The architecture of the spliceosomal U4/U6.U5 tri–snRNP

Thi Hoang Duong Nguyen1, Wojciech P. Galej1, Xiao-chen Bai1, Christos G. Savva1, Andrew J. Newman1, Sjors H. W. Scheres1 & Kiyoshi Nagai1

U4/U6.U5 tri–snRNP is a 1.5-megadalton pre–assembled spliceosomal complex comprising U5 small nuclear RNA (snRNA), extensively base–paired U4/U6 snRNAs and more than 30 proteins, including the key components Prp8, Brr2 and Snu114. The tri–snRNP combines with a precursor messenger RNA substrate bound to U1 and U2 small nuclear ribonucleoprotein particles (snRNPs), and transforms into a catalytically active spliceosome after extensive compositional and conformational changes triggered by unwinding of the U4 and U6 (U4/U6) snRNAs. Here we use cryo–electron microscopy single–particle reconstruction of Saccharomyces cerevisiae tri–snRNP at 5.9 Å resolution to reveal the essentially complete organization of its RNA and protein components. The single–stranded region of U4 snRNA between its 3′ stem–loop and the U4/U6 snRNA stem 1 is loaded into the Brr2 helicase active site ready for unwinding. Snu114 and the amino–terminal domain of Prp8 position U5 snRNA to insert its loop 1, which aligns the exons for splicing, into the Prp8 active site cavity. The structure provides crucial insights into the activation process and the active site of the spliceosome.

The protein–coding sequences of most eukaryotic genes are interrupted by non-coding segments called introns. Introns are removed from precursor mRNAs (pre-mRNAs) and the flanking coding segments (exons) are spliced together to form mRNAs by two successive trans–esterification reactions within a dynamic multi–megadalton protein–RNA complex known as the spliceosome. This complex comprises five canonical subunits, namely U1, U2, U4, U5 and U6 snRNPs, and numerous non–snRNP factors. Each snRNP contains an snRNA, seven Sm or LSm proteins, and a number of snRNP–specific proteins. During the initial stages of spliceosome assembly, U1 and U2 snRNPs recognize the pre–mRNA 5′ splice site and branch point, forming pre–spliceosomal A complex. The subsequent binding of the pre–assembled U4/U6.U5 tri–snRNP allows formation of the fully assembled spliceosomal B complex, which is converted to the catalytically active B* complex through extensive structural and compositional remodelling. During this process, U4 and U6 snRNAs, which are extensively base–paired in tri–snRNP, are unwound, U1 and U4 snRNPs are released, and many new proteins join the spliceosome1,2. This leads to the formation of a highly structured RNA network between U2, U5 and U6 snRNAs and the 5′ splice site and branch point sequences in the pre–mRNA. The extensively base–paired U2–U6 snRNAs harbour catalytic magnesium ions and position the branch point and 5′ splice site for the first trans–esterification reaction, which produces exon 1 and lariat intron–exon 2 intermediates. Further remodelling to C complex enables U5 snRNA loop 1 to align exons 1 and 2 for nucleophilic attack of exon 1 at the 3′ splice site, yielding spliced mRNA and lariat intron products3,4. Finally, the spliceosome is disassembled before the next round of splicing.

U4/U6.U5 tri–snRNP is the largest pre–assembled spliceosomal complex, containing U5 snRNA, extensively base–paired U4/U6 snRNAs and over 30 proteins5,6 (Extended Data Table 1). Three key proteins, Prp8, Brr2 and Snu114, have crucial roles in activation of the spliceosome and formation of the active site1. Prp8 forms crosslinks with 4-thiouridine introduced at key positions within U5 and U6 snRNAs and the substrate pre–mRNA, showing that Prp8 is involved in substrate positioning and closely associated with the catalytic RNA core7,8. Brr2 helicase, the activity of which is regulated by the GTPase Snu114 (refs 11–13), catalyses the unwinding of the U4/U6 snRNA duplex14,15. Interactions between tri–snRNP proteins have been investigated by yeast two–hybrid and in vitro binding assays16,17. Electron cryo–microscopy (cryoEM) reconstruction of crosslinked human tri–snRNP at 21 Å resolution revealed a tetrahedral overall shape with no clear domain separation18. Negative stain microscopy of crosslinked yeast tri–snRNP revealed a triangular shape with maximum dimension of 30–34 nm (ref. 19). The highly biased orientation of tri–snRNP on carbon films precluded full three–dimensional analysis while the projection structure revealed three extruding domains termed head, foot and arm; the arm domain adopts variable positions with respect to the rest. Some key proteins were localized within the projection structure using genetically introduced tags19. Brr2 and U4/U6 snRNPs were attributed to the head and arm domains, respectively. On the basis of this, it was proposed that Brr2 may engage with U4/U6 snRNAs for unwinding when Snu114—mapped in the hinge region—brings the arm and head domains closer19.

The development of high–speed direct electron detectors20,21 and powerful maximum likelihood algorithms for classification and particle alignment22 have made it possible to determine the structure of macromolecular assemblies at near–atomic resolution by cryoEM23. By applying these new methods we obtained a map of native unstained yeast tri–snRNP at an overall resolution of 5.9 Å in which protein 2–helices and RNA double helices are readily discernible. This enabled us to fit the double–stranded helices of U5 snRNA and U4/U6 snRNAs as well as previously determined crystal structures or homology models of nearly all the proteins. The structure accounts for a wealth of biochemical and genetic data from yeast and human spliceosomes, and suggests a possible mechanism for B complex formation and the activation of the spliceosome.
CryoEM of the tri-snRNP complex

U4/U6,U5 tri-snRNP was purified from yeast by a gentle procedure without crosslinking, and the sample was subjected to cryoEM analyses (Methods and Extended Data Fig. 1). Using a combination of statistical classification and movie processing\(^2\) (Extended Data Fig. 2), we obtained a density map with an overall resolution of 5.9 Å by the ‘gold standard’ Fourier shell correlation (FSC)\(^5\) criterion with local resolution ranging from 5.0 Å to 20 Å (Extended Data Fig. 3 and 4; Methods). The map revealed clear densities for double-stranded RNA, with protein helices appearing as long tubes and \(\beta\)-sheets as flat densities (Supplementary Video 1). The density for the LSm proteins in the flexible arm domain became clearer after using a new multi-body refinement method (Methods and Extended Data Fig. 3).

Overall structure

Yeast U4/U6,U5 tri-snRNP has an overall Y-shape with a maximum dimension of approximately 300 Å (Fig. 1). The large domain of Prp8 (residues 885–1,824), consisting of the reverse transcriptase-like (RT), linker and type II endonuclease-like domains\(^1\), is located near the centre of the assembly and its crystal structure was fitted into the map as a rigid body\(^10\) (Fig. 2 and Extended Data Fig. 5a). The orientation of the RNaseH-like domain with respect to the large domain is inverted in tri-snRNP as compared with the Prp8–Aar2 complex\(^10\) (Fig. 2d). Three segments of clear double-stranded RNA density extending from Prp8 to the foot domain are assigned to co-axially stacked stems I and II of U5 snRNA (Fig. 3; Extended Data Fig. 6) connected to the U5 Sm core (Extended Data Fig. 5c). Snu114 shows a significant sequence similarity to eukaryotic translation elongation factor 2 (EF2)\(^11,26\) comprising domains I–V (Fig. 4 and Extended Data Figs 5c and 7). Homology models of each domain of Snu114 (residues 120–1,088) were fitted individually into the density adjacent to Prp8 and U5 snRNA, revealing a contact between domain III of Snu114 and the RT domain of Prp8 (Fig. 4). The structure of the N-terminal domain of Prp8 is still unknown. The N-terminal helix of the RT domain of Prp8 (RT\(\alpha_1\))\(^10\) extends further in tri-snRNP towards a bundle of four long helices. Another cluster of long helices, which makes close contact with the co-axially stacked stems I and II of U5 snRNA, is found in the vicinity (Fig. 3b and Extended Data Fig. 6c). The region containing residues 420–542 of Prp8 is known to interact with the N-terminal half of Snu114 (ref. 27), and 4-thiouridine introduced at C79 of U5 snRNA crosslinks with both Prp8 and Snu114 (ref. 28), suggesting that the density adjacent to Snu114 and U5 snRNA is part of the Prp8 N terminus. At the tip of stem I the density assigned to the U5 loop I extends towards the RT thumb/X domain of Prp8 and makes close contact with a thioredoxin-like fold of Dib1 (Fig. 2c and Extended Data Fig. 5f).

---

**Figure 1** Overview of the U4/U6.U5 tri-snRNP structure with its protein and RNA components modelled into cryo-EM density. a, Front view, facing concave surface; b, back view; c, top view. d, 2D class average showing the different domains of tri-snRNP: head, body, arm and foot.

**Figure 2** Prp8 in tri-snRNP. a, Domain organization of Prp8. The structure of the N-terminal domain (residues 1–884) is unknown. RT, reverse transcriptase-like domain; X, thumb/X; L, linker; E, endonuclease-like domain; RH, RNaseH-like domain; JM, Jab1/MPN domain. b, The large domain of Prp8 is located at the centre of tri-snRNP. The Jab1/MPN domain is bound to Br2 (refs 31, 32). c, Loop 1 of U5 snRNA is inserted into the active site cavity and in contact with Dib1. d, Prp8 in the Prp8–Aar2 complex\(^10\) is shown with its large domain in the same orientation as in b. In tri-snRNP, the RNaseH domain is inverted while the Jab1/MPN domain in complex with Br2 is located at the opposite end of the large domain.
(ref. 29). This is in good agreement with the fact that a 16-kDa protein is crosslinked to 4-thiouridine incorporated at U97 in the U5 snRNA loop I (ref. 28). The binding of Dib1 is further stabilized by the N-terminal helices of Prp8 (Fig. 2c).

Brr2 forms a stable complex with the Jab1/MPN domain of Prp8 (refs 30–32), and its characteristic shape was recognized in the less-ordered head domain (Figs 1 and 5 and Extended Data Fig. 5b). Although this part of the map is lower in resolution, the individual domains of yeast Brr2 were fitted into the density together with the Jab1/MPN domain31. This revealed a widening of the gap between the two RecA domains of the N-terminal cassette (Fig. 5c). Co-axially stacked stems I and II of U4/U6 snRNAs and the 5′ stem–loop of

---

Figure 3 | The snRNA components of U4/U6.U5 tri-snRNP. a, Secondary structures of U4/U6 and U5 snRNAs. b, Double-stranded regions of U4/U6 and U5 snRNAs modelled into the cryo-EM map. SL, stem–loop.

Figure 4 | Structure of Snu114 in tri-snRNP. a, Location of Snu114 in the U4/U6.U5 tri-snRNP. b, Arrangement of domains (I–V) in Snu114 (see Extended Data Fig. 7). c, Domain arrangement in EF-G bound to the ribosome44. d, The interface between the N-terminal domain of Prp8 and Snu114. Some of the uninterpreted density at the interface may be attributed to the unmodelled switch I loop. e, The interaction of the switch region of EF-G with the sarcin–ricin loop44 for GTPase activation.
Density sandwiched between the Prp4 WD40 domain and U4/U6 stem II, two long helices lying along U4/U6 stem II and a number of connected helices nearby probably belong to Prp3 (Fig. 1 and Extended Data Fig. 5i). The U4 core domain is wedged between the tandem helicase cassettes of Br2 (Figs 1a, c and 5b and Extended Data Fig. 6a). The 3′ stem–loop of U4 snRNA contacts the helix–loop–helix domain of the N-terminal helicase cassette (Fig. 5b), which contains several lysine/arginine residues close to the RNA backbone of the 3′ stem–loop of U4 snRNA. On the basis of the previous labelling data19, U6 LSm proteins are fitted into the flexible arm region in the multi-body refined map (Fig. 1a and Extended Data Figs 3c and 5l).

A striking elongated curved α-solenoid density bridging the RNaseH-like domain of Prp8 and the WD40 domain of Prp4 is assigned to the tetratricopeptide (TPR) motifs of Prp6 (Fig. 1b, c and Extended Data Fig. 5d). Prp6 is required for the accumulation of tri-snRNP40 and is proposed to act as a bridge between U5 and U4/U6 snRNPs46,49. Prp6 contains up to 19 predicted TPR motifs, each comprising a helix–loop–helix motif49 and 37 connected idealized poly-alanine helices were built into the map. Nine canonical tandem TPR motifs at the C terminus of the protein form a highly curved α-helical solenoid-like structure, which contacts Snu13, U4 snRNA 5′ stem-loop and the Prp4 WD40 domain in tri-snRNP (Fig. 1b, c). This is consistent with the fact that antibodies against the C-terminal fragment of human Prp6 immunoprecipitate U5 snRNP but not tri-snRNP, as the C-terminal domain in our structure is in close contact with U4/U6 snRNP, which presumably occludes the epitope39.

**Central role of Prp8 in tri-snRNP assembly**

Our single-particle cryoEM reconstruction of yeast U4/U6.U5 tri-snRNP has revealed a nearly complete organization of its RNA and protein components, although some densities remain unassigned, and Snu66, Snu23, Prp38 and possibly sub-stoichiometric Spp381 are yet to be located (Extended Data Table 1 and Extended Data Figs 4f and 8d, e). Prp8 positioned at the centre of the assembly functions as a hub of protein–protein and protein–RNA interactions, holding the whole assembly together (Figs 1 and 2b). In yeast, a stable Prp8–Snu114–Aar2–U5 core domain complex is imported into the nucleus40, where Br2 replaces Aar2. The Jab1/MPN and RNaseH-like domains, held tightly onto the Prp8 large domain by Aar2 (Fig. 2d), are released in tri-snRNP wherein the Jab1/MPN domain forms a stable complex with Br2 as in the crystal structure31,32 (Figs 1b and 2b and Extended Data Fig. 5b). The tri-snRNP structure provides the first glimpse of the interaction between Snu114, the U5 core domain and the N-terminal domain of Prp8 which holds the co-axially stacked stems I and II, and variable stem–loop of U5 snRNA (Fig. 3). On the opposite side of Prp8, Snu13 and Prp31 firmly bound to 5′ stem–loop of U4 snRNA34, the U4 Sm protein assembly, the Br2–Jab1/MPN domain complex, the Prp3–Prp4 complex, the RNaseH domain of Prp8, and the U4/U6 snRNA duplex assemble together (Figs 1 and 2).

Prp8 has a surface on which are exposed the 5′ splice-site-binding ACAGAGA sequence of U6 snRNA, the U6 sequences which pair with U2 snRNA, and U5 snRNA loop I which interacts with exon 1 and exon 2. This surface is partly occluded by a highly conserved protein, Dib1 (Figs 2c and 6), suggesting its potential role in regulating the incorporation of RNA components into the active site cavity during spliceosome assembly and activation. When U4 and U6 snRNAs are unwound, releasing U4 snRNA together with Snu13, Prp31, Prp3 and Prp4 (ref. 3) from the spliceosome, Br2 stably bound to the Jab1/MPN domain of Prp8 is no longer held in place and could be repositioned during catalysis and spliceosome disassembly (Fig. 5 and Extended Data Fig. 8d).

**Br2 mode of action during activation**

The unwinding of the U4/U6 snRNA duplex is an essential step in spliceosomal activation and is catalysed by Br2 (refs 14, 15). Like

---

**Figure 5 | Br2 mode of unwinding.** a, Domain organization of Br2 N-terminal helicase cassette (NHC). WH, winged helix; HLH, helix–loop–helix; FN3, fibronectin3-like domains. The inactive C-terminal helicase cassette (CHC) has the same domain organization. b, U4/U6 di-snRNP and its interaction with Br2 in tri-snRNP. The domains of Br2 NHC are coloured as in a. The single-stranded RNA between U4/U6 stem I and U4 snRNA 3′ stem–loop is already loaded in the active site of Br2. When the Hel308 structure is overlaid onto the NHC of Br2, its 10-nucleotide DNA substrate coincides with the density in the Br2 active site, which extends to U4 snRNA 3′ stem–loop (red dotted line). The helix–loop–helix domain of Br2 interacts with U4 snRNA 3′ stem–loop (inset). c, Superposition of the RecA1 domain of Br2 in the crystal structure (Protein Data Bank (PDB) accession 4BGD, in grey) and in tri-snRNP (domains coloured as in a) shows the opening of the gap between the RecA1 and RecA2 domains (indicated by the red arrow) to accommodate the RNA substrate.
other Ski2-like helicases, Brr2 unwinds any RNA duplex with 3’ overhangs. The U4/U6 snRNA duplex has 3’ overhangs on both ends (Fig. 3a and Extended Data Fig. 6a) and it has been suggested that Brr2 binds to the single-stranded region of U4 snRNA and translocates along U4 snRNA. Our structure shows that Brr2 is pre-loaded into the active site cavity by superposition of the EBS1 stem of group II intron (PDB 3IGI) and stem I of the U5 snRNA.

The structural resemblance between the group II intron active site and the catalytic RNA core of the spliceosome endorsed the hypothesis that they evolved from a common ancestor. On the basis of the similarity of the domain architecture between the group II intron encoded protein (IEP) and Prp8, we proposed that Prp8 evolved from IEP and recruited more domains and interacting proteins to assemble spliceosomal snRNAs, which derived from fragmented group II intron. We placed the catalytic core of group II intron RNA in the tri-snRNP structure by superimposing its exon binding stem–loop with the extra density in the Brr2 active site and six additional nucleotides can be accommodated in the density extending further to the 5’ end of the U4 snRNA 3’ stem–loop (Fig. 5b).

**Role of Snu114 in spliceosome activation**

Snu114 shows substantial sequence similarity to EF2 (Fig. 4 and Extended Data Fig. 7), suggesting that it might induce conformational change in the spliceosome upon GTP binding or hydrolysis and regulate spliceosomal activation. EF-G, the bacterial counterpart of EF2, enters the ribosome in the GTP-bound form. Its GTPase is activated when switch regions I and II are remodelled upon interacting with the saring–ricin loop of 23S rRNA (Fig. 4e) and GTP hydrolysis leads to translocation. The activation process of the spliceosome has not been dissected in detail and it is not known at what stage GTP is hydrolysed or how Snu114 GTPase is activated. Snu114 and EF2 share highly similar switch I and II sequences, including the critical His residue, which in EF-G places a water molecule adjacent to the γ-phosphate. Unassigned density connecting the junction between stems I and II of U5 snRNA and the switch I and II loops coincides with the position of the saring–ricin loop (Fig. 4d). This is likely to be the N-terminal domain of Prp8, which may have a role in the activation of GTP hydrolysis.

Before the unwinding of the U4/U6 duplex, the 5’ splice site sequence pairs with the ACAGAGA sequence in U6 snRNA. The U4-cs1 cold-sensitive mutation, which extends U4/U6 stem I at the restrictive temperature and sequesters the ACAGAGA box from the 5’ splice site, stalls the spliceosome before unwinding (Extended Data Fig. 6a). A suppressor of U4-cs1 has a duplication of the ACAGAGA sequence in U6 snRNA. This shows that pairing of the ACAGAGA sequence with the 5’ splice site is a checkpoint to ensure proper assembly of complex B before the unwinding of the U4/U6 snRNA duplex. Notably, suppressors of U4-cs1 in Prp8 form three clusters on the surface of the large domain of Prp8 (Fig. 6a and Extended Data Table 2). In tri-snRNP, one of these clusters is located at the interface between the RT domain and domain III of Snu114, and another is at the interface with Prp31 and the junction between the RT and N-terminal domains of Prp8, showing that this checkpoint can be bypassed when these subunit interfaces are tampered with (Fig. 6a). This suggests that the interactions between these components undergo allosteric changes, which possibly couple the guanine-nucleotide binding state of Snu114 and the pairing between 5’ splice site and the ACAGAGA sequence to the activation of the U4/U6 duplex unwinding. Understanding the activation process will require extensive interplay between structural and biochemical work—the tri-snRNP structure provides an important structural framework for further investigation of this process.

The structural resemblance between the group II intron active site and the catalytic RNA core of the spliceosome endorsed the hypothesis that they evolved from a common ancestor. On the basis of the similarity of the domain architecture between the group II intron encoded protein (IEP) and Prp8, we proposed that Prp8 evolved from IEP and recruited more domains and interacting proteins to assemble spliceosomal snRNAs, which derived from fragmented group II intron. We placed the catalytic core of group II intron RNA in the tri-snRNP structure by superimposing its exon binding stem–loop with the extra density in the Brr2 active site and six additional nucleotides can be accommodated in the density extending further to the 5’ end of the U4 snRNA 3’ stem–loop (Fig. 5b).
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 24 February; accepted 6 May 2015. Published online 24 June 2015.

1. Will, C. L. & Lührmann, R. Spliceosome structure and function. Cold Spring Harb. Perspect. Biol. 3, a003707 (2011).
2. Chan, S. P. & Cheng, S. C. The Prp19-associated complex is required for specifying interactions of U5 and U6 with pre-mRNA during spliceosome activation. J. Biol. Chem. 280, 31190–31199 (2005).
3. Fabrizio, P. et al. The evolutionarily conserved core design of the catalytic active site of the spliceosome. Mol. Cell. 36, 693–698 (2009).
4. Fica, S. M. et al. RNA catalyses nuclear pre-mRNA splicing. Nature 503, 229–234 (2013).
5. Newman, A. J. & Norman, C. U5 snRNA interacts with exon sequences at 5’ and 3’ splice sites. Cell 68, 743–754 (1992).
6. Somellie, E. J. & Steitz, J. A. The U5 and U6 small nuclear RNAs as active site components of the spliceosome. Science 262, 1989–1996 (1993).
7. Stevens, S. W. Evolutionary conservation of the spliceosome. Nature 418, 647–652 (2002).
8. Hahn, D., Kudla, G., Tollery, D. & Beggs, J. D. Brr2p-mediated conformational rearrangements in the spliceosome during activation and substrate repositioning. Genes Dev. 26, 942–952 (2012).
9. Bourick, K., Nehring, S. & Hovcen, K. P. Structural basis for DNA duplex separation by a superfamily-2 helicase. Nature Struct. Mol. Biol. 14, 647–652 (2007).
10. Tourigny, D. S., Fernández, I. S., Kelley, A. C. & Ramakrishnan, V. Elongation factor G bound to the ribosome in an intermediate state of translocation. Science 340, 2349–2350 (2013).
11. Lin, J. Gagnon, M. G., Bulkley, D. & Steitz, T. A. Conformational changes of elongation factor G on the ribosome during tRNA translocation. Cell 160, 219–227 (2015).
12. Kuhn, A. N. & Brow, D. A. Suspectors of a cold-sensitive mutation in yeast U4 snRNA define five domains in the splicing factor Prp8 that influence splicing activation. Genetics 155, 1667–1682 (2000).
13. Li, Z. & Brow, D. A. A spontaneous duplication in U6 spliced RNA uncouples the early and late functions of the ACA box element in vivo. RNA 5, 879–894 (1999).
14. Toor, N., Keating, K. S., Taylor, S. D. & Pyle, A. M. Crystal structure of a self-spliced group II intron. Science 320, 77–82 (2008).
15. Fica, S. M., Mefford, M. A., Piccirilli, J. J. & Staley, J. E. P. Evidence for a group II intron-like catalytic triplex in the spliceosome. Nature Struct. Mol. Biol. 21, 464–471 (2014).
16. Sharp, P. A. Five easy pieces. Science 254, 663 (1991).

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

Acknowledgements We thank S. Chen, G. McMullan, J. Grimmert and T. Darling for smooth running of the EM and computing facilities; P. da Fonseca, N. Unwin, S. Sanchez Fernández, A. Amunts, P. Embacher, M. P. Fernandez, W. P. G. for their continuing support and encouragements. T.H.D.N. was supported in part by a Hechel Smith Research Studentship. X.-c.B. was supported by a European Union Marie Curie Fellowship. The project was supported by the Medical Research Council (MC_U010614030 to K.N. and MC_UP_A025_1013 to S.H.W.S.).

Author Contributions T.H.D.N. developed the purification procedure for yeast tri-snRNP, prepared EM grids, collected all EM images, processed data, calculated the map and fitted model into the map. W.P.G. built most of the unknown components and made essential contributions to sequence analysis, homology modelling and model fitting. X.-c.B. helped T.H.D.N. with map computation and map calculation. C.O.S. guided T.H.D.N. with EM sample preparation and data collection. A.J.N. produced the Br2 TAPS-tagged strain and contributed to the project through his knowledge of yeast spliceosome. T.H.D.N. and W.P.G prepared all illustrations. T.H.D.N. prepared the video. S.H.W.S. carried out multi-body refinement and oversaw the EM analysis. K.N. initiated and orchestrated the project. T.H.D.N., A.N., K.N. and I.S. wrote the results and wrote the paper with crucial contribution from all other authors.

Author Information 3D cryo-EM maps of yeast U4/U6/U5 tri-snRNPs have been deposited with the Electron Microscopy Database under accession numbers EMD-2966. 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 T.H.D.N. (knguyen@mrc-lmb.cam.ac.uk) or K.N. (km@mrc-lmb.cam.ac.uk).

Abstract We present a cryo-EM reconstruction of the yeast spliceosome. Our high-resolution map shows the organization of the spliceosome in its native state, with a combination of previously unknown components and known structural elements. We identify five major domains that influence spliceosome activation. These domains form a superfamily-2-type helicase that is essential for spliceosome rearrangements during activation and substrate repositioning. We establish the role of an evolutionarily conserved protein in GTP cleavage and RNA unwinding in the spliceosome. We also identify a previously unknown type of SnRNP and propose a model for spliceosome formation. We conclude that the spliceosome is a dynamic assembly of preformed splicing factors that are recruited in an ATP-dependent manner to the pre-mRNA in the cytoplasm.
**METHODS**

**Statistics.** No statistical methods were used to predetermine sample size.

**Brr2-TAPS tagging for yeast U4/U6, U5 snRNP purification.** Primers specific for 55 nucleotides of the C terminus and 3’ UTR of Brr2 were used to PCR-amplify the TAPS-tag cassette together with the KanMX6 gene from pFA6a-TAPS-KanMX5, a modified version of pFA6a-TAP-kanMX6 in which the calmodulin-binding peptide tag is replaced by two tandem copies of the StreptII tag.

The PCR product was used to transform yeast strain BCY123 [MATa pep4::HIS3 pbr1::LEU2 bar1::HIS6 yks2::GAL11-10-GAL4 can1 ade2 trpl ura3 his3 leu2-3, 12] by homologous recombination, selecting for G418-resistance. C-terminal TAPS-tagging of Brr2 was confirmed by PCR analysis of genomic DNA and DNA sequencing.

**Sample preparation.** The Brr2-TAPS-tagged yeast cells (72 l) were grown in YEPD medium to OD600 of 3.5, harvested and resuspended in lysis buffer (100 mM HEPES KOH pH 8.0, 200 mM KC1, 2 mM Mg(OAc)2, and 10% w/v glycerol).

The cells were frozen and lysed by a Freezer Mill 6870 (SPEX CertiPrep). The crude lysate was centrifuged at 45,000 r.p.m. for 1 h. The resulting supernatants were incubated with IgG sepharose overnight at 4°C. The resin was washed with TAPS wash buffer (20 mM HEPES KOH pH 7.9, 150 mM KC1, 1 mM Mg(OAc)2, and 0.1% NP40) and incubated with TAPS wash buffer in the presence of TEV protease at 4°C overnight. The flow-through was collected and incubated with Streptactin resin (GE Healthcare) for 3 h. The resin was washed with TAPS wash buffer and particles were eluted with Strept elution buffer (20 mM HEPES KOH pH 7.9, 150 mM KC1, 1 mM Mg(OAc)2, 0.1% NP40, 5 mM desthiobiotin).

The eluate was subsequently applied to a 10–30% v/v glycerol gradient centrifuged at 210,000g at 4°C in a SW60 rotor. The fraction of the gradient were analysed by SDS–PAGE for protein composition. Glycerol was removed from the peak fractions containing tri-snRNP by dialysis against B150 buffer (20 mM HEPES KOH pH 7.9, 150 mM KC1, 1 mM Mg(OAc)2) before EM sample preparation (Extended Data Fig. 1a and 1b).

**Electron microscopy.** For cryo-EM analysis, 3.5 μl of the tri-snRNP sample was applied to Quantifoil R2/2 or R1.2/1.3 grids which were previously coated with 6-nm-thick layer of homemade carbon film and glow-discharged (Extended Data Fig. 1c). The grids were blotted for 2 s at 4°C and plunged into liquid ethane using an FEI Vitrobot MKIII. The grids were loaded onto a Tecnai F30 Polara transmission electron microscope operated at 300 kV. Images were collected manually in low-dose mode at a calibrated magnification of 79,096×. The micrographs were recorded on either a Falcon II or a (ultra back-thinned) Falcon III detector at the same calibrated pixel size of 1.77 Å in movie mode at 17 frames s−1. A data set was extracted from the micrographs using a soft mask (with a 5-pixel soft edge) for not only the entire model of the tri-snRNP sample was collected (Extended Data Fig. 4a) at 100,000× magnification in a 3°–4° tilt range.

All refinements used gold standard Fourier shell correlation (FSC) calculations39 and reported resolutions are based on the FSC = 0.143 criterion. The FSC curves are calculated using a soft spherical mask (Extended Data Fig. 4d). Prior to visualization, all maps were corrected for the modulation transfer function of the detector and sharpened by applying a negative B-factor.

**Local resolution analyses.** Local resolution analyses were performed by Resmap40 and compared with those calculated by us for each protein/RNA component of the map (Extended Data Fig. 4a–c). For the latter calculations, FSC curves are calculated using a soft spherical mask (with a 30-pixel fall-off) around each protein/RNA component of interest. Convolution effects of the masks on the FSC curves were corrected using high-resolution noise substitution32. Resolution was estimated at FSC = 0.143. These calculations were performed for each of the following components: Prp8 large domain, Prp8 RNaseH domain, Prp8 Jab1/MPN domain, Prp8 N-terminal domain, Br2, U4 Sm with U4 snRNA 3’ stem–loop, U5 Sm with Sm site, Snu114, Dib1, Prp6, Prp3, Snu13, Prp31, LSm, U5 snRNA and U4/U6 snRNA. Extended Data Fig. 4d shows some representative curves from these calculations.

FSC curves of model versus map were calculated using the Xnipp package36 and the reported resolutions were based on the FSC = 0.25 criterion. FSC curves of model versus map were calculated for not only the entire model of all components but also different parts of the maps. The model of each modelled component was extracted from the tri-snRNP map using a soft mask (with a 5-pixel soft edge) surrounding the component. A model of each map was created by the program pdb2mrc within the EMAN package57. Some proteins/domains that are close to each other were grouped together for these calculations, including Prp8 N-terminal domain/Dib1, Br1/Prp6/Prp1/MPN domain and Prp3/Prp4. The model was visualized using Chimera53 and the 10.5 Å resolution map was used for further refinement. Further rigid-body fitting was performed in Coot40. The model for the U4 snRNA component of Prp3 is available at the Yeast Genome Center (http://www.yeastrc.org), which contains its structure predictions41. The model with the highest Mammoth Confidence Metric (MCM) score was selected for fitting. For Prp4, the protein sequence was input into Robetta Beta Full-chain Protein Structure (http://robetta.bakerlab.org), which...
yielded a model for the C-terminal part of Prp4 based on the structure of the WDR5 protein62 (PDB 3MXK). Double-stranded RNA helices and idealized polyalanine helices were built into the masked map in Coot when possible. U4 snRNA 3’ stem–loop was modelled based on the structure of the human Prp31–15.5K–U4 snRNA complex53 (PDB 2OZB) using ModeRNA modelling tool54. Yeast Snu13 structure44 (PDB 2ALE) was fitted into the map. U4 snRNA 3’ stem–loop partial model was adapted from the structure of the human U4 snRNP core domain55 (PDB 4WZJ). The short and long forms of U5 snRNAs are present in our sample (Extended Data Fig. 1b) but no density for 3’ stem–loop was observed, presumably because 3’ stem–loop attached to the Sm site with a long single-stranded stretch is disordered or the particle population with the long U5 snRNA is classified out during classification. U5 snRNP Sm core with only the Sm site was also adapted from the human U4 snRNP core domain. The LSm proteins56 (PDB 4M77) were placed in the low-resolution arm region of the map with the flat surface of the LSm complex facing the entrance side of U6 snRNA. The register of the LSm proteins cannot be accurately determined.

A map with the flat surface of the LSm complex facing the entrance side of U6 snRNA. The register of the LSm proteins cannot be accurately determined. Human Dib1 structure29 (PDB 1QGV) was used for fitting. Extended Data Table 1 and Extended Data Fig. 4f summarize all the details of tri-snRNP components and modelling in our study. The active site cavity of Prp8 was described previously10 and defined by crosslinks with crucial elements of U5 snRNA, U6 snRNA and pre-mRNA 9 and suppressors of defective splice site mutations68–74. U4–csl mutants have been described68–75. Extended Data Table 2 summarizes all the U4–csl mutants and their locations in tri-snRNP.

Map and model visualization. Maps were visualized in Chimera58. Map segmentation was performed in Chimera using each of the fitted models and the ’zone-masking’ function (Fig. 1 and Extended Data Figs 5 and 6b, d). The LSm protein density was obtained from multi-body refinement and low-pass filtered to 20 Å. For all the remaining components, the sharpened tri-snRNP map (B = 214 Å) low-pass filtered to 5.9 Å was used. Figures were generated using either Chimera or PyMOL (http://www.pymol.org) and the video was made in Chimera.

ATP assays. Purified tri-snRNP from glycerol gradient (~25 nmL) was incubated at 30°C for 30 min in the presence of either no nucleotide or with each of the following nucleotide combinations: ATP, ATP/GTP, ATP/GDP, ADP and AMP/NNP (1 mM each). The samples (10 μL) were loaded onto a native agarose gel (0.5% in TB buffer supplemented with 1 mM MgCl2) and run at 75 V at 4°C for 2.5 h. The gel was stained with ethidium bromide for 1 h before being imaged by a Syngene UV imager (Extended Data Fig. 8a). For negative staining, the sample was also treated similarly and stained with 2% uranyl acetate before imaging on a Tecnai T12 transmission electron microscope operated at 120 kV (Extended Data Fig. 8b, c).

Schreieck, A. et al. RNA polymerase II termination involves C-terminal domain tyrosine dephosphorylation by CPF subunit Glc7. Nature Struct. Mol. Biol. 21, 175–179 (2014).

Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. J. Struct. Biol. 142, 334–347 (2003).

Eilmund, E., Eilmund, D. & Bengio, S. PRIME: probabilistic initial 3D model generation for single-particle cryo-electron microscopy. Structure 21, 1299–1306 (2013).

Kuhn, A. N., Li, Z. & Brow, D. A. Splicing factor Prp8 governs U4/U6 RNA unwinding during activation of the spliceosome. Mol. Cell 3, 65–75 (1999).

Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. Nature Methods 11, 63–65 (2014).

Chen, S. et al. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron microscopy. Ultramicroscopy 135, 24–35 (2013).

Scheres, S. H. W., Nuñez-Ramirez, R., Sorzano, C. O. S., Carazo, J. M. & Marabini, R. Image processing for electron microscopy single-particle analysis using Xmipp. Nature Protocols 3, 977–990 (2008).

Tang, G. et al. EMAN2: an extensive image processing suite for electron microscopy. J. Struct. Biol. 157, 38–46 (2007).

Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. J. Struct. Biol. 157, 281–287 (2007).

Ensmley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. Acta Crystallogr. D 66, 486–501 (2010).

Zhang, Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9, 40 (2008).

Malmström, L. et al. Superfamily assignments for the yeast proteome through integration of structure prediction with the gene ontology. PLoS Biol. 5, e76 (2007).

Wu, X. H., Chen, R. C., Gao, Y. & Wu, Y. D. The effect of Asp-His-Ser/Thr-Trp tetrad on the thermostability of WD40-repeat proteins. Biochemistry 45, 10237–10245 (2010).

Rother, M. et al. ModeRNA server: an online tool for modeling RNA 3D structures. Bioinformatics 27, 2441–2442 (2011).

Dobbs, H. C. et al. Analysis of pre-mRNA and pre-rRNA processing factor Snu13p structure and mutants. Biochem. Biophys. Res. Commun. 360, 557–562 (2007).

Leung, A. K., Nagai, K. & Li, J. Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis. Nature 473, 536–539 (2011).

Chanfreau, Q., Elela, S. A., Ares, M. Jr & Guthrie, C. Alternative 3’-end processing of U5 snRNA by RNase III. Genetics 166, 2741–2751 (1997).

Zhou, L. et al. Crystal structures of the LSm complex bound to the 3’-end sequence of U6 small nuclear RNA. Nature 506, 116–120 (2014).

Quercy, C. C. & Konarska, M. M. Suppression of multiple substrate mutations by spliceosomal prp8 alleles suggests functional correlations with ribosomal ambiguity mutants. Mol. Cell 14, 343–354 (2004).

Urmen, J. G. & Guthrie, C. Mutagenesis of the yeast gene PRP8 reveals domains governing the specificity and fidelity of 3’ splice site selection. Genetics 143, 723–739 (1996).

Liu, L., Quercy, C. C. & Konarska, M. M. Opposing classes of prp8 alleles modulate the transition between the catalytic steps of pre-mRNA splicing. Nature Struct. Mol. Biol. 14, 519–526 (2007).

Dagher, S. F. & Fu, X. D. Evidence for a role of Sky1p-mediated phosphorylation in 3’ splice site recognition involving both Prp8 and Prp17/Slu4. RNA 7, 1284–1297 (2001).

Ben-Yehuda, S. et al. Extensive genetic interactions between PRP8 and PRP17/CDC40; two yeast genes involved in pre-mRNA splicing and cell-cycle progression. Genetics 154, 61–71 (2000).

Collins, C. A. & Guthrie, C. Allele-specific genetic interactions between Prp8 and RNA active site residues suggest a function for Prp8 at the catalytic core of the spliceosome. Genetics 13, 1970–1982 (1999).

Sratecka, M., Reyes, J. L. & Konarska, M. M. Functional interactions of Prp8 with both splice sites at the spliceosomal catalytic center. Genetics 13, 1963–1993 (1999).
Extended Data Figure 1 | U4/U6.U5 tri-snRNP sample used for this study. a, Coomassie-blue-stained SDS–PAGE gel showing protein composition of the purified tri-snRNP. U5-, U4/U6- and tri-snRNP-specific proteins are labelled in blue, red and teal, respectively. Sm proteins present in both U5 and U4/U6 are in black. b, Toluidine-blue-stained denaturing acrylamide (9%) gel showing RNA compositions. c, Electron cryo-micrograph of tri-snRNP where the carbon-coated grid was discharged in N-amylamine. d, e, Reference-free two-dimensional class averages of a data set collected on a grid discharged in air and N-amylamine, respectively.
A total of 367,327 particles were subjected to reference-free 2D classification. A subset of 347,241 particles from good 2D classes was selected for 3D classification using an initial model obtained from SIMPLE-PRIME, which was low-pass filtered to 60 Å. The data were divided into four 3D classes, two of which (a total of 179,079 particles) showed better features and were combined for refinement. This resulted in a 7.6 Å reconstruction. To further improve the reconstruction, these particles were subjected to beam-induced motion correction (particle polishing). Refinement of these polished particles with a soft mask around the rigid part of the map (as indicated by the red envelope) yielded a 5.9 Å reconstruction while refinement with a mask around the whole map yielded a 6.4 Å reconstruction. The polished particles were also subject to further 3D classification with a finer angular sampling of 1.8°. The most populated class (47,674 particles), which also has the best rotational accuracy, was refined with a soft mask around the whole density. This resulted in a 7.0 Å reconstruction. In this study, the 5.9 Å reconstruction was used for subsequent biological interpretation. All steps were performed in RELION unless otherwise stated.
Extended Data Figure 3 | CryoEM maps and tilt-pair validation. a, CryoEM density of the whole tri-snRNP at 5.9 Å resolution by ‘gold standard’ Fourier shell correlation (FSC) of 0.143 criterion at two different contour levels. The high contour map (gold) shows well-resolved densities for protein and RNA helices and flat densities for β-sheets. The low contour map (silver) shows densities for the more flexible head and arm. The map was sharpened by a B-factor of −214 Å² and low-pass filtered to 5.9 Å as determined by RELION. b, The unsharpened full map of tri-snRNP. c, The map resulting from multi-body refinement, in which tri-snRNP is divided into four parts: the head, body, arm and foot. This resulted in better density for the arm domain (indicated by red circles), which is at 20 Å resolution. d, Tilt-pair validation plot for tri-snRNP. This was obtained from 1,196 particles from 32 micrograph pairs, imaged at 0° and 10° tilt angles. The position of each dot represents the direction and the amount of tilting for a particle pair in polar coordinates. Blue dots correspond to in-plane tilt transformations; red and purple dots correspond to out-of-plane tilt transformations. Blue dots cluster in the same region of the plot at a tilt angle of approximately 10° as indicated by the red circle.
Overall map resolution and some examples of local map resolution

Examples of model vs map FSC curves
Extended Data Figure 4 | Resolution estimation of tri-snRNP map. a, Local resolution of the tri-snRNP map estimated by ResMap using the colour scheme shown in panel c. b, Local resolution of the tri-snRNP map calculated by ‘gold-standard’ FSC. For each component of the map that we modelled protein/RNA components, a soft mask (with a 30-pixel soft edge) surrounding the region of interest was prepared and used for FSC calculations. Convolution effects of the masks on the FSC curves were corrected using high-resolution noise substitution55. Resolution was estimated at FSC = 0.143. Local resolution for the unmodelled region of the map (in red) was not estimated. c, Local resolution of model versus map. The map of each modelled component was extracted from the map using a soft mask (with a 5-pixel soft edge) surrounding the component. The model was converted into density by EMAN57. FSC of model versus map was calculated using Xmipp56. The map is coloured according to resolution estimates based on a FSC threshold of 0.25. The lower resolution estimates from the FSC of model versus map compared to the estimates from ResMap and the gold-standard FSCs are explained by the nature of our models. Because of the limited resolution of our map, we did not perform full atomic refinement, but placed known crystal structures and homology models as rigid bodies in the map. d, Gold-standard FSC curves for the whole tri-snRNP map and some of its components calculated as described in b. e, FSC curves of model versus map for the whole model and some of the components. f, The full tri-snRNP map in which portions of the structure produced from crystal structures, homology modelling and de novo building or unmodelled are coloured as indicated.
Extended Data Figure 5  | Fitting of protein components into tri-snRNP map.  

a, Prp8(885–2,413) crystal structure (PDB 4I43, green) and additional helices built de novo assigned to the N terminus of Prp8 (blue). b, Brr2–Jab1/MPN complex (PDB 4BGD). c, Snu114 homology model based on EF2 (ref. 26). d, The Prp6 TPR motifs built into the tri-snRNP map. e, U5 Sm proteins (grey) with Sm site (blue) based on the human U4 Sm structure (PDB 4WZJ). f, Dib1 (ref. 29) (PDB 1QGV). g, (i) Prp31. (ii), Comparison between the crystal structure of human Prp31(78–333) (ref. 33) (PDB 2OZB, grey) and that in tri-snRNP (yellow and blue). The coiled-coil domain (yellow) rotates by 60° in tri-snRNP with respect to the Nop domain (grey). Additional helices (blue) that extend from the N and C termini were built. h, U4 Sm proteins with part of U4 snRNA (blue) based on the human U4 Sm structure. i, Prp3 model. The ferredoxin-like domain was obtained from homology modelling while the extra helices were built de novo. j, Prp4 WD40 homology model with the extra helices built de novo. k, Snu13 (ref. 64) (PDB 2ALE). l, U6 LSm proteins (PDB 4M77).
Extended Data Figure 6 | Fitting of the RNA components in tri-snRNP map. a, c, The sequences and predicted secondary structures of U4/U6 snRNA and the long version of U5 snRNA, respectively. b, d, The maps of the fitted parts of U4/U6 snRNA and U5 snRNA, respectively. Unmodelled density assigned to U5 snRNA is also shown in d.
Extended Data Figure 7 | Sequence alignment of yeast and human Snu114 with yeast and human elongation factor 2 (EF-2). The secondary structures of our homology model for yeast Snu114 and the yeast EF-2 (ref. 26) (PDB 1N0V) are shown on the top and bottom of the alignment, respectively. Important sequence elements are also shown. The greyscale shading indicates the level of sequence conservation. A higher level of conservation is shown in a darker shade.
Extended Data Figure 8 | The effect of ATP on Brr2-TAPS purified tri-snRNP. a, Ethidium-bromide-stained native agarose gel (0.5%) showing the effects of ATP addition to Brr2-TAPS purified tri-snRNP used in this study. Upon ATP addition either without or with GTP/GDP, tri-snRNP fell apart (lanes 1–4). Under the same conditions, the addition of ADP or the non-hydrolysable ATP-analogue, AMPPNP, had no effects on the complex (lanes 5, 6). b, c. The effect of ATP addition observed by negative stain microscopy. When ATP was not present, tri-snRNP particles could be observed. When ATP was added to the sample before grid preparations, tri-snRNP particles fell apart as observed by many small components on the micrograph rather than tri-snRNP particles. d, Tri-snRNP model where U4/U6 snRNP proteins are not shown. In tri-snRNP, Brr2–Prp8Jab complex is loosely associated to the remaining U5 snRNP components including Prp8Larg, Prp8RCHIP, Prp8Nterm, Snu114, Dib1, U5 Sm proteins and U5 snRNA. After U4/U6 snRNA unwinding by Brr2, Brr2–Prp8Jab could be repositioned within the spliceosome. e, A schematic showing the arrangement of tri-snRNP protein and RNA components.
### Extended Data Table 1 | Components and modelling of yeast U4/U6.U5 tri-snRNP

| Protein | Component | M.W. | Domain | Residues | PDB Code |
|---------|-----------|------|--------|----------|----------|
| **Prp8** | N-terminal domain | 1-884 | α-helices modelled |
| | RT-like | 885-1251 |
| | Thumb/X | 1257-1375 |
| | Linker | 1376-1649 |
| | Endonuclease | 1653-1824 |
| | RNaseH-like | 1839-1029 |
| | Jabb/MPN | 2150-2396 |
| | 48GD |
| **Br2** | N-terminal domain | 1-639 | not modelled |
| | N-terminal helicase cassette | 442-478 |
| | C-terminal helicase cassette | 478-1309 |
| | 48GD |
| **Snu14** | N-terminal domain | 1-114 | not modelled |
| | G domain | 120-443 |
| | domain II | 446-580 |
| | domain III | 603-671 |
| | domain IV | 675-853 |
| | domain V | 856-990 |
| | homology model (1N0V) |
| **Db1** | threonine like | not modelled |
| **SmB** | 196 | 22,403 | Sm fold |
| **SmD3** | 110 | 11,229 | Sm fold |
| **SmD1** | 146 | 16,288 | Sm fold |
| **SmD2** | 110 | 12,856 | Sm fold |
| **SmE** | 94 | 10,373 | Sm fold |
| **SmF** | 96 | 9,659 | Sm fold |
| **SmG** | 77 | 8,479 | Sm fold |
| **Us snRNA-L** | Loop 1 | 92-102 | not modelled |
| | Stem 1 | 84-91;103-110 | A-form double helix |
| | IL 1 | 75-83;111-113 | not modelled |
| | VSL | 41-74 | A-form double helix |
| | Stem 2 | 28-40;114-125 | A-form double helix |
| | IL 2 | 13-27;126-135 | not modelled |
| | Stem 3 | 4-12;136-144 | A-form double helix |
| | 3’SL | 185-212 | not modelled |
| **Snu13** | 126 | 13,570 | 2ALE |
| **Prp31** | N-terminal domain | α-helices modelled |
| | coiled-coil domain | 40ZB (human model) |
| | Nap domain | α-helices modelled |
| **prp3** | N-terminal domain | α-helices modelled |
| | Ferredoxin-like domain | model obtained from yeast genome center |
| **prp4** | N-terminal domain | α-helices modelled |
| | β-propeller domain | 166-465 |
| | 3MXX from Roberts prediction |
| **SmB** | 196 | 22,403 | Sm fold |
| **SmD3** | 110 | 11,229 | Sm fold |
| **SmD1** | 146 | 16,288 | Sm fold |
| **SmD2** | 110 | 12,856 | Sm fold |
| **SmE** | 94 | 10,373 | Sm fold |
| **SmF** | 96 | 9,659 | Sm fold |
| **SmG** | 77 | 8,479 | Sm fold |
| **Us snRNA** | Stem I | 57-64 | A-form double helix |
| | Stem II | 1-17 | A-form double helix |
| | 5’SL | 20-53 | homology model (20ZB) |
| | central domain | 65-80 | partial homology model (2P6R) |
| | 3’SL | 91-142 | partial model (2W2Z: human model) |
| **U6 snRNA** | 5’SL | 1-25 | not modelled |
| | Stem I | 59-62 | A-form double helix |
| | Stem II | 64-60 | A-form double helix |
| **Prp6** | 899 | 104,234 | TPR domain | 1-191 |
| | 192-899 | not modelled |
| **Snu66** | 587 | 66,426 | not modelled |
| **Prp38** | 242 | 27,957 | not modelled |
| **Snu23** | 194 | 22,582 | C2H2 zinc finger-like |
| | not modelled |
| **Snp381** | 291 | 32,764 | not modelled |

See Methods for details.
## Extended Data Table 2 | U4-cs1 suppressors

| region       | mutations                                                                 | domains        | locations                        | contact                   |
|--------------|----------------------------------------------------------------------------|----------------|----------------------------------|---------------------------|
| region a     | R236G<br> L261P<br> L280P                                                  | N-terminal domain | unknown                          |                           |
| region b     | KG11R<br> E624G<br> N643S<br> V644A<br> D651G or N<br> H659P<br> K684E | N-terminal domain | unknown                          |                           |
| region c     | E788G or V<br> N796S<br> W856R<br> E860K<br> Q861R                        | N-terminal domain | unknown                          |                           |
| cluster 1    | DT094A or N or V<br> M1095T<br> V1098D<br> N1099K<br> I1104M<br> R1105L | RT domain       | loop in 4 stranded<br> β-sheet   | Interface with Prp31     |
| cluster 2    | P1191L or S or T<br> D1192Y<br> N1194D                                    | within loop following α12 | within loop following α12 | interface with Snu14 domain III |
| cluster 3    | L1624M<br> L1634F<br> L1641F<br> T1685I<br> P1688L or R<br> A1754V<br> N1809D | endonuclease     | top surface<br> top surface<br> top surface<br> top surface<br> side surface<br> side surface |                           |
| region f     | F1851L<br> V1860D or N<br> T1861P<br> V1862A or D or Y<br> I1875T       | RNaseH          | on inner surface<br> β-finger    |                           |

All suppressor mutants are described in Kuhn and Brow\(^{37}\) and Kuhn et al.\(^{75}\).