Published in final edited form as:
Nature. 2017 February 16; 542(7641): 377–380. doi:10.1038/nature21078.

Structure of a spliceosome remodelled for exon ligation

Sebastian M. Fica, **Chris Oubridge**, **Wojciech P. Galej**[#], **Max E. Wilkinson**, **Xiao-Chen Bai**, **Andrew J. Newman**, and **Kiyoshi Nagai**
MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, United Kingdom

Abstract

The spliceosome excises introns from pre-mRNAs in two sequential transesterifications – branching and exon ligation[1] – catalysed at a single catalytic metal site in U6 snRNA[2,3]. The recent structures of the spliceosomal C complex[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 nucleophilic 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[4]. Here we present the 3.8Å cryo-EM structure of a *Saccharomyces cerevisiae* spliceosome stalled after Prp16-mediated remodelling but prior to 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 degrees compared to complex C and is stabilized into a new position by Prp17, Cef1, and the reoriented Prp8 RNaseH 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 RNaseH 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[6,7]. Together with the C complex structure[4], our new C* complex structure reveals the two major conformations of the spliceosome during the catalytic stages of splicing.

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).
[#]Present address: EMBL GRENOBLE, 71 avenue des Martyrs, CS 90181, 38042 Grenoble Cedex 9, France

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.C.B. collected and processed EM data, C.O., S.M.F., W.P.G. & M.E.W. carried out model building and C.O. refined and finalised the PDB file. S.M.F., C.O., W.P.G., M.E.W., & 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. & K.N. and finalised with input from all authors. K.N. initiated and coordinated the spliceosome project.

The authors declare no competing financial interests.

Europe PMC Funders Author Manuscripts

The spliceosome converts from branching to exon ligation through the ATP-dependent activity of Prp16, a DEAH box helicase[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[7]. These factors are essential for splicing of pre-mRNAs with long distances between the branch point (BP) and the 3'-splice site (3'SS)[12] and for correct 3'SS selection[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'SS 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 C* complex[14]; thus purified spliceosomes contain predominantly lariat-intermediate (Extended Data Fig. 1). We obtained a 3.8Å cryoEM reconstruction of C* complex (Extended Data Figs 1,2) into which we modelled 40 components (Extended Data Tables 1,2; Extended Data Figs 3,4).

Prp8, Snu114 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 NTC and 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 C complex. The N-terminal end of Syf1 interacts with U2 snRNP, therefore this rotation disrupts the interface between the Prp8 RNaseH domain and the U2 snRNP observed in C complex. Consequently, the Prp8 RNaseH 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 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 C complex, consistent with biochemical and genetic evidence[15]. Density consistent with the presence of Mg^{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[2]. As in C complex, the 5'-exon is base-paired with loop 1 of U5 snRNA in agreement with genetic analysis and cross-linking experiments[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'SS (Fig. 2c, Extended Data Fig. 6). Prp16-induced remodelling results in a dramatic rotation of the branch helix by approximately 75 degrees around the hinge at A30 of U2 snRNA (Fig. 2b). The BP moves away from the catalytic centre (approximately 20Å), creating sufficient space for the 3'SS to dock at the catalytic Mg^{2+} site (Fig. 2a,c). This movement disrupts the non-Watson-Crick interactions of the BP adenosine (A70) with the branch helix[4] and reorganises the interactions of the 5'-splice site (5'SS) 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(+2). In C complex, U(+2) forms a base-triple with G37 of U2 snRNA and C67 of the intron[4,5]

whereas in C* complex U(+2) forms a non-canonical base-pair with A51 of U6 snRNA (Fig. 2d,e), consistent with cross-linking in human spliceosomes[17]. Interestingly, mutations at both U(+2) and A51 impair exon ligation[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-γ' interaction)[20]. Thus A51 and intron U(+2) may interact with the last nucleotide G(-1) of the intron. It is noteworthy that all three positions (A70, G(+1), U(+2)/A51), whose mutations lead to second step defects, are aligned towards the active site, strongly suggesting a path for the 3'SS (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[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 branch helix rotates substantially between the two states (Fig. 2b). In 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 (Fig. 3ab)[4,5]. After branching, Prp16 promotes dissociation of Cwc25, Yju2 and Isy1 and binding of Prp18, Slu7 and Prp22. In C complex the U2 Sm core domain interacts with the RNaseH domain[4,5] but this interaction is disrupted completely in C* complex when Prp17 wedges between the U2 Sm ring and the Prp8 RNase H domain (Fig 3a; Extended Data Fig. 5). In C* complex, the RNaseH domain of Prp8 is rotated by about 80° with respect to the Large domain (Extended Data Fig. 5d,e). In this orientation the β-finger of the RNaseH domain crosses the minor groove of the branch helix and reaches Cef1 (Fig. 3a, d). Prp17 binds across the interface between the β-finger and Cef1 stabilising this interaction. A long α-helix bridges the Prp8 RNaseH 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 RNaseH domain observed in C* would clash with Isy1 and Cwc25, explaining how Prp16-dependent dissociation of Isy1 enables the RNaseH 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 RNaseH 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 RNaseH domain onto the endonuclease domain of Prp8, thus stabilizing the rotated conformation of the RNaseH domain and the binding of Prp18 in C* (Extended Data Fig. 7). Indeed the region of Slu7 that binds the RNaseH 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 reorganises 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'SS by interacting with the last nucleotide of the intron like the equivalent nucleotide in group II introns[20] (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'SS UAG sequence, the ACAGAGA sequence and the 5'SS intron sequence for exon-ligation has been suggested previously[19]. However, mutational studies of the 5'SS and 3'SS have not led to a clear base-pairing scheme[19,25] 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 3'OH group of the 5'-exon close to M1 such that it can act as a nucleophile when the phosphate group at the 3'SS is bound to M1 and M2 (Extended Data Figure 6; refs 2, 26). Slu7 and Prp18 are dispensable for exon ligation when the distance between the BP and the 3'SS is less than 9 nucleotides[11]. Interestingly, in our structure 3 nucleotides of the intron downstream of the BP are visible and an additional 6 nucleotides would be sufficient to fold back and reach the catalytic Mg^{2+} site (Extended Data Fig. 8; ref. 12). When the distance to the 3'SS is less than 9 nucleotides, the 3'SS 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 present[13,27].

The DEAH ATPase Prp22, which promotes exon ligation[7,28], binds on top of the Prp8 Large domain (Fig. 4a) near where Prp16 binds in C complex, consistent with a mutually exclusive interaction with the spliceosome[10]. Compared to the crystal structure of the homologous Prp43 ATPase in the ADP-Mg-bound form[29], 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 cross-links 17 nucleotides downstream of the exon-exon junction. Indeed the distance between the catalytic centre of the spliceosome and Prp22 in our structure can be spanned with 16-17 nucleotides[6]. 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.

Methods

Pre-mRNA substrate preparation

Spliceosomes were assembled on a modified *UBC4* pre-mRNA substrate[31] containing a deoxyguanosine at the 3'-splice site UAG sequence (UAdG)[14] synthesized by ligation from a long 5' piece ending 11 nucleotides before the 3'SS and a short 3' oligonucleotide containing the dG modification and a 3'-Cy5 fluorophore (purchased from Dharmacon [GE Healthcare]). The 5' piece was generated by run-off transcription from a DNA template containing 3xMS2 stem loops[32] 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 described[14] 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)[4] were grown in a 120 L fermenter, and splicing extract was prepared using the liquid nitrogen method, essentially as described[33,34]. *In vitro* splicing reactions were assembled using pre-mRNA substrate pre-bound to MS2-MBP fusion protein, as previously described[4]. ATP was depleted 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 amylose-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-Mg^{2+} on beads, washed, and eluted with 12 mM maltose. The eluted total spliceosomes (Extended Data Fig. 1) were subsequently further purified via the Step II tag on Slu7 using Streptactin affinity resin (GE) in buffer K-100 and eluted with desthiobiotin, essentially as described[35]. The Streptactin eluate was cross-linked with 0.5 mM BS3, concentrated and buffer-exchanged with 20 mM HEPES KOH pH 7.9, 100 mM KCl, 1 mM $MgCl_2$ 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 integral core components such as Snu114 (Extended Data Fig. 1), while Prp16 is barely detectable, consistent with the purified complex being in a post-Prp16 state.

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 – 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: 1571 micrographs were collected at the LMB on an FEI Titan Krios; 2025 micrographs were collected at the Diamond Light Source (DLS) BioImaging facility on an X-FEG FEI Titan Krios. Both transmission electron microscopes were operated at 300 kV and images collected using a Gatan K2 summit direct electron detector and a GIF Quantum energy filter (slit width 20 eV). Images at the LMB were collected in super-resolution counting mode at 1.25 e/pixel / s and a calibrated pixel size of 1.43 Å/pixel.; a total dose of 40 e / $Å^2$ over 16 s and a defocus range of 0.5 – 4.5 μm were used. Images at the DLS were collected in counting mode at 2 frames / s and a calibrated pixel size of 1.025 Å; a total dose of 42 e / $Å^2$ 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 MOTIONCORR[36] 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 reconstruction[4]. 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 significantly 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 RELION[38] was then applied to all micrographs from both datasets. 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 $Å^2$ 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 subtraction[39]. 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 smaller subsets 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 (ED 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 allowed 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 criterion[40]. 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 curves[41]. 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; ref. 41).

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 C complex[4] were docked into the C* map, 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 structure[42] replaced the positions of Yju2 and Cwc25 seen in C complex and allowed building of the intron, including the branch linkage and 3

nucleotides downstream of the branch adenosine. A β-propellar domain buttressing the branch helix was assigned to Prp17. A homology model produced using SWISS-MODEL[43], based on the structure of the WD40 domain of ribosomal assembly protein 4 (PDB 5FL8), was docked based on loop sizes and occasionally visible side chains, and then manually rebuilt. Unassigned density remained around the Prp8 RNaseH-like domain. On the face projecting away from the core a characteristic shape matched the crystal structure of Prp18 (ref. 23), which could be readily docked into the map. Alpha-helical density on the top of the Prp8 RNaseH domain was assigned to Slu7 based on secondary structure predictions using the GeneSilico Metaserver[44]. Additional meandering density descending into the interface between the RNaseH and Endonuclease domains of Prp8 and passing through Prp18 as well as a helix tucked underneath the Prp8 Endonuclease domain and extended loop regions adjacent to the Endonuclease domain were also assigned to Slu7 based on secondary structure predictions, although limited resolution invites caution about the precise register of these regions. All Slu7 regions were built *de novo*.

The majority of the model building described above was for the core of the spliceosome where the resolution was uniformly between 3.5 – 5.0 Å (Extended Data Fig. 4). For the periphery of the complex, the resolution was more heterogeneous, ranging from 4.5 to 8 Å. The map obtained by focused classification of the U2 snRNP region allowed us to dock the structures of the U2 Sm ring, Msl2, and Lea1, as well as the U2 stems IIb, IIc, IV and V from the C complex structure[4]; the docked structures were then manually adjusted in Coot[45]. For Prp22 a homology model based on the crystal structure of Prp43 (ref. 29) was obtained using SWISS-MODEL[43] and then individual secondary structure elements were docked and rebuilt in Coot[45]. Finally, two copies of the Prp19 WD40 domain as well as the Prp19 coiled-coils from the C complex structure could be docked into the Prp19-focussed map.

With the exception of the Prp19 modules all models were manually adjusted in order to obtain the best fit to the cryo-EM density. The model was refined using REFMAC 5.8 (ref. 46) with secondary structure restraints generated in PROSMART[47] and RNA base-pair and stacking restraints generated in LIBG[48]. Refinement was also carried out in Phenix[49]. Extended Data Table 1 summarizes refinement statistics and PBD and EMDB accession codes.

Map visualisation

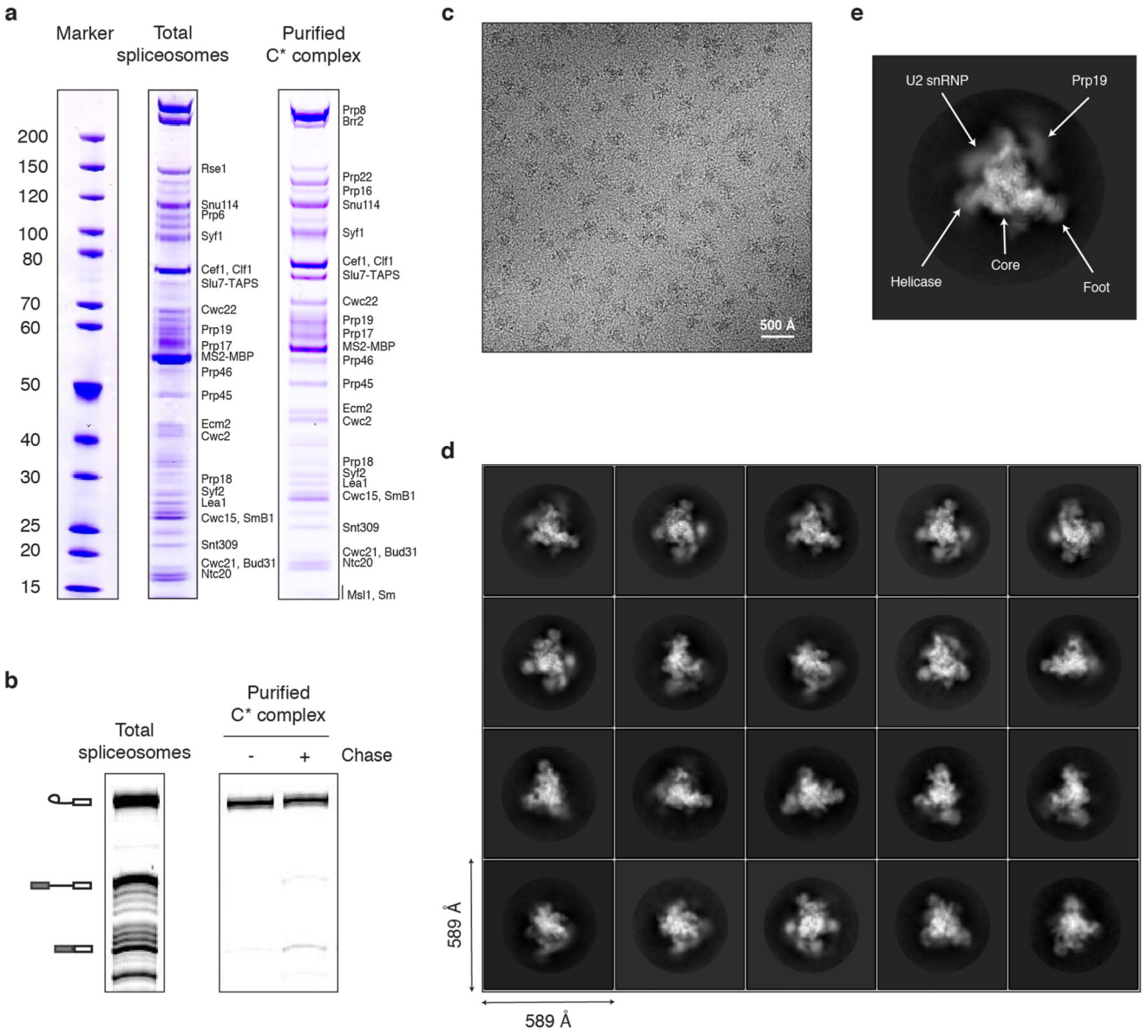
Maps were visualised in Chimera[50] and figures were prepared using PyMOL (http://www.pymol.org).

Data Availability

The cryo-EM maps have 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 code 5MPS (core) and 5MQ0 (core+Prp22+U2 snRNP).

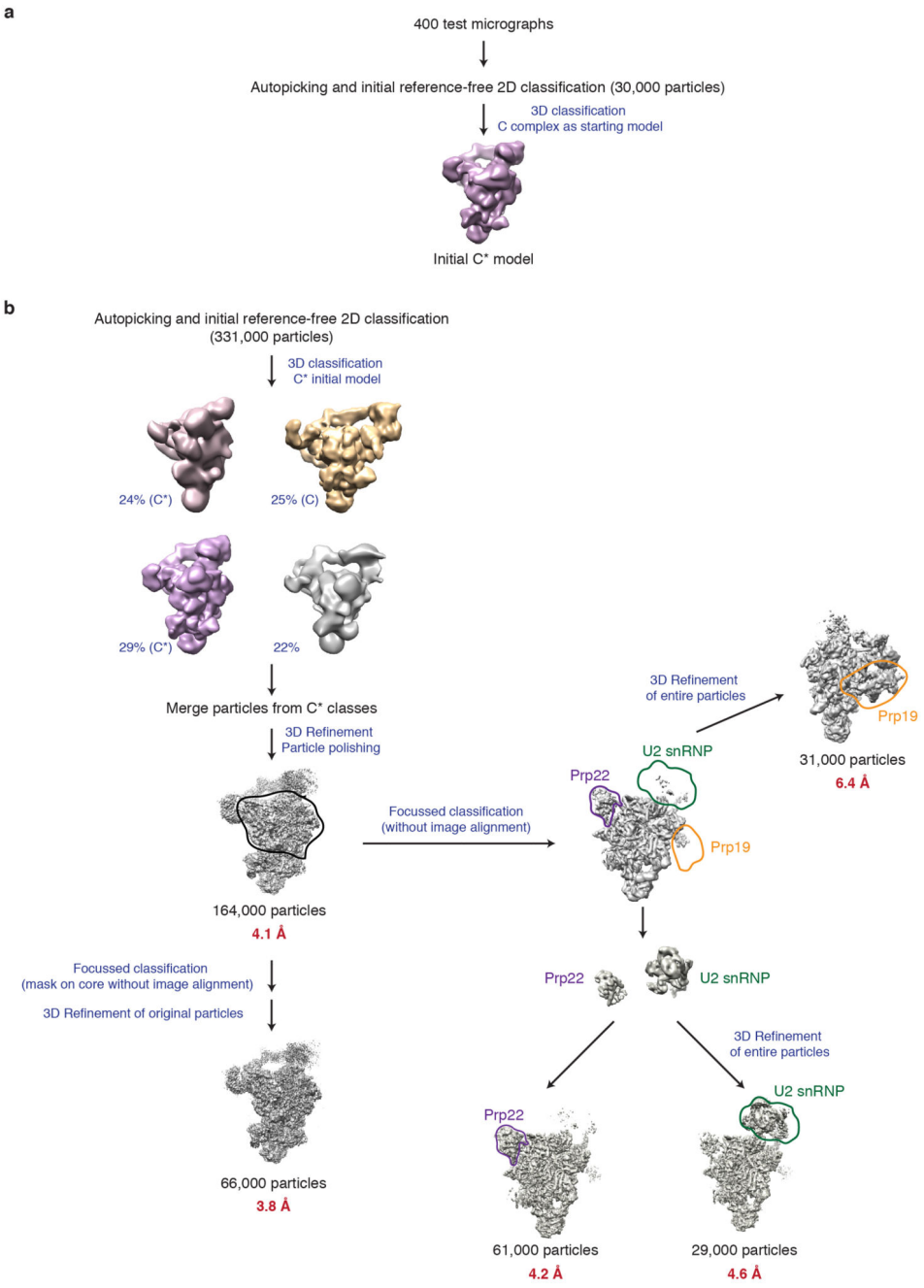
Europe PMC Funders Author Manuscripts

Extended Data



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 significantly 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[8]. **b,** The purified C* complex contains mostly lariat-intermediate and catalyses exon ligation with low efficiency when incubated in the presence of Mg^{2+}. The identity of the major species, inferred by size and migration pattern, is indicated by 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[39]. **e,** Image of a highly abundant C* complex 2D class average illustrating the major domains of the complex.

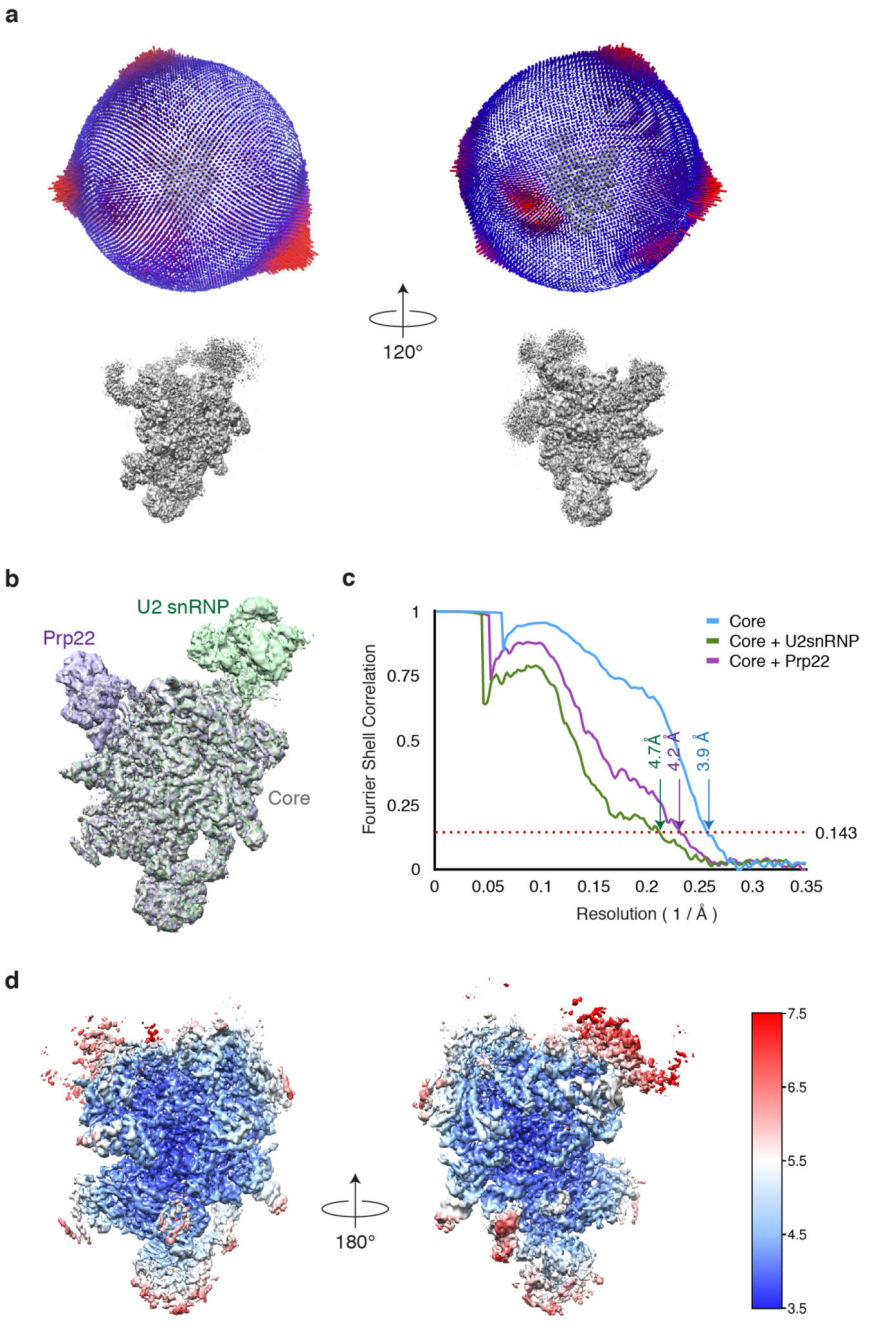


Extended Data Figure 2. Data processing workflow.

a, Method used to obtain an initial model of the C* complex for 3D classification in RELION[39]. **b,** Scheme for 3D classification and refinement. The mask used to obtain the overall 3.8Å map excluded the Clf1 and Syf1 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 RELION[39].

Europe PMC Funders Author Manuscripts

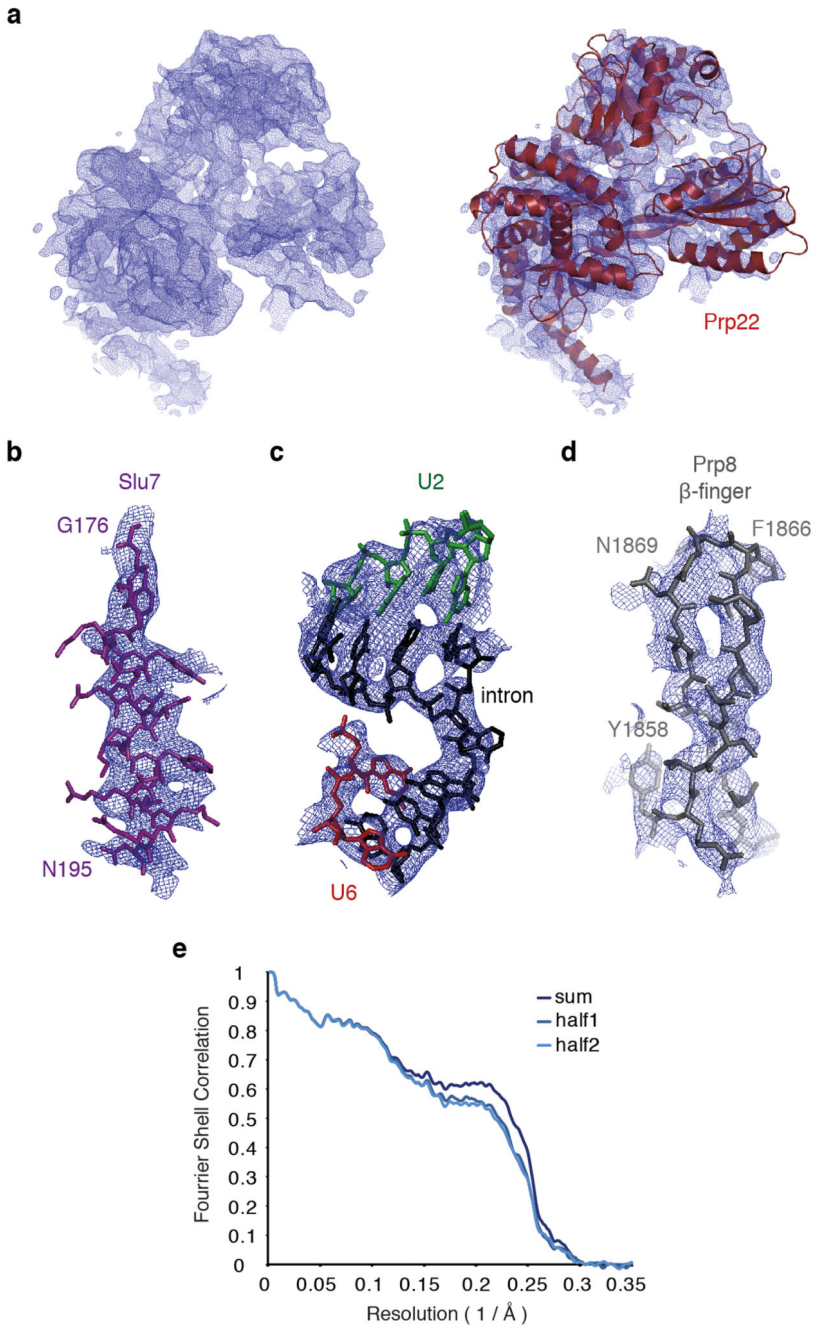
Europe PMC Funders Author Manuscripts



Extended Data Figure 3. Angular distribution and FSC curves for the C* reconstructions.

a, Angular distribution for the 3.8Å 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[50]. **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).

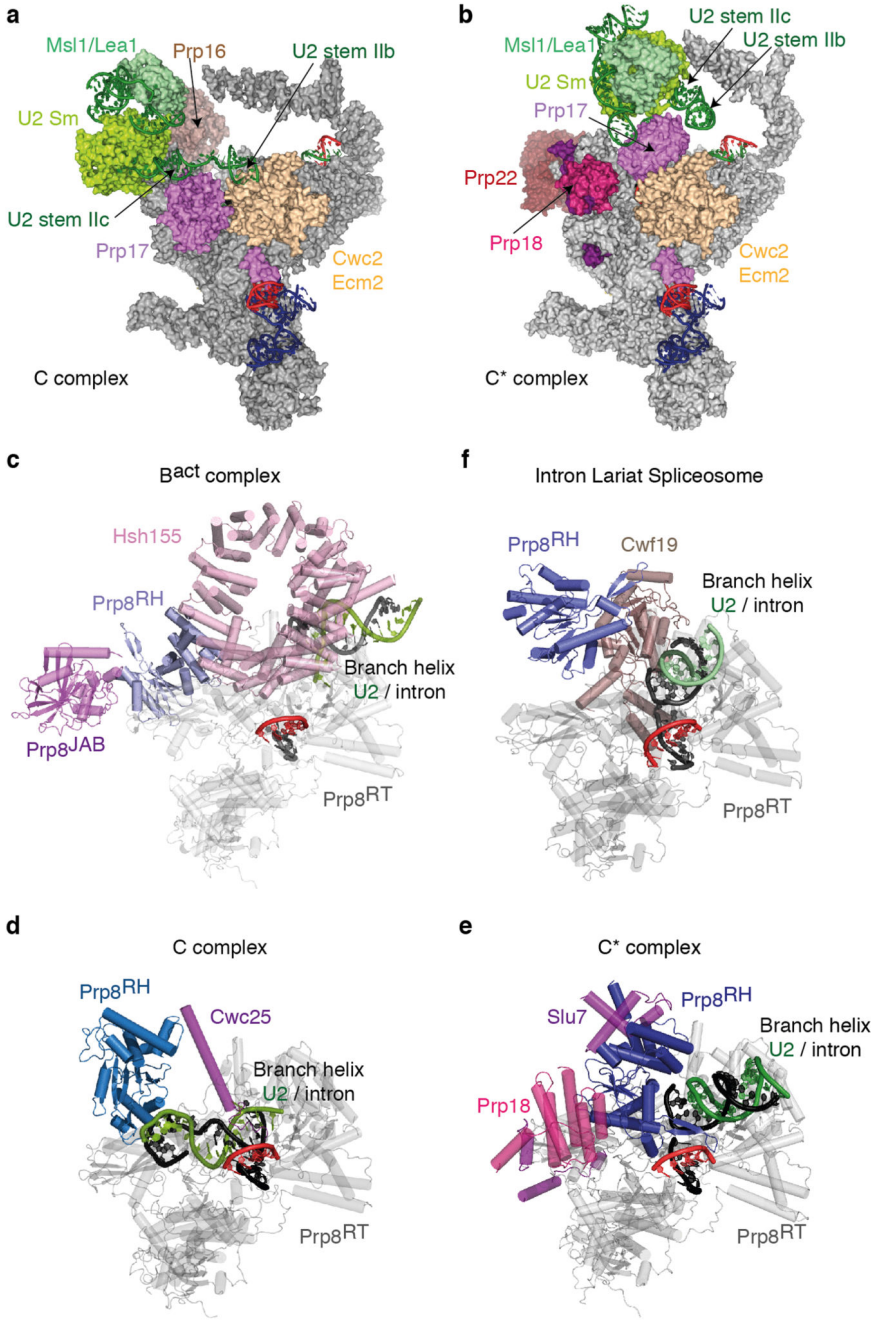
Europe PMC Funders Author Manuscripts



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.

Europe PMC Funders Author Manuscripts

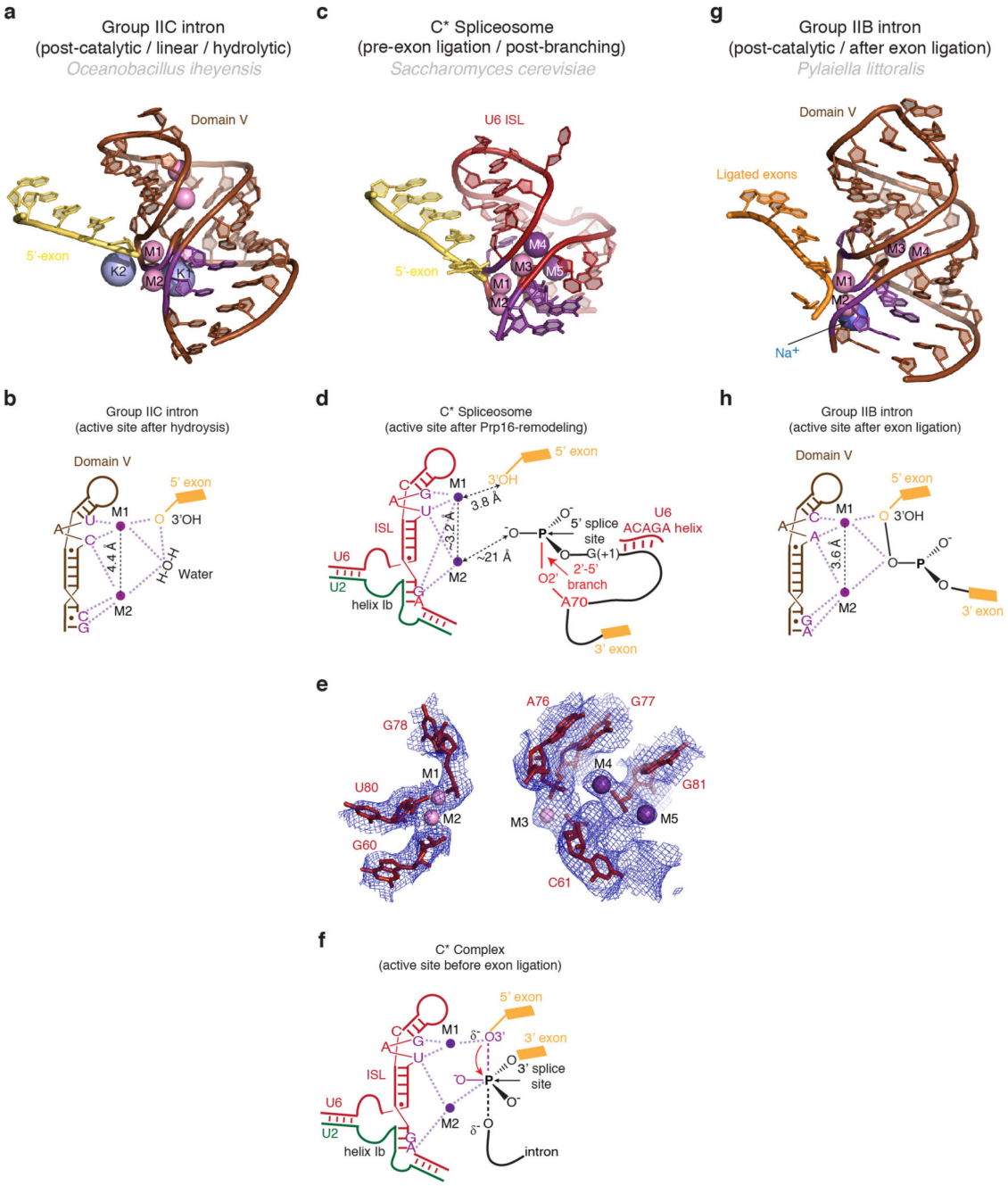


Extended Data Figure 5. U2 snRNP rearrangement between C and C* complexes and repositioning of the Prp8 RNaseH-like domain during splicing.

a-b, Movement of the U2 snRNP domain between 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**, Brr2 was omitted for clarity. **c,** Prp8 RNaseH domain conformation in the B^{act} complex (PDB 5GM6; ref. 52). Note how Hsh155 sequesters the branch helix away from the RNaseH domain, which projects its β-

hairpin into solvent and is stabilized by the Prp8 Jab1/MPN domain and Brr2 (not shown). **d-e,** RNaseH conformation in C (**d**) and C* (**e**) complexes. Note that the RNaseH 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,** RNaseH conformation in the *S. pombe* ILS. Note that the de-branching specific factor Cwf19 is now wedged between the RNaseH domain and the branch helix. All structures were aligned on the Prp8 endonuclease domain ($Prp8^{EN}$), shown in grey; complex-specific factors are coloured in magenta shades; $Prp8^{RH}$, Prp8 RNaseH-like domain.

Europe PMC Funders Author Manuscripts

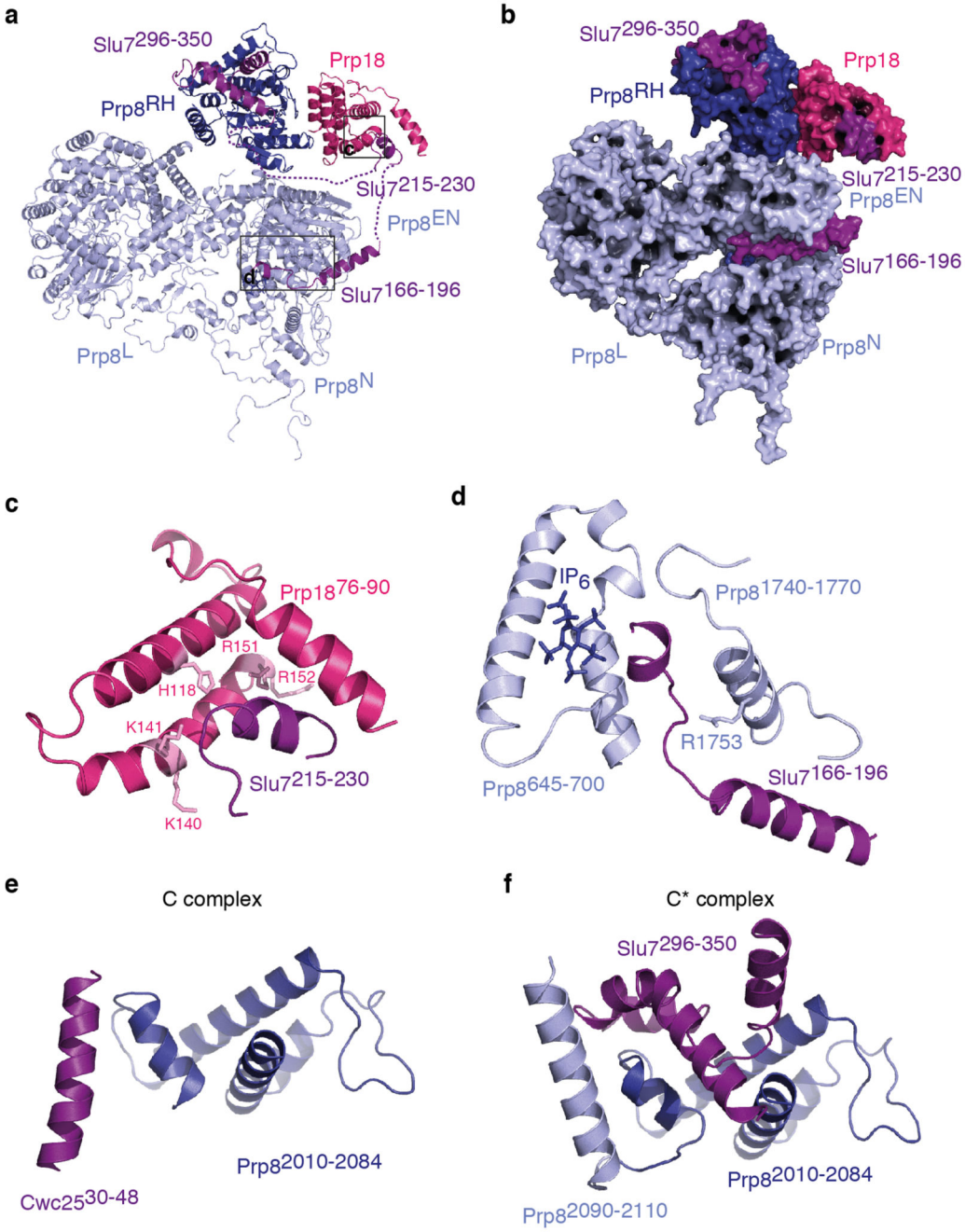


Extended Data Figure 6. Metals in the RNA core of the C* complex.

a,b, Structure (**a**) and schematic representation (**b**) of the active site of a group IIC 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-d,** Structure of the RNA at the active site of 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 structure[53]. 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.

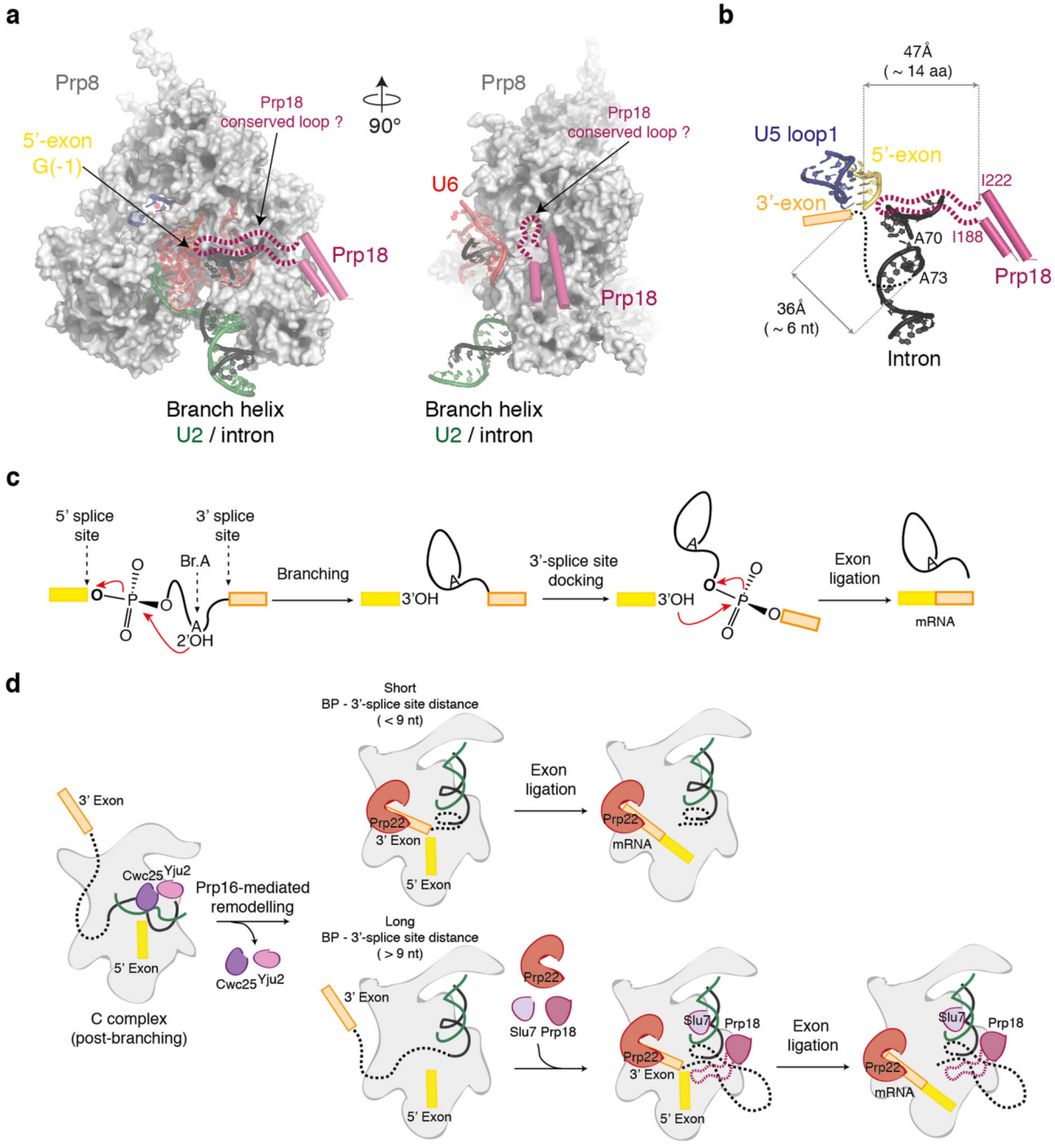
Europe PMC Funders Author Manuscripts



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 chain continuity inferred based on regions of weak density. Note how Slu7 latches the Prp8 RNaseH domain onto the Prp8 Endonuclease ($Prp8^{EN}$) and N-terminal domains ($Prp8^{N}$). **c,** Interaction between Slu7 and Prp18. The Slu7 helix was modelled based on secondary structure predictions and previously reported genetic interactions[10,24]. **d,** Interactions between Slu7 and Prp8. The R1753A allele impairs exon ligation and interacts genetically with Slu7 (ref. 55). IP_6,

inositol hexakisphosphate. **e-f,** Exchange of Cwc25 in 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 C complex; indeed, this helix is not visible in 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

Europe PMC Funders Author Manuscripts

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 6 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'SS docking and exon ligation. For pre-mRNAs with short BP to 3'SS distances Prp16 action could be sufficient to allow docking of the 3'SS. For longer branch point (BP) to 3'SS distances Prp18 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'SS 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		
$FSC_{average}$	0.801		
R factor	0.322		
Validation[3]			
Molprobity score	2.23 (100th percentile)		
Clashscore, all atoms	7.29 (100th percentile)		
Good rotamers (%)	89.6		

Europe PMC Funders Author Manuscripts

	Core	Core+Prp22	Core+U2 snRNP
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	5MQ0	5MQ0
EMDB 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 (ref. 55).

Europe PMC Funders Author Manuscripts

Extended Data Table 2

Summary of components modelled into the complex C* 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*names**
U5 snRNP	Prp8	N-terminal	1-870	101,767	127-357; 367-428; 456-870	5LJ5	Docked & rebuilt			
		Large	871-1827	111,525	871-1826	5LJ5	Docked & rebuilt	3.6 – 4.8	A	220K/*Spp42*
		RNaseH	1828-2085	29,453	1834-2085	5LJ5	Docked & rebuilt			
		Jab1/MPN	2086-2413	36,812	2086-2111	5LJ5	Docked & rebuilt			
	Snu114		1008	114,041	73-997	5LJ5	Docked & rebuilt	3.8 – 5.0	C	116K/*Cwf10*
	SmB		196	22,403	4-55; 75-102	5LJ5	Docked	4.4 – 6.0	b	SmB/*SmB*
	SmD3		110	11,229	4-85	5LJ5	Docked	4.2 – 5.4	d	SmD3/*SmD3*
	SmD1		146	16,288	1-48; 76-109	5LJ5	Docked	4.8 – 6.2	h	SmD1/*SmD1*
	SmD2		110	12,856	15-108	5LJ5	Docked	5.0 – 6.8	j	SmD2/*SmD2*
	SmF		94	10,373	12-83	5LJ5	Docked	5.6 – 7.6	f	SmF/*SmF*
	SmE		96	9,659	10-64; 73-92	5LJ5	Docked	5.2 – 7.0	e	SmE/*SmE*
	SmG		77	8,479	2-46; 53-76	5LJ5	Docked	4.4 – 6.4	g	SmG/*SmG*
	U5 snRNA-L		214	68,847	4-53; 62-145; 167-173	5LJ5	Docked & rebuilt	3.6 – 6.2	5	U5 snRNA
U2 snRNP	Msl1		111	12,830	28-111	5LJ5	Docked	~6 – 10	Y	U2-B"
	Lea1		238	27,193	1-167	5LJ5	Docked	~6 – 10	W	U2-A'
	SmB		196	22,403	4-55; 75-102	5LJ5	Docked	~6 – 10	k	SmB/*SmB*
	SmD3		110	11,229	4-85	5LJ5	Docked	~6 – 10	n	SmD3/*SmD3*
	SmD1		146	16,288	1-48; 76-106	5LJ5	Docked	~6 – 10	I	SmD1/*SmD1*

Europe PMC Funders Author Manuscripts

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*names**
	SmD2		110	12,856	15-108	5LJ5	Docked	~6 – 10	m	SmD2/*SmD2*
	SmF		94	10,373	12-83	5LJ5	Docked	~6 – 10	q	SmF/*SmF*
	SmE		96	9,659	10-64; 73-92	5LJ5	Docked	~6 – 10	p	SmE/*SmE*
	SmG		77	8,479	2-46; 53-76	5LJ5	Docked	~6 – 10	r	SmG/*SmG*
	U2 snRNA		1175	363,824	1-49	5LJ5	Docked & rebuilt	3.8 – 5.8	2	U2 snRNA
					54-150	5LJ5	Docked & adjusted	~6 – 10	2	U2 snRNA
					1089-1169	5LJ5	Docked & adjusted	~6 – 10	2	U2 snRNA
U6	U6 snRNA		112	36,088	1-10; 16-104	5LJ5	Docked & rebuilt	3.6 – 6.2	6	U6 snRNA
NTC	Prp19	U-box	1-51	5,713	1-51	5LJ5	Docked	~10 – 15	t,u,v,w	PRPF19/*Cwf8*
		Coiled-coil	52-143	10,247	78-143	5LJ5	Docked	~10 – 15		
		WD40	144-503	40,646	171-501	5LJ5	Docked	~15 – 20		
	Snt309		175	20,709	12-174	5LJ5	Docked	~15 – 20	s	BCAS2/*Cwf7*
	Syf1	Periphery	859	100,229	21-405	5LJ5	Idealised alpha helices	~8 – 12	T	SYF1/*Cwf3*
		Core			407-811	5LJ5	Idealised alpha helices	5.6 – 8.4		
	Syf2		215	24,803	95-115	5GMK	Docked & rebuilt	4.0 – 4.8	y	SYF2/*Syf2*
					142-186	5GMK	Docked & rebuilt			SYF2/*Syf2*
					194-210	5GMK	Docked & rebuilt			SYF2/*Syf2*
	Clf1		1-687	82,463	36-556	5LJ5	Docked & rebuilt	3.8 – 8.2	S	CRNKLI/*Cwf4*
	Cef1		1-590	97,767	5-107; 144-251	5LJ5	Docked & rebuilt	3.8 – 4.6	O	CDC5L/*Cdc5*
NTR	Prp45		379	42,483	31-225	5LJ5	Docked & rebuilt	3.6 – 4.8	K	SNW1/*Prp45*
	Prp46		451	50,700	110-451	5LJ5	Docked & rebuilt	3.6 – 4.4	J	PLRG1/*Prp5*

Europe PMC Funders Author Manuscripts

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*names**
	Ecm2		364	40,925	4-324	5LJ5	Docked & rebuilt	3.8 – 5.6	N	RBM22/*Cwf5*
	Cwc2		339	38,431	3-254	5LJ5	Docked & rebuilt	3.8 – 4.6	M	RBM22/*Cwf2*
	Cwc15		175	19,935	4-42; 126-175	5LJ5	Docked & rebuilt	3.6 - 7.6	P	CWC15/*Cwf15*
	Bud31		157	18,447	2-156	5LJ5	Docked & rebuilt	3.8 – 4.4	L	BUD31/*Cwf14*
Splicing Factors	Prp18		251	28,377	76-188; 222-247	1DVK	Docked & rebuilt	4.2 – 6.2	a	PRPF18/*Prp18*
	Prp17		455	52,048	51-73; 154-455	5GMK 5FL8	Homology modelled & rebuilt	4.0 – 5.2	0	CDC40/*Prp17*
	Slu7		382	44,637	166-196		De novo	3.8 – 5.8	c	hSLU7/*Slu7*
					215-230					
					296-350					
	Cwc21	N-terminal	1-64	7,057	2-50	5LJ5	Docked & rebuilt	3.6 – 6.8	R	SRRM2/*Cwf21*
		Coiled-coil	65-135	8,724	58-116	5LJ5	Docked & rebuilt			
	Cwc22	MIF4G	1-288	33,187	11-262	5LJ5	Docked & rebuilt	3.8 – 7.4	H	CWC22/*Cwf22*
		MA3	289-577	34,125	289-481	5LJ5	Docked & rebuilt			
Helicase	Prp22		1,124	130,014	476-1124	3KX2	Homology modelled & adjusted	~4.5 – 7	V	DHX8/*Prp22*
Substrate	5'-exon		20	6,683	(-16) - (-1)	5LJ5	Docked & rebuilt	3.6 – 4.2	E	
	Intron		95	30,405	1-16; 56-73	3JB9	Homology modelled & rebuilt	3.8 – 5.0	I	
Unknown	X		66				Idealised alpha helices	~4.5 – 6	X	

Acknowledgements

We thank S. Scheres for his help and advice on data collection and processing; C. Savva, S. Chen, K. R. Vinothkumar, G. McMullan, J. Grimmett and T. Darling 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 Skłodowska-Curie fellowships, M.E.W was supported by a Rutherford Memorial Cambridge Scholarship.

References

1. Wahl MC, Will CL, Lührmann R. The Spliceosome: Design principles of a dynamic RNP machine. Cell. 2009; 136:701–718. [PubMed: 19239890]
2. Fica SM, et al. RNA catalyses nuclear pre-mRNA splicing. Nature. 2013; 503:229–234. [PubMed: 24196718]
3. Madhani HD, Guthrie C. A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. Cell. 1992; 71:803–817. [PubMed: 1423631]
4. Galej WP, et al. Cryo-EM structure of the spliceosome immediately after branching. Nature. 2016; 537:197–201. [PubMed: 27459055]
5. Wan R, Yan C, Bai R, Huang G, Shi Y. Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution. Science. 2016; 353:895–904. [PubMed: 27445308]
6. Schwer B. A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. Mol Cell. 2008; 30:743–754. [PubMed: 18570877]
7. Semlow DR, Blanco MR, Walter NG, Staley JP. Spliceosomal DEAH-Box ATPases Remodel Pre-mRNA to Activate Alternative Splice Sites. Cell. 2016; 164:985–998. [PubMed: 26919433]
8. Schwer B, Guthrie C. Prp16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. Nature. 1991; 349:494–499. [PubMed: 1825134]
9. Tseng CK, Liu HL, Cheng SC. DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. RNA. 2010; 17:145–154. [PubMed: 21098140]
10. James S-A, Turner W, Schwer B. How Slu7 and Prp18 cooperate in the second step of yeast pre-mRNA splicing. RNA. 2002; 8:1068–1077. [PubMed: 12212850]
11. Ohrt T, et al. Molecular dissection of step 2 catalysis of yeast pre-mRNA splicing investigated in a purified system. RNA. 2013; 19:902–915. [PubMed: 23685439]
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. 1996; 2:707–717. [PubMed: 8756413]
13. Chua K, Reed R. The RNA splicing factor hSlu7 is required for correct 3' splice-site choice. Nature. 1999; 402:207–210. [PubMed: 10647016]
14. Moore MJ, Sharp PA. Site-specific modification of pre-mRNA: the 2'-hydroxyl groups at the splice sites. Science. 1992; 256:992–997. [PubMed: 1589782]
15. Fica SM, Mefford MA, Piccirilli JA, Staley JP. Evidence for a group II intron-like catalytic triplex in the spliceosome. Nat Struct Mol Biol. 2014; 21:464–471. [PubMed: 24747940]
16. Newman AJ, Norman C. U5 snRNA interacts with exon sequences at 5' and 3' splice sites. Cell. 1992; 68:743–754. [PubMed: 1739979]
17. Sontheimer EJ, Steitz JA. The U5 and U6 small nuclear RNAs as active site components of the spliceosome. Science. 1993; 262:1989–1996. [PubMed: 8266094]
18. Siatecka M, Reyes JL, Konarska MM. Functional interactions of Prp8 with both splice sites at the spliceosomal catalytic center. Genes Dev. 1999; 13:1983–1993. [PubMed: 10444596]
19. Collins CA, Guthrie C. Genetic interactions between the 5' and 3' splice site consensus sequences and U6 snRNA during the second catalytic step of pre-mRNA splicing. RNA. 2001; 7:1845–1854. [PubMed: 11780639]

20. Jacquier A, Michel F. Base-pairing interactions involving the 5' and 3'-terminal nucleotides of group II self-splicing introns. J Mol Biol. 1990; 213:437–447. [PubMed: 2191139]
21. Konarska MM, Vilardell J, Query CC. Repositioning of the Reaction Intermediate within the Catalytic Center of the Spliceosome. Mol Cell. 2006; 21:543–553. [PubMed: 16483935]
22. Villa T, Guthrie C. The Isy1p component of the NineTeen complex interacts with the ATPase Prp16p to regulate the fidelity of pre-mRNA splicing. Genes Dev. 2005; 19:1894–1904. [PubMed: 16103217]
23. Jiang J, Horowitz DS, Xu RM. Crystal structure of the functional domain of the splicing factor Prp18. Proc Natl Acad Sci USA. 2000; 97:3022–3027. [PubMed: 10737784]
24. Zhang X, Schwer B. Functional and physical interaction between the yeast splicing factors Slu7 and Prp18. Nucleic Acids Res. 1997; 25:2146–2152. [PubMed: 9153314]
25. Luukkonen BG, Séraphin B. The role of branchpoint-3' splice site spacing and interaction between intron terminal nucleotides in 3' splice site selection in Saccharomyces cerevisiae. EMBO J. 1997; 16:779–792. [PubMed: 9049307]
26. Steitz TA, Steitz JA. A general two-metal-ion mechanism for catalytic RNA. Proc Natl Acad Sci USA. 1993; 90:6498–6502. [PubMed: 8341661]
27. Frank D, Guthrie C. An essential splicing factor, SLU7, mediates 3' splice site choice in yeast. Genes Dev. 1992; 6:2112–2124. [PubMed: 1427075]
28. Schwer B, Gross CH. Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. EMBO J. 1998; 17:2086–2094. [PubMed: 9524130]
29. He Y, Andersen GR, Nielsen KH. Structural basis for the function of DEAH helicases. EMBO Rep. 2010; 11:180–186. [PubMed: 20168331]
30. Hilliker AK, Mefford MA, Staley JP. U2 toggles iteratively between the stem IIa and stem IIc conformations to promote pre-mRNA splicing. Genes Dev. 2007; 21:821–834. [PubMed: 17403782]
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. 2010; 472:31–40. [PubMed: 20580958]
32. Zhou Z, Licklider LJ, Gygi SP, Reed R. Comprehensive proteomic analysis of the human spliceosome. Nature. 2002; 419:182–185. [PubMed: 12226669]
33. Umen JG, Guthrie C. A novel role for a U5 snRNP protein in 3' splice site selection. Genes Dev. 1995; 9:855–868. [PubMed: 7535718]
34. Lin RJ, Newman AJ, Cheng SC, Abelson J. Yeast mRNA splicing in vitro. J Biol Chem. 1985; 260:14780–14792. [PubMed: 2997224]
35. Nguyen THD, et al. Cryo-EM structure of the yeastU4/U6.U5 tri-snRNP at 3.7Å resolution. Nature. 2016; 530:298–302. [PubMed: 26829225]
36. Li X, et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nat Methods. 2013; 10:584–590. [PubMed: 23644547]
37. Rohou A, Grigorieff N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J Struct Biol. 2015; 192:216–221. [PubMed: 26278980]
38. Scheres SHW. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol. 2012; 180:519–530. [PubMed: 23000701]
39. Scheres SH. Processing of Structurally Heterogeneous Cryo-EM Data in RELION. Methods Enzymol. 2016; 579:125–157. [PubMed: 27572726]
40. Scheres SHW, Chen S. Prevention of overfitting in cryo-EM structure determination. Nature Methods. 2012; 9:853–854. [PubMed: 22842542]
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. 2013; 135:24–35. [PubMed: 23872039]
42. Yan C, et al. Structure of a yeast spliceosome at 3.6-angstrom resolution. Science. 2015; 349:1182–1191. [PubMed: 26292707]
43. Biasini M, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 2014; 42:W252–258. [PubMed: 24782522]

44. Kozlowski LP, Bujnicki JM. MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. BMC Bioinformatics. 2012; 13:111. [PubMed: 22624656]
45. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010; 66:486–501. [PubMed: 20383002]
46. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr. 1997; 53:240–255. [PubMed: 15299926]
47. Nicholls RA, Fischer M, McNicholas S, Murshudov GN. Conformation-independent structural comparison of macromolecules with ProSMART. Acta Crystallogr D Biol Crystallogr. 2014; 70:2487–2499. [PubMed: 25195761]
48. Brown A, et al. Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. Acta Crystallogr D Biol Crystallogr. 2015; 71:136–153. [PubMed: 25615868]
49. Zwart PA, et al. Automated structure solution with the PHENIX suite. Methods Mol Biol. 2008; 426:419–435. [PubMed: 18542881]
50. Goddard TD, Huang CC, Ferrin TE. Visualizing density maps with UCSF Chimera. J Struct Biol. 2007; 157:281–287. [PubMed: 16963278]
51. Yan C, Wan R, Bai R, Huang G, Shi Y. Structure of a yeast activated spliceosome at 3.5 Å resolution. Science. 2016; 353:904–911. [PubMed: 27445306]
52. Marcia M, Pyle AM. Visualizing group II intron catalysis through the stages of splicing. Cell. 2012; 151:497–507. [PubMed: 23101623]
53. Robart AR, Chan RT, Peters JK, Rajashankar KR, Toor N. Crystal structure of a eukaryotic group II intron lariat. Nature. 2014; 514:193–197. [PubMed: 25252982]
54. Aronova A, Bacíková D, Crotti LB, Horowitz DS, Schwer B. Functional interactions between Prp8, Prp18, Slu7, and U5 snRNA during the second step of pre-mRNA splicing. RNA. 2007; 13:1437–1444. [PubMed: 17626844]
55. Davis IW, Murray LW, Richardson JS, Richardson DC. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Res. 2004; 32:W615–W619. [PubMed: 15215462]

Europe PMC Funders Author Manuscripts

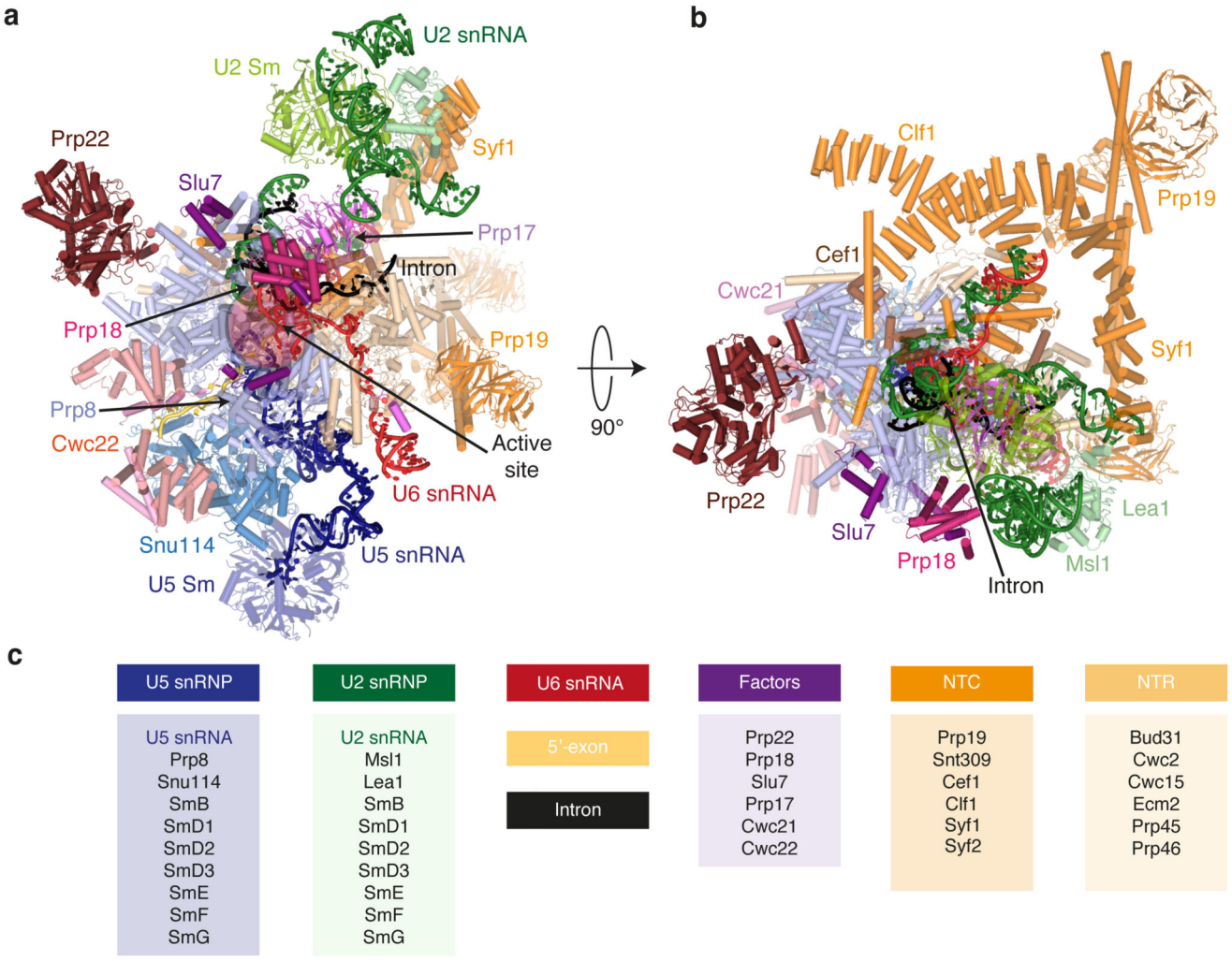


Figure 1. Subunit organization of the C* spliceosome.

a-b, Orthogonal views of the complex coloured by subunit identity. **c,** List of modelled subunits grouped into functional sub-complexes.

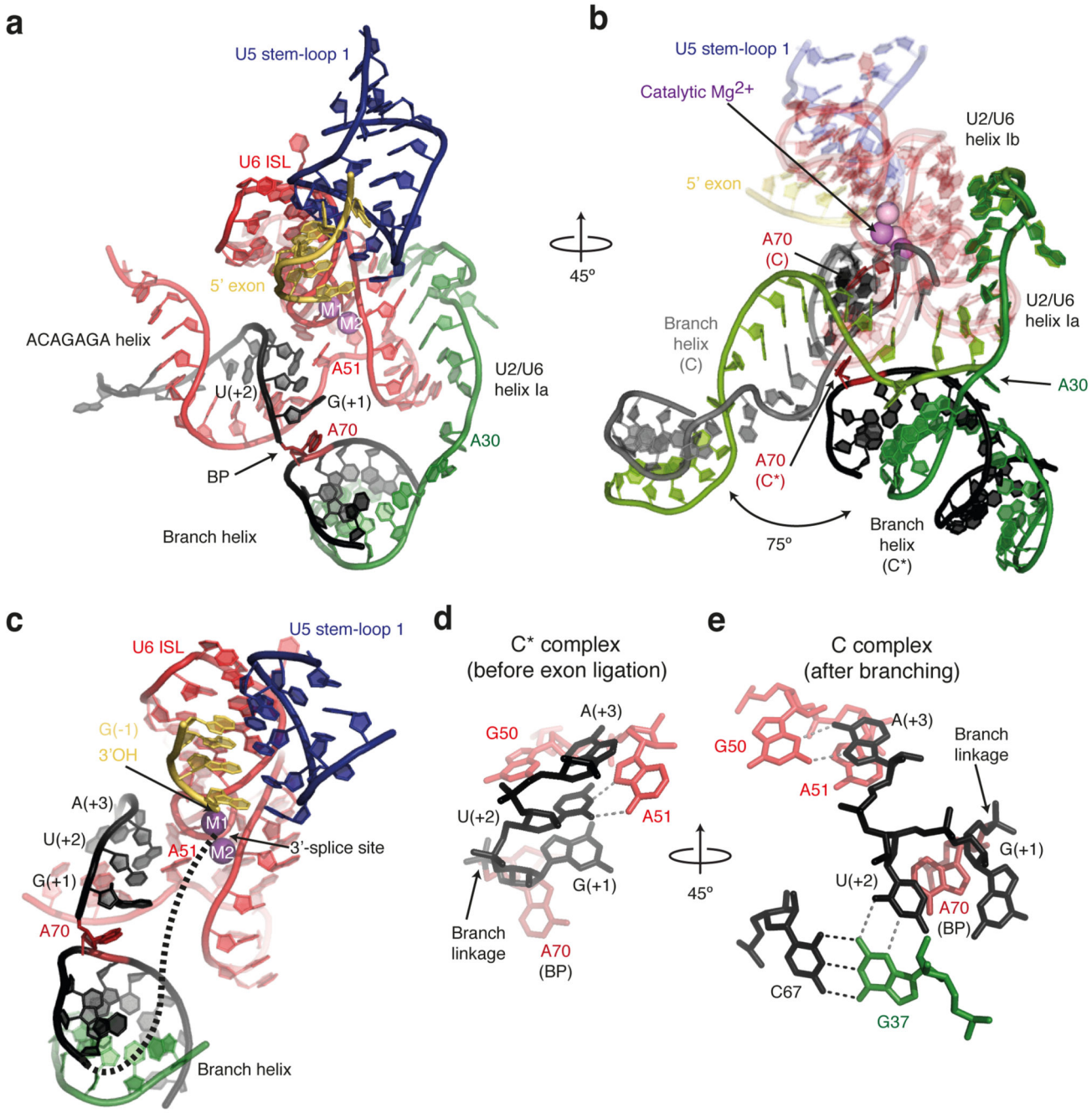


Figure 2. Architecture of the RNA catalytic core in C* complex.

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 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 C complex.

Europe PMC Funders Author Manuscripts

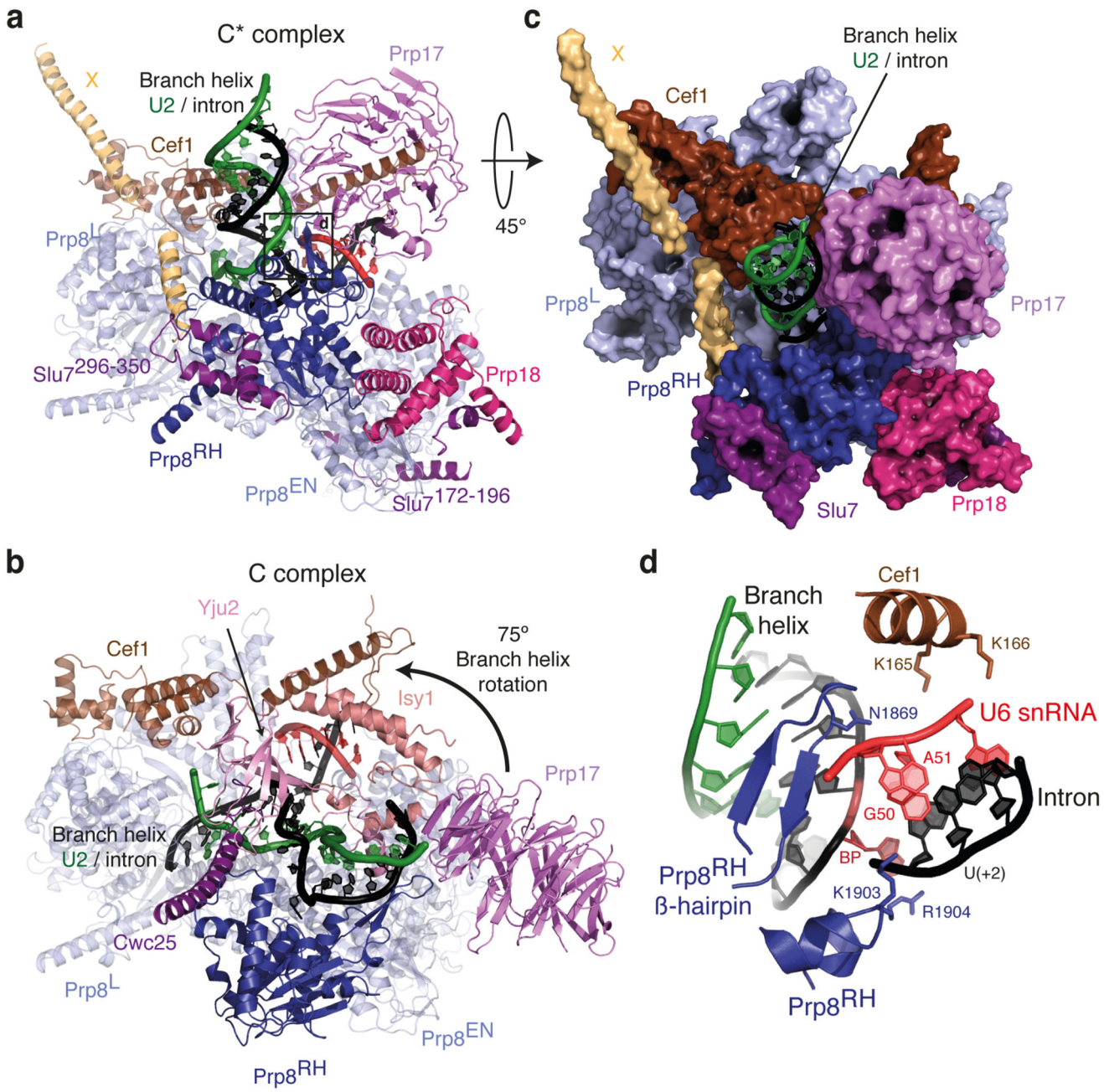


Figure 3. Proteins stabilise the repositioned branch helix.

a, Key proteins that interact with the branch helix in C* complex. Prp8^{RH}, Prp8 RNaseH-like domain; Prp8^{EN}, Prp8 Endonuclease domain; Prp8^{L}, Prp8 Large domain. **b**, Key proteins that interact with the branch helix in C complex. Prp17 has undergone a dramatic repositioning between C and C* complexes. **c,** Surface representation of proteins that promote an undocked conformation of the branch helix. Slu7 and Prp18 bind on the surface of the Prp8 RNaseH domain. **d,** The Prp8 β-hairpin stabilises the reorganised interactions between the 5'-splice site and the ACAGAGA sequence of U6 snRNA.

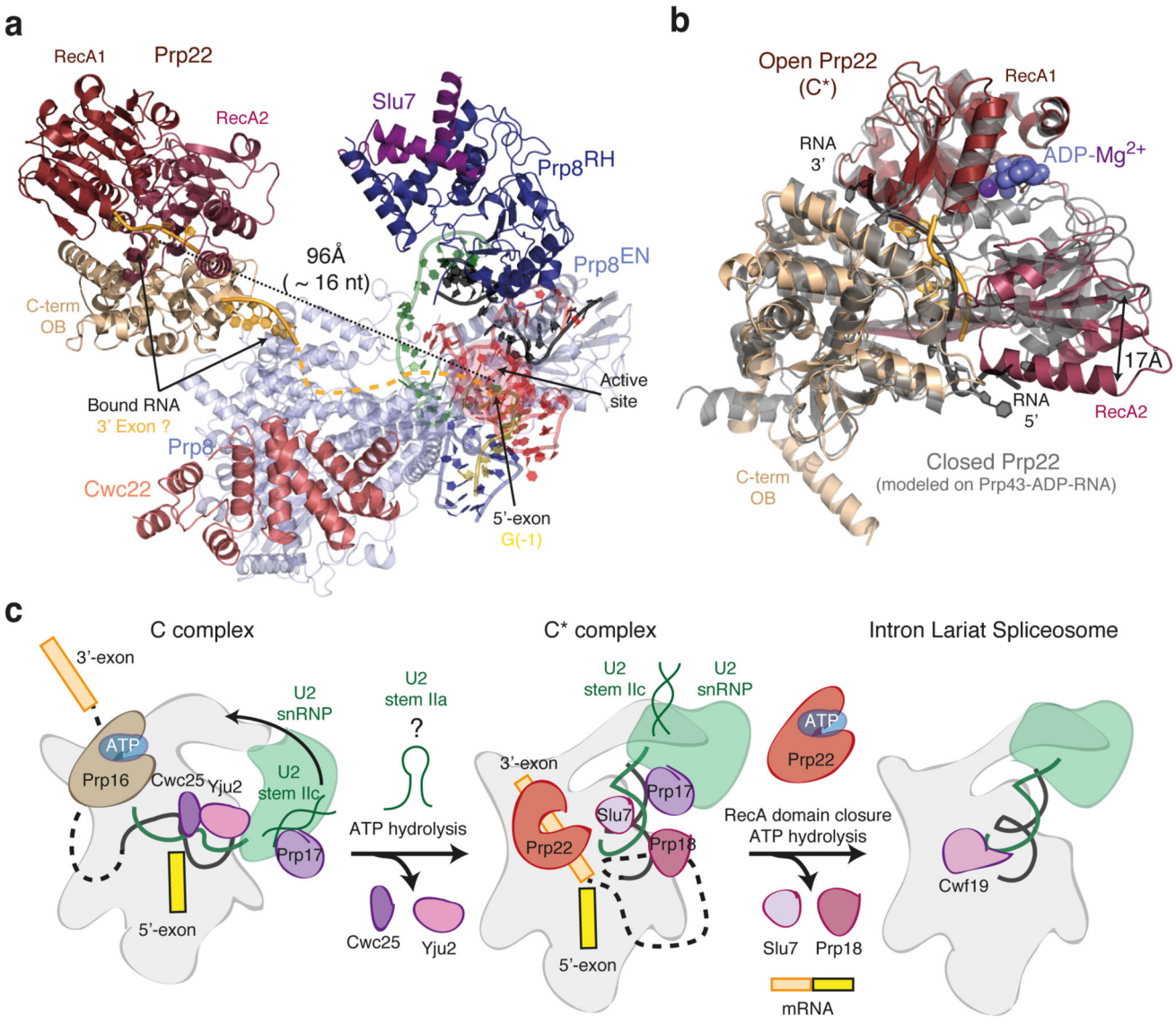


Figure 4. Prp22 and ATP-mediated transitions at the catalytic stage of splicing.

a, Position of the helicase Prp22 in the C* complex. Focussed classification allowed us to obtain an improved map for Prp22 into which individual secondary structural elements could be modelled (Extended Data Fig. 4). Helicase domains are coloured individually and RNA (shown in pale orange) was modelled in density observed at the entrance and in the centre of the helicase RNA channel. An RNA path (shown as dashed line) can be modelled from the active site of the spliceosome to Prp22. The distance from the end of the 5'-exon to the centre of the Prp22 RNA channel could be spanned by 16 nucleotides. **b,** Open conformation of Prp22 observed in C*. The model of the closed conformation, based on the structure of Prp43 (PDB 3KX2; ref. 29), is in grey. **c,** Model for Prp16-mediated remodelling. U2 stem IIa could form at an intermediate stage in remodelling to allow a binding switch of Prp17 from stem IIc to the branch helix (cf. ref. 30). Following domain closure and ATP binding, Prp22 would translocate the 3'-exon and release the mRNA.