



Supplementary Materials for

Mechanism of 5' splice site transfer for human spliceosome activation

Clément Charenton*†, Max E. Wilkinson*†, Kiyoshi Nagai*

*Corresponding author. Email: ccharent@mrc-lmb.cam.ac.uk (C.C.); mwilkin@mrc-lmb.cam.ac.uk (M.E.W.); kn@mrc-lmb.cam.ac.uk (K.N.)

†These authors contributed equally to this work.

Published 11 April 2019 on *Science* First Release
DOI: 10.1126/science.aax3289

This PDF file includes:

Materials and Methods
Figs. S1 to S11
Tables S1 and S2
Caption for Movie S1
Captions for Supplementary Files S1 and S2
References

Other Supplementary Material for this manuscript includes the following:

(available at www.sciencemag.org/cgi/content/full/science.aax3289/DC1)

Movie S1 (.mp4)
Supplementary Files S1 and S2 (.pse)

Materials and Methods

Cloning, expression and purification of recombinant proteins

The coding sequence of human Prp28 (DDX23) was purchased from Dharmacon and sub-cloned into a pET28a vector containing N-terminal hexa-histidine and StrepII (His₆-S_{II}) tags. The DEAD-box motif (residues 549 – 552) was mutated to AAAD by the quick-change method. This His₆-S_{II}-Prp28^{AAAD} coding sequence was sub-cloned into the pENTCQ vector between the GAL-GAP promoter and the SV40 3'UTR sequence and the expression cassette was transferred into pRS Dest426 destination plasmid by LR Clonase-mediated recombination. Protein expression was carried out in *S. cerevisiae* strain BCY123 (MATa pep4::HIS3 prb1::LEU2 bar1::HIS6 lys2::GAL1/10-GAL4 can1 ade2 trp1 ura3 his3 leu2-3,112) essentially as described (27). Prp28^{AAAD} was purified by Ni-NTA followed by size exclusion chromatography (Superdex© 200 26/60) in a final buffer containing 50 mM Tris-HCl pH 8, 500 mM NaCl, 5% Glycerol and 5 mM 2-mercaptoethanol, concentrated to ~5-10 mg/mL, flash frozen in liquid nitrogen and stored at -80°C (fig. S1A).

Splicing substrate preparation

An SP72 vector containing the AdML pre-mRNA sequence with three MS2 loops at the 3' end was purchased from Edigene and was used to generate an *in vitro* transcription template by PCR. After gel purification of the PCR product, *in vitro* transcription was carried out for 3h at 37°C using in-house produced T7 RNA polymerase. The transcription product was then purified by 6% denaturing polyacrylamide gel electrophoresis (PAGE), 3'-end labelled with fluorescein as described (28) and capped using the Vaccinia capping system kit (NEB).

HeLa nuclear extract preparation and *in vitro* splicing

Each batch of HeLa S3 nuclear extract was prepared from ~25 L of HeLa S3 cells grown at 37°C and 8% CO₂ in roller bottles to a final density of ~1,500,000 cells/mL. Nuclear extract and S100 fraction were prepared as described in (29), using a 30 mL Dounce homogenizer, dialyzed twice for 90 min against Buffer D100 (20 mM HEPES-KOH, pH 7.9, 100 mM KCl, 20% glycerol, 0.2 mM EDTA) before flash freezing in liquid nitrogen and storage at -80°C.

In vitro splicing assays were carried out for 30 min at 30°C using 50 µL splicing reaction mix composed of 20% of HeLa S3 nuclear extract, 20% of HeLa S3 S100, 2 mM ATP, 10 mM creatine phosphate, 3 mM of MgCl₂ and 10 nM of AdML pre-mRNA and in the presence of increasing concentrations of Prp28^{AAAD} (0, 25, 50, 100 ng/µL final). Each splicing reaction was quenched by addition of 30 µL of stop mix (1 mg/mL proteinase K, 50 mM EDTA, 0.1% SDS). RNA was purified by phenol-chloroform extraction followed by ethanol precipitation. Purified RNA was run on a 14% denaturing PAGE and splicing efficiency was monitored by visualizing fluorescein signal using a Typhoon scanner (fig. S1B).

Pre-B complex preparation

18x 10mL splicing reactions containing 20% of HeLa S3 nuclear extract, 20% of HeLa S3 S-100, 2 mM ATP, 10 mM creatine phosphate and 3 mM of MgCl₂ were incubated in the presence of 50 ng/µL His₆-S_{II}-Prp28^{AAAD} for 30 min at 30°C before addition of the AdML-3xMS2 substrate pre-bound to MS2-MBP at a final concentration of 10 nM. Spliceosomes were then assembled for 30 min at 30°C. Pre-B spliceosomes were purified by glycerol cushion, amylose affinity and Strep affinity purification essentially as described (6). Briefly, each 10 mL of splicing reaction was diluted by addition of 10 mL of dilution buffer DK40 (20 mM HEPES-KOH pH 7.9, 40 mM KCl, 0.25 mM EDTA, 1% glycerol, 1 mM DTT, 0.03% IGEPAL). The resulting diluted splicing reactions were ultra-centrifuged at 28,000rpm on top of an 8.5mL 40% glycerol cushion (20 mM HEPES-KOH pH 7.9, 50 mM KCl, 0.25 mM EDTA, 40% glycerol, 1 mM DTT, 0.02% IGEPAL).

The cushions were then harvested, diluted in 5 mL of DK50 (20 mM HEPES-KOH pH 7.9, 50 mM KCl, 0.25 mM EDTA, 1% glycerol, 1 mM DTT, 0.03% IGEPAL) and incubated with amylose resin high flow (NEB) overnight at 4°C. After extensive washing with WK50 (20 mM HEPES-KOH pH 7.9, 50 mM KCl, 0.25 mM EDTA, 1% glycerol, 1 mM DTT, 0.01% IGEPAL) the complexes assembled on the pre-mRNA were eluted from the resin by addition of WK50 supplemented with 12 mM of maltose. The fractions containing the complexes, as judged by fluorescein signal, were pooled and incubated several hours at 4°C in the presence of StrepTactin Sepharose high performance resin (GE Healthcare). After extensive washing of the StrepTactin resin using WK50 the Strep-tagged complexes were eluted by addition of WK50 supplemented with 5 mM D-Desthiobiotin (Santa Cruz Biotechnology).

Purified complexes were then subjected to BS3 (Thermo Fischer) crosslinking (1 mM final) for 1h on ice. Remaining free BS3 was quenched for 15 min on ice by addition of ammonium bicarbonate (50 mM final). Finally, the sample was concentrated using 100k MWCO Amicon filters, dialysed 3h against a buffer containing 20 mM HEPES-KOH, pH 7.9, 50 mM KCl, 0.2 mM EDTA and 1 mM DTT and used immediately for cryo-EM grid preparation. Cryo-EM grids were prepared on a FEI Vitrobot Mark III at 4°C and 100% humidity. 3 μ L of concentrated sample were applied to glow-discharged (20s using Edward plasma cleaner) Cu 300 R2/2 holey carbon grids (Quantifoil) coated with a \sim 60 Å thick homemade continuous carbon film. After 30s the grids were blotted for 2-3 s using in-house prepared blotting paper and immediately plunged into 93K liquid ethane.

Cryo-EM data acquisition

Five data sets and a total of 27,931 movies were collected using EPU on various Titan Krios microscopes (Thermo Fisher) all equipped with energy filters (slit width of 20 eV) and K2 detectors operated in counting mode. Dataset 1 (4,627 movies) was collected on Astbury Biostructure Laboratory (University of Leeds) Krios 2 using a magnification of 130,000x equivalent to a “relative” pixel size of 1.025 Å.pix⁻¹ with defocus ranging from -1.5 μ m to -3.5 μ m with an exposure time of 8, fractionated in 40 frames and a flux of 7.5 e⁻/Å²/sec equivalent to a total fluence per micrograph of 57 e⁻/Å². Dataset 2 (6,869 movies) was collected on LMB Krios 3 using a magnification of 130,000x equivalent to a “relative” pixel size of 1.022 Å.pix⁻¹ with defocus ranging from -1 μ m to -3 μ m with an exposure time of 6s fractionated in 40 frames and a flux of 8.7 e⁻/Å²/sec equivalent to a total fluence per micrograph of 50 e⁻/Å². Dataset 3 (7,303 movies) was collected on University of Cambridge Biochemistry Department Krios using a magnification of 130,000x equivalent to a “relative” pixel size of 1.020 Å.pix⁻¹ with defocus ranging from -1.5 μ m to -3 μ m with an exposure time of 12s fractionated in 40 frames and a flux of 4.8 e⁻/Å²/sec equivalent to a total fluence per micrograph of 55.4 e⁻/Å². Dataset 4 (6,409 movies) was collected on LMB Krios 3 using a magnification of 130,000x equivalent to a “relative” pixel size of 1.022 Å.pix⁻¹ with defocus ranging from -1 μ m to -3 μ m with an exposure time of 6s fractionated in 40 frames and a flux of 8.47 e⁻/Å²/sec equivalent to a total fluence per micrograph of 48.6 e⁻/Å². Dataset 5 was collected on a sample where BS3 crosslinking was performed prior to addition of StrepTactin resin. 2,723 movies were collected on LMB Krios 3 using a magnification of 130,000x equivalent to a “relative” pixel size of 1.022 Å.pix⁻¹ with defocus ranging from -1 μ m to -3 μ m with an exposure time of 6s fractionated in 40 frames and a flux of 7.67 e⁻/Å²/sec equivalent to a total fluence per micrograph of 44 e⁻/Å².

Image processing

Movies were corrected for motion using the RELION 3.0 implementation of MotionCor2, with 5x5 patches and applying dose-weighting to individual frames. CTF parameters were estimated

using Gctf (30). Resolution is reported based on the gold-standard Fourier shell correlation (FSC) (0.143 criterion) as described (31). Particle picking was carried out for datasets 1 – 4 using the deep learning program crYOLO (32). We manually picked 800 – 1,200 particles per dataset from 30 – 50 micrographs and trained a model for each dataset using the YOLO network with 2x2 patches. Further data processing was performed using RELION 3.0 (33).

Datasets 2, 3, and 4 were used to produce the high-resolution tri-snRNP reconstruction. The detailed data processing workflow is outlined in fig. S2. Datasets were processed separately. We started with a total of 1,147,653 picked particles. Particles were first extracted with a 512 x 512 box with 2-fold binning and badly picked particles were discarded either by using reference-free 2D classification then 3D classification with human tri-snRNP as a starting model (EMD-6581), or just with 3D classification. Particles in the best 3D classes were then re-extracted with a 420 x 420 box without binning and refined with a soft mask around the whole tri-snRNP. Each reconstruction was then subjected to multiple rounds of CTF refinement and Bayesian polishing until resolution no longer improved. For dataset 2, two rounds were performed, whereas for dataset 4 CTF refinement did not improve the resolution and only one round of Bayesian polishing was performed. Dataset 3 was merged with dataset 2 to improve the resolution for CTF refinement before being separated again for Bayesian polishing. Note that although dataset 3 had a 1.020 Å.pix⁻¹ pixel size compared to 1.022 Å.pix⁻¹ for datasets 2 and 4, these pixel sizes were close enough that we did not scale these particles before merging. Particles from datasets 2, 3, and 4 were then merged to a total of 585,488 particles, refined to 2.94 Å resolution, then subjected to multi-body refinement (34). We defined five bodies, from largest to smallest: *Core*, encompassing Prp8 (excluding the RNaseH and Jab1 domains), Snu114, U5 snRNA, Sad1, Dim1, Prp28 NTD, Brr2 NTD, Prp6 NTD, and Prp31 coiled-coil domain; *Brr2*, encompassing Brr2 helicase domain, Prp8 Jab1 domain, and Prp28 N-terminal helix; *Foot*, encompassing U5 Sm ring, U5-40k, and Snu114; *U4*, encompassing the U4 Sm ring, Snu66, U4/U6 stems I and III, Prp8 RNaseH domain, and RBM42; and finally *Prp4 lobe*, encompassing Prp4, Prp3, Snu13, Prp31 Nop domain, Prp6 helical repeats, and U4/U6 stem II. The multi-body refinement excluded the U6 LSm ring and Prp28 RecA domains. Multi-body refinement gave a uniform reconstruction of tri-snRNP with more of the complex at higher resolution (fig. S4). We used focused classification without alignment to improve density for the U6 LSm ring, U5-40K, and Prp28 RecA2 domain.

All five datasets were used for cryo-EM analysis of pre-B complex. The detailed data processing workflow is outlined in fig. S3. We started with a total of 1,477,221 picked particles. Datasets 2, 3, and 4 were initially processed as detailed above. Dataset 1 was similarly processed, but a pre-B reference was included in the initial 3D classification. For dataset 5, we pre-enriched for pre-B complex by picking particles using Gautomatch (30) using pre-B templates generated by 2D classification without alignment of pre-B particles isolated during processing of the other four datasets. Weak density for U1 and U2 snRNPs in the good “tri-snRNP” particles initially selected after 2D and 3D classification indicates that our tri-snRNP particles also include some pre-B particles which align on the tri-snRNP. Our 420 x 420 box for tri-snRNP analysis was not suited for analyzing these pre-B particles, as it was centered on the tri-snRNP core and excluded density for U2 snRNP. To analyze the pre-B particles, we re-extracted tri-snRNP particles with a larger 600 x 600 pixel box, scaled down to 360 x 360 pixels to reduce computational demand (giving a scaled pixel size of 1.703 Å.pix⁻¹), and applied a shift of (52, -40, 5) pixels to move the box center to the center of pre-B complex. For dataset 1, which was 1.025 Å.pix⁻¹ compared to 1.022 Å.pix⁻¹ for most of the datasets, we extracted with a 658 x 658 pixel box, scaled to 396 x 396 then windowed to 360 x 360 to give the same 1.703 Å.pix⁻¹ pixel size. Multiple rounds of 3D

classification without alignment then allowed isolation of particles with stronger density for U1 and U2 snRNP. We applied a 20 – 30 Å resolution limit during classification to focus on these large-scale structural differences. After merging all datasets and selection of 117,284 particles with strong U2 or U1 snRNP density, we obtained a reconstruction of pre-B complex at 4.03 Å resolution as calculated using a mask encompassing all of tri-snRNP, U1 snRNP, and U2 snRNP. To visualise the “tri-snRNP core” of this complex at higher resolution, we reverted to the equivalent high-quality (polished and CTF-refined) tri-snRNP particles from datasets 2, 3, and 4 and refined this 86,146 particle subset to 3.28 Å resolution. Presence of a peptide corresponding to SF3A1 within this tri-snRNP, which was not visible in the overall tri-snRNP reconstruction, confirms that this subset of tri-snRNP particles is *bona fide* pre-B complex. To improve U1 and U2 snRNPs further, we performed another round of 3D classification without alignment on the datasets 1 – 5 pre-B particles. This produced one class of 18,974 particles with stably bound U1 snRNP, corresponding to 2.3% of our original good particles, or 16% of *bona fide* pre-B. These refined to 6.07 Å resolution (determined with a soft mask around the whole complex), and after subtracting signal for the rest of pre-B, the U1 snRNP component refined to approximately 28 Å resolution, sufficient for unambiguous docking of the U1 snRNP crystal structure (35,36). The same classification also identified 23,113 particles where U2 snRNP was more stably bound to tri-snRNP due to a contact between SF3B3 and Prp4 kinase. Focused refinement with a mask around SF3b produced a reconstruction of SF3b and U6 LSm ring at approximately 27 Å resolution, allowing unambiguous docking of the U6 LSm and SF3b, and approximate positioning of SF3a and the U2 Sm core into adjacent density. To identify the weak density, we eventually assigned as Prp4 kinase near the Prp6 helical repeats, we performed multiple rounds of focused 3D classification without alignment using a spherical mask around the density. For the first round we used a regularization parameter (T) of 10 and a resolution limit of 10 Å. For the next round we removed the resolution limit and tried T values between 10 and 100. Only with T = 100 could we reliably classify out particles with stably bound Prp4 kinase. The region was too small and particle numbers too limiting to allow focused refinement, so to improve the density further we performed a third round of classification with T = 200, producing a class with 8293 particles with smoothly defined density and secondary structure consistent with approximately 8 Å resolution.

Model Building and refinement

Model building was carried out in COOT (37). A list of protein and RNA components included in the model is given in Table S2. We started by building the tri-snRNP structure into a composite map from the multi-body refinement. Existing crystal or cryo-EM structures of human spliceosomal proteins including Prp8, Snu114, U5-40K (38), Brr2 helicase domain (PDB 4F91), Dim1 (PDB 1QGV), U4 snRNP Sm domain (PDB 4WZJ), Prp31 in complex with U4 snRNA stem loop (PDB 2OZB), and Prp28 RecA1 domain (PDB 4NHO) were docked into density and manually adjusted and extended. The Sm ring from the U4 Sm domain crystal structure was also docked into density for the U5 Sm and adjusted. Density for the U5 and U6 snRNAs, and the parts of U4 snRNA not accounted for by the above structures, were built *de novo* by first fitting idealised A-form helices, adjusting the fit, then linking the strands. This included U4/U6 “stem III,” which was not identified in previous structures. Loop 1 of U5 snRNA was disordered, probably due to the lack of a docked 5'exon, so was left unmodeled.

Homology models for Prp3, Prp4, and Snu13 were generated with SWISS-MODEL (<https://swissmodel.expasy.org>) using structures from the yeast B-complex spliceosome (PDB 5NRL) as templates, and were docked into the map. Homology models were also made with SWISS-MODEL for Sad1 and the N-terminal plug and PWI domain of Brr2 using crystal

structures of their fungal homologs (PDB 4MSX, 4F91 and 4RVQ) as templates. The characteristic fold of an RNA-recognition motif (RRM) domain was visible between U4/U6 stems II and III binding a single-strand segment of U4 snRNA in a canonical mode (39). Since this density was also visible in preparations of tri-snRNP alone (10) we looked for candidate tri-snRNP proteins with RRM domains. We made a homology model of the only clear candidate, RBM42, using Phyre2 (40) with U1-70K (PDB 5UZ5) as a template, and the side-chains in the resultant model were well supported by the density. All homology models were then manually adjusted to fit the density.

Several proteins or domains were built entirely *de novo*. The remainder of the Brr2 N-terminus linking the plug and PWI domains was observed to meander around most of the tri-snRNP, and was assigned on the basis of density and secondary structure prediction (41). The structure of Prp6 diverges significantly from its ortholog in the yeast tri-snRNP (8) so was built entirely *de novo*. High-resolution density around U5 snRNA loop 1 was assigned to the N-terminus of Prp28 based on the fit of sidechains to density and its connectivity to the lower-resolution density for the Prp28 helicase domain. Sporadic high-resolution densities around RBM42 and the U4 Sm domain were assigned to Snu66 and built *de novo* with the aid of secondary structure predictions and some conservation of contacts in the yeast and human B-complex structures (PDB 5NRL, 6AH0, 5O9Z). High resolution density connecting U4/U6 stem III and U4 Sm ring could be assigned to the C-terminus of tri-snRNP protein 27K and was built *de novo*.

Some proteins in the tri-snRNP are not well resolved, including U5-40K, the LSm ring, and the RecA 2 domain of Prp28. 3D classification without alignment in RELION could produce improved densities for these regions. For U5-40K classification was followed by focused refinement of the entire tri-snRNP foot, producing a map that allowed orientation of the U5-40K WD40 domain by the position of the N-terminal helix of Prp8 bound on one face. This model was docked from human P complex (PDB 6DQV). A homology model for the LSm ring was prepared using SWISS-MODEL based on the yeast structure (PDB 4M7A), and although not precisely determined, we chose an orientation that maximised the cross-correlation to the density in Chimera. The second RecA lobe of Prp28 is very flexible in tri-snRNP and so has very weak density. In the previous medium-resolution study of the human tri-snRNP (10), the RecA2 lobe was positioned using crosslinks. Using the same crosslinking data, we chose a similar position, but moved it further from Prp8 into the weak density we could observe.

Coordinates for the tri-snRNP were refined in real space using phenix.real_space_refine in PHENIX (42) into a composite map from multi-body refinement, applying secondary structure, rotamer, base-pairing, base-stacking, ligand, and metal coordination restraints. Restraints for inositol hexaphosphate were generated with JLigand (43). For the 2,2,7-trimethyl guanosine cap of U4 snRNA, restraints for the methyl groups were generated using phenix.elbow and added manually to existing restraints for GDP from the CCP4 monomer library. A data link in PHENIX was used to connect the cap to residue 1 of U4 snRNA, applying tetrahedral restraints to the bridging phosphate. RBM42 was additionally restrained using U1A (PDB 1URN) as a reference model. The peripheral domains U5-40K, Prp28 RecA domains, and the LSm ring were not refined. The final tri-snRNP model contains 35 proteins and 3 snRNAs.

To build a model for pre-B complex, we docked our tri-snRNP model into the tri-snRNP domain of our pre-B map. To check for any differences in the tri-snRNP domain, we calculated the difference in density between all of our particles, and the particles corresponding to pure pre-B complex. While most of the density was identical, we noticed a well-resolved alpha helix between Dim1 and the Prp8 large domain that connected to lower resolution density extending over the U4

snRNA stem-loop and Prp4. This density is found in pre-B particles but not tri-snRNP particles. Since this density was also visible but unassigned in human B complex (PDB 5O9Z, 6AH0) we looked at previously obtained cross-linking and mass-spectrometry data for human B complex (19). We noticed SF3A1 forms a pattern of crosslinks extending C-terminally from the U2 snRNP first to Prp4, then Prp31, then Prp8 and Prp6, then finally to U5-40K, with these crosslinks exactly matching the path of this unassigned density. Moreover, SF3A1 also crosslinks with the U4 snRNA (44). On the basis of these crosslinks, secondary structure prediction, and limited high-resolution density, we assigned one helix of SF3A1 unambiguously, and modelled the remainder of this C-terminal extension using idealised poly-alanine helices. Another difference was movement of the Prp28 RecA2 domain, which was mobile and distant from the RecA1 domain in the tri-snRNP. In our pre-B density we found that this domain was close to the RecA1 domain, although still similarly mobile. We moved the RecA2 domain closer, but stress that its position is approximate. We also noticed that the LSm ring is moved in pre-B relative to tri-snRNP. Finally, we noticed a pre-B specific low resolution density present between the Prp6 helical repeats, the U4/U6 duplex and Prp8 RNaseH domain, which after classification we could improve the density for enough to allow assignment as the core domain of Prp4 kinase.

The remainder of the pre-B density is accounted for by the U1 and U2 snRNPs, and the relocalized LSm ring. We prepared a composite U1 snRNP model based on existing crystal structures and docked it into low-resolution density near Prp28. Briefly, the low-resolution crystal structure of the entire U1 snRNP complex 3CW1 (35) was used as a reference to align the high-resolution crystal structures of the U1 Sm ring, 5'SS/U1 duplex, U1C U1-70K (fragment 2-61) (PDB 4PJO), U1-70K RRM (fragment 62-202) (4PKD) and U1-A RRM (4PKD) (36). The 5'SS was then mutated to match the AdML pre-mRNA.

To achieve a reasonable composite model for the human U2 snRNP, we made a homology model of SF3a from the crystal structure of yeast SF3a (PDB 4DGW), extracted coordinates for the U2 snRNA 3' region, Sm ring, U2A', and U2B'' components from the cryo-EM structure of human P complex (38) and aligned both to U2 snRNP in yeast B complex, giving a model for the U2 snRNP 3'-domain. For the U2 snRNP 5'-domain we extracted coordinates for SF3b, U2 snRNA 5' region, and the pre-mRNA branchpoint sequence from the cryo-EM structure of human B^{act} complex (PDB 6FF4) and performed minor rebuilding into EMDB map 4255. The pre-mRNA sequence was mutated to match AdML pre-mRNA. For the second beta-propeller of SF3B3 we rigid-body docked the corresponding region from crystal structures (PDB 5IFE). For residues 393 – 462 of SF3A3 we docked a homology model built with SWISS-MODEL based on yeast B complex (PDB 5NRL) and adjusted to the density. The 5'-domain was then re-refined in real space using phenix.real_space_refine with base pairing and metal coordination restraints into EMDB map 4255. A homology model for the SF3B4-SF3B2 complex was made with SWISS-MODEL based on the yeast Hsh49-Cus1 crystal structure (PDB 5LSL) and docked into a peripheral density at the analogous position to that assigned in yeast B complex (PDB 5NRL). The model for the U2 snRNP 5' domain was docked into a focus-refined map for SF3B. Our SF3B density was not consistent with the presence as in B^{act} of the SF3B6 and SRRM1 proteins, which were therefore removed from the model. The U2 3' domain is highly mobile and adopts different mobile positions relative to the 5' domain in B complex (PDB 5NRL; 6AHD) and B^{act} (EMD-4255). We observed weak density for the 3' domain near the 5' domain, but not with sufficient signal to unambiguously orient the domain. We therefore tried aligning to the relative orientation observed in B or B^{act} complexes and found that the B-complex orientation produced the best fit to the density. In yeast B complex, Hsh155 (orthologous to SF3B1) contacts the LSm ring next to U2/U6 helix II (PDB 5NRL). We

docked this interface into our SF3B density and saw that density next to SF3B1 could indeed be accounted for by the LSm ring and U2/U6 helix II, and so docked structures of the human orthologs into this density. Finally, we shifted the entire U2 snRNP/LSm model back into density for an overall pre-B map, so that its final position would approximate its average position in pre-B complex. The final pre-B complex model contains 65 proteins, 5 snRNAs, and the pre-mRNA. Figures were generated with PyMOL (<http://www.pymol.org>), UCSF Chimera (45) and ChimeraX.

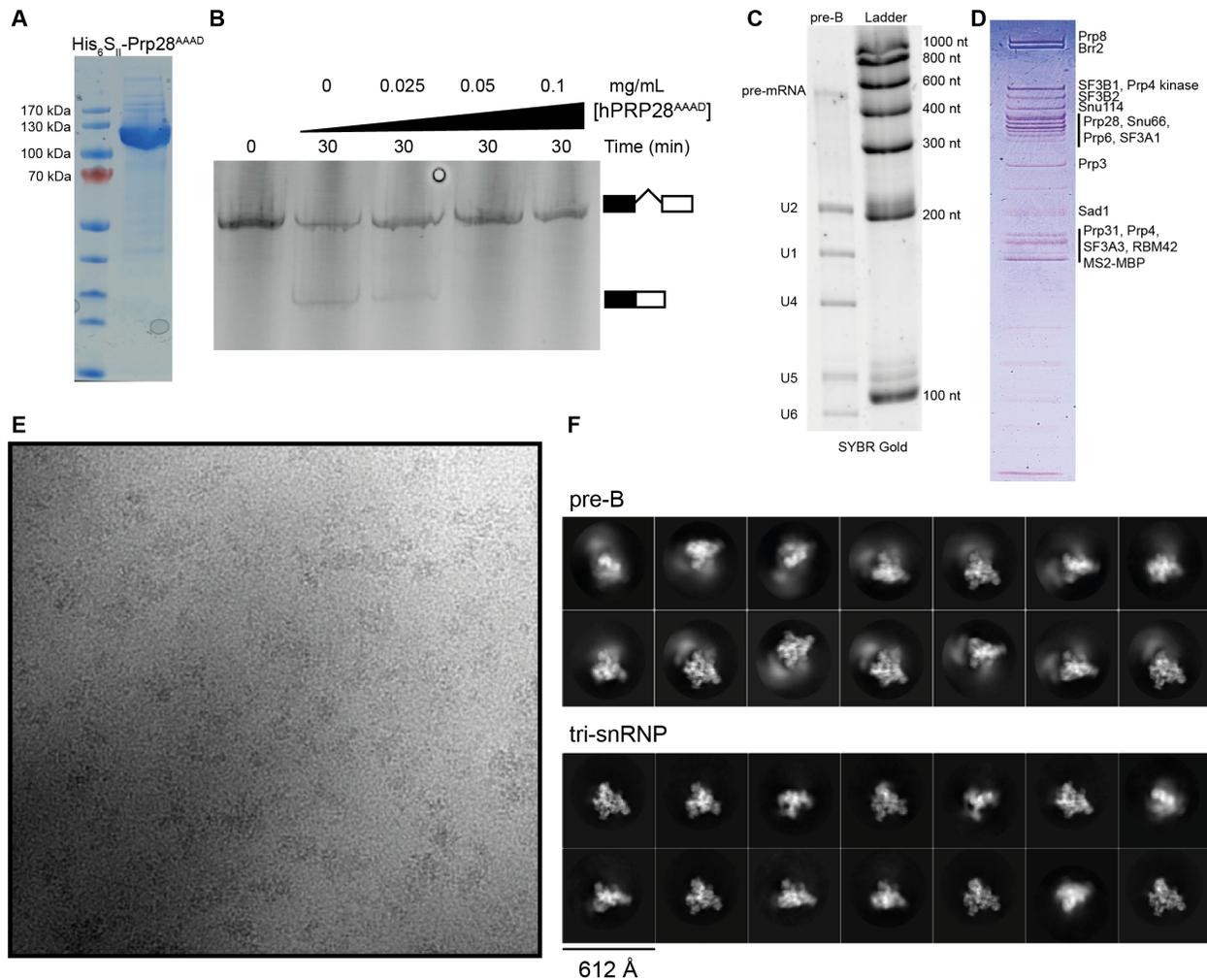


Fig. S1. Pre-B complex and tri-snRNP purification and preliminary characterization

(A) Purified His₆-S_{II}-Prp28^{AAAD} used to stall the spliceosomes at the pre-B stage. (B) Small-scale splicing assay showing that the addition of Prp28^{AAAD} impairs splicing. (C) SYBR gold-stained 14% Urea-PAGE showing the RNA content of the purified sample. (D) Coomassie-stained 4 – 12 % Bis-Tris gel showing the protein content of the purified sample. (E) Representative micrograph. (F) 2D classes for the pre-B (top) and tri-snRNP (bottom) complexes calculated from the same micrographs, illustrating the compositional heterogeneity of the sample.

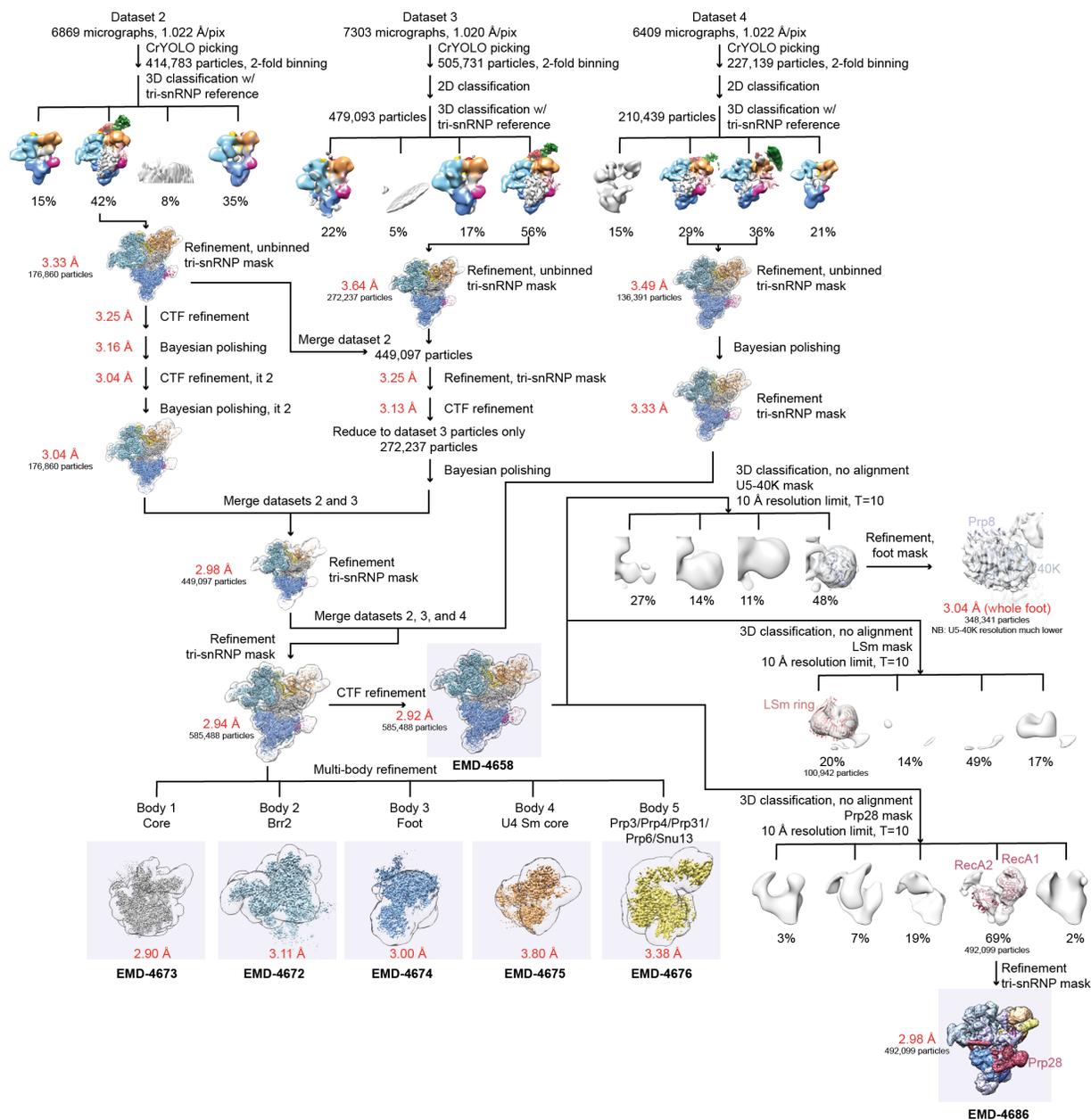


Fig. S2. Cryo-EM data processing for the tri-snRNP. All processing steps were carried out in RELION 3.0. Deposited densities are depicted on a light blue background. The masks used for postprocessing to calculate gold-standard resolutions (red text) are shown in outline.

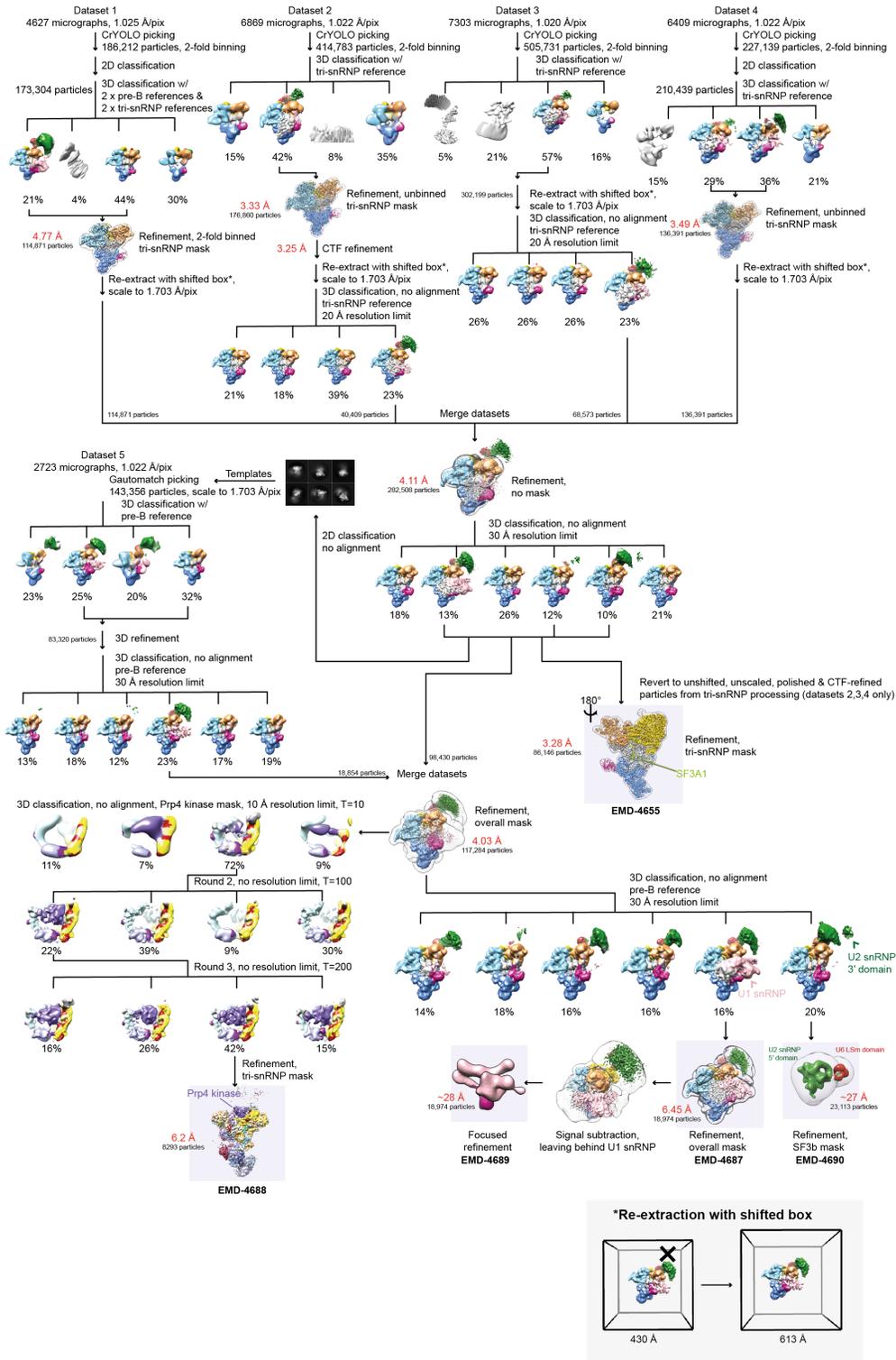


Fig. S3. Cryo-EM data processing for the pre-B complex. All processing steps were carried out in RELION 3.0 (32). Deposited densities are depicted on a light blue background. The masks used for postprocessing to calculate gold-standard resolutions (red text) are shown in outline. Inset: re-extraction with a shifted box allows reconstruction of U2 snRNP without loss of signal.

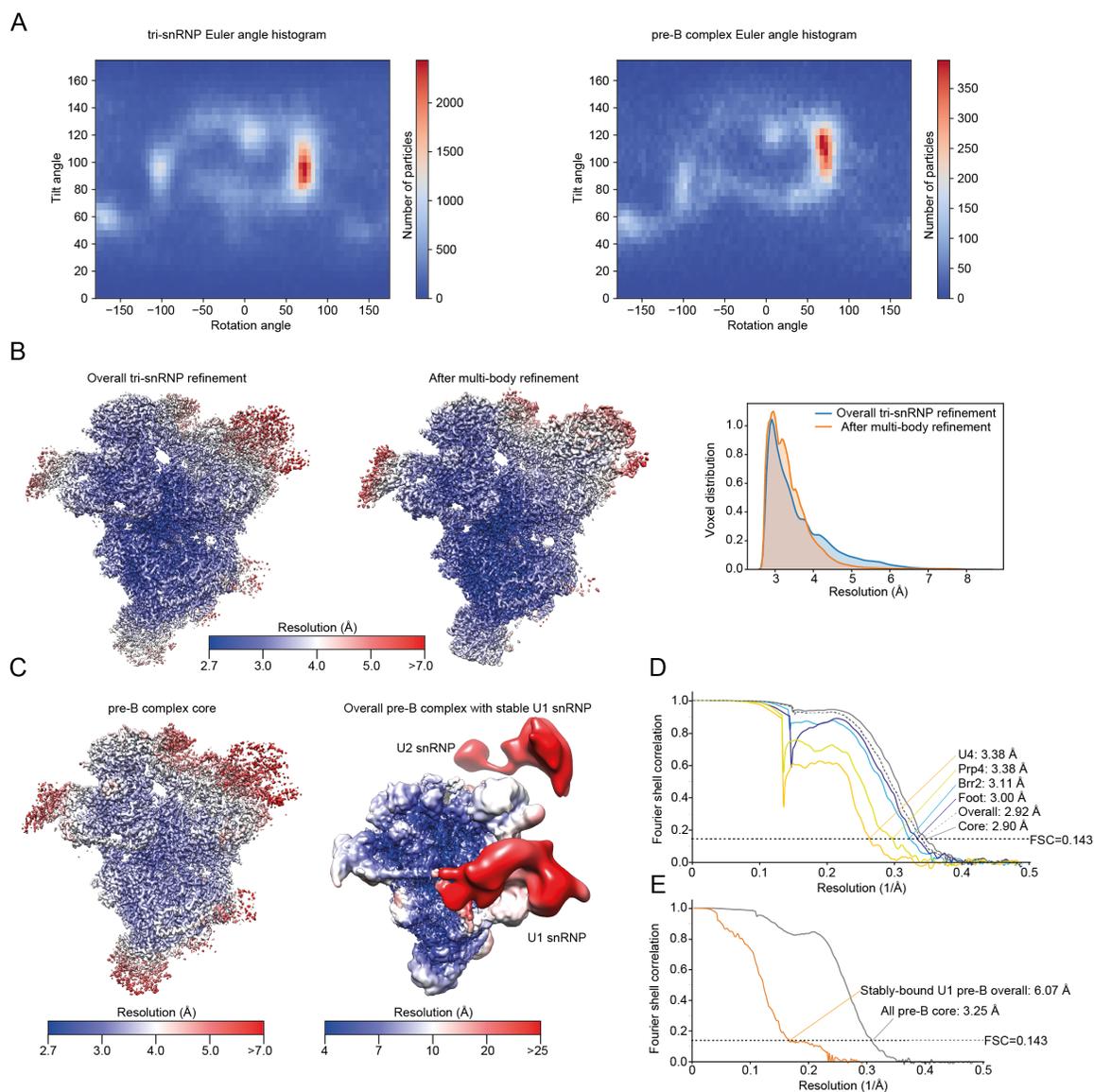


Fig. S4. Assessment of the cryo-EM reconstructions. (A) Euler angle distributions for the overall tri-snRNP reconstruction (left) and core pre-B complex reconstruction (right). (B) Local resolution for tri-snRNP after refinement with a soft mask over the whole complex (left) and after multi-body refinement (mid). For depiction of the multi-body refinement, at each voxel the body with the highest calculated local resolution is displayed. The local resolution distribution for voxels inside the tri-snRNP is depicted on the right, showing that after multi-body refinement the tri-snRNP has a more homogeneous, higher resolution. (C) Local resolution for the pre-B complex refined with a mask around the tri-snRNP (left) and the pre-B complex class with stable U1 snRNP (right). Local resolution in (B) and (C) is calculated using RELION 3.0 (32). (D) Fourier-shell correlation (FSC) curves for overall tri-snRNP (dashed line grey) and each body (solid lines; U4: orange, Prp4: yellow, Brr2: light blue, foot: dark blue, core: grey) defined for the multi-body refinement. The FSC=0.143 criterion, used to estimate the resolution, is depicted on the chart. (E) FSC curves for pre-B core and overall stably-bound U1 pre-B complex subclass.

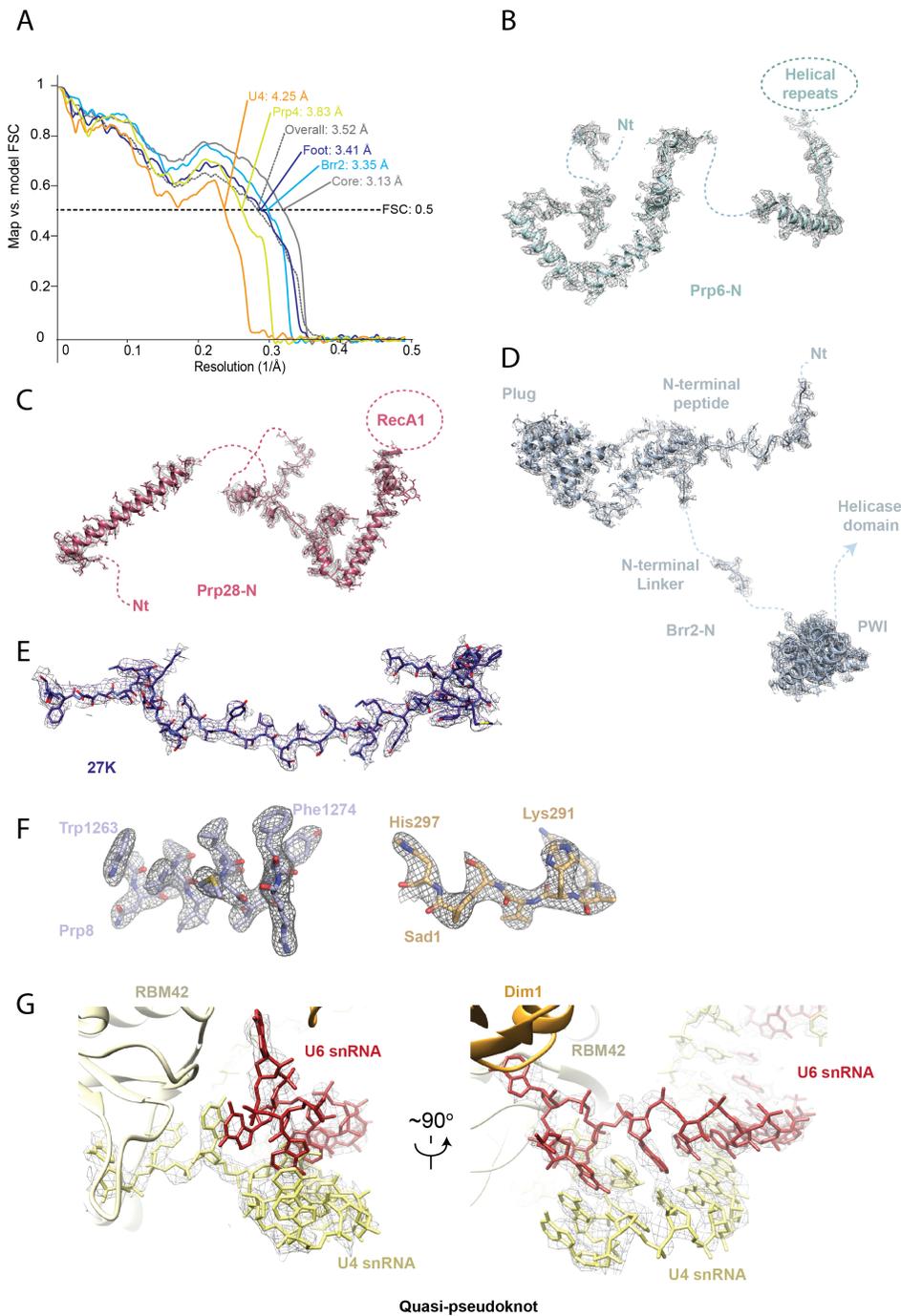


Fig. S5. Quality of the tri-snRNP model. (A) Map vs. model FSC for overall tri-snRNP (dashed grey line) and each body (solid lines; U4: orange, Prp4: yellow, Brr2: light blue, foot: dark blue, core: grey) defined for the multi-body refinement. (B) Density for Prp6 N-terminal domain. (C) Density for Prp28 N-terminal domain. (D) Density for Brr2 N-terminal. (E) Density for 27K C-terminus. (F) Density for Prp8 and Sad1. (G) Density for U4/U6 quasi-pseudoknot in two orientations.

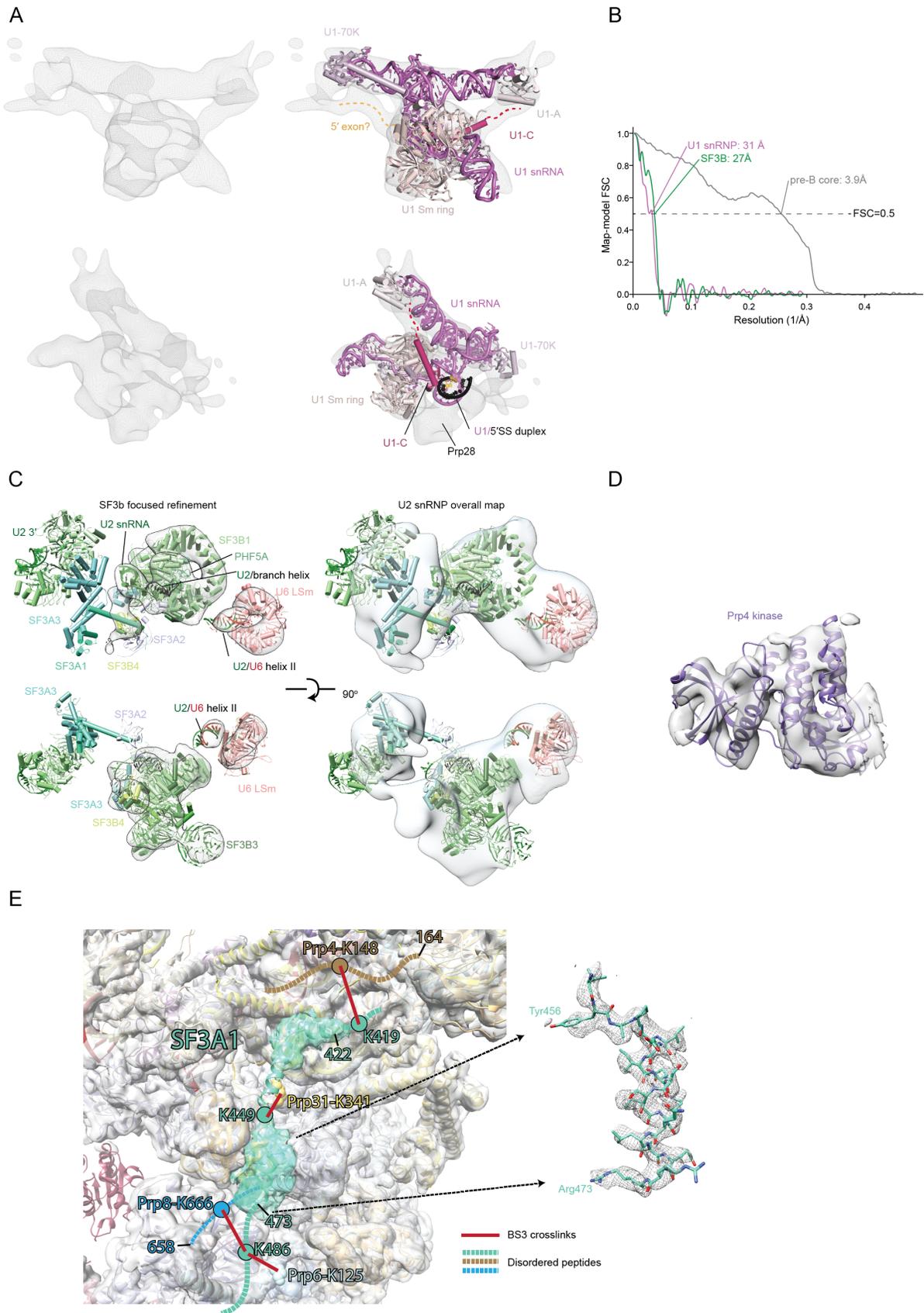


Fig. S6. Quality of the pre-B model. (A) Density for the U1 snRNP and fit of the model into the density. The atomic model used for the U1 snRNP was made based on crystal structures of the whole U1 snRNP (34) and subcomplexes (35). (B) Map vs. model FSC for pre-B core (grey line), U1 snRNP (pink line) and SF3B (green line). (C) Density for the U2-snRNP and fit of the model into the density after focused refinement on SF3B (left) and the overall pre-B map (right). (D) Density for Prp4 kinase in the pre-B complex after three rounds of focused classification. (E) Building of the C-terminal region of SF3A1 bound to core pre-B. Crosslinks observed in human B-complex (19) are depicted as red solid lines, disordered peptides as dashed lines (Prp8: blue, SF3A1: blue/green, Prp31: yellow, Prp4: brown). Low-threshold density corresponding to SF3A1 is displayed in blue/green, density corresponding to pre-B core is displayed in grey. High-resolution density corresponding to SF3A1 helix 458-473, as observed in our core pre-B map, assessing the correct attribution of this density to SF3A1.

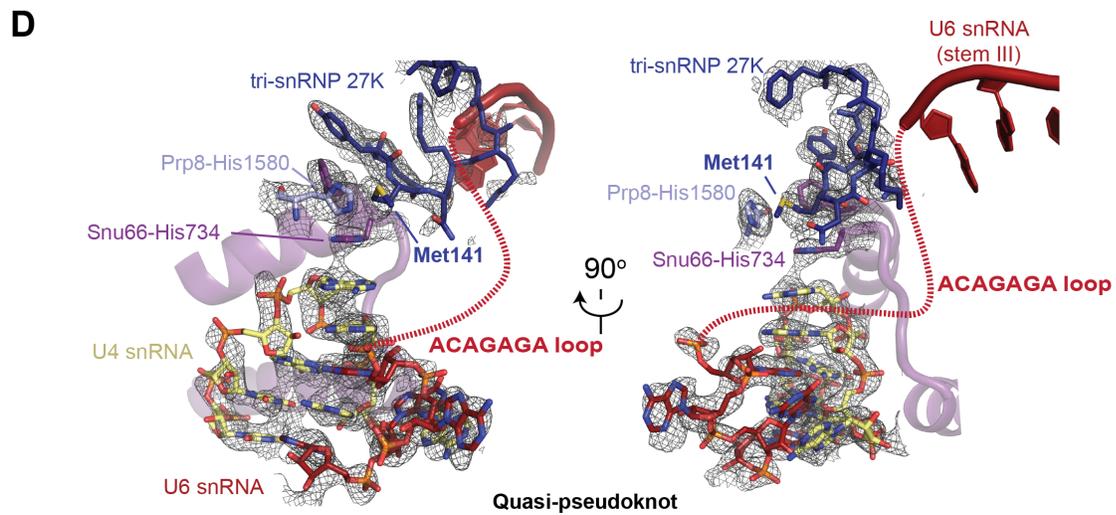
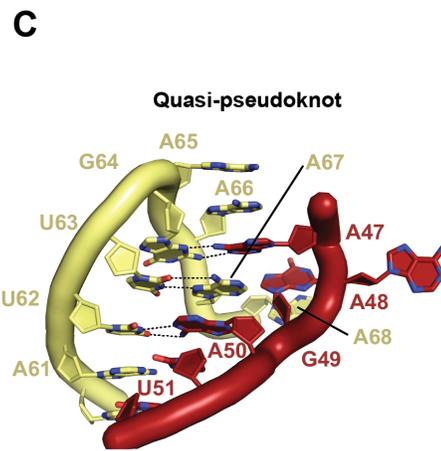
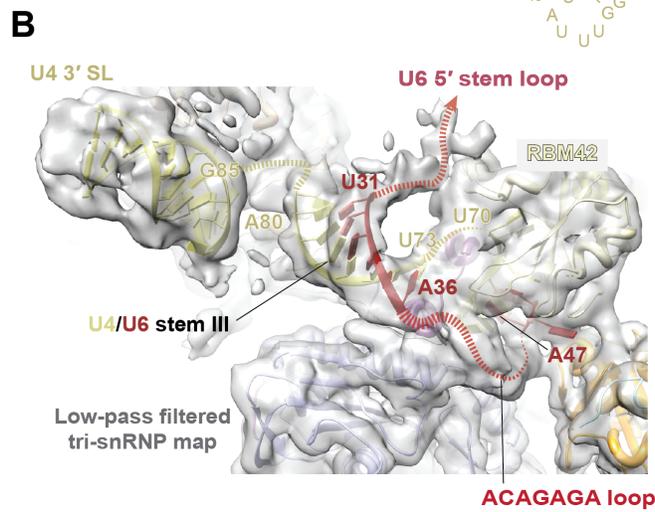
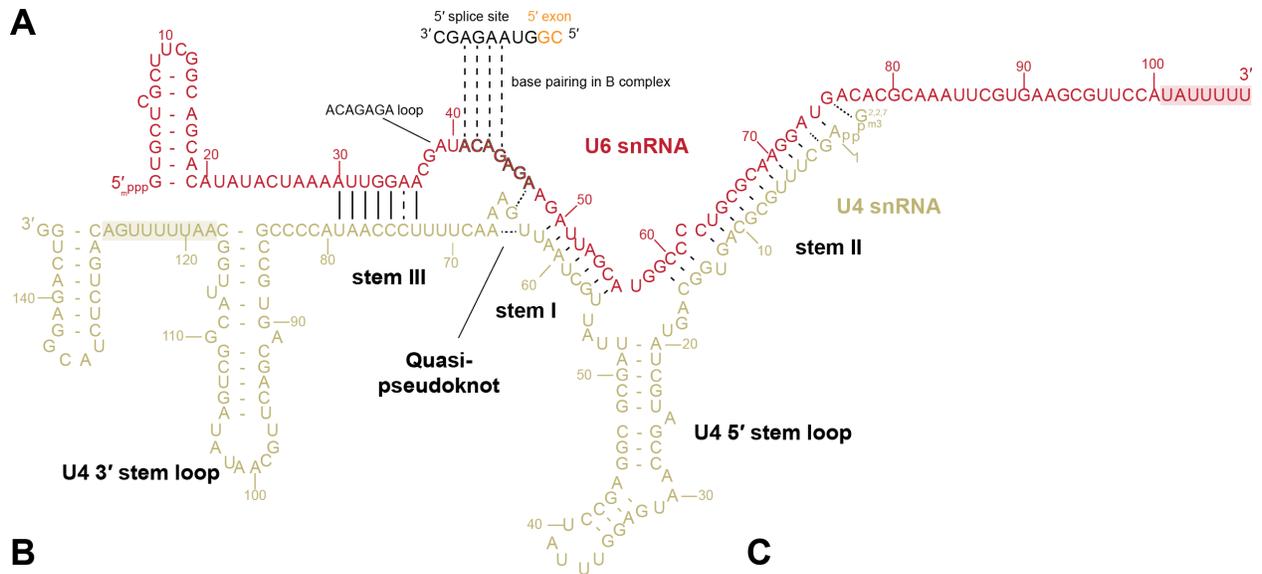


Fig. S7. Organization of the U4 region in tri-snRNP and pre-B. (A) Organization of the U4/U6 duplex in tri-snRNP and pre-B. Note the ACAGAGA box, which pairs with the first nucleotides of the intron at the B-complex stage, is looped out between U4/U6 quasi-pseudoknot that cap stem I and the newly discovered stem III. (B) Density corresponding to U4/U6 stem III in tri-snRNP and pre-B. Note the clear connectivity to the U4 snRNA and U6 snRNA in the low-passed filtered map. This density was incorrectly attributed to the U6 5' SL in the low-resolution structure of pre-B (13). The formation of the U4/U6 stem III is also supported by conservation (18). (C) Organization of the quasi-pseudoknot. (D) Density for the interactions between tri-snRNP 27K, Snu66, Prp8 and the U4/U6 quasi-pseudoknot. Tri-snRNP 27K methionine 141 (labelled) corresponds to the M141T dominant allele in *C. elegans* SNRNP-27k (20). In worms, M141 mutation to threonine or alanine activates cryptic 5' splice sites that lack complementarity to the U6 ACAGAGA box, while other mutations (e.g. to glycine or phenylalanine) are homozygous lethal. In the human tri-snRNP, Met141 packs against His1580 of Prp8 and His734 of Snu66, which in turn stacks on the quasi-pseudoknot. This packing may stabilize positioning of the 27K C-terminal loop, which in turn buttresses the ACAGAGA loop as it projects off U4/U6 stem III.

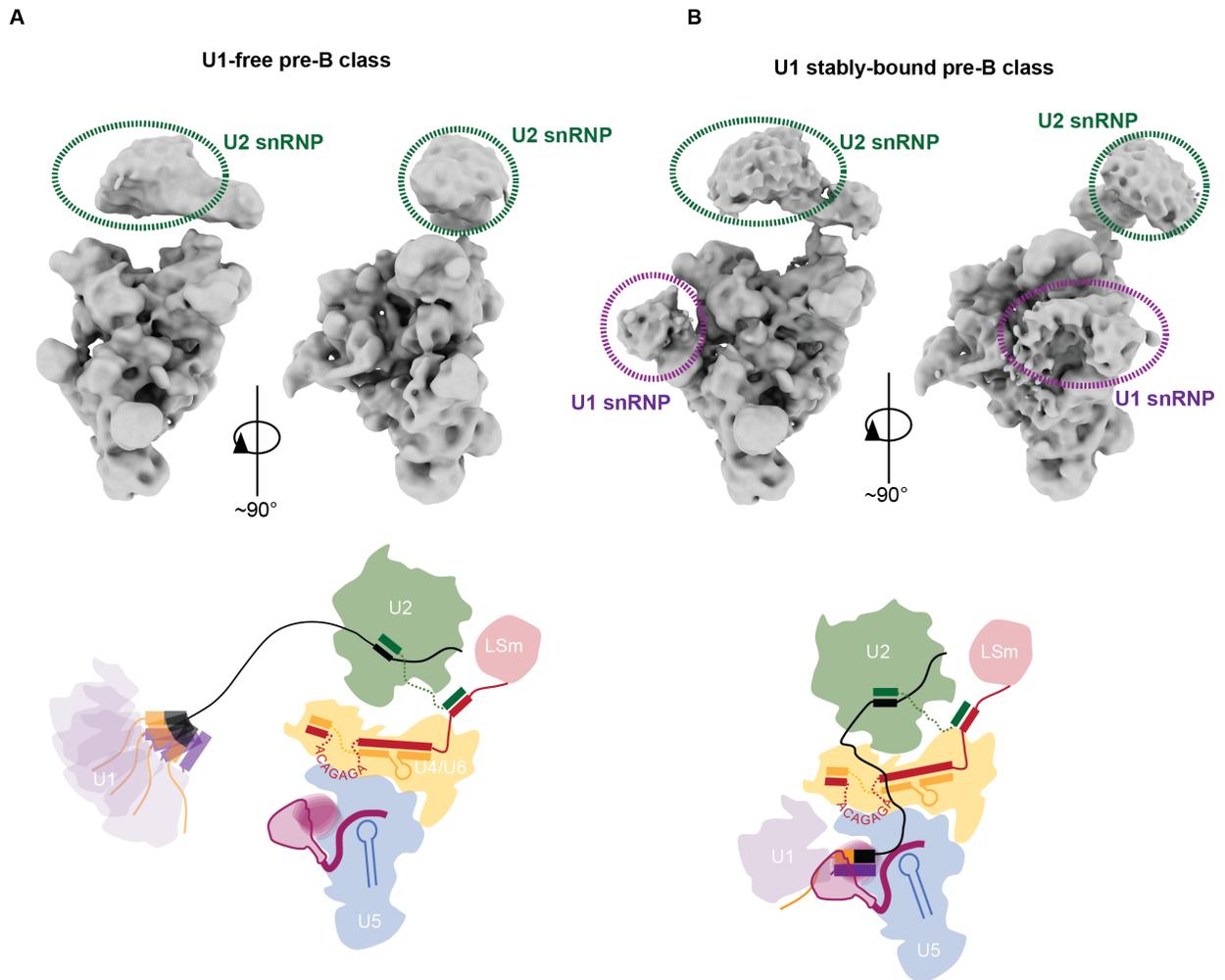


Fig. S8. Dynamics of the pre-B complex. (A) U1-free and (B) U1 stably-bound pre-B complex maps. In pre-B complex the U1 snRNP is only clearly located in a sub-population of particles, indicating that its interactions with tri-snRNP are transient and it is mainly anchored to the complex via the pre-mRNA substrate and U2 snRNP.

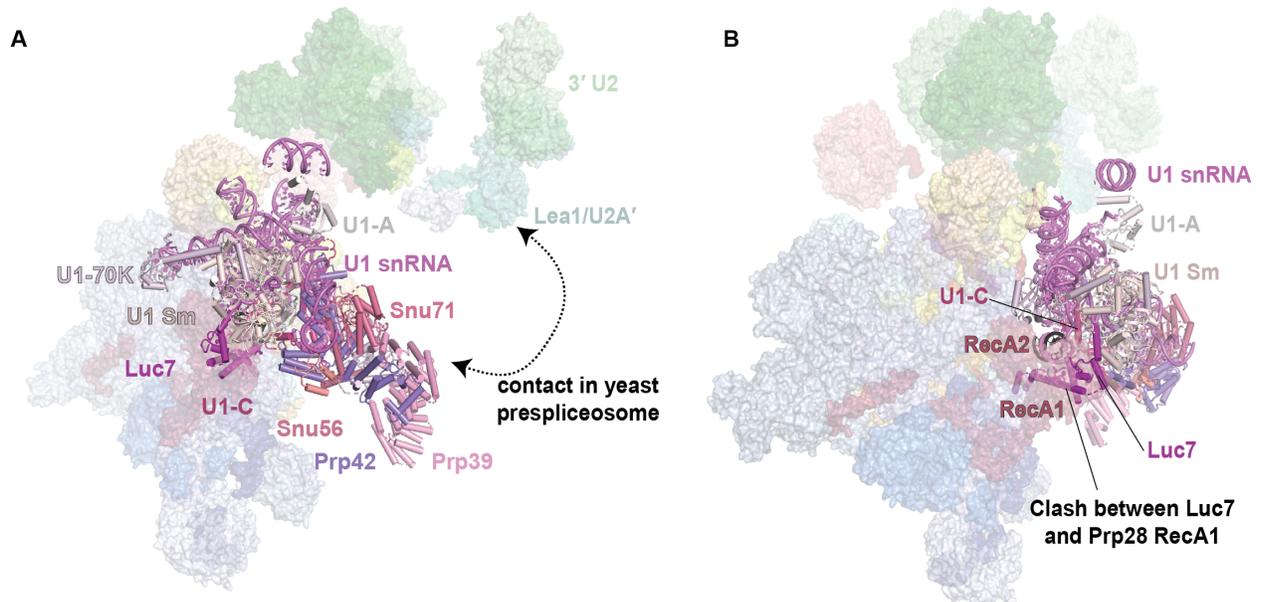
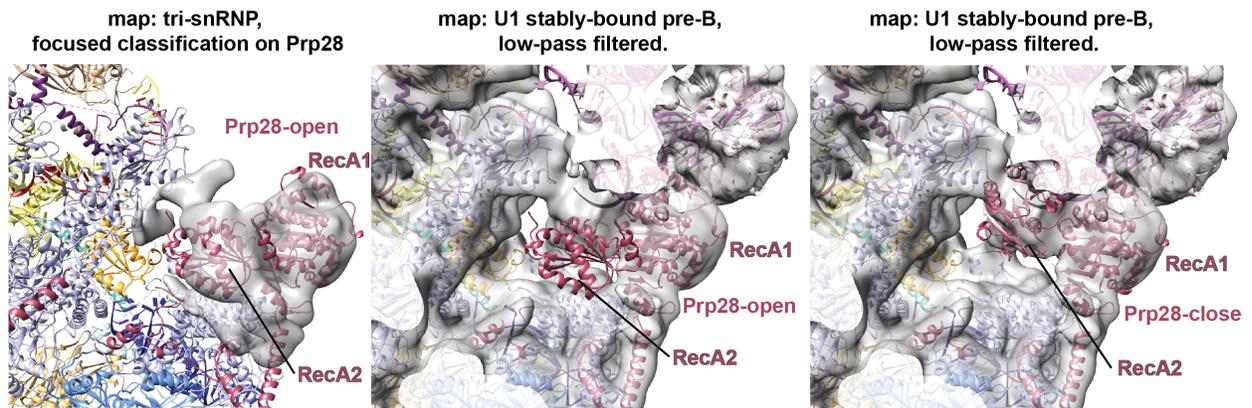


Fig. S9. Overlay of yeast U1 snRNP, as observed in yeast prespliceosome, and human U1 snRNP, as observed in human pre-B complex. Yeast U1 snRNP (represented as cartoon) (PDB 6G90) Sm ring was aligned onto human Sm ring as observed in our pre-B structure (represented as transparent surfaces). Note that, in human pre-B, the U1 snRNP orientation is compatible with the role of auxiliary factor such as Prp39 in bridging U1 with U2. Note also, Luc7 which contacts the phosphate backbone of the 5'SS as observed in yeast A complex clashes extensively with Prp28 RecA1 suggesting that Luc7 dissociates upon U1 snRNP docking.

A



B

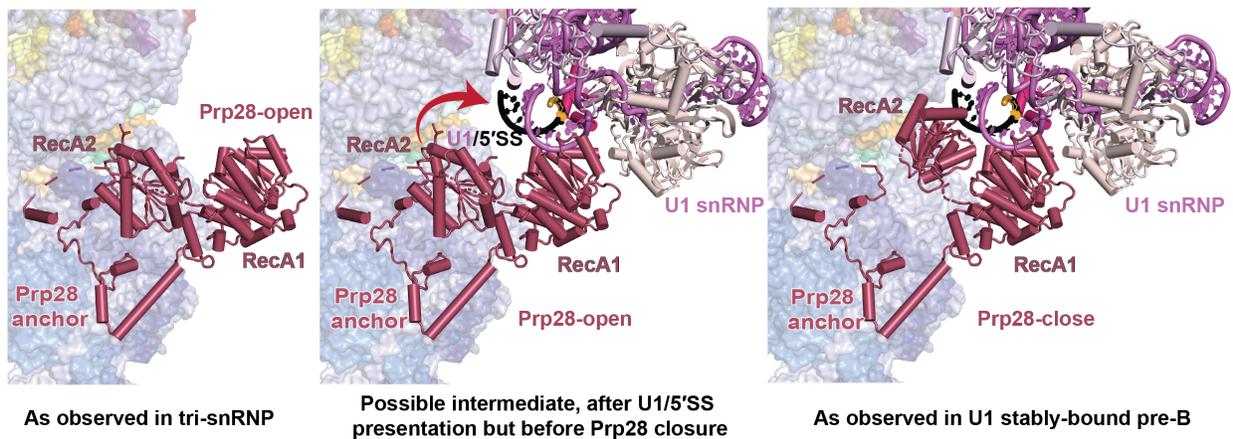


Fig. S10. Position of Prp28 RecA2 in the presence or absence of U1. (A) Comparison of tri-snRNP (focused classification on Prp28) map (left panel) and stably-bound U1 pre-B map low pass filtered (middle and right panels). Compare the average positions of the RecA2 domain of Prp28 in the absence (left panel) or in presence of the U1/5'SS duplex (right panel). (B) Model for Prp28 closure during 5'SS delivery. Note that the middle panel represent an intermediate that is not observed but is extrapolated from the conformation of Prp28 in the absence of U1 and the position of U1 in pre-B.

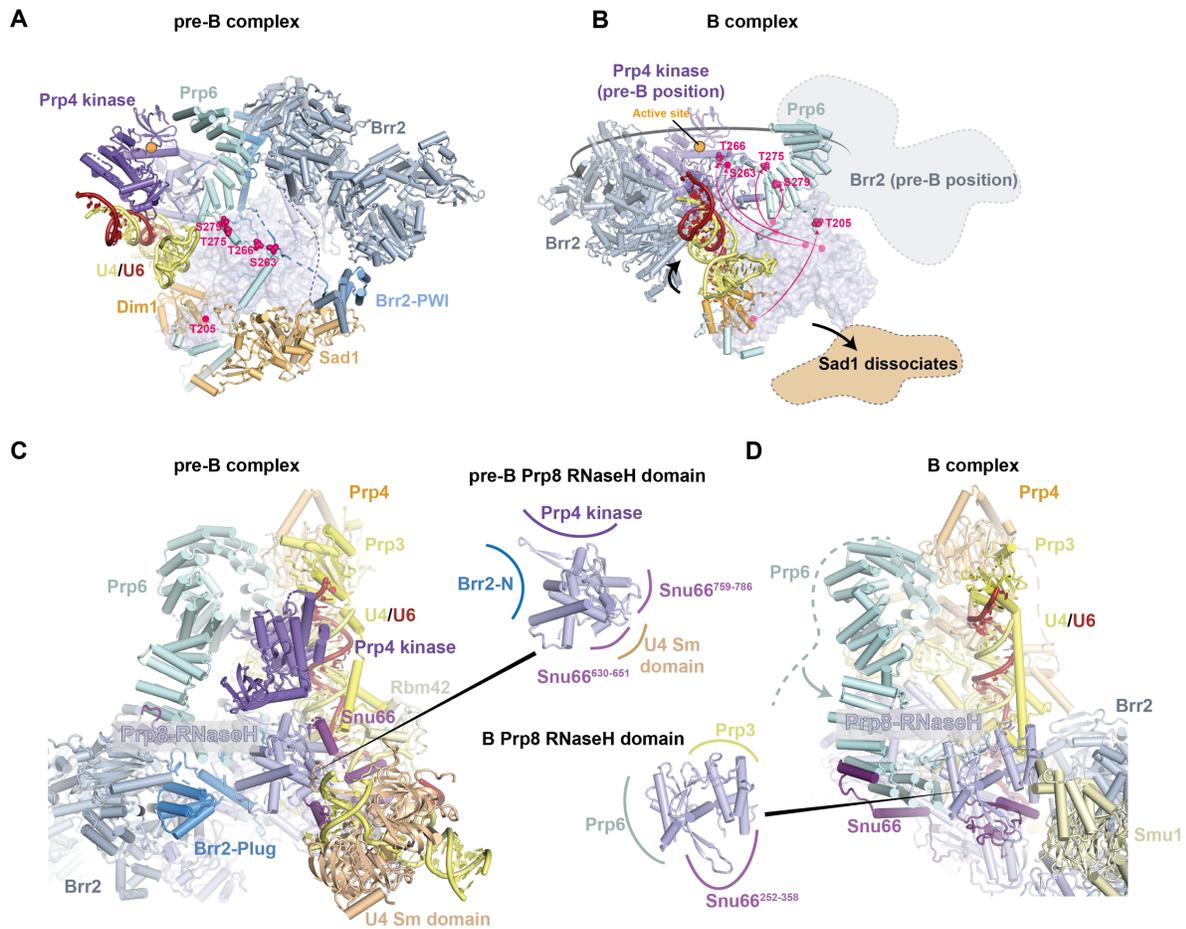


Fig. S11. Prp4 kinase in the pre-B to B transition. The same regions of pre-B (A and C) and B-complex (B and D) are aligned on Prp8 Large domain. (A) Pre-B stage. Prp4 kinase active site is represented as an orange circle. Prp4 kinase phosphorylation targets on Prp6 N-terminus are depicted as magenta spheres. (B) B-complex stage. Prp4 kinase, as observed in pre-B, is overlaid on B-complex, its active site is represented as an orange circle. Prp4 kinase phosphorylation targets on Prp6 N-terminus are depicted as magenta spheres. (C) Pre-B, top view and summary of Prp8 RNaseH domain interactions at the pre-B stage. (D) B-complex, top view and summary of Prp8 RNaseH domain interactions at the B-complex stage.

Table S1. Refinement statistics

EM data collection				
Microscope model		FEI Titan Krios		
Detector model		Gatan K2 Summit		
Number of datasets		5		
Number of micrographs collected		27931		
Frames per micrograph		40		
Data collection software		EPU (Thermo Fisher)		
Magnification		130000x		
Voltage (kV)		300		
Electron fluence ($e^- \text{Å}^{-2}$)		44–57		
Pixel size (Å)		1.022		
Defocus range (μm)		1–3.5		
Energy filter slit width (eV)		20		
Symmetry imposed		none		
3D reconstruction	U5.U4/U6 tri-snRNP	Pre-B spliceosome core	Pre-B, SF3B focused refinement	Pre-B, U1 snRNP focused refinement
EMDB map entry	EMD-4658	EMD-4665	EMD-4690	EMD-4689
PDB coordinate entry	6QW6	6QX9	6QX9	6QX9
Initial particle number	1,147,653	1,477,221	1,477,221	1,477,221
Final refined particle number	585,488	86,146	23,113	18,974
Resolution (Å)				
-Masked, at FSC = 0.143	2.92	3.25	28	28
-Masked, at FSC = 0.5	3.38	3.90	28	28
-Unmasked, at FSC = 0.143	3.38	7.53	19	20
-Unmasked, at FSC = 0.5	3.94	4.01	26	26
Local resolution (median and 90 th percentile range within solvent mask)	3.34 Å, 2.80–5.44 Å	3.77 Å, 3.06–6.53 Å	Not determined	Not determined
Map sharpening B-factor (Å ²)	Composite map after multibody refine: 3.23 Å, 2.77–4.43 Å -56	-43	Not sharpened	Not sharpened
Real space refinement and model validation statistics				
Model resolution (Å) (at FSC = 0.5)	3.5 (all atoms) 3.1 (multibody Core) 3.4 (multibody Brt2) 3.4 (multibody Foot) 4.2 (multibody U4) 3.8 (multibody Prp4)	3.9	27	31
Model composition				
Non-hydrogen atoms (# refined)	89,018 (81,105)	137,390 (789)	38,999 (21,987)	137,390 (0)
Protein residues (# refined)	10,339 (9355)	15,577 (97)	4532 (2670)	15,577 (0)
RNA bases (# refined)	272 (267)	567 (0)	128 (49)	567 (0)
Refinement method (PHENIX)	Real-space refinement	Only refined around SF3A1 CTD	Real-space refinement of SF3B into EMD-4255	Not refined
Map CC (around atoms)	0.7638	0.7143	0.8206	
Refined model B factors (Å ²) (median and 90 th percentile range)	All: 56.7, 17.4–151.6 RNA: 109.4, 41.1–246.1	All: 50.8, 35.6–64.2 RNA: -	All: 76.7, 43.1–165.4 RNA: 143.9, 55.3–261.1	
RMS deviations				
Bond lengths (Å)	0.009	0.005	0.010	
Bond angles (°)	1.105	0.624	1.034	
Validation ^a				
Molprobity score	2.02	1.52	1.79	
All-atom clashscore	8.98	5.8	6.37	
Rotamer outliers (%)	1.66	0	0.98	
C-beta deviations	1	0	0	
Ramachandran plot				
Outliers (%)	0.09	0	0.08	
Allowed (%)	5.35	4.30	6.79	
Favoured (%)	94.56	96.70	93.13	
RNA validation				
Correct sugar pickers (%)	97	NA	96	
Good backbone conformations (%)	74	NA	71	
EMRinger score	1.52	2.3	0.85	
CaBLAM outliers (%)	2.7	0	4.2	

^a The model was validated using the MolProbity server (<http://molprobity.biochem.duke.edu>)

Table S2. Modelling

Sub-complex	Protein/RNA	Domains	Total residues	Residues modelled	Model template (PDB ID)	Modelling	Chain ID	<i>S.cerevisiae</i> name	Mass spec EMPAI	Uniprot ID
U5	PRP8	N-terminal	1-658	25-51; 58-73; 84-362; 369-658	6QDV	docked and rebuilt	5A	Prp8	9.5	Q6P2G9
		Large	679-1766	679-1355; 1363-1766	6QDV	docked and rebuilt				
		RNaseH	1767-2019	1767-2019	6QDV	docked and rebuilt				
	BRR2	Jab1/MpN	2067-2335	2067-2335	6QDV	docked and rebuilt	5B	Brr2	10.2	O75643
		N-terminal 1	1-106	36-97	-	de novo				
		Plug	107-180	110-180	5DCA	docked and rebuilt				
		N-terminal linker	181-256	195-223; 238-243	-	de novo				
		PWI	257-391	257-364	4RVQ	homology modelled				
		NC clamp	392-474	405-474	5DCA	docked and rebuilt				
	SNU114	N-terminal helicase	475-1299	475-1299	4F91	docked and rebuilt	5C	Snu114	20.3	Q15029
		C-terminal helicase	1300-2136	1300-2125	4F91	docked and rebuilt				
	U5-40K		357	105-956	6QDV	docked and rebuilt	5D	-	7.5	Q96DI7
	SmB		240	6-50; 61-88	4WZJ	docked and rebuilt	5b	SmB	-	P14678
	SmD1		119	2-82	4WZJ	docked and rebuilt	51	SmD1	-	P62314
	SmD2		118	10-77; 87-116	4WZJ	docked and rebuilt	52	SmD2	-	P62316
SmD3		126	1-84	4WZJ	docked and rebuilt	53	SmD3	-	P62318	
SmF		86	3-75	4WZJ	docked and rebuilt	5f	SmF	-	P62306	
SmE		92	14-90	4WZJ	docked and rebuilt	5e	SmE	-	P62304	
SmG		76	3-76	4WZJ	docked and rebuilt	5g	SmG	-	P62308	
U5 snRNA		117	3-38; 44-78; 85-117	6QDV	docked and rebuilt	5	U5 snRNA	-	-	
PRP6	N-terminal	1-285	18-28; 72-165; 236-285	-	de novo	5J	Prp6	9.4	O94906	
	Helical repeats	286-941	286-533; 536-587; 590-937	-	de novo					
	N-terminal	1-380	158-203; 241-255; 270-380	-	de novo					
DDX23/hPRP28	Helicase domain 1	381-626	381-626	4NHO	rigid body fitted	5X	Prp28	10	Q8BU08	
	Helicase domain 2	627-820	627-694; 703-723; 728-806	4NHO	rigid body fitted					
DIM1		142	2-142	10GV	docked and rebuilt	5D	Dib1	11	P83876	
U4/U6 di-snRNP	PRP31		499	59-74; 89-255; 266-333; 341-390	2OZB	docked and rebuilt	4C	Prp31	2.3	Q8WMY3
	PRP4		522	164-522	5NRL	homology modelled, rebuilt	4B	Prp4	4.8	O43172
	PRP3		683	385-395; 412-427; 434-475; 493-605; 627-682	5NRL	homology modelled, rebuilt	4A	Prp3	5.9	O43395
	SNU13		128	6-128	5NRL	homology modelled, rebuilt	4D	Snu13	3.9	P55769
	SmB		240	8-89	4WZJ	docked and rebuilt	4b	SmB	-	P14678
	SmD1		119	2-82	4WZJ	docked and rebuilt	41	SmD1	-	P62314
	SmD2		118	22-113	4WZJ	docked and rebuilt	42	SmD2	-	P62316
	SmD3		126	3-85	4WZJ	docked and rebuilt	43	SmD3	-	P62318
	SmF		86	4-75	4WZJ	docked and rebuilt	4f	SmF	-	P62306
	SmE		92	14-89	4WZJ	docked and rebuilt	4e	SmE	-	P62304
	SmG		76	3-76	4WZJ	docked and rebuilt	4g	SmG	-	P62308
	LSM2		95	1-95	4M7D	rigid body fitted	62	LSM2	-	Q9Y333
	LSM3		102	14-98	4M7D	rigid body fitted	63	LSM3	-	P62310
	LSM4		139	2-74	4M7D	rigid body fitted	64	LSM4	-	Q9Y420
	LSM5		91	7-88	4M7D	rigid body fitted	65	LSM5	-	Q9Y4Y9
LSM6		80	6-77	4M7D	rigid body fitted	66	LSM6	-	P62312	
LSM7		103	11-87	4M7D	rigid body fitted	67	LSM7	-	Q9UK45	
LSM8		96	2-96	4M7D	rigid body fitted	68	LSM8	-	O95777	
U4 snRNA	5' domain	1-84	1-52; 55-70; 73-80	2OZB	docked and rebuilt, de novo	4	U4 snRNA	-	-	
	3' domain	85-145	85-96; 105-133; 138-145	4WZJ	docked and rebuilt					
U6 snRNA		106	31-36; 47-77; 103-106 (preB 85-95)	-	de novo	6	U6 snRNA	-	-	
tri-snRNP proteins	RBM42		480	369-474	5UZ5	homology modelled, manually adjusted	R	-	3.2	Q9BTD8
	27K		155	110-151	-	de novo	X	-	-	Gel bands
	SNU66		800	550-576; 631-651; 699-759; 766-777; 782-786	-	de novo	S	Snu66	8.9	O43290
	SAD1		565	100-555	4MSX	homology modelled, manually adjusted	U	Sad1	12	Q53G59
U1 snRNP	SmB		240	6-91	4PJO, 3CW1	rigid body fitted	1b	SmB	-	P14678
	SmD1		119	2-82	4PJO, 3CW1	rigid body fitted	11	SmD1	-	P62314
	SmD2		118	11-77; 88-114	4PJO, 3CW1	rigid body fitted	12	SmD2	-	P62316
	SmD3		126	3-83	4PJO, 3CW1	rigid body fitted	13	SmD3	-	P62318
	SmF		86	2-75	4PJO, 3CW1	rigid body fitted	1f	SmF	-	P62306
	SmE		92	14-90	4PJO, 3CW1	rigid body fitted	1e	SmE	-	P62304
	SmG		76	4-76	4PJO, 3CW1	rigid body fitted	1g	SmG	-	P62308
	U1-A		282	6-103	4PKD, 3CW1	rigid body fitted	1A	Mud1	-	P09012
	U1-C		159	2-51	4PJO, 3CW1	rigid body fitted	1C	Snp1	-	P09234
	U1-70K		437	2-202	4PJO, 4PKD, 3CW1	rigid body fitted	1K	Yhc1	5.5	P08621
U1 snRNA		164	1-164	4PJO, 3CW1	rigid body fitted	1	U1 snRNA	-	-	
U2 snRNP	SmB		240	6-87	6QDV	rigid body fitted	2b	SmB	-	P14678
	SmD1		119	1-80	6QDV	rigid body fitted	21	SmD1	-	P62314
	SmD2		118	13-78; 87-115	6QDV	rigid body fitted	22	SmD2	-	P62316
	SmD3		126	2-84	6QDV	rigid body fitted	23	SmD3	-	P62318
	SmF		86	4-75	6QDV	rigid body fitted	2f	SmF	-	P62306
	SmE		92	11-91	6QDV	rigid body fitted	2e	SmE	-	P62304
	SmG		76	4-76	6QDV	rigid body fitted	2g	SmG	-	P62308
	SF3B1		1304	398-415; 457-1304	6FF4	rebuild in EMDB 4225 rigid body fitted	B1	Hsh155	8.4	O75533
	SF3B2		895	458-602; 604-667	6FF4	rebuild in EMDB 4225 rigid body fitted	B2	Cus1	4.4	Q13435
	SF3B3		1217	1-645; 663-691; 695-830; 834-1068; 1078-1217	6FF4, 5IFE	rebuild in EMDB 4225 rigid body fitted	B3	Rse1	7	O15393
	SF3B4		424	12-89	6FF4	rebuild in EMDB 4225 rigid body fitted	B4	Hsh49	2.4	Q15427
	SF3B5		86	12-80	6FF4	rebuild in EMDB 4225 rigid body fitted	B5	Ysf3	12	Q9BWJ5
	PHF5A		110	2-101	6FF4	rebuild in EMDB 4225 rigid body fitted	BP	Rds3	9.4	Q7RTV0
	SF3A1		793	160-282; 422-447 (polyAla); 455-473	4DGW	homology modeled rigid body fitted	A1	Prp21	5	O15459
	SF3A2		464	41-85; 104-126; 133-209	4DGW	homology modeled rigid body fitted	A2	Prp11	2.2	Q15428
SF3A3		501	1-229; 279-362; 393-461	4DGW, 5NRL	homology modeled rigid body fitted	A3	Prp9	8.2	Q12874	
U2A'		255	2-163	6QDV	rigid body fitted	2A	Lea1	3.7	P09861	
U2B'		225	3-94	6QDV	rigid body fitted	2B	Msl1	-	P08579	
U2 snRNA		189	3-13; 32-65; 96-109; 147-184	6FF4, 6QDV	rebuild in EMDB 4225 rigid body fitted	2	U2 snRNA	-	-	
Pre-B specific	Prp4 Kinase		1007	657-1007	4IAN	rigid body fitted	K	-	1.75	Q13523
Substrate	pre-mRNA	branch sequence		88-102	6FF4	rebuild in EMDB 4225 rigid body fitted	I	-	-	-
		5'ss		-2 - 8	4PJO, 3CW1	rigid body fitted	I	-	-	-

Movie S1. Model for 5' splice site transfer during the transition from pre-B to B complex.

Transcription of the narration: “Here is the human tri-snRNP”, “Pre-B complex is formed when U1 and U2 snRNPs bind. U2 in green binds stably while U1 in purple is more transient and docks between the RecA domains of Prp28 helicase”, “Upon binding, RecA2 clamps around U1 snRNA, promoting release of the 5'SS as U1 and Prp28 dissociate”, “The 5'SS anneals to the U6 ACAGAGA-box shown in red”, “This exposes the Brr2 binding site on U4, in yellow”, “A concerted set of conformational rearrangements result in Brr2 binding U4 snRNA to form the B complex spliceosome, primed for catalytic activation.”

Supplementary File S1. PyMOL session for human tri-snRNP (6QW6.pse).

Supplementary File S2. PyMOL session for human pre-B (6QX9.pse).

References and Notes

1. C. Plaschka, A. J. Newman, K. Nagai, Structural basis of nuclear pre-mRNA splicing: Lessons from yeast. *Cold Spring Harb. Perspect. Biol.* 10.1101/cshperspect.a032391 (2019). [doi:10.1101/cshperspect.a032391](https://doi.org/10.1101/cshperspect.a032391) [Medline](#)
2. C. Yan, R. Wan, Y. Shi, Molecular mechanisms of pre-mRNA splicing through structural biology. *Cold Spring Harb. Perspect. Biol.* **11**, a032409 (2019). [doi:10.1101/cshperspect.a032409](https://doi.org/10.1101/cshperspect.a032409) [Medline](#)
3. B. Kastner, C. L. Will, H. Stark, R. Lührmann, Structural Insights into Nuclear pre-mRNA Splicing in Higher Eukaryotes. *Cold Spring Harb. Perspect. Biol.* 10.1101/cshperspect.a032417 (2019). [doi:10.1101/cshperspect.a032417](https://doi.org/10.1101/cshperspect.a032417) [Medline](#)
4. T. A. Steitz, J. A. Steitz, A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6498–6502 (1993). [doi:10.1073/pnas.90.14.6498](https://doi.org/10.1073/pnas.90.14.6498) [Medline](#)
5. S. M. Fica, N. Tuttle, T. Novak, N. S. Li, J. Lu, P. Koodathingal, Q. Dai, J. P. Staley, J. A. Piccirilli, RNA catalyses nuclear pre-mRNA splicing. *Nature* **503**, 229–234 (2013). [doi:10.1038/nature12734](https://doi.org/10.1038/nature12734) [Medline](#)
6. C. Plaschka, P. C. Lin, C. Charenton, K. Nagai, Prespliceosome structure provides insights into spliceosome assembly and regulation. *Nature* **559**, 419–422 (2018). [doi:10.1038/s41586-018-0323-8](https://doi.org/10.1038/s41586-018-0323-8) [Medline](#)
7. T. H. Nguyen, W. P. Galej, X. C. Bai, C. G. Savva, A. J. Newman, S. H. Scheres, K. Nagai, The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature* **523**, 47–52 (2015). [doi:10.1038/nature14548](https://doi.org/10.1038/nature14548) [Medline](#)
8. T. H. D. Nguyen, W. P. Galej, X. C. Bai, C. Oubridge, A. J. Newman, S. H. W. Scheres, K. Nagai, Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. *Nature* **530**, 298–302 (2016). [doi:10.1038/nature16940](https://doi.org/10.1038/nature16940) [Medline](#)
9. R. Wan, C. Yan, R. Bai, L. Wang, M. Huang, C. C. Wong, Y. Shi, The 3.8 Å structure of the U4/U6.U5 tri-snRNP: Insights into spliceosome assembly and catalysis. *Science* **351**, 466–475 (2016). [doi:10.1126/science.aad6466](https://doi.org/10.1126/science.aad6466) [Medline](#)
10. D. E. Agafonov, B. Kastner, O. Dybkov, R. V. Hofele, W. T. Liu, H. Urlaub, R. Lührmann, H. Stark, Molecular architecture of the human U4/U6.U5 tri-snRNP. *Science* **351**, 1416–1420 (2016). [doi:10.1126/science.aad2085](https://doi.org/10.1126/science.aad2085) [Medline](#)
11. C. Boesler, N. Rigo, M. M. Anokhina, M. J. Tauchert, D. E. Agafonov, B. Kastner, H. Urlaub, R. Ficner, C. L. Will, R. Lührmann, A spliceosome intermediate with loosely associated tri-snRNP accumulates in the absence of Prp28 ATPase activity. *Nat. Commun.* **7**, 11997 (2016). [doi:10.1038/ncomms11997](https://doi.org/10.1038/ncomms11997) [Medline](#)
12. R. Bai, R. Wan, C. Yan, J. Lei, Y. Shi, Structures of the fully assembled *Saccharomyces cerevisiae* spliceosome before activation. *Science* **360**, 1423–1429 (2018). [doi:10.1126/science.aau0325](https://doi.org/10.1126/science.aau0325) [Medline](#)
13. X. Zhan, C. Yan, X. Zhang, J. Lei, Y. Shi, Structures of the human pre-catalytic spliceosome and its precursor spliceosome. *Cell Res.* **28**, 1129–1140 (2018). [doi:10.1038/s41422-018-0094-7](https://doi.org/10.1038/s41422-018-0094-7) [Medline](#)

14. J. P. Staley, C. Guthrie, An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. *Mol. Cell* **3**, 55–64 (1999). [doi:10.1016/S1097-2765\(00\)80174-4](https://doi.org/10.1016/S1097-2765(00)80174-4) [Medline](#)
15. B. Lagerbauer, T. Achsel, R. Lührmann, The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplexes in vitro. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4188–4192 (1998). [doi:10.1073/pnas.95.8.4188](https://doi.org/10.1073/pnas.95.8.4188) [Medline](#)
16. P. L. Raghunathan, C. Guthrie, RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Curr. Biol.* **8**, 847–855 (1998). [doi:10.1016/S0960-9822\(07\)00345-4](https://doi.org/10.1016/S0960-9822(07)00345-4) [Medline](#)
17. C. Plaschka, P. C. Lin, K. Nagai, Structure of a pre-catalytic spliceosome. *Nature* **546**, 617–621 (2017). [doi:10.1038/nature22799](https://doi.org/10.1038/nature22799) [Medline](#)
18. G. Jakab, A. Mougín, M. Kis, T. Pollák, M. Antal, C. Branlant, F. Solymosy, Chlamydomonas U2, U4 and U6 snRNAs. An evolutionary conserved putative third interaction between U4 and U6 snRNAs which has a counterpart in the U4atac-U6atac snRNA duplex. *Biochimie* **79**, 387–395 (1997). [doi:10.1016/S0300-9084\(97\)86148-2](https://doi.org/10.1016/S0300-9084(97)86148-2) [Medline](#)
19. K. Bertram, D. E. Agafonov, O. Dybkov, D. Haselbach, M. N. Leelaram, C. L. Will, H. Urlaub, B. Kastner, R. Lührmann, H. Stark, Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation. *Cell* **170**, 701–713.e11 (2017). [doi:10.1016/j.cell.2017.07.011](https://doi.org/10.1016/j.cell.2017.07.011) [Medline](#)
20. A. M. Zahler, L. E. Rogel, M. L. Glover, S. Yitiz, J. M. Ragle, S. Katzman, SNRP-27, the *C. elegans* homolog of the tri-snRNP 27K protein, has a role in 5' splice site positioning in the spliceosome. *RNA* **24**, 1314–1325 (2018). [doi:10.1261/rna.066878.118](https://doi.org/10.1261/rna.066878.118) [Medline](#)
21. A. K. Leung, K. Nagai, J. Li, Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis. *Nature* **473**, 536–539 (2011). [doi:10.1038/nature09956](https://doi.org/10.1038/nature09956) [Medline](#)
22. E. Absmeier, J. Wollenhaupt, S. Mozaffari-Jovin, C. Becke, C. T. Lee, M. Preussner, F. Heyd, H. Urlaub, R. Lührmann, K. F. Santos, M. C. Wahl, The large N-terminal region of the Brr2 RNA helicase guides productive spliceosome activation. *Genes Dev.* **29**, 2576–2587 (2015). [Medline](#)
23. A. Jacewicz, B. Schwer, P. Smith, S. Shuman, Crystal structure, mutational analysis and RNA-dependent ATPase activity of the yeast DEAD-box pre-mRNA splicing factor Prp28. *Nucleic Acids Res.* **42**, 12885–12898 (2014). [doi:10.1093/nar/gku930](https://doi.org/10.1093/nar/gku930) [Medline](#)
24. M. Schneider, H. H. Hsiao, C. L. Will, R. Giet, H. Urlaub, R. Lührmann, Human PRP4 kinase is required for stable tri-snRNP association during spliceosomal B complex formation. *Nat. Struct. Mol. Biol.* **17**, 216–221 (2010). [doi:10.1038/nsmb.1718](https://doi.org/10.1038/nsmb.1718) [Medline](#)
25. A. L. Mallam, M. Del Campo, B. Gilman, D. J. Sidote, A. M. Lambowitz, Structural basis for RNA-duplex recognition and unwinding by the DEAD-box helicase Mss116p. *Nature* **490**, 121–125 (2012). [doi:10.1038/nature11402](https://doi.org/10.1038/nature11402) [Medline](#)
26. T. Sengoku, O. Nureki, A. Nakamura, S. Kobayashi, S. Yokoyama, Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa. *Cell* **125**, 287–300 (2006). [doi:10.1016/j.cell.2006.01.054](https://doi.org/10.1016/j.cell.2006.01.054) [Medline](#)

27. W. P. Galej, C. Oubridge, A. J. Newman, K. Nagai, Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature* **493**, 638–643 (2013). [doi:10.1038/nature11843](https://doi.org/10.1038/nature11843) [Medline](#)
28. T. P. Wu, K. C. Ruan, W. Y. Liu, A fluorescence-labeling method for sequencing small RNA on polyacrylamide gel. *Nucleic Acids Res.* **24**, 3472–3473 (1996). [doi:10.1093/nar/24.17.3472](https://doi.org/10.1093/nar/24.17.3472) [Medline](#)
29. A. Mayeda, A. R. Krainer, Preparation of HeLa cell nuclear and cytosolic S100 extracts for in vitro splicing. *Methods Mol. Biol.* **118**, 309–314 (1999). [Medline](#)
30. K. Zhang, Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016). [doi:10.1016/j.jsb.2015.11.003](https://doi.org/10.1016/j.jsb.2015.11.003) [Medline](#)
31. S. Chen, G. McMullan, A. R. Faruqi, G. N. Murshudov, J. M. Short, S. H. Scheres, R. Henderson, High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013). [doi:10.1016/j.ultramic.2013.06.004](https://doi.org/10.1016/j.ultramic.2013.06.004) [Medline](#)
32. T. Wagner, F. Merino, M. Stabrin, T. Moriya, C. Gatsogiannis, S. Raunser, SPHIRE-crYOLO: A fast and well-centering automated particle picker for cryo-EM. *BioRxiv* 356584 [Preprint]. 26 June 2018. <https://doi.org/10.1101/356584>
33. J. Zivanov, T. Nakane, B. O. Forsberg, D. Kimanius, W. J. H. Hagen, E. Lindahl, S. H. W. Scheres, New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018). [doi:10.7554/eLife.42166](https://doi.org/10.7554/eLife.42166) [Medline](#)
34. T. Nakane, D. Kimanius, E. Lindahl, S. H. W. Scheres, Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife* **7**, e36861 (2018). [doi:10.7554/eLife.36861](https://doi.org/10.7554/eLife.36861) [Medline](#)
35. D. A. Pomeranz Krummel, C. Oubridge, A. K. Leung, J. Li, K. Nagai, Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature* **458**, 475–480 (2009). [doi:10.1038/nature07851](https://doi.org/10.1038/nature07851) [Medline](#)
36. Y. Kondo, C. Oubridge, A. M. van Roon, K. Nagai, Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5' splice site recognition. *eLife* **4**, e04986 (2015). [doi:10.7554/eLife.04986](https://doi.org/10.7554/eLife.04986) [Medline](#)
37. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010). [doi:10.1107/S0907444910007493](https://doi.org/10.1107/S0907444910007493) [Medline](#)
38. S. M. Fica, C. Oubridge, M. E. Wilkinson, A. J. Newman, K. Nagai, A human postcatalytic spliceosome structure reveals essential roles of metazoan factors for exon ligation. *Science* **363**, 710–714 (2019). [doi:10.1126/science.aaw5569](https://doi.org/10.1126/science.aaw5569) [Medline](#)
39. C. Oubridge, N. Ito, P. R. Evans, C. H. Teo, K. Nagai, Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**, 432–438 (1994). [doi:10.1038/372432a0](https://doi.org/10.1038/372432a0) [Medline](#)
40. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. E. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015). [doi:10.1038/nprot.2015.053](https://doi.org/10.1038/nprot.2015.053) [Medline](#)

41. I. Korneta, M. Magnus, J. M. Bujnicki, Structural bioinformatics of the human spliceosomal proteome. *Nucleic Acids Res.* **40**, 7046–7065 (2012). [doi:10.1093/nar/gks347](https://doi.org/10.1093/nar/gks347) [Medline](#)
42. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010). [doi:10.1107/S09074444909052925](https://doi.org/10.1107/S09074444909052925) [Medline](#)
43. A. A. Lebedev, P. Young, M. N. Isupov, O. V. Moroz, A. A. Vagin, G. N. Murshudov, JLigand: A graphical tool for the CCP4 template-restraint library. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 431–440 (2012). [doi:10.1107/S090744491200251X](https://doi.org/10.1107/S090744491200251X) [Medline](#)
44. S. Sharma, S. P. Wongpalee, A. Vashisht, J. A. Wohlschlegel, D. L. Black, Stem-loop 4 of U1 snRNA is essential for splicing and interacts with the U2 snRNP-specific SF3A1 protein during spliceosome assembly. *Genes Dev.* **28**, 2518–2531 (2014). [doi:10.1101/gad.248625.114](https://doi.org/10.1101/gad.248625.114) [Medline](#)
45. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. G. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004). [doi:10.1002/jcc.20084](https://doi.org/10.1002/jcc.20084) [Medline](#)