The spliceosome excises introns from pre-mRNAs in two sequential transesterifications—branching and exon ligation—catalysed at a single catalytic metal site in U6 small nuclear RNA (snRNA)\textsuperscript{2,3}. Recently reported structures of the spliceosomal C complex\textsuperscript{4,5} with the cleaved 5′ exon and lariat-3′-exon bound to the catalytic centre revealed that branching-specific factors such as Cwc25 lock the branch helix into position for nucleolytic attack of the branch adenosine at the 5′ splice site. Furthermore, the ATPase Prp16 is positioned to bind and translocate the intron downstream of the branch point to destabilize branching-specific factors and release the branch helix from the active site\textsuperscript{6}. Here we present, at 3.8 Å resolution, the cryo-electron microscopy structure of a *Saccharomyces cerevisiae* spliceosome stalled after Prp16-mediated remodelling but before exon ligation. While the U6 snRNA catalytic core remains firmly held in the active site cavity of Prp8 by proteins common to both steps, the branch helix has rotated by 75° compared to the C complex and is stabilized in a new position by Prp17, Cef1 and the reoriented Prp8 RNase H-like domain. This rotation of the branch helix removes the branch adenosine from the catalytic core, creates a space for 3′ exon docking, and restructures the pairing of the 5′ splice site with the U6 snRNA ACAGAGA region. Slu7 and Prp18, which promote exon ligation, bind together to the Prp8 RNase H-like domain. The ATPase Prp22, bound to Prp8 in place of Prp16, could interact with the 3′ exon, suggesting a possible basis for mRNA release after exon ligation\textsuperscript{6,7}. Together with the structure of the C complex\textsuperscript{3}, our structure of the C* complex reveals two major conformations of the spliceosome during the catalytic stages of splicing.

The spliceosome converts from branching to exon ligation through the ATP-dependent activity of Prp16, a DEAH box helicase\textsuperscript{8}. Prp16 action destabilizes the branching factors Cwc25 and Yju2 (refs 4,9) and creates strong binding sites for the exon ligation factors Slu7 and Prp18 (refs 10,11), while promoting 3′ exon docking\textsuperscript{2}. These factors are essential for splicing of pre-mRNAs with long distances between the branch point and the 3′ splice site\textsuperscript{12} and for correct 3′ splice site selection\textsuperscript{13}, but their precise roles are unknown.

To understand the mechanism of Prp16-mediated spliceosomal remodelling, we assembled spliceosomes in *S. cerevisiae* extracts on a pre-mRNA substrate containing a deoxy-guanosine at the 3′ splice site UAG sequence and purified them via an affinity-tag on Slu7 (Methods). With this substrate the spliceosomes stall after Prp16-dependent remodelling and before exon ligation, forming the C* complex\textsuperscript{14}, thus purified spliceosomes contain predominantly lariat intermediates (Extended Data Fig. 1). We obtained a cryoEM reconstruction of C* complex at a resolution of 3.8 Å (Extended Data Figs 1, 2), into which we modelled 40 components (Extended Data Tables 1, 2, Extended Data Figs 3, 4 and Supplementary Information).

Prp8, Snu14 and the U5 Sm core domain form the foot domain in both C* and C complexes (Fig. 1), which functions as an assembly platform for most of the Prp19 complex (NCT) and Prp19-related (NTR) components including Cwc2, Bud31 and Ecm2. The N-terminal end of Clf1 is anchored by Cef1, Syf2 and the U2/U6 snRNAs exiting the active site (Fig. 1b). The HAT repeats of Clf1 and Syf1 together form a large arch which is rotated considerably in C* relative to the C complex. The N-terminal end of Syf1 interacts with the U2 small nuclear ribonucleoprotein (snRNP), therefore this rotation disrupts the interface between the Prp8 RNase H-like (RH) domain and the U2 snRNP observed in C complex. Consequently, the Prp8 RH domain rotates inward (Extended Data Fig. 5) and the Prp17 WD40 domain moves into the body of the complex. U2 stem IIc swings outwards and no longer interacts with Cwc2/Ecm2 (Extended Data Fig. 5). Unlike in the C complex, no density is visible for Brr2.

The RNA structure in the core of the spliceosome remains remarkably unchanged during the C to C* transition, with the exception of the branch helix (Fig. 2a, b). The U2/U6 catalytic triplex adopts the same configuration as in the C complex, consistent with biochemical and genetic evidence\textsuperscript{15}. Density consistent with the presence of Mg\textsuperscript{2+} ions is observed adjacent to the phosphate oxygen ligands for catalytic metal ions M1 and M2 identified by metal rescue studies (Extended Data Fig. 6), providing further evidence for a single active site for both catalytic steps\textsuperscript{2}. As in the C complex, the 5′ exon is base-paired with loop 1 of U5 snRNA in agreement with genetic analysis and crosslinking experiments\textsuperscript{16,17}, and the 3′-OH of the last 5′-exon nucleotide (G(−1)) lies close to the M1 site and is ready to act as a nucleophile for the incoming 3′ splice site (Fig. 2c, Extended Data Fig. 6). Prp16-induced remodelling results in a dramatic rotation of the branch helix by approximately 75° around the hinge at A30 of U2 snRNA (Fig. 2b). The branch point moves away from the catalytic centre (approximately 20 Å), creating sufficient space for the 3′ splice site to dock at the catalytic Mg\textsuperscript{2+} site (Fig. 2a, c). This movement disrupts the non-Watson–Crick interactions of the branch point adenosine (A70) with the branch helix\textsuperscript{3} and reorganizes the interactions of the 5′ splice site with the U6 snRNA ACAGAGA sequence (Fig. 2d, e). The A70 base is packed against the ribose of the first intron nucleotide G(1), while the G(1) base stacks with the base of U(1) (Fig. 2e). In the C complex, U(1) forms a base triple with G37 of U2 snRNA and C67 of the intron\textsuperscript{2}, whereas in the C* complex U(1) forms a non-canonical base-pair with A51 of U6 snRNA (Fig. 2d, e), consistent with crosslinking in human spliceosomes\textsuperscript{17}. Notably, mutations at both U(1) and A51 impair exon ligation\textsuperscript{18,19}. Indeed, in group II introns the nucleotide equivalent to A51 (adjacent to the two nucleotides involved in triplex formation) base-pairs with the last nucleotide of the intron (−γ′ interaction)\textsuperscript{20}. Thus A51 and intron U(1) may interact with the last nucleotide G(−1) of the intron. It is noteworthy that all three positions (A70, G(1), U(1)/A51), whose mutations lead to second step defects, are aligned towards the active site, strongly suggesting a path for the 3′ splice site (Fig. 2c). Finally, the Hoogsteen base-pair between A(−3) of the intron and G50 of U6 snRNA (Fig. 2e) no longer forms in C* (Fig. 2d), consistent with genetic evidence that this interaction must be disrupted during the Prp16 rearrangement\textsuperscript{21}.

The proteins common to both C and C* restrain the catalytic RNA core (U6 snRNA ISL and helices Ia and Ib) onto Prp8, whereas the
Figure 1 | Subunit organization of the C* spliceosome. a, b, Orthogonal views of the complex coloured by subunit identity. c, List of modelled subunits grouped into functional sub-complexes.

Figure 2 | Architecture of the RNA catalytic core in the C* complex. a, Key RNA structures at the active site. The branch helix has undocked from the catalytic Mg\(^{2+}\) site. BP, branch point; ISL, internal stem loop; M1 and M2, catalytic metal ions. b, Rotated view showing superposition of the RNA catalytic core for the C (PDB, 5LJ5; ref. 4) and C* spliceosomes. C elements are coloured in light shades. Note substantial rotation of the branch helix between the two complexes. c, The branch point (BP) and 5′ splice site nucleotides align in a path to the catalytic Mg\(^{2+}\) site. A possible intron path guiding the 3′ splice site to the Mg\(^{2+}\) site is shown as a dashed line. d, Watson–Crick–Hoogsteen interaction between U(+2) and A51 of U6 snRNA. e, Different interactions of U(+2) and A(+3) observed in the C complex.

branch helix rotates substantially between the two states (Fig. 2b). In the C complex, the branch helix is locked into the branching conformation predominantly by Cwc25, Yju2 and Isy1, such that A70 is inserted into the catalytic centre\(^4,5\) but this interaction is disrupted completely in the C* complex when Prp17 wedges between the U2 Sm ring and the Prp8 RH domain (Fig. 3a, Extended Data Fig. 5). In the C* complex, the RH domain of Prp8 is rotated by about 80° with respect to the large domain (Extended Data Fig. 5d, e). In this orientation, the 3′-finger of the RH domain crosses the minor groove of the branch helix and reaches Cef1 (Fig. 3a, d). Prp17 binds across the interface between the 3′-finger and Cef1, stabilizing this interaction. A long α-helix bridges the Prp8 RH domain and the Cef1 Myb domain and reaches the C terminus of Syf1 (Fig. 3c). The direction and sequence of this helix are uncertain from the current map. These interactions lock the branch helix in a conformation predisposed for exon ligation. Prp17 and the rotated position of the RH domain observed in C* would clash with Isy1 and Cwc25, explaining how Prp16-dependent dissociation of Isy1 enables the RH domain to rotate, promoting the exon-ligation configuration. Consistently, deletion of Isy1 suppresses a Prp16 mutation that impairs remodelling of the spliceosome\(^22\). The helical domain\(^23\) of Prp18 is bound to the Prp8 RH domain opposite from the branch helix binding face (Fig. 3a). Slu7 meanders from the binding site of its predicted globular region towards the foot of the complex, interacts with Prp18, and latches the RH domain onto the endonuclease domain of Prp8, thus stabilizing the rotated conformation of the RH domain and the binding of Prp18 in C* (Extended Data Fig. 7). Indeed, the region of Slu7 that binds the RH domain is essential for yeast viability\(^24\).

Our C* complex structure provides important insight into the organization of the active site during exon ligation even though the 3′ exon is not yet docked. As discussed above the rotation of the branch helix not only creates a space for the 3′ exon at the catalytic metal binding site but also reorganizes the interaction between the U6 snRNA ACAGAGA sequence and the 5′ end of the intron. An important
outcome of Prp16 action is repositioning A51 of U6 snRNA so that it could position the 3′ splice site by interacting with the last nucleotide of the intron like the equivalent nucleotide in group II introns20 (Fig. 2c). If this is the case, it is possible that the penultimate nucleotide A(−2) and its preceding nucleotide Y(−3) may also interact with the first intron nucleotide G(+1) and the branch point adenosine (A70). An interaction between the 3′ splice site UAG sequence, the ACAGAGA sequence and the 5′ splice site intron sequence for exon-ligation has been suggested previously19. However, mutational studies of the 5′ splice site and 3′ splice site have not led to a clear base-pairing scheme19,25 suggesting that these interactions may involve non-canonical base pairing. The base pairing between the 5′ exon and loop 1 of U5 snRNA (Fig. 2c) places the 5′-OH group of the 5′ exon close to M1 such that it can act as a nucleophile when the phosphate group at the 3′ splice site is bound to M1 and M2 (Extended Data Fig. 6; refs 2,26). Slu7 and Prp18 are dispensable for exon ligation when the distance between the branch point and the 3′ splice site is less than nine nucleotides11. Notably, in our structure, three nucleotides of the intron downstream of the branch point are visible and an additional six nucleotides would be sufficient to fold back and reach the catalytic Mg2+ site (Extended Data Fig. 8; ref. 12). When the distance to the 3′ splice site is less than nine nucleotides, the 3′ splice site could easily reach the catalytic centre and allow the 3′ exon to dock. When the distance is much greater, the entropic cost of docking the 3′ exon would be greater and Slu7 and Prp18 could become indispensable in guiding the path of the intron (Extended Data Fig. 8). Indeed, mutation of Slu7 impairs splicing at distal 3′ splice sites without affecting proximal 3′ splice sites, when two competing sites are present13,27.

The DEAH ATPase Prp22, which promotes exon ligation7,28, binds on top of the Prp8 large domain (Fig. 4a) near where Prp16 binds in the C complex, consistent with a mutually exclusive interaction with the spliceosome10. Compared to the crystal structure of the homologous Prp43 ATPas in the ADP–Mg-bound form29, Prp22 in our C′ complex is in an open conformation with a wider separation of the RecA1 and RecA2 domains (Fig. 4b). Density attributable to RNA is present in the Prp22 active site (Fig. 4b) and Prp22 crosslinks 17 nucleotides downstream of the exon–exon junction4. Indeed, the distance between the catalytic centre of the spliceosome and Prp22 in our structure can be spanned by 16–17 nucleotides. Thus Prp22 could bind to the 3′ exon emerging from the core to promote mRNA release and dissociation of Slu7, Prp18 and Cwc22 after exon ligation.

Our C′ complex structure elucidates the structural consequences of Prp16 activity (Fig. 4c), reveals a large rotation of the branch helix which creates a space for 3′ exon docking at the catalytic centre in the exon ligation conformation of the spliceosome, and provides a structural framework for investigating 3′ splice site selection and ligated exon release.

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 15 November 2016; accepted 4 January 2017.
Published online 11 January 2017.

1. Wahl, M. C., Will, C. L. & Lührmann, R. The spliceosome: design principles of a dynamic RNP machine. Cell 136, 701–718 (2009).
2. Fica, S. M. et al. RNA catalyses nuclear pre-mRNA splicing. Nature 503, 229–234 (2013).
3. Madhani, H. H. & 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).
4. Galej, W. P. et al. Cryo-EM structure of the spliceosome immediately after branching. *Nature* **537**, 197–201 (2016).
5. Wan, R., Yan, C., Bai, R., Huang, G. & Shi, Y. Structure of a yeast catalytic step I spliceosome at 3 Å resolution. *Science* **353**, 895–904 (2016).
6. Schwer, B. A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. *Mol. Cell* **30**, 743–754 (2008).
7. Semilow, D. R., Blanco, M. R., Walter, N. G. & Staley, J. F. Spliceosomal DEAH-box ATPases remodel pre-mRNA to activate alternative splice sites. *Cell* **164**, 985–998 (2016).
8. Schwer, B. & Guthrie, C. PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. *Nature* **349**, 494–499 (1991).
9. Tseng, C. K., Liu, H. L. & Cheng, S. C. DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. *RNA* **17**, 145–154 (2011).
10. Jiang, J., Horowitz, D. S. & Xu, R. M. Crystal structure of the functional domain of the splicing factor Prp18. *Proc. Natl Acad. Sci. USA* **97**, 3022–3027 (2000).
11. 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).
12. Brys, A. & Schwer, B. Requirement for SLU7 in yeast pre-mRNA splicing is dictated by the distance between the branchpoint and the 3′ splice site. *RNA* **2**, 707–717 (1996).
13. Moore, M. J. & Sharp, P. A. Site-specific modification of pre-mRNA: the 2′-hydroxyl groups at the splice sites. *Science* **256**, 992–997 (1992).
14. 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).
15. Newman, A. J. & Norman, C. U5 snRNA interacts with exon sequences at 5′ and 3′ splice sites. *Cell* **68**, 743–754 (1992).
16. Konarska, M. M., Vilardell, J. & Query, C. C. Repositioning of the reaction intermediate within the catalytic center of the spliceosome. *Mol. Cell* **21**, 543–553 (2006).
17. Steitz, T. A. & Steitz, J. A. A general two-metal-ion mechanism for catalytic RNA. *Science* **211**, 2022–2027 (1981).
18. Gall, J. E. & Guthrie, C. Splicing in yeast: a complex and dynamic process. *Nature* **402**, 207–210 (1999).
19. Luukkonen, B. G. & Séraphin, B. The role of branchpoint-3′ hydroxyl groups at the splice sites. *EMBO J.* **16**, 779–792 (1997).
20. Schwer, B. & Gross, C. H. Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. *EMBO J.* **15**, 2086–2094 (1996).
21. Hilliker, A. K., Mefford, M. A. & Staley, J. P. U2 toggles iteratively between the stem Ila and stem lIC conformations to promote pre-mRNA splicing. *Genes Dev.* **21**, 821–834 (2007).

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

**Acknowledgements** We thank S. Scheres for his help and advice on data collection and processing; C. Sawwa, S. Chen, K. R. Vinothkumar, Q. McMullan, J. Grimmett and T. Darlington for smooth running of the EM and computing facilities; the staff at Diamond Light Source (DLS) for help with data collection; the mass spectrometry facility for help with protein identification, P. Emsley and G. Murshudov for help and advice with model building and refinement; the members of the spliceosome group for help and advice throughout the project. We thank J. Löwe, V. Ramakrishnan, D. Barford and R. Henderson for their continuing support, C. Plaschka, P. C. Lin and L. Strittmatter for critical reading of the manuscript and J. Vilardell for a generous gift of reagent. The project was supported by the Medical Research Council (MC_U105184330) and European Research Council Advanced Grant (693087 - SPLICE3D). S.M.F. was supported by EMBO and Marie Sklodowska-Curie fellowships, M.E.W was supported by a Rutherford Memorial Cambridge Scholarship.

**Author Contributions** S.M.F. and W.P.G. established the method of C* complex preparation. S.M.F. prepared the EM sample and grids, S.M.F., W.P.G., M.E.W. and X.G.B. collected and processed EM data, C.O., S.M.F., W.P.G. and M.E.W. carried out model building and C.O. refined and finalized the PDB file. S.M.F., C.O., W.P.G., M.E.W. and K.N. analysed the structure. A.J.N. prepared the substrate and contributed to the project through his knowledge and experience on yeast splicing. Manuscript was initially written by S.M.F. and finalized with input from all authors. K.N. initiated and coordinated the spliceosome 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. Correspondence and requests for materials should be addressed to S.M.F. (sfica@mrc-lmb.cam.ac.uk) and K.N. (kn@mrc-lmb.cam.ac.uk).
METHODS
Pre-mRNA substrate preparation. Spliceosomes were assembled on a modified UBC4 pre-mRNA substrate containing a deoxyguanosine at the 3′ splice site UAG sequence (UAAG)14 synthesized by ligation from a long 5′ piece ending 11 nucleotides before the 3′ splice site and a short 3′ oligonucleotide containing the dG modification and a 3′ C5 fluorophore (purchased from Dharmacco). The 5′ piece was generated by run-off transcription from a DNA template containing 3′ MS2 stem loops15 at the 5′ end of the UBC4 sequence followed by the hepatitis delta virus ribozyme sequence. Following run-off transcription, ribozyme cleavage was induced to obtain a precise 3′ end for ligation. The 5′ and 3′ pieces were joined by splint-mediated ligation using T4 DNA ligase, essentially as described24, and the ligated full-length pre-mRNA was gel-purified before use.

Spliceosome purification. Yeast containing a TAPS affinity tag on endogenous Slu7 (Slu7-TAPS)16 were grown in a 120 l fermenter, and splicing extract was processed separately up to and including the particle polishing step (see below). Image processing. For cryo-EM analysis, Quantifoil R 1.2/1.3 Cu 400 mesh grids were coated with a 6–7-nm-thick layer of homemade carbon film and glow discharged. After applying 3.5 μl of the sample, the grids were blotted for 3–5 s and vitrified in liquid ethane in an FEI Vitrobot MKIII, at 100% humidity at 4 °C. Discharging of the grids was induced to obtain a precise 3′ end for ligation. The 5′ and 3′ pieces were joined by splint-mediated ligation using T4 DNA ligase, essentially as described24. In vitro splicing reactions were assembled using pre-mRNA substrate pre-bound to MS2–MBP fusion protein, as previously described21. ATP depletion was achieved by addition of 2 mM glucose following splicing to minimize Prp22 activity and promote its association with the spliceosome. The resulting spliceosomes were bound to amyllose-resin in buffer K–100 (20 mM HEPES KOH pH 7.9, 100 mM KCl, 0.25 mM EDTA, 5% glycerol, 0.025% NP-40), incubated with 2 mM ATP-Mg2+ on beads, washed, and eluted with 12 mM maltose. The eluted total spliceosomes (Extended Data Fig. 1) were subsequently further purified via the Strept II tag on Slu7 using Strept-Tactin affinity resin (GE) in buffer K–100 and eluted with desthiobiotin, essentially as described25. The Strept-Tactin eluate was crosslinked with 0.5 mM BS3, concentrated and buffer-exchanged with 20 mM HEPES KOH pH 7.9, 100 mM KCl, 1 mM MgCl2 and used for cryo-EM studies. This procedure, combined with the second step block induced by the deoxy-G at the 3′ splice site, should enrich for spliceosomes that have undergone Prp16-mediated remodelling. Indeed, analysis of protein components by gel electrophoresis and subsequent mass spectrometry shows that Prp22 and Slu7-TAPS are present at near-stoichiometric levels to Prp8, the U2 snRNP and the Prp19 module.

Electron microscopy. For cryo-EM analysis, Quantifoil R 1.2/1.3 Cu 400 mesh grids were coated with a 6–7-nm-thick layer of homemade carbon film and glow discharged. After applying 3.5 μl of the sample, the grids were blotted for 3–5 s and vitrified in liquid ethane in an FEI Vitrobot MKIII, at 100% humidity at 4 °C. Grids were imaged on two separate microscopes: 1,571 micrographs were collected at the DLS and 8,746 micrographs were collected at the LMB. Transmission electron microscopes were operated at 300kV and images were collected using a Gatan K2 summit direct electron detector and a GIF Quantum energy filter (silt width 20 eV). Images at the LMB were collected in super-resolution counting mode at 1.25 e− per pixel per second and a calibrated pixel size of 1.43 Å per pixel; a total dose of 40 e− per Å2 over 16 and a defocus range of 0.5–4.5 μm were used. Images at the DLS were collected in counting mode at 2 frames per second and a calibrated pixel size of 1.025 Å; a total dose of 42 e− per Å2 over 14 s and a defocus range of 0.5–3.5 μm were used.

Image processing. Raw micrographs collected at the LMB and the DLS were processed separately up to and including the particle polishing step (see below). Micrographs were subjected to whole-frame drift correction in MOTIONCORR46 followed by contrast transfer function (CTF) parameter estimation in CTFFIND4 (ref. 37). All subsequent processing steps were performed using RELION 1.4 (ref. 38).

An initial subset of 400 micrographs collected at the LMB was subjected to automated particle picking using 2D class averages obtained in RELION from particles used for the C complex reconstruction4. The selected particles were used for initial reference-free 2D classification and the resulting 30,000 particles were subjected to 3D classification using an initial 3D reference obtained by low-pass filtering (60 Å) the reconstruction of the C complex (EMD–4055; Extended Data Fig. 2a). This procedure produced a subclass substantially different from the starting C complex model, which was used as a starting C model (Extended Data Fig. 2a). The automated particle-picking algorithm in RELION48 was then applied to all micrographs from both data sets. 3D classification followed by 3D refinement and particle-based beam-induced motion correction and radiation-damage weighting (particle polishing) was performed separately for the two datasets. The resulting particles were combined from the two datasets and scaled to 1.43 Å pixel size, yielding a total of 164,912 particles, which were subjected to global 3D classification using a soft mask around the core of the complex and finer angular sampling of 1.8° and local searches of 10° (Extended Data Fig. 2b). The subset of 65,824 particles produced by this procedure was used for 3D refinement and produced a final reconstruction at 3.8 Å overall resolution and estimated accuracies of rotations of 1.3° (Extended Data Fig. 3).

Weak density observed at three peripheral regions of the map corresponding to Prp22, the U2 snRNP and the Prp19 module was improved by focused classification without signal subtraction39. A mask was applied to the region of interest, particles were 3D classified without image alignment, and the best class was selected for further refinement of the original (unmasked) particles. This resulted in small crystals of the original particles, in which Prp22 and the U2 snRNP adopt a more homogeneous conformation (Extended Data Figs 2 and 3). 3D refinement of the 61,000 Prp22-selected particles resulted in a map at overall 4.2 Å resolution, where individual secondary structure elements corresponding to a DEAH helicase are clearly visible for Prp22 (Extended Data Fig. 4a). 3D refinement of 29,000 U2 snRNP-selected particles produced a map at overall 4.7 Å resolution, in which RNA and the U2 Sm ring density are clearly distinguishable, while 31,000 Prp19-containing particles yielded a map at 6.4 Å resolution that revealed docking of the Prp19 module from the C complex structure. All reported resolutions are based on the gold-standard Fourier shell correlation (FSC) = 0.143 criterion40. 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 curves41. Prior to visualization, all maps were corrected for the modulation transfer function of the detector. Local resolution was estimated using Relion 2.0 (S. Scheres, unpublished).

Model building. A list of protein and RNA components included in the model is given in Extended Data Table 2. Initially, known structures of S. cerevisiae Prp8, Snu114, the U5 Sm ring, U5 snRNA, U6 snRNA, part of U2 snRNA, the 5′ exon and NTR and NTC components from the C complex4 were docked into the C* structure, accounting for the majority of the protein and RNA density in the core of the complex. Density for the repositioned branch helix similar to the one previously observed in the ILS structure42 replaced the positions of Ytu2 and Cwc25 seen in C complex and allowed building of the intron, including the branch linkage and three nucleotides downstream of the branch adenosine. A β-propeller domain buttressing the branch helix was assigned to Prp17. A homology model produced using SWISS-MODEL43, based on the structure of the WD40 domain of ribosomal assembly protein 4 (PDB 5FL8), was docked based on local characteristics and side chains, and then manually rebuilt. Unassigned density remained around the Prp8 RH domain. On the face projecting away from the core a characteristic helical density was observed to Prp22, the U2 snRNP and the Prp19 module.

Data availability. The cryo-EM map has been deposited in the Electron Microscopy Data Bank with accession codes EMD–3539 (core), EMD–3541 (core + Prp22) and EMD–3542 (core + U2 snRNP). The coordinates of the atomic models have been deposited in the Protein Data Bank under accession codes 5MP5 (core) and 5MQ0 (core + Prp22 + U2 snRNP).
31. 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).
32. Zhou, Z., Licklider, L. J., Gygi, S. P. & Reed, R. Comprehensive proteomic 
analysis of the human spliceosome. Nature 419, 182–185 (2002).
33. Umen, J. G. & Guthrie, C. A novel role for a U5 snRNP protein in 3’ splice site 
selection. Genes Dev. 9, 855–868 (1995).
34. Lin, R. J., Newman, A. J., Cheng, S. C. & Abelson, J. Yeast mRNA splicing in vitro. 
J. Biol. Chem. 260, 14780–14792 (1985).
35. 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).
36. 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).
37. Rohou, A. & Grigorieff, N. CTFIND4: Fast and accurate defocus estimation from 
electron micrographs. J. Struct. Biol. 192, 216–221 (2015).
38. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM 
structure determination. J. Struct. Biol. 180, 519–530 (2012).
39. Scheres, S. H. Processing of structurally heterogeneous cryo-EM data in 
RELION. Methods Enzymol. 579, 125–157 (2016).
40. Scheres, S. H. W. & Chen, S. Prevention of overfitting in cryo-EM structure 
determination. Nat. Methods 9, 853–854 (2012).
41. 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).
42. Yan, C. et al. Structure of a yeast spliceosome at 3.6-angstrom resolution. 
Science 349, 1182–1191 (2015).
43. Biasini, M. et al. SWISS-MODEL: modelling protein tertiary and quaternary 
structure using evolutionary information. Nucleic Acids Res. 42, W252–W258 
(2014).
44. Kozlowski, L. P. & Bujnicki, J. M. MetaDisorder: a meta-server for the prediction 
of intrinsic disorder in proteins. BMC Bioinformatics 13, 111 (2012).
45. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of 
Coot. Acta Crystallogr. D 66, 486–501 (2010).
46. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular 
structures by the maximum-likelihood method. Acta Crystallogr. D 53, 240–255 
(1997).
47. Nicholls, R. A., Fischer, M., McNicholas, S. & Murshudov, G. N. Conformation-
independent structural comparison of macromolecules with ProSMART. Acta 
Crystallogr. D 70, 2487–2499 (2014).
48. Brown, A. et al. tools for macromolecular model building and refinement into 
electron cryo-microscopy reconstructions. Acta Crystallogr. D 71, 136–153 
(2015).
49. Zwart, P. H. et al. Automated structure solution with the PHENIX suite. Methods 
Mol. Biol. 426, 419–435 (2008).
50. Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF 
Chimera. J. Struct. Biol. 157, 281–287 (2007).
51. 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).
52. Marcia, M. & Pyle, A. M. Visualizing group II intron catalysis through the stages 
of splicing. Cell 151, 497–507 (2012).
53. 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).
54. Aronova, A., Baciková, D., Crotti, L. B., Horowitz, D. S. & Schwer, B. Functional 
interactions between Prp8, Prp18, Slu7, and U5 snRNA during the second step 
of pre-mRNA splicing. RNA 13, 1437–1444 (2007).
55. 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 | Purification and cryo-EM imaging of the C* spliceosome. a, Protein composition of the purified C* complex. Note that Prp16 is strongly de-enriched compared to Prp22, consistent with the majority of the purified complexes being in a post-Prp16 conformation, as Prp16 dissociates upon ATP hydrolysis. b, The purified C* complex contains mostly lariat intermediates and catalyses exon ligation with low efficiency when incubated in the presence of Mg²⁺. The identity of the major species, inferred by size and migration pattern, is indicated by the cartoon on the left. c, Representative electron micrograph of the C* complex sample collected at 3 μm defocus. d, Representative 2D class averages of the C* complex obtained in RELION. e, Image of a highly abundant C* complex 2D class average illustrating the major domains of the complex.
Extended Data Figure 2 | Data processing workflow. a, Method used to obtain an initial model of the C* complex for 3D classification in RELION39. b, Scheme for 3D classification and refinement. The mask used to obtain the overall 3.8 Å map excluded the Clfl1 and Syfl arch regions as well as the Prp22 and U2 snRNP regions. Note that focused classification was performed without signal subtraction. All nominal resolutions reported were obtained during post-processing in RELION39.
Extended Data Figure 3 | Angular distribution and FSC curves for the C*-reconstructions. a, Angular distribution for the 3.8 Å map of the core region. Note the presence of several orthogonal views. b, Overall reconstructions obtained after classification with masks on the core (grey), the core with Prp22 (magenta), and the core with the U2 snRNP (green). The three maps were superposed and aligned on the core using Chimera. c, Gold-standard FSC curves for the three maps shown in b. d, Local resolution for the core map, calculated using RELION 2.0 (ref. 39).
Extended Data Figure 4 | Fit of the model built into the C* map.
a, Experimental density for Prp22 and fitting of the model into the density. b–d, Fitting of the model into the experimental density for key regions of the C* map. e, 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 the two half maps. Excellent correlation up to the 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  U2 snRNP rearrangement between C and C* complexes and repositioning of the Prp8 RH domain during splicing. 
a, b, Movement of the U2 snRNP domain between the C complex (a) and the C* complex (b). Note that U2 stem IIa switches from an interaction with Prp17 in C complex to a position adjacent to the U2 Sm ring in C*; the binding of U2 stem IIb by Ecm2/Cwc2 is disrupted and Prp17 changes its binding surface on Ecm2/Cwc2. In a, Br2 was omitted for clarity. 
c, Prp8 RH domain conformation in the B act complex (PDB 5GM6; ref. 51). Note how Hsh155 sequesters the branch helix away from the RH domain, which projects its β-hairpin into solvent and is stabilized by the Prp8 Jab1/MPN domain and Br2 (not shown). d, e, RH conformation in the C (d) and C* (e) complexes. Note that the RH domain undergoes a dramatic inward rotation towards the body of the complex and is stabilized in alternative conformations by factors specific for branching (C) or exon ligation (C*). f, RH conformation in the Schizosaccharomyces pombe ILS. Note that the de-branching specific factor Cwf19 is now wedged between the RH domain and the branch helix. All structures were aligned on the Prp8 endonuclease domain (Prp8EN), shown in grey; complex-specific factors are coloured in magenta shades; Prp8RH, Prp8 RH domain.
Extended Data Figure 6 | Metals in the RNA core of the C* complex.

a, b, Structure (a) and schematic representation (b) of the active site of a group IIIC intron trapped in the post-catalytic state in the presence of Mg\(^{2+}\) and K\(^{+}\) (PDB 4FAR, ref. 52). The 5′-exon 3′ hydroxyl interacts with M1, while a water molecule bridges the two catalytic metals. Two additional non-catalytic Mg\(^{2+}\) and two K\(^{+}\) close to the active site are also shown. c–e, Structure of the RNA at the active site of the spliceosomal C* complex, with putative metal binding (c), schematic of catalytic metal binding (M1 and M2) (d), and comparison of the putative metal binding model with the EM density (e). Note conservation of the metal binding residues compared to the group II intron and proximity of the cleaved G(−1) 3′ hydroxyl to M1. Besides the two catalytic Mg\(^{2+}\), additional divalent and monovalent metals were observed in the group IIB structure53. Density observed at analogous position in C* complex may be attributable to a Mg\(^{2+}\) (M3) and two K\(^{+}\) (M4 and M5). f, Proposed interactions between U6 snRNA and the two catalytic Mg\(^{2+}\) during the transition state for exon ligation, as inferred from biochemistry (ref. 2). g, h, Structure (g) and schematic (h) of the RNA core of a group IIB intron in a post-catalytic configuration, following both branching and exon ligation (PDB 4R0D, ref. 53). Residues that position the catalytic metals are shown in magenta.
Extended Data Figure 7 | Structure and interactions of Slu7 and Prp18.

a, b. Overall arrangement of modelled regions for Slu7 and Prp18 in cartoon (a) and surface (b) representation. The dashed lines in (a) represent inferred chain continuity based on regions of weak density. Note how Slu7 latches the Prp8 RH domain onto the Prp8 endonuclease (Prp8EN) and N-terminal domains (Prp8N).

c. Interaction between Slu7 and Prp18. The Slu7 helix was modelled on the basis of secondary structure predictions and previously reported genetic interactions.  

d. Interactions between Slu7 and Prp8. The R1753A allele impairs exon ligation and interacts genetically with Slu7 (ref. 54). IP6, inositol hexakisphosphate.

e, f. Exchange of Cwc25 in the C complex (e) for Slu7 in the C* complex (f) on Prp8. Note that Slu7 binding is stabilized in C* by an α-helix of Prp8 (residues 2090–2110) that would clash with Cwc25 in the C complex; indeed, this helix is not visible in the C complex and only becomes ordered in C*.
Extended Data Figure 8 | Model for the role of Prp18/Slu7 in 3′ exon docking. a, Putative model for insertion of Prp18 conserved region (188–222) into the catalytic core of the C* complex. Orthogonal views of Prp8 cradling the catalytic RNA core are shown. Note accessible channel facing the location of Prp18. The conserved region, which is missing from the crystal structure and is not visible in our cryo-EM map, is shown as a dotted line. b, Distances between key elements involved in exon ligation visible in our C* map. The possible path of the intron between the last visible residue of the intron (A73) and the 5′ exon G(−1) is shown as a dotted black line. Note that six nucleotides (nt) of A-form RNA would be sufficient to reach the 5′ exon. Roughly 28 amino acids (aa) of fully extended protein would be sufficient to reach the 5′ exon; the Prp18 conserved region is 34 amino acids in length. c, Steps in pre-mRNA splicing. BrA, branch point adenosine. d, Cartoon model for 3′ splice site docking and exon ligation. For pre-mRNAs with a short distance from branch point (BP) to 3′ splice site, Prp16 action could be sufficient to allow docking of the 3′ splice site. For longer distances from branch point to 3′ splice site, Prp16 and Slu7 could become indispensable to guide the 3′ splice site to the active site. Indeed, for the UBC4 intron, Prp16 activity is not sufficient for 3′ splice site docking, which requires Slu7/Prp18 (ref. 7).
## Extended Data Table 1 | Cryo-EM data collection and refinement statistics

|                          | Core    | Core+Prp22 | Core+U2 snRNP |
|--------------------------|---------|------------|--------------|
| **Data collection**      |         |            |              |
| Microscope               | FEI Titan Krios | FEI Titan Krios | FEI Titan Krios |
| Voltage (kV)             | 300     | 300        | 300          |
| Electron dose (e Å\(^{-2}\)) | 40 or 42 | 40 or 42   | 40 or 42     |
| Detector                 | Gatan K2 Summit | Gatan K2 Summit | Gatan K2 Summit |
| Pixel (Å)                | 1.43    | 1.43       | 1.43         |
| Defocus range (μm)       | 0.5-4.5 | 0.5-4.5    | 0.5-4.5      |
| **Reconstruction (Relion)** |       |            |              |
| Particles                | 65,824  | 61,107     | 29,527       |
| Box edge (pixels)        | 412     | 412        | 412          |
| Accuracy of rotations (*)| 1.37    | 1.68       | 1.94         |
| Accuracy of translations (pixels) | 0.75 | 1.04 | 1.26 |
| Map sharpening B-factor (Å\(^{-2}\)) | -49 | -58 | -30 |
| Final resolution (Å)     | 3.85    | 4.17       | 4.60         |
| **Model composition\(^1\)** |         |            |              |
| Protein residues         | 6848    | 7496       | 7635         |
| RNA bases                | 339     | 339        | 474          |
| Ligands                  | 13      | 13         | 13           |
| **Refinement\(^2\)**    |         |            |              |
| Resolution (Å)           | 3.85    |            |              |
| F\(^3\)C\(_{1/2}\)R programmed | 0.801 |            |              |
| R factor                 | 0.322   |            |              |
| **Validation\(^3\)**    |         |            |              |
| Molprobity score         | 2.23 (100\(^{th}\) percentile) |            |              |
| Clashscore, all atoms    | 7.29 (100\(^{th}\) percentile) |            |              |
| Good rotamers (%)        | 89.6    |            |              |
| **Ramachandran plot\(^1\)** |         |            |              |
| Favoured (%)             | 90.67   |            |              |
| Outliers (%)             | 0.89    |            |              |
| **RNA validation\(^3\)** |         |            |              |
| Correct sugar puckers (%)| 97.05   |            |              |
| Good backbone conformations (%) | 60.2 |            |              |
| **Data Deposition**      |         |            |              |
| PDB ID                   | 5MPS    | 5M00       | 5M00         |
| EMD/DB ID                | EMD-3539 | EMD-3541   | EMD-3542     |

\(^1\)Excluding the Prp19 module.

\(^2\)Refinements were performed only for the Core region.

\(^3\)As determined by Molprobity\(^5\).
### Extended Data Table 2 | Summary of components modelled into the C* complex map

#### Proteins and RNA included in the model

| Sub-complexes | Protein/RNA | Domains | Total residues | M.W. (Da) | Modelled | Modelling Template (PDB ID) | Modelling | Resolution | Chain ID | Human/S. pombe
|---------------|-------------|---------|----------------|-----------|----------|--------------------------|-----------|------------|----------|-----------------|
| US snRNP      | Prp8        | N-terminal | 1-670         | 191,767   | Docked   | Docked/ rebuilt           | SL5J      | 3.6 – 4.8 | A         | 22OR/Osp42*     |
|               |             | Large    | 671-1027      | 111,525   | Docked   | Docked/ rebuilt           | SL5J      | 3.6 – 4.8 | A         | 22OR/Osp42*     |
|               |             | Rhsa1    | 1828-2085     | 29,653    | Docked   | Docked/ rebuilt           | SL5J      | 3.6 – 4.8 | A         | 22OR/Osp42*     |
|               |             | Jab1/MPN | 2006-2413     | 36,612    | Docked   | Docked/ rebuilt           | SL5J      | 3.6 – 4.8 | A         | 22OR/Osp42*     |
|               | Snu14       |          | 1008          | 114,041   | Docked   | Docked/ rebuilt           | SL5J      | 3.8 – 5.0 | C         | 116C/Caw160    |
|               | SmN         |          | 196           | 22,403    | Docked   | Docked/ rebuilt           | SL5J      | 4.4 – 6.0 | b         | SmN/SmN         |
|               | SmD3        |          | 110           | 11,229    | Docked   | Docked/ rebuilt           | SL5J      | 4.2 – 5.4 | d         | SmD3/SmD3       |
|               | SmD1        |          | 146           | 16,288    | Docked   | Docked/ rebuilt           | SL5J      | 4.8 – 6.2 | h         | SmD1/SmD1       |
|               | SmD2        |          | 110           | 12,896    | Docked   | Docked/ rebuilt           | SL5J      | 5.0 – 6.8 | j         | SmD2/SmD2       |
|               | SmD         |          | 94            | 10,373    | Docked   | Docked/ rebuilt           | SL5J      | 5.6 – 7.6 | f         | SmD/SmD         |
|               | SmE         |          | 96            | 9,859     | Docked   | Docked/ rebuilt           | SL5J      | 5.2 – 7.0 | e         | SmE/SmE         |
|               | SmG         |          | 77            | 8,479     | Docked   | Docked/ rebuilt           | SL5J      | 4.4 – 6.4 | g         | SmG/SmG         |
| US snRNA-L     |             |          | 214           | 68,847    | Docked   | Docked/ rebuilt           | SL5J      | 3.6 – 6.2 | s         | US snRNA       |
| US snRNA-L     |             |          | 1175          | 363,024   | Docked   | Docked/ rebuilt           | SL5J      | 3.6 – 5.8 | 2         | US snRNA       |
| US snRNA-L     |             |          | 112           | 38,068    | Docked   | Docked/ rebuilt           | SL5J      | 3.6 – 5.0 | 2         | US snRNA       |
| NTC           | Pip19       | U-box    | 1-51          | 5,713     | Docked   | Docked/ rebuilt           | SL5J      | –10 – 15  | t.u.w.     | PPF19/Snf4     |
|               |             | Coiled-coil | 52,143      | 10,247    | Docked   | Docked/ rebuilt           | SL5J      | –10 – 15  | PPF19/Snf4    |
|               |             | WD40     | 144-503       | 40,646    | Docked   | Docked/ rebuilt           | SL5J      | –15 – 20  | PPF19/Snf4    |
| Syt1          |             | Periphery/ Core | 859       | 120,229   | Docked   | Docked/ rebuilt           | SL5J      | –15 – 20  | s         | CAA2/Snf7      |
| Cfl1          |             |          | 1,687         | 82,663    | Docked   | Docked/ rebuilt           | SL5J      | 3.8 – 8.2 | S         | CML5/Cfl4      |
|               |             |          | 1,690         | 77,777    | Docked   | Docked/ rebuilt           | SL5J      | 3.8 – 8.6 | O         | CDC4/Cfl3      |
|               |              |          | 1,590         | 77,777    | Docked   | Docked/ rebuilt           | SL5J      | 3.8 – 6.9 | 0         | CDC5/Cfl5      |
| NTR           | Pip45       |          | 379           | 42,483    | Docked   | Docked/ rebuilt           | SL5J      | 3.6 – 4.8 | K         | SWIP1/Pip45    |
|               | Pip20       |          | 401           | 53,720    | Docked   | Docked/ rebuilt           | SL5J      | 3.6 – 4.4 | J         | PIP2/Pip20     |
|               | Emc2        |          | 364           | 49,625    | Docked   | Docked/ rebuilt           | SL5J      | 3.8 – 5.6 | N         | RBM22/Cof2     |
|               | Cac2        |          | 339           | 38,431    | Docked   | Docked/ rebuilt           | SL5J      | 3.6 – 4.6 | M         | RBM22/Cof2     |
|               | Cen15       |          | 175           | 19,835    | Docked   | Docked/ rebuilt           | SL5J      | 3.0 – 7.6 | P         | CWC15/Cen15    |
|              |              |          | 157           | 18,447    | Docked   | Docked/ rebuilt           | SL5J      | 3.8 – 4.4 | L         | BGC31/Cfl4      |
| Splicing Factors | Pip18      |          | 251           | 28,377    | Docked   | Docked/ rebuilt           | SL5J      | 4.2 – 6.2 | a         | PPI18/Pip18    |
|               | Pip17       |          | 456           | 52,048    | Docked   | Docked/ rebuilt           | SL5J      | 4.0 – 5.2 | o         | CDC4/Pip17     |
|               |             |          | 352           | 44,507    | Docked   | Docked/ rebuilt           | SL5J      | 3.8 – 5.5 | c         | RNU7/357       |
|               | Cen21       | N-terminal | 1-64         | 7,967     | Docked   | Docked/ rebuilt           | SL5J      | 3.6 – 6.6 | R         | SFNMS/Cen21    |
|               |             | Coiled-coil | 65,135      | 8,724     | Docked   | Docked/ rebuilt           | SL5J      | 3.8 – 6.8 | H         | CWC22/Cen22    |
|               | Cen22       | MIF4G     | 1,888         | 33,187    | Docked   | Docked/ rebuilt           | SL5J      | 3.6 – 7.4 | H         | CWC22/Cen22    |
| Helicase       | Pip22       |          | 1,124         | 130,014   | Docked   | Docked/ rebuilt           | SL5J      | 3.6 – 4.2 | E         | PIP22/Pip22    |
| Substrate     | S-ace1      |          | 20            | 5,952     | Docked   | Docked/ rebuilt           | SL5J      | 3.6 – 4.2 | E         | PIP22/Pip22    |
|               | Intron      |          | 90            | 30,405    | Docked   | Docked/ rebuilt           | SL5J      | 3.5 – 5.0 | I         | PIP22/Pip22    |
| Unknown       | X           |          | 90            |           | Docked   | Docked/ rebuilt           | SL5J      | 4.5 – 6.0 | X         | PIP22/Pip22    |

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.