extends towards the Prp16 helicase, suggesting a plausible mechanism of remodelling before exon ligation.

The spliceosome is a dynamic molecular machine1,2 that catalyses pre-mRNA splicing in two sequential transesterifications analogous to group II intron self-splicing3. The major spliceosomal components—U1, U2, U4/U6, and U5 small nuclear ribonucleoprotein particles (snRNPs), and the two large NineTeen and NineTeen Related (NTC and NTR) protein complexes—assemble de novo on pre-mRNA substrates in an ordered manner4–6. Initially U1 and U2 snRNPs recognize the 5′-splice site (5′SS) and branch point sequences of pre-mRNA; subsequently the pre-assembled U4/U6.U5 tri-snRNP is recruited to form the fully assembled spliceosome (complex B). During catalytic activation Prp28 helicase displaces the 5′SS from U1 snRNP and allows it to base-pair with the U6 snRNA ACAGAGA sequence7,8. Brr2 helicase unwinds the U4/U6 snRNA duplex to release U4 snRNA and its associated protein9,10, allowing recruitment of the NTC and NTR complexes. The resulting complex Bact is then remodelled to complex B*, which recruits step-one-specific factors Yju2 and Cwc25. These factors stabilize a network of RNA interactions comprising U2, U5 and U6 snRNAs, which position the pre-mRNA 5′SS and branch point sequences for catalysis of the first transesterification (branching) producing 5′-exon and lariat intron–3′-exon intermediates. The resulting complex C is further remodelled to complex C* in which the 5′- and 3′-exons are aligned on U5 snRNA loop 1 to produce spliced mRNA and lariat intron products via the second transesterification (exon ligation)11,12. The spliced mRNA is released and the remaining intron lariat spliceosome (ILS) is disassembled, recycling the snRNPs for new rounds of splicing.

During this splicing cycle DEcD/H-box helicases are recruited to the spliceosome at specific steps to remodel RNA–RNA interactions and induce binding or release of auxiliary factors13,14. Specifically, after branching, the step-one factors Yju2 and Cwc25 are released by Prp16 helicase and Prp18–Slu7 and Prp22 are recruited to produce catalytically active complex C*(ref. 13). Following exon ligation, the spliced mRNA is released by Prp22 helicase15 and the residual ILS is disassembled by Prp43 helicase6,17.

Here we describe the cryo-electron microscopy (cryo-EM) structure of the spliceosome captured immediately after branching. This structure provides insight into recognition and positioning of the 5′SS and branch point at the active site, elucidates how proteins stabilize the architecture of the catalytic RNA core, and provides a molecular basis to understand the functions of RNA helicases and auxiliary factors in remodelling the spliceosome.

Overview of the structure

Spliceosomes from the yeast Saccharomyces cerevisiae were assembled on UBC4 pre-mRNA substrate16 with a mutation of the 3′-splice site (3′SS) sequence UAGAG to UACAC, and purified via an affinity-tag on Slu7 or Prp18 (Methods). The purified spliceosomes contained predominantly lariat intron–3′-exon intermediates (Extended Data Fig. 1), indicating that the purified spliceosomes represent complex C. We obtained a cryo-EM reconstruction at 3.8 Å overall resolution to...
ArticlerSeArcH

UCAAAGAGAUUU

is readily attributable to Prp19, Cef1 and Snt309 based on its distinct more clearly visualized. The weak density observed in the latter class allowed us to select subsets of particles (core + helicase, core + Prp19) (Extended Data Fig. 2), in which less well-ordered components can be more clearly visualized. The weak density observed in the latter class is readily attributable to Prp19, Cef1 and Snt309 based on its distinct

As in U4/U6.U5 tri-snRNAP19,20, the Large domain of Prp8 (ref. 21) forms the foundation of the assembly together with the stable foot unit, comprising GTP-bound Snu114 and the N-terminal domain of Prp8, firmly gripping the U5 snRNA (Fig. 2a, b). Prp8 has undergone a large structural change including a 30° rotation of the foot with respect to the Large domain when compared to U4/U6.U5 tri-snRNAP19 (Extended Data Fig. 7). U4 snRNA and its associated proteins have been released upon unwinding of the U4/U6 duplex by Br2 (ref. 6). The 3′-domain of U2 snRNA comprising Msl1(U2B′)25, Lea1(U2A′) and the Sm core domain bridges the Prp8 RNaseH-like domain and the N-terminal HAT (Half-a-TPR)-repeat domain of Syf1 (Fig. 2a). Isy1 and Cef1 dock with the N-terminal and reverse transcriptase (RT)-like domains of Prp8 (ref. 21), respectively, and anchor the N-terminal end of Clf1 together with Prp45/Prp46 (Fig. 2c, d). These interactions support the HAT-repeat arches of Syf1 and Clf1 suspended over the Large domain of Prp8. The 5′ part of U2 snRNA and the 3′ part of U6 snRNA run side-by-side from the active site forming nine consecutive base-pairs extending towards the centre of the Syf1 HAT-repeat arch (Fig. 2a–e). Bud31 anchors the 5′-stem of U6 snRNA to the N-terminal domain of Prp8 (Fig. 2c). Cwc2 is wedged between Bud31, Ecm2 and Prp45 and guides the path of U6 snRNA22 (Fig. 2c). U2 snRNA downstream of the branch helix extends from the active site towards the 3′-domain of U2 snRN, forming two stems bridging the U2 Sm ring with Ecm2/Cwc2 and the main body of the complex (Fig. 2d, e). Density for two RNA helices emanating from the U2 Sm ring is consistent with a stem–loop IIb/stem IIc arrangement and the catalytically competent conformation of the active site22,23 (Fig. 2f). The C-terminal region of Cwc2 forms a coiled-coil that interacts with Snu114 (ref. 25) (Fig. 2a) while the N-terminal half of Cwc2 extends towards Prp8 and points into the U5 snRNA stem minor groove.

Two large regions of weak density extend from the well-ordered core of the complex (Extended Data Fig. 1e). Focused classification allowed us to select subsets of particles (core + helicase, core + Prp19) (Extended Data Fig. 2), in which less well-ordered components can be more clearly visualized. The weak density observed in the latter class is readily attributable to Prp19, Cef1 and Snt309 based on its distinct

Figure 2 | Overview of the core structure. a, Prp8 and its central role in organizing the entire assembly (SII denotes U2/U6 stem II), b, RNA only in the same orientation as in a (ISL, U6 snRNA internal stem–loop; 5′SL, U6 snRNA 5′ stem–loop; SL1, U5 snRNA stem-loop 1; VSL, U5 snRNA variable stem–loop; S3, U5 snRNA Stem III). c, Ecm2, Cwc2 and Bud31 binding to the 5′ end of the U6 snRNA. d, Top view of the complex. e, RNA only in the same orientation as in d. f, Secondary structure diagram for the 3′ end of U2 snRNA. Prp8B, Prp8A and Prp8C denote N-terminal, Large and RNaseH-like domains of Prp8.

Figure 3 | Structure of the RNA catalytic core. a, Key RNA elements at the active site. BP, branch point; ISL, internal stem–loop; M1 and M2, catalytic metal ion one and two. b, Orthogonal view illustrating the branch helix and helices Ia and Ib of U2/U6 snRNA duplex. c, The branch helix and 5′-exon with the 2′–5′ phosphodiester linkage (red arrow). d, Intricate RNA interactions at the active site (dotted lines indicate base triples; dot and star indicate G-U wobble and other non-canonical base-pairs). e, Base triple interaction between the branch helix and 5′-splice site. f, A network of interactions in the branch helix. g, Hoogsteen base-pair between intron A(+3) and G50 of U6 snRNA.
shape first observed in ILS, but the weaker density in complex C suggests these proteins are more loosely attached to the core than in ILS. A large lobe corresponding to a DEAH helicase in contact with Cwc25 is observed near the intron exit channel, downstream of the branch point. Although its limited resolution does not allow us to build a model de novo, the density is of sufficient quality to fit a DEAH box helicase model unambiguously (Extended Data Fig. 6; Extended Data Table 2) and it has been interpreted as Prp16 as it contacts Cwc25. An even larger domain is observed in contact with the DEAH helicase domain. The structure of Br2 helicase coupled to the Jab1/MPN domain of Prp8 (ref. 27) can be docked into this density, consistent with an interaction between Prp16 and Br2 (ref. 28).

**Active site**

The map shows that the phosphodiester bond at the 5′SS is cleaved and the 5′-phosphate of the first intron nucleotide G(+1) forms a 2′–5′ phosphodiester linkage with the branch point adenosine (A70), in agreement with the RNA analysis (Extended Data Figs 1b and 4b). The key RNA elements assemble around the active site harbouring the magnesium ion binding sites (Fig. 3). The 3′OH of the 5′-exon remains close to the 5′-phosphate of G(+1) such that the normal 5′–3′ phosphodiester linkage at the 5′SS could be restored with minimal structural alteration (Fig. 3c). The adenine base of branch point A70 is bulged out from the branch helix and its N1 and 6-amino group remain close to the 5′OH of A70 to project an interaction between Prp16 and Brr2 (ref. 28).

The map shows that the phosphodiester bond at the 5′-exon 3′-exon junction is ligated and the 5′-phosphate of G(+1) remains close to Mg2+ ions (Extended Data Fig. 5). The 5′-exon 3′OH and the 5′-phosphate of G(+1) remain close to M1, while U6 snRNA metal ligands have repositioned slightly, in agreement with the previously observed repositioning of the branch in structures of a branched group II intron. Nonetheless, the branch helix remains ‘locked’ at the catalytic Mg2+ site, in contrast to its ‘undocked’ configuration observed in the ILS structure, where it swings away from the ACAGAGA helix by 90° (ref. 26; Extended Data Fig. 5).

The intron downstream of the 5′SS GUAUGU sequence exits the active site near Cwc2, Ecm2, Clf1, Cef1 and Isy1 (Fig. 2), re-enters the spliceosome and runs side-by-side with U2 snRNA in the opposite direction through a channel between the Prp8 Endonuclease and RNaseH-like domains (Extended Data Fig. 7). The intron then forms the branch helix with the GUAUGU sequence of U2 snRNA in proximity to the catalytic Mg2+ site (Fig. 3b, d) and exits the active site through a channel made by the linker and RT-like domains of Prp8 (Fig. 2).

**Roles of proteins around the active site**

The RNA network at the active centre, comprising U2, U5 and U6 snRNAs and RNA substrate, is stabilized by a number of proteins (Figs 1, 2 and 4). The catalytic RNA core is surrounded by the linker and Cwc22.

---

**Figure 4 | Proteins at the active site.**

a, 5′-exon channel formed between the Large and N-terminal domains of Prp8, Cwc21 and Cwc22. b, 5′-exon–U5 loop 1 interaction surrounded by Prp8. Th/X denotes thumb domain X of Prp8 (residues 1,300–1,375). c, Interactions between the 5′-exon, the N-terminal (purple) and Large (blue) domains of Prp8, and Ytu2 (green). Interactions involving protein main and side chains are shown by solid and dotted lines. d, Components surrounding U6 internal stem–loop. e, Prp8 and Cef1 (Myb1 domain) stabilize the catalytic triplex. HB denotes helix bundle of the RT-like domain (residues 750–870). f, Structure of the catalytic triplex.
point for branching. These spliceosomal factors are reminiscent of ribosomal proteins L27 and L16, which penetrate into the peptidyl transferase active site and stabilize tRNA binding.

**Remodelling of the spliceosome**

The intron downstream of the branch point emerges from the exit channel formed by the Prp8 RT-like and linker domains and the α-finger, and projects towards Prp16 (Fig. 6a). Twelve nucleotides could span the distance between the last ordered intron nucleotide (branch point + 6) and the substrate RNA entry site of Prp16, consistent with Prp16 crosslinking to 4-thiouridine introduced 18 nucleotides downstream of the branch point. Prp16 translocates 3′→5′ towards the branch point along the intron upon ATP hydrolysis. Prp16 would thus pull the branch helix out of its pocket and hence destabilize the binding of Yju2 and Cwc25 (Fig. 6b). The undocked branch helix would allow the 3′- exon to enter the active site and bind to U5 snRNA loop 1 (refs 11, 12). Consistent with this, destabilization of the branch helix by Isy1 deletion suppresses splicing defects caused by Prp16 mutations. The step-two factors Prp18 and Slu7 are likely to dock into the space vacated by the branch helix/Yju2/Cwc25 to stabilize the 3′- exon into the active site as Slu7 and Prp18 are in direct contact with the 3′- SS bound to U5 snRNA loop 1 before exon ligation (Fig. 6b). Prp22 binds the 3′- exon at position +17 (ref. 15). Translocation of Prp22 on the 3′- exon in the 3′→5′ direction towards the active centre would displace Prp18–Slu7, releasing the mRNA. In our structure, the density assigned to Prp16 is in direct contact with Cwc25 (Fig. 6a), consistent with Cwc25 stabilizing Prp16 binding to the spliceosome before branching. We propose that the branch helix and 3′- exon confer specificity for auxiliary factors such as Cwc25/Yju2, Slu7–Prp18, which may act as adaptors that determine the identity of the next DEAH box helicase to remodel the active site.

The structure of the *Schizosaccharomyces pombe* spliceosomal complex contains a lariat intron but not 5′- exon or the spliced mRNA. The catalytic RNA core is surrounded by a similar set of NTC and NTR proteins but the structure lacks step-one or step-two factors, suggesting this corresponds to a post-splicing ILS. Instead Cwf19, a homologue of the debranching enzyme co-factor Drn1 (ref. 50), intrudes between the Large and RNaseH-like domains of Prp8, occupying the binding sites for Isy1, Cwc25, and Yju2 found in our complex C. Cwf19 marks the ILS complex for disassembly by displacing the branch helix, which rotates by 90° in ILS with respect to complex C (Fig. 5c, Extended Data Fig. 7).

A pronounced conformational change between ILS and complex C is a large rotation of the NTC (Extended Data Fig. 7d). In ILS the N terminus of Syf1 moves away from the core, promoting undocking of U2 snRNP. In complex C, the position of U2 snRNP is stabilized by the formation of stem IIC and binding of Prp19. U2 snRNP is in direct contact with the RNaseH-like domain of Prp8, which holds Cwc25 in place. This network of interactions suggests that binding of Prp19 and formation of stem IIC in U2 snRNA may have an allosteric effect on the positioning of the branch helix via step-one factors. Extended arches of Syf1 and Clf1 may have a role in communicating the signal over long distance.

Our spliceosomal complex C structure reveals the active configuration of the catalytic core, elucidating the arrangement of the RNA helicase domains of Prp8 (refs 19, 21) on one side and by NTC proteins (Prp45, Prp46, Isy1 and Cef1) on the other side, which together stabilize the catalytic RNA core for branching. Remarkable stacking of Prp8 Tyr671 and Tyr1620 against bases at positions G(−5) and A(−6) stabilizes the 5′- exon:U5 snRNA loop 1 pairing (Fig. 4c). Corresponding view in *S. pombe* spliceosomal complex shows marked repositioning of the branch helix and its further stabilization by debranching co-factor Cwf19. A close-up view of step-one factors interacting with the branch helix and the helix bundle domains of Prp8 and N terminal and Large domains of Prp8 runs across the major groove of U6 ISL, which is positioned in a pocket formed by Prp8 and Clf1, and the interactions are sealed by the extended N terminus of Cwc25 (Fig. 4d). Cef1 stabilizes the U2/U6 catalytic triplex, suggesting this corresponds to a post-splicing ILS. Instead Cwf19, a homologue of the debranching enzyme co-factor Drn1 (ref. 50), intrudes between the Large and RNaseH-like domains of Prp8, occupying the binding sites for Isy1, Cwc25, and Yju2 found in our complex C. Cwf19 marks the ILS complex for disassembly by displacing the branch helix, which rotates by 90° in ILS with respect to complex C (Fig. 5c, Extended Data Fig. 7).

A pronounced conformational change between ILS and complex C is a large rotation of the NTC (Extended Data Fig. 7d). In ILS the N terminus of Syf1 moves away from the core, promoting undocking of U2 snRNP. In complex C, the position of U2 snRNP is stabilized by the formation of stem IIC and binding of Prp19. U2 snRNP is in direct contact with the RNaseH-like domain of Prp8, which holds Cwc25 in place. This network of interactions suggests that binding of Prp19 and formation of stem IIC in U2 snRNA may have an allosteric effect on the positioning of the branch helix via step-one factors. Extended arches of Syf1 and Clf1 may have a role in communicating the signal over long distance.

Our spliceosomal complex C structure reveals the active configuration of the catalytic core, elucidating the arrangement of the RNA helicase domains of Prp8 and N terminal and Large domains of Prp8 runs across the major groove of U6 ISL, which is positioned in a pocket formed by Prp8 and Clf1, and the interactions are sealed by the extended N terminus of Cwc25 (Fig. 4d). Cef1 stabilizes the U2/U6 catalytic triplex, suggesting this corresponds to a post-splicing ILS. Instead Cwf19, a homologue of the debranching enzyme co-factor Drn1 (ref. 50), intrudes between the Large and RNaseH-like domains of Prp8, occupying the binding sites for Isy1, Cwc25, and Yju2 found in our complex C. Cwf19 marks the ILS complex for disassembly by displacing the branch helix, which rotates by 90° in ILS with respect to complex C (Fig. 5c, Extended Data Fig. 7).

A pronounced conformational change between ILS and complex C is a large rotation of the NTC (Extended Data Fig. 7d). In ILS the N terminus of Syf1 moves away from the core, promoting undocking of U2 snRNP. In complex C, the position of U2 snRNP is stabilized by the formation of stem IIC and binding of Prp19. U2 snRNP is in direct contact with the RNaseH-like domain of Prp8, which holds Cwc25 in place. This network of interactions suggests that binding of Prp19 and formation of stem IIC in U2 snRNA may have an allosteric effect on the positioning of the branch helix via step-one factors. Extended arches of Syf1 and Clf1 may have a role in communicating the signal over long distance.

Our spliceosomal complex C structure reveals the active configuration of the catalytic core, elucidating the arrangement of the RNA helicase domains of Prp8 and N terminal and Large domains of Prp8 runs across the major groove of U6 ISL, which is positioned in a pocket formed by Prp8 and Clf1, and the interactions are sealed by the extended N terminus of Cwc25 (Fig. 4d). Cef1 stabilizes the U2/U6 catalytic triplex, suggesting this corresponds to a post-splicing ILS. Instead Cwf19, a homologue of the debranching enzyme co-factor Drn1 (ref. 50), intrudes between the Large and RNaseH-like domains of Prp8, occupying the binding sites for Isy1, Cwc25, and Yju2 found in our complex C. Cwf19 marks the ILS complex for disassembly by displacing the branch helix, which rotates by 90° in ILS with respect to complex C (Fig. 5c, Extended Data Fig. 7).

A pronounced conformational change between ILS and complex C is a large rotation of the NTC (Extended Data Fig. 7d). In ILS the N terminus of Syf1 moves away from the core, promoting undocking of U2 snRNP. In complex C, the position of U2 snRNP is stabilized by the formation of stem IIC and binding of Prp19. U2 snRNP is in direct contact with the RNaseH-like domain of Prp8, which holds Cwc25 in place. This network of interactions suggests that binding of Prp19 and formation of stem IIC in U2 snRNA may have an allosteric effect on the positioning of the branch helix via step-one factors. Extended arches of Syf1 and Clf1 may have a role in communicating the signal over long distance.

Our spliceosomal complex C structure reveals the active configuration of the catalytic core, elucidating the arrangement of the RNA helicase domains of Prp8 and N terminal and Large domains of Prp8 runs across the major groove of U6 ISL, which is positioned in a pocket formed by Prp8 and Clf1, and the interactions are sealed by the extended N terminus of Cwc25 (Fig. 4d). Cef1 stabilizes the U2/U6 catalytic triplex, suggesting this corresponds to a post-splicing ILS. Instead Cwf19, a homologue of the debranching enzyme co-factor Drn1 (ref. 50), intrudes between the Large and RNaseH-like domains of Prp8, occupying the binding sites for Isy1, Cwc25, and Yju2 found in our complex C. Cwf19 marks the ILS complex for disassembly by displacing the branch helix, which rotates by 90° in ILS with respect to complex C (Fig. 5c, Extended Data Fig. 7).

A pronounced conformational change between ILS and complex C is a large rotation of the NTC (Extended Data Fig. 7d). In ILS the N terminus of Syf1 moves away from the core, promoting undocking of U2 snRNP. In complex C, the position of U2 snRNP is stabilized by the formation of stem IIC and binding of Prp19. U2 snRNP is in direct contact with the RNaseH-like domain of Prp8, which holds Cwc25 in place. This network of interactions suggests that binding of Prp19 and formation of stem IIC in U2 snRNA may have an allosteric effect on the positioning of the branch helix via step-one factors. Extended arches of Syf1 and Clf1 may have a role in communicating the signal over long distance.
substrate and its interaction with proteins. The structure accounts for a large body of biochemical and genetic data and provides crucial insights into substrate docking and catalysis and the role of DEAH helices and auxiliary factors in spliceosome remodelling.

**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 9 June; accepted 19 July 2016.**

1. Will, C. L. & Lührmann, R. Spliceosome structure and function. Cold Spring Harb. Perspect. Biol. 3, a003707 (2011).
2. Burge, C. B., Tuschl, T. & Sharp, P. A. in The RNA World II (eds Gesteland, R. F., Dattilo, J. T., Dineen, J., Gesteland, R. F. & Tautz, D.) 367–391 (Cold Spring Harbor Laboratory Press, 2019).
3. Lambowitz, A. M. & Zimmerly, S. RNA helicase activity defines the spliceosome. Cell 35, 593–608 (2009).
4. Lesser, C. F. & Guthrie, C. Mutations in U6 snRNA that alter splice site selection. Mol. Cell 32, 699–707 (2008).
5. Ohi, M. D. & Gould, K. L. Characterization of interactions among the Cef1p-Mcm10 complex and the spliceosome. Mol. Cell 35, 349–360 (2009).
6. Kandels-Lewis, S. & Séraphin, B. Involvement of U6 snRNA in 5′ splice site selection. Genes Dev 3, 1659–1671 (1989).
7. Nyborg, M., Steitz, J. A. & Brown, D. L. U1 snRNA remodeling during spliceosome assembly. Mol. Cell 53, 1059–1070 (2014).
8. Lesser, C. F. & Guthrie, C. Mutations in U6 snRNA that alter splice site selection. Mol. Cell 32, 699–707 (2008).
9. Raghunathan, P. L. & Guthrie, C. RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEAH-box splicing factor Bre2. Curr. Biol. 8, 847–855 (1998).
10. Lüggerbauer, L., Achtel, T. & Lührmann, R. The human U5-200kD DEAH-box protein unwindnig in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Bre2. Curr. Biol. 8, 847–855 (1998).
11. Grainger, R. J., Barrass, J. D., Jacquier, A., Rain, J.-C. & Beggs, J. D. Physical and genetic interactions of yeast Cwc21p, an ortholog of human SRm300/SRRM2, within the spliceosome promotes multiple steps in splicing. Mol. Cell. Biol. 25, 1451–1467 (2005).
12. Fica, S. M., Mei, H. L. & Cheng, S. C.-E. DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. RNA 17, 145–154 (2011).
13. Schmitt, M. E., Warkocki, Z. & Guthrie, C. A conformational rearrangement in the spliceosome is required for NTR-mediated spliceosome disassembly with DEAH-box ATPases Prp2, Prp16, and Prp22. Mol. Cell. Biol. 31, 514–523 (2015).
14. Garvey, S. M. et al. A homolog of lariat-debranching enzyme modulates turnover of branched RNA. RNA 20, 1330–1345 (2014).
15. Acknowledgements We thank K. Nguyen, X.-C. Bai and S. Scheres for their help with advice and processing of the manuscript. K. N. and Y. F. Chiu, Y.-F. and A. R. Robart, A. R., Chan, R. T., Peters, J. K., Rajashankar, K. R. & Toor, N. Crystal structure of a group II intron-like catalytic triplex in the spliceosome. Nat. Struct. Mol. Biol. 6, 529–533 (2009).
16. Rybak, D. M. & Guthrie, C. The Isy1p component of the NineTeen complex interacts with Prp19p-associated splicing complex. Mol. Cell. Biol. 28, 279–282 (2002).
17. Tsai, R.-T. & 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).
18. Fica, S. M., Melford, 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).
19. Steitz, T. A. & Steitz, J. A. A general two-metal-ion mechanism for catalytic RNA. Proc. Natl Acad. Sci. USA 90, 6489–6502 (1993).
20. Fica, S. M. et al. RNA catalyses nuclear pre-mRNA splicing. Nature 503, 219–234 (2013).
21. 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).
22. Buchwald, G., Schüssieller, S., Basquin, C., Le Hir, H. & Conti, E. Crystal structure of the human eIF4F1-CWC22 complex shows how a DEAD-box protein is inhibited by a MIF4G domain. Proc. Natl Acad. Sci. USA 110, E4611–E4618 (2013).
23. 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).
24. Fica, S. M., Melford, 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).
25. Steitz, T. A. & Steitz, J. A. A general two-metal-ion mechanism for catalytic RNA. Proc. Natl Acad. Sci. USA 90, 6489–6502 (1993).
26. Fica, S. M. et al. RNA catalyses nuclear pre-mRNA splicing. Nature 503, 219–234 (2013).
27. Robart, A. R., Chan, R. T., Peters, J. K., Rajashankar, K. R. & Toor, N. Crystal structure of a eukaryotic group II intron lariat. Nature 514, 193–197 (2014).
28. Chiu, Y.-F. & Cheng, S. C.-E. CryoEM structures of two spliceosomal complexes: starter and dessert at the spliceosome feast. Curr. Opin. Struct. Biol. 36, 48–57 (2016).
29. Garvey, S. M. et al. A homolog of lariat-debranching enzyme modulates turnover of branched RNA. RNA 20, 1330–1345 (2014).

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

**Author Contributions** W.P.G., M.E.W. and S.M.F. prepared the sample and grids, and processed EM data. W.P.G., M.E.W. and S.M.F. collected EM data. W.P.G., M.E.W. and C.D. carried out model building and refinement. W.P.G., M.E.W., S.M.F. and K.N. analyzed the structure. A.N.J. contributed to the project through his knowledge and experience on yeast spliceosing. Manuscript was written by W.P.G., M.E.W. and K.N. and finalized with input from all authors. K.N. initiated and orchestrated the spliceosome project.

**Article Information** The cryo-EM maps have been deposited in the Electron Microscopy Data Bank with accession codes EMD-4055, EMD-4056, EMD-4057, EMD-4058 and EMD-4059. The coordinates of the atomic models have been deposited in the Protein Data Bank under accession code 5LJ3 (core of the complex) and 5L5J (overall structure). 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 this paper. Correspondence and requests for materials should be addressed to W.P.G. (wpizable@rmc-lmb.cam.ac.uk) or K.N. (kn@rmc-lmb.cam.ac.uk).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
METHODS

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

Prp18-HA and Slu7-TAPS tagging. SLU7-TAPS homology recombination cassettes were generated by PCR from pFA6a-Taps-kanMX6, a modified version of pFA6a-TAP-kanMX6 in which the Calmodulin-binding peptide tag is replaced by two amino-terminal copies of the StreptII tag17. The PCR product was used to transform yeast strain YSC1 (MATa leu2-3,112 trp1-289 ura3-52 Prp19-HA) selecting for G418-resistance. PRP18-3xHA kanMX6 cassette was transformed into BY4741 strain (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0) and selected as above. Integration of the cassettes was confirmed by PCR and western blotting.

Sample preparation. The Prp18-HA or Slu7-TAPS yeast strains were grown in a 120 L fermenter, and splicing extract was prepared using liquid nitrogen method18 essentially as previously described. A DNA template for in vitro transcription was generated by addition of 2 × MS2 stem loops19 to the 5′-end of the UBC4 pre-mRNA sequence20, in which the 3′-splice site sequence UAGAG was mutated to UACAC. Pre-mRNA substrate was generated by run-off transcription from a plasmid DNA template and labelled at the 3′-end with fluorescein-5-thiosebacic acid21. *In vitro* splicing reactions were assembled using pre-mRNA substrate pre-bound to MS2-MBP fusion protein as previously described. The resulting spliceosomes were observed by amyllose resin in HE-75 (20 mM HEPES KOH pH 7.8, 75 mM KCl, 0.25 mM Mg EDTA, 5% glycerol, 0.01% NP-40) and eluted with 12 mM Maltose. The sample was subsequently immobilized on either anti-HA-agarose (for Prp18-HA yeast extract) or Streptacin resin (for Slu7-TAPS yeast extract) in HE-100 (20 mM HEPES KOH pH 7.8, 100 mM KCl, 0.25 mM Mg EDTA, 5% glycerol, 0.01% NP-40) and eluted with either HA peptide (for anti-HA-agarose) or des-thiobiosticin (for Streptacin resin), essentially as described. The eluate was finally dialysed against HE-75 buffer (without glycerol and NP-40) for EM sample preparation. Analysis of fluorescently labelled RNA showed that pre-mRNA is converted to the lariat intron−3′-exon intermediate in our sample and hence it is referred to as complex C (Extended Data Fig 1b).

Our experimental set-up was designed to purify step-two complexes after Prp16 action, however the presence of step-one factors in the structure and configuration of the active site clearly indicate that the complex has not undergone Prp16-mediated remodelling. It has been shown previously that in low salt conditions Prp18, Slu7 and Prp16 associate with complex B8 and C. Analysis of protein components by gel electrophoresis and subsequent mass spectrometry shows that Prp16 as well as Prp22 are present, in agreement with the previous results (Extended Data Fig 1a; Extended Data Table 2)13,43.

Electron microscopy. For cryo-EM analysis, Quantifoil R2/2 Cu 400 mesh grids were coated with a 5–7 nm-thick layer of homemade carbon film and glow-discharged. After applying 3 μl of the sample, the grids were blotted for 2.5–3 s and vitrified in liquid ethane in FEI Vitrobot MKIII, at 100% humidity at 4 °C. Grids were loaded into an FEI Titan Krios transmission electron microscope operated at 300 kV in counting mode at 1.25 frames s−1 and a calibrated pixel size of 1.43 Å. A total dose of 40 e− Å−2 over 16 s and a defocus range of 0.5–4.5 μm were used.

Image processing. A total of 2213 micrographs were subjected to whole-frame drift correction in MOTIONCORR20 followed by contrast transfer function (CTF) parameter estimation in CTFFIND4 (ref. 57). All subsequent processing steps were done using RELION58 unless otherwise stated. An initial subset of 5,000 particles was selected manually and subjected to reference-free 2D classification. Resulting 2D class averages were low-pass filtered to 20 Å and used as templates for subsequent automated particle picking within RELION58. A total of 247,603 particles were selected after initial reference-free 2D classification and subjected to 3D classification (Extended Data Fig 2). An initial 3D reference was prepared by scaling and low-pass filtering (60 Å) the reconstruction of the intron−lariat complex (EMD-6413). A subset of 93,106 particles was selected after 3D classification. Particle-based beam-induced motion correction and radiation-damage weighting (particle polishing) followed by 3D refinement resulted in a final resolution at 3.8 Å overall resolution and estimated accuracies of rotations of 1.1° (Extended Data Fig 3).

Very weak density observed at two peripheral regions of the map corresponds to Brz2/Prp16 (helicase module) and Prp19/Cef1/Snt309 (Prp19 module). We used focused classification with signal extraction to improve the resolution of these regions. The region of interest was masked out and the projection of the remaining map was subtracted from the experimental particles using angular assignment from the last iteration of the 3D auto-refine run. Subtracted particles were 3D classified without image alignment and the best classes were selected for further refinement of the original (not subtracted) particles. This resulted in a smaller subset of the original particles, in which Brz2/Prp16 and Prp19/Cef1/Snt309 were more homogeneous and the density is improved in regions (Extended Data Figs 2 and 3). 3D refinement of the selected 29,210 Prp19-selected particles resulted in a map at overall 5.1 Å resolution, while 15,872 of the helicase-containing particles yielded a map at 10 Å resolution. For the global classification approach we generated a soft mask around the core of the complex and classified polished particles with finer angular sampling of 1.8° and local searches of 10°. The resulting two major classes of 37K and 47K particles were superimposed and the best classes were selected. They revealed a considerable change of the U2 snRNP and Syt1 HAT arch correlated with the presence of WD40 domain near the stem IIC and IIB region of U2 snRNA. This WD40 domain belongs to Ptp17 or Prp19, but the local resolution did not allow us to make an unambiguous assignment. All reported resolutions are based on the gold-standard Fourier shell correlation (FSC) = 0.143 criterion54. FSC curves were calculated using soft spherical masks and high-resolution noise substitution was used to correct for convolution effects of the masks on the FSC curves54. Prior to visualization, all maps were corrected for the modulation transfer function of the detector. Local resolution was estimated using RESMAP59.

Model building. A list of protein and RNA components included in the model is given in Extended Data Table 2. Building started by docking known structures of S. cerevisiae Prp8, Snu14, US snRNase, Cwc2 (ref. 64) and Bud31 (ref. 65) into the map. Homology models for Cef1, Prp45, Prp46, Ecma2 and Cwc15 were built with SWISS-MODEL60, using structures from the *S. pombe* intron−lariat spliceosome61 as templates, and were docked into the map. This accounted for the majority of the protein density in the core, allowing building of the intron, U6 snRNA and U2 snRNA. RNA extending from the loop 1 of U5 snRNA was assigned to nucleotides −1 to −16 of the 5′-exon as previously predicted. A model for the NTD of Cwc22 was built using SWISS-MODEL based on the structure of the human Cwc22/EF4AAIII complex33 and docked near Snu14. Clear density near the NTD of Cwc22 was interpreted as the MA3 domain at the C terminus of Cwc22; this domain was built de novo. A coiled-coil was found contacting domain IV of Snu14. Based on an unpublished NMR structure from Arabidopsis thaliana (PDB ID: 2E62) and biochemical data62 we assigned this density to the CTD of Cwc22. Weak density was observed connecting this coiled-coil to a peptide contacting the 5′-exon. We therefore assigned this peptide as the N terminus of Cwc21. Unassigned density remained near the branch-point helix. Based on secondary structure prediction62 we assigned a portion of this density to Yju2 and were able to build its NTD de novo; our assignment was supported by clear density for a zinc atom coordinated by four conserved cysteines. The remainder of the density could then be assigned to the N termini of Cwc25 and Isy1.

The majority of the model building described above was for the core of the spliceosome where the resolution was uniformly between 3.5−4.5 Å (Extended Data Fig 4). For the periphery of the complex, the resolution was more heterogeneous, ranging from 4 to 20 Å. Clear features of the periphery were two large proteins with extended architectures. One of these proteins started in the core and projected outwards to the periphery. At the core, side-chains were easily visible for this protein and allowed an assignment as the N terminus of Cfi1. Towards the C terminus of Cfi1 the resolution only allowed building of idealised poly-α-helical helices, which were then assigned sequence based on secondary structure predictions67. For the other extended protein, few side-chains were visible but helices could be distinguished. This protein was generally built as poly-α-helical helices, and based on secondary structure predictions62 was assigned as Syt1. A second Sm ring at medium-resolution was found in the map and was assigned as the U2 snRNA Sm ring. Homology models for the U2 snRNP proteins Lea1 and Ms1 were generated using SWISS-MODEL60 based on the structure of the human U2B−U2A−U2 snRNA complex63 and were docked into the adjacent density. The portion of the U2 snRNA in contact with Ms1 was most consistent with the previously proposed stem IV + stem V architecture and was built based on the secondary structure prediction69. Two RNA double helices were observed bridging the U2 Sm ring to Ecma2 and were assigned as stems IIB and IIC of the U2 snRNA. Using 3D classification, we found that some of the particles contained a large lobe of extra density connected to the RT-like and RNAseH-like domains of Prp8 (see above). Although we could not resolve secondary structure in this region, we could perfectly dock the crystal structure of Brr2 and the Jab1/MPN domain of Prp8 (ref. 27). The remainder of the density could then well accommodate an I-TASSER60 homology model of Prp16 based on the crystal structure of Prp43 (ref. 71). Weak density connected to Cfi1 and Syt1 had the characteristic shape of Prp19−SnT309−Cef1 (ref. 26). Focused classification in this region could improve the density enough to resolve the U-box dimers and thus dock a homology model of these proteins. Finally, three copies of the Prp19 WD40 domain crystal structure could be docked into very weak density adjacent to the Prp19 coiled-coils. With the exception of the helicase and Prp19 modules all models were manually rebuilt in order to obtain the best fit to the cryo-EM density. The model was refined using REFMAC5 (ref. 72).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
with secondary structure restraints generated in PROSMART\textsuperscript{25} and RNA base-pair and stacking restraints generated in LIBG\textsuperscript{14}. Extended Data Table 1 summarizes refinement statistics and PBD and EMD accession codes.

**Map visualization.** Maps were visualized in Chimera\textsuperscript{75} and figures were prepared using PyMOL (http://www.pymol.org).

51. Schreieck, A. et al. RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPT subunit Gic7. Nat. Struct. Mol. Biol. 21, 175–179 (2014).
52. Lin, R. J., Newman, A. J., Cheng, S. C. & Abelson, J. Yeast mRNA splicing in vitro. J. Biol. Chem. 260, 14780–14792 (1985).
53. Zhou, Z., Licklider, L. J., Gygi, S. P. & Reed, R. Comprehensive proteomic analysis of the human spliceosome. Nature 419, 182–185 (2002).
54. Hardin, J. W., Warnasooriya, C., Kondo, Y., Nagai, K. & Rueda, D. Assembly and dynamics of the U4/U6 di-snRNP by single-molecule FRET. Nucleic Acids Res. 43, 10963–10974 (2015).
55. Nguyen, T. H. D. et al. The architecture of the spliceosomal U4/U6.U5 tri-snRNP. Nature 523, 47–52 (2015).
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. CTFIND4: Fast and accurate defocus estimation from electron micrographs. J. Struct. Biol. 192, 216–221 (2015).
58. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. 180, 519–530 (2012).
59. Scheres, S. H. W. Semi-automated selection of cryo-EM particles in RELION-3. J. Struct. Biol. 189, 114–122 (2015).
60. 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, 1485 (2015).
61. Scheres, S. H. W. & Chen, S. Prevention of overfitting in cryo-EM structure determination. Nat. Methods 9, 853–854 (2012).
62. Chen, S. et al. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. Ultramicroscopy 135, 24–35 (2013).
63. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. Nat. Methods 11, 63–65 (2014).
64. Lu, P. et al. Structure of the mRNA splicing complex component Cwc2: insights into RNA recognition. Biochem. J. 441, 591–597 (2012).
65. van Roon, A.-M. M. et al. \textsuperscript{113}Cd NMR experiments reveal an unusual metal cluster in the solution structure of the yeast splicing protein Bud31p. Angew. Chem. Int. Edn Engl. 54, 4861–4864 (2015).
66. Biasini, M. et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 42, W252–W258 (2014).
67. Kurowski, M. A. & Bujnicki, J. M. GeneSilico protein structure prediction meta-server. Nucleic Acids Res. 31, 3305–3307 (2003).
68. Price, S. R., Evans, P. R. & Nagai, K. Crystal structure of the spliceosomal U28"-U24" protein complex bound to a fragment of U2 small nuclear RNA. J. Biol. Chem. 283, 455–460 (2008).
69. Kudla, G., Granneman, S., Hahn, D., Beggs, J. D. & Tollervey, D. Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast. Proc. Natl Acad. Sci. USA 108, 10010–10015 (2011).
70. Yang, J. et al. The I-TASSER Suite: protein structure and function prediction. Nat. Methods 12, 7–8 (2015).
71. Walbott, H. et al. Prp43p contains a processive helicase structural architecture with a specific regulatory domain. EMBO J. 29, 2194–2204 (2010).
72. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecules by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. 53, 240–255 (1997).
73. Nicholls, R. A., Fischer, M., McNicholas, S. & Murshudov, G. N. Conformation-independent structural comparison of macromolecules with ProSMART. Acta Crystallogr. D Biol. Crystallogr. 70, 2487–2499 (2014).
74. Brown, A. et al. Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. Acta Crystallogr. D Biol. Crystallogr. 71, 136–153 (2015).
75. Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. J. Struct. Biol. 157, 281–287 (2007).
76. Marcia, M. & Pyle, A. M. Visualizing group II intron catalysis through the stages of splicing. Cell 151, 497–507 (2012).
77. Le Hir, H., Saulière, J. & Wang, Z. The exon junction complex as a node of post-transcriptional networks. Nat. Rev. Mol. Cell Biol. 17, 41–54 (2016).
78. Bessonov, S., Anokhina, M., Will, C. L., Urlaub, H. & Lührmann, R. Isolation of an active step 1 spliceosome and composition of its RNP core. Nature 452, 846–850 (2008).
79. Alexandrov, A., Colognori, D., Shu, M.-D. & Steitz, J. A. Human spliceosomal protein CWC22 plays a role in coupling splicing to exon junction complex deconstruction and nonsense-mediated decay. Proc. Natl Acad. Sci. USA 109, 21313–21318 (2012).
80. Steckelberg, A.-L., Boehm, V., Gromadzka, A. M. & Gehring, N. H. CWC22 connects pre-mRNA splicing and exon junction complex assembly. Cell Reports 2, 454–461 (2012).
81. Barbosa, I. et al. Human CWC22 escorts the helicase eIF4AIII to spliceosomes and promotes exon junction complex assembly. Nat. Struct. Mol. Biol. 19, 983–990 (2012).
82. Bono, F., Ebert, J., Lorentzen, E. & Conti, E. The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. Cell 151, 2194–2204 (2012).
83. Andersen, C. B. F. Crystallographic and model-based analysis of the exon junction complex. Nature 441, 713–725 (2006).
84. Davis, I. W., Murray, L. W., Richardson, J. S. & Richardson, D. C. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Res. 32, W615–W619 (2004).
Extended Data Figure 1 | Biochemical characterization of the complex and initial cryo-EM analysis. a, SDS–PAGE analysis of the purified sample. Protein identities were confirmed by mass spectrometry analysis. Protein labels are coloured according to sub-complex identity (dark blue, U5 snRNP; light blue, helicase module; orange, NTC; yellow, NTR; green, U2 snRNP; purple, splicing factors; grey, not found in density). b, Analysis of the fluorescently labelled substrate in the sample by denaturing PAGE, showing conversion of linear pre-mRNA (time point 0') into branched lariat-intron intermediate (time point 30'), which is a predominant species in the purified sample (C complex). The two hairpins on the right depict the 2 × MS2 stem–loops attached to the 5' end of the UBC4 pre-mRNA substrate for affinity purification. c, A typical cryo-EM micrograph collected on an FEI Titan Krios microscope operated at 300 kV and detected with a Gatan K2 Summit camera. d, Reference-free 2D classification results. e, Detail of a single class average with major domains labelled.
Extended Data Figure 2 | Overview of the data processing scheme used in this study. Iterative 2D classification, template selection and automated particle picking resulted in 248K particles which were classified in 3D with a scaled and low-pass-filtered model of ILS (EMDB-6413) as a reference. The best class was refined to 3.8 Å resolution overall. Focused classification allowed us to obtain two other maps with improved quality of the peripheral regions (Prp19 and helicase modules, EMD-4056 and EMD-4057). Classification of the core complex with fine angular sampling and local searches revealed a subtle movement of the U2 snRNP which correlates with the appearance of the extra density, interpreted as a WD40 domain which belongs to Prp17 or Prp19.
Extended Data Figure 3 | Global and local resolution analysis. 

a, Two orthogonal sections through the map showing variation in the local resolution as estimated by Resmap. 
b, An overall map of the core complex. 
c, Gold-standard FSC plots for three maps used in this study. 
d, Map of the core complex with a helicase module. 
e, A map of the core complex with Prp19 module.
Extended Data Figure 4 | Examples of cryo-EM density at the core of the complex with atomic models built in. a, U5 snRNA loop 1 with 5′-exon bound. b, The active site with exon, intron, U2 and U6 snRNAs. c, Two helices of the Prp8 reverse transcriptase thumb/X domain, showing a clear helical pitch and excellent densities for the side chains. d, Fourier Shell Correlation between model and the map and cross-validation of the model fitting. (The original atom positions have been randomly displaced up to 0.5 Å and refined with restraints against the half1 map only. FSC was calculated for two half maps. Excellent correlation up to high resolution between the model and the half2 map (which was not used in refinement) cross-validates the model for overfitting.)
Extended Data Figure 5 | Metal binding by the catalytic core of C complex. a, b. Structure (a) and schematic representation (b) of the active site of a group IIC intron trapped in the pre-catalytic state in the presence of Ca^{2+} (PDB 4FAQ, ref. 76). The 5′ splice site scissile phosphate is aligned with the two metals bound at the core in a catalytic configuration, as shown in b. Note that, in this pre-catalytic structure, the group II domain VI is not present and therefore the structure does not contain the bulged adenosine nucleophile required for the branching reaction. As a result, the nucleophile is a water molecule, rather than the 2′-OH of the branch site adenosine found in spliceosomal introns. c, d, e. Structure of the RNA at the active site of spliceosomal C complex, showing the overall architecture (c), schematic of metal binding (d), and comparison of the model with the EM density (e). Note conservation of the metal binding residues compared to the group II intron (compare with ref. 36) and proximity of the cleaved G(−1)–G(+1) bond to putative M1. f, Proposed interactions between U6 snRNA and the two catalytic Mg^{2+} during the transition state for branching, as inferred from biochemistry. g, h. Structure (g) and schematic (h) of the RNA core of the U2.U6.U5 ILS complex in a post-catalytic configuration (PDB 3JB9, ref. 26), probably following release of the mRNA. The two Mg^{2+} are shown as modelled in the coordinates deposited by the authors of the ILS structure (PDB 3JB9, ref. 26). In the ILS structure M1 and M2 are further apart (7.2 Å) than in most other structures of RNAs that coordinate catalytic metals (usually 3.9–5 Å); nonetheless, the ligands modelled for M1 and M2 are consistent with the ligands identified biochemically for the two catalytic Mg^{2+} necessary for splicing (compare PDB 3JB9 and 4R0D with the data in refs 34,36). Note that the branch helix is undocked from the U6 snRNA metal binding site and G(+1) is far away from the two Mg^{2+} at the core. The substrate and snRNAs are colour-coded while residues that position the catalytic metals are shown in magenta.
Extended Data Figure 6 | Examples of the structures of isolated components. De novo-built proteins are shown in cartoon form, along with a secondary structure diagram for the novel zinc-finger fold of Yju2. Proteins that were modelled into low-resolution regions by rigid-body docking of crystal structures or homology models (Prp19 module, Brr2, Prp16, Prp8(Jab1/MPN)) are shown in their cryo-EM densities.
Extended Data Figure 7 | Conformational changes between U4/U6.U5 tri-snRNP, complex C and intron–lariat spliceosome. a, Rearrangement of the RNaseH-like domain with respect to the main body of Prp8 in all three complexes. b, α-Finger (1,575–1,598) contacting the key RNA and proteins in a context-dependent manner. c, Prp8 N-terminal domain movements along with Prp8 residues 1,406–1,436 transiently docking on top of the 5′-exon and Cwc21 in complex C, stabilizing the 5′-exon and interdomain contacts in Prp8. d, Conformational rearrangements between complex C and S. pombe ILS26 showing a coupled movement of the U2 snRNP, Syf1 and Prp19.
Extended Data Figure 8 | Implications for deposition of the exon–junction complex. In higher eukaryotes exon–junction complexes (EJCs) are deposited 20–24 nucleotides (nt) upstream of splice junctions, and form a binding platform for factors involved in nuclear export, translation, alternative splicing and nonsense-mediated mRNA decay. The core EJC components eIF4AIII, MAGOH and Y14 are found in human B and C complexes. Cwc22 is required for eIF4AIII recruitment to spliceosomes and holds it in an open, inactive conformation.

a, Crystal structure of the eIF4AIII–Cwc22 complex docked onto the spliceosomal C complex via superposition on Cwc22. b, Crystal structure of the core EJC superimposed on the previous model via the second RecA domain of eIF4AIII. c, The 5′-exon exiting the channel at the interface between the Prp8 Large and N-terminal domains is positioned perfectly for the deposition of the EJC, explaining how the Cwc22 MIF4G domain is involved in determining the distance of EJC deposition from the splice junction.
Extended Data Table 1 | Cryo-EM data collection and refinement statistics

|                         | Core | Core+Prp19 | Core+helicase |
|-------------------------|------|------------|--------------|
| **Data collection**     |      |            |              |
| Microscope              | FEI Titan Krios | FEI Titan Krios | FEI Titan Krios |
| Voltage (kV)            | 300  | 300        | 300          |
| Electron dose (e Å$^{-2}$) | 40   | 40         | 40           |
| Detector                | Gatan K2 Summit | Gatan K2 Summit | Gatan K2 Summit |
| Pixel size (Å)          | 1.43 | 1.43       | 1.43         |
| Defocus Range (µm)      | 0.5-4.0 | 0.5-4.0 | 0.5-4.0      |
| **Reconstruction (Relion)** |     |            |              |
| Particles               | 93 106 | 29 210    | 15 872       |
| Box size (pix)          | 412  | 412        | 412          |
| Accuracy of rotations (°) | 1.13 | 1.13       | 1.51         |
| Accuracy of translations (pix) | 0.64 | 0.96       | 1.30         |
| Map sharpening B-factor (Å$^{-2}$) | -57 | -17        | -350         |
| Final resolution (Å)    | 3.75 | 5.08       | 9.78         |
| **Model composition**   |      |            |              |
| Protein residues        | 7447 |            | 11978$^\dagger$ |
| RNA bases               | 458  |            | 458          |
| Ligands                 | 10   |            | 10           |
| **Refinement (Refmac)** |      |            |              |
| Resolution (Å)          | 3.8  |            |              |
| FSCaverage              | 0.82 |            |              |
| R factor                | 0.32 |            |              |
| **R.m.s deviations**    |      |            |              |
| Bond lengths (Å)        | 0.007 |          |              |
| Bond angles (°)         | 1.25 |            |              |
| **Validation$^\ddagger$** |    |            |              |
| Molprobity score        | 2.5 (98th percentile) |          |              |
| Clashscore, all atoms   | 5.3 (100th percentile) |          |              |
| Good rotamers (%)       | 80   |            |              |
| **Ramachandran plot**   |      |            |              |
| Favoured (%)            | 90.84 |          |              |
| Outliers (%)            | 1.16 |            |              |
| **RNA validation$^\ddagger$** | |            |              |
| Correct sugar puckers (%) | 95   |            |              |
| Good backbone conformations (%) | 60   |            |              |
| **Deposition**           |      |            |              |
| PDB ID                  | 5LJ3 | 5LJ5$^\ddagger$ | 5LJ5$^\ddagger$ |
| EMBD ID                 | EMBD-4055 | EMBD-4056 | EMBD-4057 |

$^*$Represents a sub-set of the whole data set (Core).
$^\ddagger$Determined by Molprobity.$^\dagger$Overall model including Prp19 and helicase modules.
Extended Data Table 2 | Summary of model building for spliceosomal complex C

| Sub-complexes | Protein/RNA | Domains | Total residues | M.W. (Da) | Modelled | Modelling template (PDB ID) | Modelling | Resolution | Chain ID | Human/ S. pombe names |
|---------------|-------------|---------|----------------|-----------|----------|-----------------------------|-----------|------------|----------|----------------------|
| US snRNP      | Prp8        | N-terminal | 1-870          | 101,767   | Docked & rebuilt | 3.4 - 5.8 | A | 220K/Smpr42 |
|               | Large       | 871-1827 | 111,525        | Docked & rebuilt | 3.6 - 6.2 |
|               | RNaseH      | 1828-2085 | 29,453         | Docked & rebuilt | 4.2 - 6.6 |
|               | Jab1/MPN    | 2086-2413 | 36,812         | Rigid docking | ~15 - 20 |
|               | Smn14       | 1008     | 114,041        | Docked & rebuilt | 3.8 - 7.2 | C | 116K/Cwff10 |
|               | SmB         | 196      | 22,403         | Docked | 4.6 - 7.2 | b | SmB/SmB |
|               | SmD3        | 110      | 11,229         | Docked | 4.4 - 7.8 | d | SmD3/SmD3 |
|               | SmD1        | 146      | 16,288         | Docked | 4.8 - 7.8 | h | SmD1/SmD1 |
|               | SmD2        | 110      | 12,856         | Docked | 5.2 - 8.0 | j | SmD2/SmD2 |
|               | SmF         | 94       | 10,373         | Docked | 5.2 - 8.0 | f | SmF/SmF |
|               | SmE         | 96       | 9,659          | Docked | 5.4 - 8.0 | e | SmE/SmE |
|               | SmG         | 77       | 8,479          | Docked | 5.0 - 7.8 | g | SmG/SmG |
| US snRNA-L    | U2 snRNA-L  | 68,847   | 4,144          | De novo | 3.8 - 7.6 | U |
| U2 snRNP      | Ms1         | 111      | 12,830         | 1A9N      | Homology modelled | 6.6 - 8.8 | Y | U2-8' |
|               | Lea1        | 238      | 27,193         | 1A9N      | Homology modelled | 5.6 - 8.6 | W | U2-A' |
|               | SmB         | 196      | 22,403         | Docked | 5.4 - 8.2 | k | SmB/SmB |
|               | SmD3        | 110      | 11,229         | Docked | 6.0 - 8.2 | n | SmD3/SmD3 |
|               | SmD1        | 146      | 16,288         | Docked | 5.0 - 8.0 | l | SmD1/SmD1 |
|               | SmD2        | 110      | 12,856         | Docked | 5.0 - 7.6 | m | SmD2/SmD2 |
|               | SmF         | 94       | 10,373         | Docked | 5.2 - 7.4 | q | SmF/SmF |
|               | SmE         | 96       | 9,659          | Docked | 5.4 - 8.0 | p | SmE/SmE |
|               | SmG         | 77       | 8,479          | Docked | 5.8 - 8.2 | r | SmG/SmG |

| U2 snRNA      | U2 snRNA    | 363,824  | 1-150          | 1089-1150 | De novo | 3.8 - 6.0 | Z |
| US            | US snRNA    | 36,088   | 2-102          | De novo | 3.5 - 6.4 | V |

| NTC           | Prp19       | U-box    | 1-51           | 5,713    | Homology modelled | ~20 | t.u.v.w | PrPF19/Cwff8 |
|               | Coiled-coil | 52-143   | 10,247         | 78-143   | Homology modelled | ~20 |
|               | WD40        | 144-503  | 40,646         | 171-501  | Docked | ~25-30 |
|               | Snt309      | 175      | 20,709         | 12-174   | Homology modelled | ~20 | s | BCAS2/Cwff7 |
|               | Syf1        | 859      | 100,229        | 21-590   | Idealised alpha helices | 4.8 - 8 | T | Syf1/Cwff3 |
|               | Cfl1        | 1-271    | 32,396         | 1-271    | Homology modelled & rebuilt | 3.8 - 6.4 | S | CRKNL1/Cwff4 |
|               | Periphery   | 272-687  | 50,067         | 277-556  | Idealised alpha helices | 5.2 - 8.8 |
|               | Cef1        | 1-191    | 21,868         | 12-191   | Homology modelled & rebuilt | 3.8 - 6.2 | O | CDC5L/Cdc5 |
|               | Middle      | 192-505  | 65,905         | -        | Not modelled | - |
|               | C-terminal  | 506-590  | 9,994          | 506-590  | Homology modelled | ~20 |
|               | Isy1        | 235      | 32,992         | 1-96     | De novo | 3.8 - 6.2 | G | Isy1/Cwff12 |

| NTR           | Prp45       | 379      | 42,483         | 32-224   | Homology modelled & rebuilt | 4.8 - 8.4 | K | Snw1/Prp45 |
|               | Prp46       | 451      | 50,700         | 111-445  | Homology modelled & rebuilt | 3.4 - 6.6 | J | PrUG1/Prp5 |
|               | Ecm2        | 362      | 40,925         | 6-324    | Homology modelled & rebuilt | 4.0 - 7.0 | N | RBM22/Cwff5 |
|               | Cwc2        | 339      | 38,431         | 3-252    | Docked & rebuilt | 3.6 - 6.0 | M | RBM22/Cwff2 |
|               | Cwc15       | 175      | 19,935         | 7-40     | Homology modelled & rebuilt | 3.6 - 7.6 | P | CWC15/Cwff15 |
|               | Bud31       | 157      | 18,447         | 2-156    | Docked & rebuilt | 3.6 - 6.8 | L | BUD31/Cwff14 |

| Splicing factors | Yju2        | 278      | 32,712         | 2-115    | De novo | 3.8 - 5.4 | D | CDC94/Cwff16 |
|                 | Cwc21       | 1-64     | 7,057          | 2-50     | De novo | 3.8 - 7.4 | R | SRRM2/Cwf21 |
|                 | N-terminal  | 65-135   | 8,724          | 64-111   | Homology modelled | 4.4 - 7.6 |
|                 | Coiled-coil | 289-577  | 34,123         | 289-481  | Homology modelled & adjusted | 4.6 - 8.2 | H | Cwc22/Cwff2 |

| Helicases       | Br2         | 2,163    | 246,185        | 442-2163 | Docked | ~13 - 20 | B | 200K/Brr2 |
|                 | Prp16       | 1,071    | 121,659        | 328-978  | Homology modelled & domains fitted | ~12 - 15 | Q | DHX38/Prp16 |

| Substrate       | S'-exon     | 20       | 6,683          | (-16)    | (-1)   | De novo | 3.4 - 6.4 | E |
|                 | Intron      | 95       | 30,405         | 1-108    | 5-76   | De novo | 3.4 - 7.2 | I |

Resolution was calculated by averaging ResMap-calculated resolution voxels over each residue using Chimera. The resolution of residues at the 5th and 95th percentile for each chain then gave the resolution range for that chain. Da, Dalton.