Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution

Thi Hoang Duong Nguyen1*, Wojciech P. Galej1*, Xiao-chen Bai1, Chris Oubridge1, Andrew J. Newman1, Sjors H. W. Scheres1 & Kiyoshi Nagai1

U4/U6. U5 tri-snRNP represents a substantial part of the spliceosome before activation. A cryo-electron microscopy structure of *Saccharomyces cerevisiae* U4/U6.U5 tri-snRNP at 3.7 Å resolution led to an essentially complete atomic model comprising 30 proteins plus U4/U6 and U5 small nuclear RNAs (snRNAs). The structure reveals striking interactions of the protein and RNA components, including extended polypeptides penetrating into subunit interfaces. The invariant ACAGAGA sequence of U6 snRNA, which base-pairs with the 5′-splice site during catalytic activation, forms a hairpin stabilized by Dib1 and Prp8 while the adjacent nucleotides interact with the exon binding loop of U5 snRNA. Snu14 harbours GTP, but its putative catalytic histidine is held away from the γ-phosphate by hydrogen bonding to a tyrosine in the amino-terminal domain of Prp8. Mutation of this histidine to alanine has no detectable effect on yeast growth. The structure provides important new insights into the spliceosome activation process leading to the formation of the catalytic centre.

Pre-messenger RNA splicing is catalysed by an intricate molecular machine called the spliceosome and proceeds by a two-step trans-esterification mechanism, analogous to group II intron self-splicing. The spliceosome is assembled on pre-mRNA by the ordered addition of small nuclear ribonucleoprotein particles (snRNPs) and numerous proteins, including the nineteen complex (NTC) and the nineteen related (NTR) complex. Initially U1 and U2 snRNPs recognize the pre-mRNA 5′-splice site (5′SS) and branch point (BP), respectively. Recruitment of U4/U6.U5 tri-snRNP produces the fully assembled but catalytically inactive complex B1. U1 snRNP is displaced from the 5′SS by Prp8 (ref. 5), and the 5′SS pairs with the NTCAGAGA sequence in U6 snRNA, and Brr2 helicase unwinds the extensively base-paired U4/U6 snRNAs to release U4 snRNA with its associated proteins. This allows U6 snRNA to base-pair with U2 snRNA generating the group II intron-like catalytic RNA core remaining after spliced mRNA is released.

Here we present an essentially complete atomic model of *S. cerevisiae* U4/U6.U5 tri-snRNP based on a cryo-EM density map at 3.7 Å overall resolution, revealing the architectural and mechanistic principles of spliceosome activation.

**Overall structure**

We collected a new data set on a Titan Krios microscope using the Gatan K2 Summit direct electron detector (Methods). The overall resolution of the tri-snRNP map was improved from 5.9 Å to 3.7 Å (Extended Data Fig. 1). Using a modified masked refinement with signal subtraction, we obtained more homogeneous 3.6, 3.7 and 4.2 Å reconstructions for the body, foot and head domains, respectively, and improved the resolution of the arm domain from 10 Å to 6–7.5 Å (Extended Data Fig. 2). The new maps enabled us to build a near-complete atomic model of the yeast tri-snRNP containing 30 proteins, U4/U6 and U5 snRNAs (Fig. 1 and Supplementary Information), revealing an amazing web of interactions between components of the complex (Extended Data Fig. 3).

**Prp8**

A complete atomic model of Prp8 is now built, except for the unstructured N terminus and inter-domain linkers. The α-helix (αRT1) at the N terminus of the reverse transcriptase (RT) domain in the crystal structure extends further and forms a helix bundle (HB) with three additional long helices appended to the RT domain (Fig. 2a, b). Residues 108–733 form a predominantly α-helical N-terminal domain.

Stems I and II of U5 snRNA are coaxially stacked and an extra variable stem protrudes from the three-way junction (Extended Data Fig. 4). A long, slightly bent C-terminal α-helix (residues 703–735) of the N-terminal domain fits into the minor groove of the co-axially stacked stems I and II, which is tightly harnessed in the major groove by a polypeptide loop (residues 535–543) protruding from the N-terminal domain (Fig. 2c). The conserved loop 1 of U5 snRNA, which aligns the exons during the second trans-esterification reaction, points...
towards the most positively charged and conserved surface of Prp8 in the thumb/linker domain, part of the active site cavity. The BP+2 nucleotide cross-links in active spliceosomes between Prp8 residues 1585–1598, on the cavity surface (C. M. Norman and A.J.N., unpublished observations). This region is disordered in the Prp8–Aar2 complex, whereas in U4/U6.U5 tri-snRNP it forms a helix-turn-helix (the α-finger) and contacts U54–U55 of U4 snRNA near the three-way junction (Fig. 2b).

The 5′-stem-loop of U6 snRNA interacts with the N-terminal domain of Prp8 and the adjacent single-stranded region pairs with the exon binding U5 snRNA loop 1 (Fig. 2d). The small highly conserved protein Dib1 (ref. 25) binds to the helix bundle and α-finger of Prp8, and a long polypeptide of Prp31. U6 snRNA forms a short stem-loop, involving part of the ACAGAGA sequence, which is sandwiched between Dib1 and the Prp8 large domain (residues 1648–1653) (Fig. 2d; Extended Data Fig. 4a).

**Snu114**

We built a near complete atomic model of Snu114 comprising five domains (D1–D5) similar to EF-G/EF-2 (refs 26, 27). The relative arrangement of D1–D3 closely resembles that of EF-G/EF-2, whereas D4 and D5 pack more compactly (Fig. 3a). The guanine nucleotide density is consistent with GTP bound via canonical interactions with surrounding residues (Fig. 3b; Extended Data Figs 3a and 5a–e). In most GTPases the glutamine residue in the switch 2 loop places a water molecule at the γ-phosphate of GTP and hydrolyses the phosphate ester. As in EF-Tu, EF-G and their eukaryotic counterparts, the catalytic glutamine residue is replaced by histidine in Snu114 (ref. 26) (Extended Data Fig. 5e). In U4/U6.U5 tri-snRNP, His218 is hydrogen-bonded to Tyr403 of Prp8, preventing the His218 side chain from rotating towards the γ-phosphate of GTP and hence keeping the GTPase inactive (Fig. 3c). In EF-G and EF-Tu, GTP is hydrolysed when this histidine is repositioned by a hydrogen-bond with a phosphate in the sarcin-ricin loop of the ribosome (Fig. 3d). The extensive interactions between Snu114 and the N-terminal domain of Prp8 are conserved between U4/U6.U5 tri-snRNP and the S. pombe ILS (Extended Data Fig. 5f). The hydrogen bond between His218 of Snu114 and Tyr403 of Prp8 is maintained by the equivalent residues in ILS (Extended Data Fig. 5d). The GTP binding site of Snu114 is at the interface with the N-terminal domain of Prp8, leaving insufficient room for U5 snRNA or any proteins to access the GTPase active site and act like the sarcin-ricin loop or as a GTPase activating protein (GAP). Since the structure suggests no obvious mechanism for Snu114 GTPase activation.
activation we investigated the function of Snu114 by mutagenesis. With the His218Arg mutation, yeast shows only a mild temperature-sensitive phenotype, confirming earlier results (Extended Data Fig. 5g), whereas the equivalent mutation in EF-Tu reduces cognate RNA-induced GTPase activity 10^5-fold. Surprisingly, yeast containing the His218Ala mutant of Snu114 shows no apparent phenotype (Extended Data Fig. 5g), while the equivalent mutation in EF-Tu reduces the rate of GDP hydrolysis more than 10^6-fold. Furthermore, mutations of Tyr403 (Tyr403Phe and Tyr403Ala) in Prp8, which hydrogen-bonds with His218 in Snu114, have no apparent phenotype (Extended Data Fig. 5h). These results raise the possibility that Snu114-bound GDP may not be hydrolysed during splicing.

The guanine nucleotide in the post-splicing ILS is interpreted as GDP, but its conformation is distinct from that of GDP in other GTPases (Extended Data Fig. 5a). In contrast, the conformation of the Snu114-bound GDP in tri-snRNP superimposes well with GDP or non-hydrolysable GTP analogues in other GTPases (Extended Data Fig. 5c). When we refined our structure of Snu114 with GDP, the result reveals extra density for the N terminus of Prp3 extending towards the C-terminal extension of Prp3. Masked classification of the arm domain further extends to form a ferredoxin-like domain, which packs against the Prp4 WD40 domain and Brr2 N-terminal domain. Prp3 folds back to form a long α-helix binding across the major groove of U4/U6 helix II (Fig. 4b; Extended Data Fig. 3d). This latter two Prp3 helices and the connecting loop interact extensively with the RNase H-like domain of Prp8 and Brr2 N-terminal domain. Prp3 further extends to form a ferredoxin-like domain, which packs against the Prp4 WD40 domain. Masked classification of the arm domain reveals extra density for the N terminus of Prp3 extending towards the LSm protein ring (Extended Data Fig. 6a, b). The 3′-end of U6 snRNA binds to the central hole of the LSm protein ring while the preceding single-stranded region binds to the ferredoxin-like domain of Prp3. The Nop and coiled-coil domains of Prp3 interact with Snu13, whereas the k-turn motif of U4 5′-stem-loop is sandwiched between Snu13 and Prp3 (refs 36, 39) (Fig. 4c). The extended polypeptide chain of Prp3 runs between the phosphate backbone of U4 5′-stem and Dib1, and forms a small domain together with Prp6 which is surrounded by the three-way RNA helix junction and the α-helix bundle of Prp8 (Fig. 4d).

The C terminus of the Prp6 TPR repeats interacts with the Prp4 WD40 domain, Snu13, Prp31 and the tip of U4 5′-stem-loop (Fig. 4e), while an extended N-terminal polypeptide of Prp6 packs against the RNase H-like domain of Prp8 and interacts with the small carboxy-terminal domain of Prp31, the Prp8 α-helix and U4/U6 snRNA three-way junction and then wraps around the Prp8 helix bundle (Fig. 4d; Extended Data Fig. 3d–f). The numerous interactions that Prp6 makes with U4/U6 snRNP components and Prp8 reflect its importance for tri-snRNP assembly.

The U4/U6 di-snRNP

The extensively base-paired U4/U6 snRNAs form a three-way helix junction (Extended Data Figs 4a, b). Snu13, bound to the k-turn motif, is wedged between the U4 5′-stem-loop and the U4/U6 snRNA helix II and packs against the Prp4 WD40 domain. Prp3 makes extensive interactions with the Prp4 WD40 domain, the basket handle-like structure and Snu13, and forms a long α-helix sitting in the minor groove of U4/U6 helix II. After forming a short α-helix, Prp3 folds back to form a long α-helix binding across the major groove of U4/U6 helix II (Fig. 4b; Extended Data Fig. 3d). This latter two Prp3 helices and the connecting loop interact extensively with the RNase H-like domain of Prp8 and Brr2 N-terminal domain. Prp3 further extends to form a ferredoxin-like domain, which packs against the Prp4 WD40 domain. Masked classification of the arm domain reveals extra density for the N terminus of Prp3 extending towards the LSm protein ring (Extended Data Fig. 6a, b). The 3′-end of U6 snRNA binds to the central hole of the LSm protein ring while the preceding single-stranded region binds to the ferredoxin-like domain of Prp3 (ref. 38). The Nop and coiled-coil domains of Prp31 interact with Snu13, whereas the k-turn motif of U4 5′-stem-loop is sandwiched between Snu13 and Prp3 (refs 36, 39) (Fig. 4c). The extended polypeptide chain of Prp31 runs between the phosphate backbone of U4 5′-stem and Dib1, and forms a small domain together with Prp6 which is surrounded by the three-way RNA helix junction and the α-helix bundle of Prp8 (Fig. 4d).

Brr2

The single-stranded region of U4 snRNA (Extended Data Figs 3b and 4a), extending from stem I, enters the active site of Brr2 N-terminal helicase cassette near the strand-separating β-hairpin and passes through the channel between the RecA1, RecA2, Ratchet and WH domains (Extended Data Fig. 7a–c). The N-terminal domain (NTD) of Brr2 extends towards U4/U6 stem I and contacts the long helix of Prp3 running along the phosphate backbone of U4 snRNA. Brr2 inserts a loop...
of the NTD into the minor groove of U4/U6 stem II (Extended Data Fig. 7b, d). These interactions may guide U4/U6 stem II during unwinding, Snu13, Prp4 WD40, Prp3 ferredoxin, and Prp31 Nop and coiled-coil domains assemble together while the long α-helices and extended polypeptides may function like elastic bands to accommodate conformational changes and partial strand separation of the U4/U6 duplex as Brr2 translocates along U4 snRNA and unwinds U4/U6 stem I (refs 16, 43). Brr2 forms a stable complex with the Jab1/MPN domain of Prp8 (ref. 42), which is attached to the RNase H-like domain of Prp8 via a long flexible linker, enabling both Brr2 and U4/U6 di-snRNP to detach from the main body of Prp8 during unwinding.

The improved map of the head domain at 4.5–5.0 Å resolution, obtained by masked refinement, enabled us to build most of the Snu66 structure as poly-Ala chains. Its N-terminal region forms a globular domain that interacts with Prp8 endonuclease-like and Brr2 N-terminal ratchet domains. This is followed by a long helix wedged between Prp8 Jab1/MPN and Brr2 N-terminal HLH domains while its C terminus wraps around Brr2, forming extensive interactions with the Brr2 C-terminal cassette (Extended Data Fig. 7e), fully consistent with yeast two-hybrid and co-immunoprecipitation assays44. Interestingly, our global classification approach showed ‘open’ and ‘closed’ conformations of the head and foot domains (Extended Data Fig. 6c–e). In the ‘closed’ conformation, the globular domain of Snu66 contacts the N-terminal domain of Prp8, which in turn interacts with Snu114.

**Insight into spliceosome activation**

A comparison of U4/U6, U5 tri-snRNP with B45 and BΔU1 complexes46 shows that U2 snRNPs docks with tri-snRNP where the LSm complex, Prp3 and Prp6 are located, while U1 snRNP sits on top of U2 snRNP (Fig. 5a). The components of NTC/NTR are also detected by mass spectrometry in complex B3. We compared the structures of our tri-snRNP and the post-splicing ILS19 by overlaying the large domain of Prp8 together with Snu114 and the U5 core domain. This shows that NTC and NTR can associate with tri-snRNP without clashing and contact U2 snRNP (Fig. 5a, b). In complex B, U2 snRNP interacts with U4/U6,U5 tri-snRNP46, but when NTC and NTR dock with tri-snRNP, U2 snRNP is passed to NTC and NTR, and U2 Sm domain and U2B′U2A′ complex associate with Aquarius(Cwf11), Syf1(Cwf3) and Isy1(Cwf12)47 as revealed in the S. pombe ILS19 (S. pombe protein names are shown in parentheses).

The S. cerevisiae U4/U6,U5 tri-snRNP and S. pombe ILS structures reveal that the foot domains of the two structures, containing the Prp8 N-terminal domain, U5 snRNA stem-loop I and Snu114, superpose very well showing that they form a stable structural unit (Extended Data Fig. 5f). Overlay of their Prp8 large domains shows that the foot domain rotates as a rigid body by 30° between the two structures, causing U5 loop 1 to move closer towards the Prp8 α-finger in the post-splicing ILS19 (Fig. 5c). NTC forms extensive interfaces with both the N-terminal and large domains of Prp8, hence the rotation of the foot domain may be caused by NTC. When the foot domain of the U4/U6,U5 tri-snRNP structure rotates by 30° (as in the post-splicing ILS) Prp8 resides 602–614 clash with Dib1 and the ACAGAGA helix, forcing Dib1 to dissociate from the large domain of Prp8 and liberating the ACAGAGA sequence to bind the 5′-splice site.

**Figure 5** | **B complex formation and activation mechanism.** a, U4/U6,U5 tri-snRNP fits into the EM envelope of human complex B45 (reproduced from ref. 45 with permission), showing that U2 snRNP binds near the LSm core domain, Prp6 and Prp3. b, Overlay of the Prp8 large domain between tri-snRNP and the ILS19 shows how NTC/NTR might bind to complex B and interact with U2 snRNP so that U2 snRNP can be passed to the NTC/NTR complex. c, A comparison of the tri-snRNP and the ILS19 structures shows rotation of the foot domain with respect to the Prp8 large domain. Upon rotation, Prp8 residues 602–614 will clash with Dib1 and ACAGAGA helix, causing them to dissociate thus liberating the ACAGAGA sequence to bind the 5′-splice site.

© 2016 Macmillan Publishers Limited. All rights reserved

**Received 3 November; accepted 18 December 2015.**

**Published online 1 February; corrected online 17 February 2016** (see full-text HTML version for details).

1. Will, C. L. & Lührmann, R. Spliceosome structure and function. Cold Spring Harb. Perspect. Biol. 3, a003707 (2011).
2. Chen, C. H. et al. Functional and physical interactions between components of the Prp19p-associated complex. Nucleic Acids Res. 30, 1029–1037 (2002).
3. Fabrizio, P. et al. The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. Mol. Cell 36, 593–608 (2009).
4. Oht, T. et al. Molecular dissection of step 2 catalysis of yeast pre-mRNA splicing investigated in a purified system. RNA 19, 902–915 (2013).
5. Slaby, J. P. & Guthrie, C. An RNA switch at the 5′ splice site requires ATP and the DEAD box protein Prp28p. Mol. Cell 3, 55–64 (1999).
6. Raghunathan, P. L. & Guthrie, C. RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. Curr. Biol. 8, 847–855 (1998).
7. Laggerbauer, B., Achsel, T. & Lührmann, R. The human U5-200KD DEAH-box protein unwinds U4/U6 RNA duplexes in vitro. Proc. Natl Acad. Sci. USA 95, 4188–4192 (1998).
Bartels, C., Urlaub, H., Lührmann, R. & Fabrizio, P. Mutagenesis suggests
Small, E. C., Leggett, S. R., Winans, A. A. & Staley, J. P. Evidence for a group II intron-like catalytic triplex in the spliceosome. Nature Struct. Mol. Biol. 21, 464–471 (2014).
Newman, A. J. & Norman, C. U5 snRNA interacts with exon sequences at 5′ and 3′ splice sites. Cell 68, 743–754 (1992).
Sonnheimer, E. J. & Stelzl, J. A. The U5 and U6 small nuclear RNAs as active site components of the spliceosome. Science 262, 1989–1996 (1993).
Stark, H. & Lührmann, R. Cryo-electron microscopy of spliceosomal components. Annu. Rev. Biophys. Biomol. Struct. 35, 435–457 (2006).
Bai, X. C., McMullan, G. & Scheres, S. H. How cryo-EM is revolutionizing structural biology. Trends Biochem. Sci. 40, 49–57 (2015).
Nguyen, T. H. et al. The architecture of the spliceosomal U4/U6/U5 tri-snRNP. Nature 523, 47–52 (2015).
Onishi, D., Ren, T., Waseda, K. G. M., L. K. & Watz, T. Structural characterization of the fusion yeast U5/U6 spliceosome complex. Proc. Natl Acad. Sci. USA 104, 3195–3200 (2007).
Chen, W. et al. Endogenous U2/U5-U6 snRNA complexes in S. pombe are intron tail snRNPs. RNA 20, 308–320 (2014).
Van, C. et al. Structure of a yeast spliceosome 3.6-angstrom resolution. Science 349, 1182–1191 (2015).
Hang, J., Wan, R., Yan, C. & Shi, Y. Structural basis of pre-mRNA splicing. Science 349, 1181–1186 (2015).
Stevens, S. W. et al. Biochemical and genetic analyses of the U5, U6, and U4/U6/U5 small nuclear ribonucleoproteins from Saccharomyces cerevisiae. RNA 7, 1543–1553 (2001).
Gottschalk, A. et al. Identification by mass spectrometry and functional analysis of novel proteins of the yeast U5/U6/U5 tri-snRNP. EMBO J. 18, 4535–4548 (1999).
Bai, X.-C., Rajendra, E., Yang, G., Shi, Y. & Scheres, S. H. W. Sampling the active site components of the spliceosome. Nature Struct. Mol. Biol. 22, 138–144 (2015).
Li, Z. & Brow, D. A. A spontaneous duplication in U6 spliceosomal RNA uncouples the early and late functions of the ACAGA element in vivo. RNA 2, 879–894 (1996).
Kuhn, A. N. & Brow, D. A. Suppressors of a cold-sensitive mutation in yeast U4 RNA define five domains in the splicing factor Prp8 that influence spliceosome activation. Genetics 155, 1667–1682 (2000).
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).

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

Acknowledgements We thank C. Savva, S. Chen, G. McMullan, J. Grimmert and T. Darling for running the electron microscopy and computing facilities, A. Brown, P. Emsley, G. Murshudov for advice and help with model building and refinement, R. O’Keefe for the ΔSnu114 yeast strain, the members of the spliceosome group for help and advice throughout the project and R. Leiro for help with data processing. We thank S. Fica for critical reading of the manuscript and J. Löwe, V. Ramakrishnan, and R. Henderson for their continuing support and encouragements. The project was supported by the Medical Research Council (MC_U105184330 to K.N. and MC_UP_A025_1013 to S.H.W.S.).

Author Contributions T.H.D.N. purified yeast tri-snRNP and prepared EM grids, T.H.D.N., W.P.G. and E.O. calculated the maps with the help of X.-C.B and S.H.W.S.; T.H.D.N., W.P.G. and X.-C.B collected all EM images. T.H.D.N. processed data and refinement, R. O’Keefe for the EMDB-8014. The coordinates of the atomic models have been deposited in the EMDB with accession code EMD-8006, EMD-8007, EMD-8008, EMD-8009, EMD-8010, EMD-8011, EMD-8012, EMD-8013 and EMD-8014. The coordinates of the atomic models have been deposited in the Protein Data Bank under accession codes 5GAN (overall), 5GA9 (head domain) and 5GAM (foot domain). 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), W.P.G (wgalej@mrc-lmb.cam.ac.uk) or K.N. (kn@mrc-lmb.cam.ac.uk).

Author Information The cryo-EM maps have been deposited in the Electron Microscopy Data Bank with accession codes EMD-8006, EMD-8007, EMD-8008, EMD-8009, EMD-8010, EMD-8011, EMD-8012, EMD-8013 and EMD-8014. The coordinates of the atomic models have been deposited in the Protein Data Bank under accession codes 5GAN (overall), 5GA9 (head domain) and 5GAM (foot domain). 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), W.P.G (wgalej@mrc-lmb.cam.ac.uk) or K.N. (kn@mrc-lmb.cam.ac.uk).
METHODS

Statistics. No statistical methods were used to predetermined sample size. Sample preparation. Tri-snRNP sample was prepared as described in our published protocol46.

Electron microscopy. Aliquots of 3.5 μl of purified yeast tri-snRNP were applied to Quantifoil Cu R1.2/1.3, 400 mesh grids, which were coated with 6 nm-thick homemade carbon film and glow-discharged in N-amylnitrile. The grids were blotted for 2 s at 4 °C, plunged into liquid ethane by an FEI Vitrobot MKIII at 100% humidity and loaded onto a Titan Krios transmission electron microscope operated at 300 kV. Zero-loss-energy images were collected manually on a Gatan K2-Summit detector in super-resolution counting mode at a calibrated magnification of 35,714× (pixel size of 1.43 Å) and a dose rate of ~2.5 electrons A⁻² per second (Extended Data Fig. 1a). We used a slit width of 20 eV on a GIF Quantum 4.3 Å reconstructions, respectively, after auto-refinement with more homogeneous

A subset of ~5-0,000 particles was picked manually, extracted using a 380² pixel box and subjected to reference-free 2D classification. Some of the resulting 2D class averages were low-pass filtered to 20 Å and used as references for automatic particle picking of the whole data set of 2,477 micrographs. The automatically picked particles were screened manually to remove false positives, aggregation and ice contamination, resulting in an initial set of 473,827 particles for reference-free 2D classification. We selected 438,602 particles from good 2D classes for the 3D classification (Extended Data Fig. 1b, c), which was run for 25 iterations, using an angular sampling of 7.5°, a regularisation parameter T of 4 and a 60 Å low-pass filtered initial model from our previous reconstruction16. A subset of 140,155 particles was selected for the first 3D auto-refinement. Particle-based beam-induced motion correction and radiation-damage weighting (particle polishing) were performed on these particles54. Auto-refinement of the polished particles resulted in a reconstruction at 3.7 Å overall resolution with an estimated angular accuracy of 1.1°.

Local resolution analysis by Resmap55 showed a range of resolution from 3.0 Å in the core to 10 Å in the arm domain and part of the head domain, indicating conformational heterogeneity within the complex. As previously observed, the four domains of the structure, particularly the head and arm domains, are flexible in our structure. We employed two classification/refinement approaches: a local approach to improve the local resolution of the domains and a global approach to allow global conformations of the domains relative to one another to be observed (Extended Data Fig. 1c). For the local approach, we used a masked refinement procedure with signal subtraction for each of the head, body and foot domains23 and a masked classification with signal subtraction followed by a masked refinement for the most flexible arm domain23. Each of the four domains only makes up a third or less of the total mass of the complex. For each domain, we subtracted projections from the remaining three domains of the reconstruction in the experimental particle images using the relative orientation of each experimental image from the last auto-refinement run of all the polished particles. This resulted in four sets of new experimental particle images that only have signal from the domain of interest. For the body, foot and head domains the subtracted experimental images were used in 3D auto-refinement with a soft mask for that domain, yielding 3.6, 3.7 and 4.2 Å reconstructions for the body, foot and head domains, respectively (Extended Data Figs 1b, 2a–c). The arm domain is too small for accurate alignments of the

Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nature Methods 10, 584–590 (2013).

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

Scheres, S. H. A Bayesian view on cryo-EM structure determination. J. Mol. Biol. 415, 406–418 (2012).

Scheres, S. H. Beam-induced motion correction for sub-megadalton cryo-EM particles. eLife 3, e03655 (2014).

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

Scheres, S. H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. Nature Methods 9, 853–854 (2012).

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

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. Acta Crystallogr. D 66, 486–501 (2010).
59. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).
60. Nicholls, R. A., Fischer, M., McNicholas, S. & Murshudov, G. N. Conformation-independent structural comparison of macromolecules with ProSMART. *Acta Crystallogr. D* **70**, 2487–2493 (2014).
61. Brown, A. et al. Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr. D* **71**, 136–153 (2015).
62. Scheres, S. H. W., Núñez-Ramírez, 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).
63. Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. *J. Struct. Biol.* **157**, 281–287 (2007).
64. Kunkel, T. A. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl Acad. Sci. USA* **82**, 488–492 (1985).
65. Turner, I. A., Norman, C. M., Churcher, M. J. & Newman, A. J. Dissection of Prp8 protein defines multiple interactions with crucial RNA sequences in the catalytic core of the spliceosome. *RNA* **12**, 375–386 (2006).
66. Frazer, L. N., Lovell, S. C. & O’Keefe, R. T. Analysis of synthetic lethality reveals genetic interactions between the GTPase Snu114p and snRNAs in the catalytic core of the *Saccharomyces cerevisiae* spliceosome. *Genetics* **183**, 497–515 (2009).
Extended Data Figure 1 | Image processing procedures. a, Representative micrograph. b, Representative 2D class averages obtained from reference-free 2D classification. c, Classification and refinement procedures used in this study.
Extended Data Figure 2 | Local and overall resolutions of tri-snRNP maps. Local resolution estimation by Resmap of a, the overall 3.7 Å map and b, maps of the head, body and foot domains obtained from masked refinements with signal subtraction. c, Gold-standard FSC curves for the overall map and the maps of the head, body and foot domains obtained from masked refinements. Their resolutions are estimated at FSC = 0.143. d–g, FSC curves of model versus map and cross-validation of model refinement by half-maps for the body, foot, head and overall maps, respectively. The red curves show FSC between the atomic model and the half-map it was refined against (half1) and the blue curves show FSC between the atomic model and the other half-map (half2) it was not refined against. The black curves show FSC between the atomic model and the sum map which the model was refined against.

© 2016 Macmillan Publishers Limited. All rights reserved
Extended Data Figure 3 | Representative EM density for different components of the map. a, Snu114 in the Foot domain with a bound GTP (magenta). The inset shows the GTP-binding pocket. b, Brr2 in the head domain with a bound single-stranded region of U4 snRNA. The inset shows the density in the RNA binding tunnel. c, Density for Prp8 large and RNase-like domains. The inset shows the density in the core of Prp8. d–f, Prp3, Prp31 and Prp6 densities, respectively, with extended polypeptides.
Extended Data Figure 4 | Secondary structure of the snRNAs in tri-snRNP. a, U4/U6 snRNA; c, U5 snRNA. The coloured nucleotides with red, green and blue background were built de novo into our EM density. The region near the ACAGAGA sequence of U6 snRNA forms a stem-loop that was not predicted previously. b, d, Representative EM density for U4/U6 snRNA duplex and U5 snRNA, respectively.
Extended Data Figure 5 | See next page for figure caption.
Extended Data Figure 5 | Interactions of Snu114 with guanine nucleotides and the N-terminal domain of Prp8 in the *S. cerevisiae* U4/U6.U5 tri-snRNP and *S. pombe* ILS complexes. **a**, Conformation of the Snu114(Cwf10)-bound GDP refined in the *S. pombe* ILS spliceosomal complex,19,20 (red, PDB 3JB9), was overlaid on GDPS found in other guanine-nucleotide binding proteins (grey, PDB coordinates: 1DAR, 2E1R, 2WRI, 1Z0I, 5CA8, 1XTQ, 4YLG, 1SF8, 5BXQ). **b**, Guanine nucleotide refined as GDP in Snu114 of the *S. cerevisiae* U4/U6.U5 tri-snRNP (blue) is overlaid on GDPS found in the PDB coordinates as in **a**. **c**, Conformation of guanine nucleotide refined as GTP in Snu114 of the *S. cerevisiae* U4/U6.U5 tri-snRNP (blue) agrees well with GTP or GTP analogues in other guanine-nucleotide binding proteins (PDB code: 2BV3, 2DY1, 2J7K, 4YW9, 1ASO, 1LF0 (grey)). **d**, Superposition of the active site of Snu114-GTP and Cwf10-GDP. **e**, Superposition of the GDP-bound EF-G (2WRI), GMP-PCP bound EF-G (4JUW) and Snu114 (*S. cerevisiae* tri-snRNP) active sites. His218 (His87 in EF-G) positions water molecule crucial for GTP hydrolysis. **f**, Comparison of Prp8 N-terminal domain, Snu114 and U5 snRNA in the *S. cerevisiae* U4/U6.U5 complex and *S. pombe* ILS complex. **g**, Growth of serial dilutions of yeast strains carrying wild-type Snu114, His218Arg or His218Ala Snu114 mutants at different temperatures. Cells were spotted on YPD plates and grown at 14 °C for 10 days, 30 °C and 37 °C for 2 days. **h**, Growth of serial dilutions of yeast strains carrying wild-type Prp8, Tyr403Phe and Tyr403Ala mutants. Cells were spotted on YPD plates and grown at 14 °C for 9 days, 30 °C for 3 days. This yeast strain does not survive at 37 °C and thus is not shown.
Extended Data Figure 6 | Conformational flexibility of tri-snRNP observed by classification. a, Different conformations of the arm domain demonstrated by the unsharpened maps of the three major classes (purple, magenta and red) obtained from masked classification of the arm domain alone followed by masked refinement with the body and arm domains. The body domain was included in the refinement because the arm domain is too small for accurate alignments. b, The sharpened map of one of the three classes with Prp3 and LSm models shown. In the improved domain maps for the arm domain, extra density for the N-terminal helix of Prp3 could be observed to extend to the LSm proteins. c, The sharpened map of the tri-snRNP and the locations of Snu66 and Prp8. d, The open and closed conformations of the head and foot domains of the tri-snRNP observed by global classification. The unsharpened maps for the two major classes obtained from global classification with finer angular sampling (1.8°) followed by 3D auto-refinement are shown. The open and closed states are indicated. e, Superposition of the unsharpened maps of the open (grey) and closed (yellow) states shown in d. The arrows indicate the rotations of the head and foot domains.
Extended Data Figure 7  |  Brr2 helicase and its U4/U6 snRNA substrate.

**a**, Domain structure of Brr2 helicase comprising the N-terminal domain and two helicase cassettes. Individual domains of N-terminal helicase cassette (NHC) are colour-coded. **b**, Extensive interactions of Brr2 with U4/U6 snRNA and Prp3. The single-stranded region of U4 snRNA extending from stem I enters the active site near the β-finger (red). **c**, 3′ stem of U4 snRNA interacts with the HLH domain of NHC. **d**, The N-terminal domain (NTD) of Brr2 interacts with a long helix of Prp3 and inserts a loop into U4/U6 Stem II. **e**, Snu66 has a long extended region that wraps around both helicase cassettes of Brr2.
Extended Data Table 1 | Summary of model building of tri-snRNP components

| Protein | Total Residues | M.W. | Modeled | Chain Name | Local Map | Human/S. pombe Names |
|---------|----------------|------|---------|------------|-----------|----------------------|
| **US snRNP** | | | | | | |
| Prp8 | 2413 | 279,299 | 110-2401 | A | 10B-735: Foot | 220K/Spp42 |
| Brr2 | 2163 | 246,125 | 364-2163 | B | 363-433: Body | 200K/Brr2 |
| Snu114 | 1008 | 114,025 | 102-989 | C | Foot | 116K/Cwf10 |
| Dib1 | 143 | 16,774 | 2-137 | D | Body | 15K/Dib1 |
| SmB | 196 | 22,403 | 4-102 | b | SmB/SmB |
| SmD3 | 110 | 11,229 | 1-109 | d | SmD3/SmD3 |
| SmD1 | 146 | 16,288 | 15-108 | h | SmD1/SmD1 |
| SmD2 | 110 | 12,856 | 4-85 | i | Foot | SmD2/SmD2 |
| SmE | 94 | 10,373 | 4-92 | e | SmE/SmE |
| SmE | 96 | 6,559 | 12-83 | f | SmE/SmE |
| SmG | 77 | 8,479 | 2-76 | g | SmG/SmG |
| **US snRNA-L** | | | | | | |
| Snu13 | 126 | 13,570 | 3-126 | K | Body | 15.5K/Snu13 |
| Prp31 | 494 | 56,305 | 43-457 | F | Body | 61K/Prp31 |
| Prp3 | 469 | 55,877 | 150-467 | G | Body | 90K/Prp3 |
| Prp4 | 465 | 52,425 | 109-465 | H | Body | 60K/Rna4 |
| **U4/U6 snRNP** | | | | | | |
| SmB | 196 | 22,403 | 4-102 | k | SmB/SmB |
| SmD1 | 146 | 16,288 | 1-118 | l | SmD1/SmD1 |
| SmD2 | 110 | 12,856 | 15-108 | m | SmD2/SmD2 |
| SmD3 | 110 | 11,229 | 4-85 | n | Head | SmD3/SmD3 |
| SmE | 94 | 10,373 | 10-92 | p | SmE/SmE |
| SmF | 96 | 6,559 | 12-83 | q | SmF/SmF |
| SmG | 77 | 8,479 | 2-76 | r | SmG/SmG |
| Lsm2 | 95 | 11,164 | 1-90 | 2 | Lsm2/Lsm2 |
| Lsm3 | 89 | 10,020 | 3-79 | 3 | Lsm3/Lsm3 |
| Lsm4 | 172 | 20,304 | 1-90 | 4 | Lsm4/Lsm4 |
| Lsm5 | 93 | 10,415 | 4-84 | 5 | Arm | Lsm5/Lsm5 |
| Lsm6 | 86 | 9,396 | 11-84 | 6 | Lsm6/Lsm6 |
| Lsm7 | 115 | 13,010 | 26-105 | 7 | Lsm7/Lsm7 |
| Lsm8 | 109 | 12,385 | 1-67 | 8 | Lsm8/Lsm8 |
| **U4 snRNA** | | | | | | |
| U4 snRNA | 160 | 51,390 | 1-152 | V | 1-67: Body | 73-152: Head |
| **U6 snRNA** | | | | | | |
| U6 snRNA | 112 | 36,088 | 1-112 | W | 1-67: Foot | 26-88: Body |
| | | | | | | |
| **tri-snRNP specific** | | | | | | |
| Prp6 | 899 | 104,234 | 155-898 | J | Body | 102K/Prp1 |
| Snu66 | 587 | 66,426 | 5-560 (poly-Ala) | E | Head | 110K/Snu66 |
| Prp38 | 242 | 27,957 | Not modeled | h | Prp38/Prp38 |
| Snu23 | 194 | 22,682 | Not modeled | h | Snu23/Snu23 |
| Spp381 | 291 | 33,764 | Not modeled | | |

© 2016 Macmillan Publishers Limited. All rights reserved
Extended Data Table 2 | Refinement, model statistics and structure/map depositions

### a. Statistics of tri-snRNP structure determination

| Data collection | | | |
|----------------|----------------|----------------|----------------|
| EM             | Titan Krios 300kV, K2 Gatan Summit | | |
| Pixel size (Å) | 1.43 | | |
| Defocus range (µm) | -0.5 to -3.5 | | |

| Reconstruction (RELION) | Overall | Body | Foot | Head |
|-------------------------|---------|------|------|------|
| Accuracy of rotations (°) | 1.13 | 1.15 | 1.73 | 2.42 |
| Accuracy of translations (pixel) | 0.65 | 0.67 | 0.89 | 1.28 |
| Final resolution (Å) | 3.7 | 3.6 | 3.7 | 4.2 |

| Refinement (REFMAC) | | | |
|---------------------|----------------|----------------|----------------|
| Refinement weight | 0.001 | 0.001 | 0.001 | 0.001 |
| Resolution limit (Å) | 3.6 | 3.6 | 3.6 | 3.6 |
| Residue numbers | 9325 | 3728 | 2186 | 2922 |
| Fourier Shell Correlation | 0.75 | 0.85 | 0.82 | 0.60 |
| R-factor (%) | 29.7 | 27.8 | 28.7 | 31.5 |
| Rms bond length (Å) | 0.0078 | 0.0073 | 0.0073 | 0.011 |
| Rms bond angle (°) | 1.27 | 1.33 | 1.38 | 1.40 |

| Ramachandran plot | | | |
|-------------------|----------------|----------------|----------------|
| Favoured | 8066 (91.4%) | 3266 (91.9%) | 1810 (90.9%) | 2531 (89.3%) |
| Allowed | 615 (6.9%) | 237 (6.7%) | 135 (7.2%) | 238 (9.3%) |
| Outliers | 152 (1.7%) | 50 (1.4%) | 37 (1.9%) | 39 (1.4%) |

| Validation by Molprobity | | | |
|--------------------------|----------------|----------------|----------------|
| Geometry score (percentile) | 2.52 (98th) | 2.41 (99th) | 2.79 (95th) | 2.62 (97th) |
| Clash score (percentile) | 7.48 (97th) | 6.78 (100th) | 11.4 (97th) | 6.82 (100th) |
| Good rotamer (%) | 94.8 | 95.7 | 93.5 | 93.2 |

### b. Deposited maps and associated coordinate files

| Maps | EMDB code | Associated PDB ID |
|------|-----------|-------------------|
| Overall map | EMD-8012 | 5GAN |
| Body map | EMD-8014 | 5GAP |
| Head map | EMD-8013 | 5GAO |
| Foot map | EMD-8011 | 5GAM |
| Global class 1 (closed state) | EMD-8007 | |
| Global class 2 (open state) | EMD-8006 | |
| Masked body/arm class 1 | EMD-8008 | |
| Masked body/arm class 2 | EMD-8009 | |
| Masked body/arm class 3 | EMD-8010 | |

© 2016 Macmillan Publishers Limited. All rights reserved