Supplementary Materials for

Molecular architecture of the human U4/U6.U5 tri-snRNP

Dmitry E. Agafonov, Berthold Kastner, Olexandr Dybkov, Romina V. Hofele, Wen-Ti Liu, Henning Urlaub,* Reinhard Lührmann,* Holger Stark*

*Corresponding author. E-mail: reinhard.luehrmann@mpi-bpc.mpg.de (R.L.); hstark1@gwdg.de (H.S.); henning.urlaub@mpibpc.mpg.de (H.U.)

Published 18 February 2016 on Science First Release
DOI: 10.1126/science.aad2085

This PDF file includes:
Materials and Methods
Figs. S1 to S16
Table S1
Captions for table S2 and movie S1
References

Other supplementary material for this manuscript includes the following:
Table S2 (Excel format)
Movie S1
Supplementary online material

Materials and Methods

Purification of human U4/U6.U5 tri-snRNP

The nuclear extract for tri-snRNP purification was prepared essentially as described by Dignam et al. (1983) (37) with the following modifications. In brief, the nuclei pellet was resuspended in buffer containing 20 mM HEPES pH 7.9, 250 mM KCl, 8.5 mM MgCl₂, 0.2 mM EDTA and 20% sucrose. The nuclei were lysed with 20 strokes of a Dounce homogenizer at 4 °C. The lysate was stirred for 40 min at 4 °C, followed by centrifugation in an SS34 rotor at 16,000 rpm for 30 min. The supernatant was additionally cleared by two consecutive centrifugations in an F14-6x250y rotor at 4 °C for 30 min at 13,000 and 14,000 rpm. 320 ml of the resulting supernatant was loaded onto a 17–30% sucrose gradient containing 20 mM HEPES pH 7.9, 150 mM KCl, 5 mM MgCl₂, 0.1 mM EDTA in a Ti-15 zonal rotor (Beckmann Coulter) and centrifuged for 20 h at 4 °C. After gradient fractionation the U4, U5 and U6 snRNAs were detected by SYBR Gold (Invitrogen) staining of gradient fractions separated on a NuPAGE 4–12% Midi Gel (Invitrogen). Fractions containing tri-snRNPs were loaded onto an H2O column with bound m₃G-cap antibody in the presence of approximately 15% sucrose from the gradient. Subsequent affinity chromatography was performed as described previously (12) in a buffer containing 20 mM HEPES pH 7.9, 150 mM KCl, 5 mM MgCl₂, 0.1 mM EDTA and 15% sucrose. Eluted particles were loaded onto a 20%–50% sucrose gradient containing 20 mM HEPES pH 7.9, 100 mM KCl, 5 mM MgCl₂ and 0.1 mM EDTA. For mild fixation of particles for cryo-EM, the gradient also contained 0%-0.26% glutaraldehyde and 1 mM–0 mM p-maleimidophenyl isocyanate (PMPI, Pierce) from top to bottom. For biochemical controls the same gradient without glutaraldehyde and without PMPI was used. Centrifugation was performed at 29,000 rpm in a Surespin 630 rotor (Sorvall) for 40 h at 4 °C. Gradients were fractionated from the bottom. Aliquots of fractions were removed immediately and analyzed by a Bradford protein assay, and the remaining material in each fraction was quenched with 100 mM aspartate. The RNA and protein compositions of purified particles from non-fixed gradients were visualized after
separation on a NuPAGE 4–12% Midi Gel by staining with SYBR Gold (RNA) and with Coomassie blue (protein). Proteins were also analyzed on 2D gels as described previously (38).

Crosslinking of tri-snRNPs and crosslink identification by mass spectrometry

Purified tri-snRNPs were cross-linked with 150 µM BS3 and purified further by centrifugation in a 20-50% sucrose gradient. Approximately 40 µg tri-snRNP pelleted by ultracentrifugation from the glycerol gradient fractions corresponding to 25S were analyzed essentially as described before (39), with the following modifications: precipitated material was dissolved in 4M Urea/50mM ammonium bicarbonate, reduced with DTT, alkylated with iodoacetamide, diluted to 1M Urea and digested with trypsin (1:20 w:w). Peptides were reverse-phase extracted and fractionated by gel filtration on Superdex Peptide PC3.2/30 column (GE HealthCare). 50 µl fractions corresponding to elution volume of 1.2-1.8 ml were analyzed on Orbitrap Fusion Tribrid (dataset 1), Q Exactive HF (dataset 2) or Q Exactive (dataset 3) mass spectrometers (Thermo Scientific). Protein-protein crosslinks were identified by pLink1.22 search engine and filtered at FDR 1% (pfind.ict.ac.cn/software/pLink) according to the recommendations of the developers (40). For simplicity, the crosslink Score is represented as a negative value of the common logarithm of the original pLink Score, that is Score=−log10(“pLink Score”). Dataset 1 was not filtered further, while the filtering criteria of Scoremax≥1 or Scoremax≥3 were applied for the datasets 2 and 3, respectively (table S2). 169 BS3-crosslinks were identified between the known structured tri-snRNP protein domains. For model building, a maximum distance of 3 nm between the Cα-atoms of the crosslinked lysines was allowed. The actual distribution of Cα-Cα distances between crosslinked residues that can be mapped in the model of the tri-snRNP is shown in fig. S16.

Electron Microscopy

Purified tri-snRNPs were allowed to adsorb on a thin film of carbon for 10 min, transferred onto a cryo-EM grid (Quantifoil 3.5/1 1µm, Jena) and then prepared under controlled environmental conditions of 4 °C and 100% humidity in a vitrification device (Vitrobot Mark IV, FEI Company, Eindhoven). Images were recorded in spot-scanning
mode (2 × 2 image frames per 3.5 µm hole in the Quantifoil carbon film) on a Falcon II direct detector under liquid-nitrogen conditions with a Titan Krios electron microscope (FEI, Eindhoven) equipped with a XFEG electron source and a Cs corrector (CEOS, Heidelberg) using 300 kV acceleration voltage. An electron dose of ~30 electrons per Å², −0.7 to −2.5 µm defocus and a nominal magnification of 74,000× were used, resulting in a final pixel size of ~2 Å. 1,150,811 particle images were initially extracted in a fully automated manner using template-independent custom-made software (CowPicker, B. Busche and H.S., unpublished data). The extracted particle images were corrected locally for the contrast-transfer function by classification and averaging and selected according to quality of power spectra (41). Initial sorting of images was performed based on CTF parameters. Only images showing Thon rings better than 7 Ångstrom were used for further processing. Additional sorting was performed applying two rounds of multivariate statistics, first without alignment and subsequently after one round of image alignment. This resulted in a total number of 636,460 particle images that were used for further processing. 3D classification in RELION 1.2 (42) was used to obtain 190,939 intact tri-snRNP particles using a previously published 3D structure as a reference (EMD-1729). A second round of 3D classification yielded the 141,109 particle images that were used for the final refinement that resulted in a 7 Å resolution structure (using the ‘gold-standard procedure’ in RELION 1.2) by applying a mask that covers the entire tri-snRNP. The structure was used for manual model building in UCSF Chimera 1.10.1 (43).

**Preparation of Figures**

Maps were visualized and movie and figures generated using UCSF Chimera (43), except figures S6, S7C, D and S15 which were generated by PyMOL Molecular Graphics System.
Supplementary Figure Legends

Figure S1. Protein and RNA composition of human U4/U6.U5 tri-snRNPs used for cryo-EM. (A) Proteins present in human tri-snRNPs and their molecular mass. The yeast (*S. cerevisiae*) nomenclature was used for all human tri-snRNP proteins also present in *S. cerevisiae*. The U5-40K, cyclophilin H (CypH), 27K and RBM42 proteins do not have counterparts in *S. cerevisiae*. Alternative names for the human tri-snRNP proteins are also listed. (B) RNA and protein composition of purified human HeLa U4/U6.U5 tri-snRNPs. Left panel: SDS-polyacrylamide gel stained with SYBR Gold (lane 1) or Coomassie (lane 2) showing the snRNA and protein composition, respectively, of purified tri-snRNPs. Proteins were identified by mass spectrometry. Right panel: RuBPS-stained, two-dimensional gel of tri-snRNP proteins separated according to (38). Proteins with a molecular weight less than ca. 25 kDa have run out of the gel. (C) RNA sequences and secondary structures of human U5 snRNA and base-paired U4 and U6 snRNAs. Red boxes indicate regions of the snRNAs that interact with the LSm2–8 and Prp3 proteins in the isolated tri-snRNP, or with the Brr2 RNA helicase in the course of U4/U6 snRNA unwinding during activation of the spliceosomal B complex. Blue boxes show the regions of U6 snRNA (its 3’ end) that base pair with the 5’ end of U2 snRNA to form the U2/U6 helix II, and with the 5’ splice site (ACAGA box), during B complex formation.
Figure S2. Computational Sorting

4812 micrographs were recorded from which 1,150,811 particle images were selected by a newly developed particle picking software that does not require reference images for particle selection. After 2D classification, sorting of particles and CTF correction, 636,460 particles were used for structure calculations. Our previously determined low-resolution structure of the tri-snRNP was used as a reference to obtain the initial 3D structure. 3D classification was performed in RELION. Roughly 30% of the particles revealed intact tri-snRNPs. Other populations revealed mainly tri-snRNPs that had lost one or more of the Sm cores or other proteins. 190,939 particles were analyzed in a second classification step and revealed tri-snRNPs that mainly differed in the conformation of the U4 Sm core which seems to be flexible. A population of 141,109 particles was used for the final refinement to obtain the 7 Å 3D structure. Shown is the structure prior to sharpening and the sharpened map that was calculated using a B factor of 250.
4,812 micrographs
picked particles: 1,150,811
after 2D classification, sorting and CTF correction: 636,460 particles

3D classification 1

30% intact
9% no U5 core
20% bad particles I
34% lower U4, U6 occupancy
7% bad particles II

30%

190,939 particles

3D classification 2

30%

U4 up-down

141,109 particles

Refine

30%

non sharpened

sharpened
Figure S3. Cryo-EM and image processing of the human tri-snRNP. (A) Raw images of vitrified tri-snRNPs imaged with a Titan Krios electron microscope at 74,000x magnification on a Falcon II direct detector. Some tri-snRNPs are encircled. The panels on the right show 2D class averages obtained after alignment and classification (upper 2 rows) and their corresponding reprojections calculated from the 3D structure (lower 2 rows). (B) Left: Euler angle distribution obtained for the final 3D reconstruction. Tri-snRNP molecules exhibit a relatively strong preference to bind to carbon film. This is reflected by the red bars of the spherical histogram. However, the presence of almost all other possible views becomes obvious from the relatively uniform distribution as indicated by the blue bars. Right: Fourier-shell-correlation function of two independently refined half maps indicate a resolution of 7 Å using the 0.143 threshold. (C) Local resolution plot obtained from calculating local Fourier-shell correlations using a kernel of 15x15x15 voxel over the entire 3D volume. Three different views of the tri-snRNP are depicted which are rotated along the vertical axis. The tri-snRNP map is best defined in the center of the particle where the resolution exceeds the average resolution obtained for the entire reconstruction. The resolution drops below 10 Å in some areas of the map. The lower resolution in the U4/U6 part of the tri-snRNP (top region) can be explained by compositional and/or conformational heterogeneity.
Figure S4. Crosslinking of the C-terminal TPR repeats of Prp6 to U4/U6 proteins and its N-terminal TPR repeats to U5 proteins Prp8 and Brr2. α-Helical regions of Prp6 TPR repeats 1–19 are indicated by cylinders. The numbers of crosslinked lysine residues (connected by stippled lines) correspond to their positions in the respective protein sequences. Intramolecular crosslinks between TPR18-19 and TPRs 9-13 of Prp6 are also indicated by stippled lines. Brr2-NHD, N-terminal helical domain of Brr2; Brr2-NC, N-terminal helicase cassette of Brr2 (see also Figs. 2 and 4).
Figure S5. Similarities and differences in the architecture of the U4/U6 snRNP in the human and yeast tri-snRNP models.

Structures of the U4/U6 snRNA and proteins in corresponding views of the human (A) and yeast (B) tri-snRNP model (9). The geometry of the base-paired U4/U6 regions of the three-way junction is essentially the same in both models. In the human tri-snRNP, the position of the LSm ring is closer to the region between the U4/U6 stem II and the single-stranded 3’end of the U6 snRNA than in the yeast tri-snRNP. In the human tri-snRNP, the ferredoxin-like domain of Prp3-CTF faces the LSm structure in such a way that the single-stranded 3’-terminal region of the U6 snRNA can be continued in the direction of the LSm doughnut (dotted purple line, panel A). The U4 Sm core structure is also positioned differently in the two tri-snRNP models (see superposition of the human and yeast tri-snRNP models in fig. S12) and the architecture of the Prp31 protein also differs. That is, the crystal structure of the arc-like form of Prp31’s coiled-coil and Nop domain fits perfectly into a similarly shaped density element in the human tri-snRNP that is bent downwards at the periphery of the tri-snRNP contacting the Sad1 protein (Figs. 2A and 5). In contrast, in the yeast tri-snRNP model, the coiled-coil domain of Prp31 is rotated by ca. 60° with respect to the Nop domain (9).
**Figure S6. Snu114 has a structure in the human tri-snRNP similar to the compact form of the elongation factor EF-2.** (A) Structural organization of Snu114 in the human tri-snRNP; the positions of the Prp8 RT and Brr2 CC domains that are closest to Snu114’s D3 to D5 domains in the tri-snRNP model are also shown. With the exception of its N-terminal 115-residue domain, the 116-kDa protein Snu114 is highly homologous to the ribosomal elongation factor EF-2/EF-G (see inset). The compact form (B) of crystallized EF-2 (16) and its extended form (C) as observed in the 80S ribosome post-translocation state (35). All structures are aligned via domains D1 and D2. A comparison of the structure of Snu114 with those of EF-2 reveals significant similarity with the compact EF-2 form. (D) The elongated form of EF-2 (35) modelled into the tri-snRNP structure, whereby domains D1 and D2 are aligned to the position of Snu114’s D1 and D2 in the tri-snRNP. If during the catalytic activation of the spliceosome Snu114 would undergo a structural change in the direction of the extended form of EF-2, then domains D4 and D5 would lie at the interface between Prp8-RT and Brr2-CC, and would clash with Brr2-CC. Thus, such a conformational change in Snu114 would only be possible if it were part of a concerted structural change in the spliceosome, including movement of Brr2 toward U4/U6 RNAs (see also main text and fig. S14).
Figure S7. Differences in the architecture of the Prp8 RH domain in the cryo EM structures of human and yeast tri-snRNPs. The structures of the Prp8 RT/En and Prp8 RH domains are shown in corresponding views of the human (A) and yeast (B) tri-snRNP model (9). (C and D) Structure of the yeast Prp8 RT/En and Prp8 RH domain as found in the crystallised complex with the Aar2 protein (44). The spatial organization of the Prp8 RT/En domain is almost identical in the two tri-snRNP structures. The positions of the RH domains relative to the axis of the respective Prp8 RT/En domains are also similar. However, the spatial orientation of the RH domain differs significantly between the two structures. In the human tri-snRNP structure, the RH domain is rotated by approximately 90° relative to its position in the crystal structure (A and C). In the yeast tri-snRNP, it must be rotated additionally by 180° about the central axis. Thus, in the human tri-snRNP the β hairpin of the RH domain points towards the RT domain, while in the yeast tri-snRNP it points in the opposite direction, i.e. towards the En domain of Prp8.
Figure S8. Prp8 interactions and crosslinks with Prp28, Dim1, Snu114 and the U5-40K protein. (A) Intermolecular crosslinks between Prp8’s RT/En, NTD1, NTD2, and NTDL domains, and Dim1 and the RecA2 domain of Prp28. Numbers indicate positions of crosslinked lysine residues in the three proteins. (B) Left: View 1 of the human tri-snRNP. Right: expanded view of boxed region, with Snu114 colored in orange. Prp8’s NTD1-lasso interacts with the D1 domain of Snu114 in the tri-snRNP in a similar manner as in the S. pombe spliceosome (10). The NTD1 lasso (in lilac) of the S. pombe spliceosome (10) fits as a rigid body into the ring-like density element (in grey) encircling part of the Snu114 D1 domain in the tri-snRNP model (shown in space filling mode). (C) Location of the WD40 domain of U5-40K. Left: View 3 of the human tri-snRNP. Right: Expanded view of boxed region. The 40K-WD40 domain fits into a peripheral knob-like density element at the bottom of the tri-snRNP, close to the internal loop 2 of U5 snRNA. There is a density element on top of the WD40 domain into which the N-terminal helix of Prp8’s NTD1 fits well, consistent with protein-protein crosslinks between both protein domains in the tri-snRNP (table S1 and S2)
Figure S9. Intermolecular crosslinks support the juxtaposition of Brr2 domains with Prp6’s N-terminal TPR repeats and Prp8’s RT, RH and Jab1 domains. Left panel: View of the human tri-snRNP from the top. The frame highlights the spatial relationships between the density elements of the cryo-EM structure with (i) the fitted crystal structures of the Brr2-NC, Prp8-Jab1/Brr2-CC and Prp8 RH and RT domains and (ii) the fitted, modelled N-terminal TPRs of Prp6 and NHD of Brr2. Right: Intermolecular crosslinks between the protein domains are shown. Numbers indicate the positions of crosslinked lysine residues (connected by stippled lines) in each protein. Although projected onto the plane of the paper, the positions in space of the crosslinked lysines correspond to their positions in the 3D structures of their respective protein domains. The maximum observed length for any crosslink was 30 Å. An internal crosslink (2034–2049) and several crosslinks from these two positions to lysines 2070 and 2244 of the Jab1 domain were detected in the ~35-residue-long tract of Prp8, which lies immediately on the N-terminal side of the crystallised Jab1 domain of Prp8 (residues 2067–2305). We therefore placed this region (small area in pink Jab1 region indicated by a dotted line) next to Jab1 (see also table S1).
Figure S10: Network of crosslinks between Brr2, Sad1, Snu114 and Prp8 RT domains. Left: View 5 of the human tri-snRNP. The frame highlights the spatial relationships between the density elements of the cryo-EM structure with (i) the fitted crystal structures of Sad1, the Brr2-CC, Brr2-PWI, Prp8-Jab1 and Prp8 RH domains, and (ii) the fitted, modelled D1 to D4 domains of Snu114. Right: Intermolecular crosslinks between the protein domains shown. Numbers indicate the positions of crosslinked lysine residues (connected by stippled lines) in each protein. Although projected onto the plane of the paper, the positions in space of the crosslinked lysines correspond to their positions in the 3D structures of their respective protein domains.
Figure S11. Brr2 is located in a similar position in native human tri-snRNPs not subjected to chemical crosslinking. (A) Raw image of non-fixed human tri-snRNPs by negative stain electron microscopy. The purified complexes were stained with uranyl formate, air dried and imaged at RT in a Philips CM12 electron microscope. (B) 21,747 tri-snRNP particles were selected for image processing. Typical class averages are shown after several rounds of alignment and classification without using a 3D model as template. (C) Upper row: class averages from negatively-stained human tri-snRNP particles that were not crosslinked. Middle row: reprojections calculated from the cryo-EM 3D structure that was determined from a GraFix treated sample. Projection angles were assigned to class averages by projection matching using the cryo-EM 3D model without Brr2 density (see D) as an initial reference. Overall, the structural features match very well. Lower row: reprojections calculated from the cryo-EM map after removing the Brr2 density. The reprojections therefore show no density for Brr2 and the areas of the missing density are marked by the white lines. We therefore conclude that the position of Brr2 within the tri-snRNP is not affected by the use of chemical crosslinking. (D) Surface view of the cryo-EM reconstruction obtained from GraFix treated tri-snRNPs. The position of Brr2 is marked with a white line. (E) Low resolution reconstruction obtained from negatively-stained images from non-crosslinked tri-snRNPs. Density for Brr2 is clearly visible even at low resolution.
Figure S12. Superposition of the 3D cryo-EM structures of the human and yeast tri-snRNPs. (A and C) Views from the top of the 3D structures of the human (red; for clarity reasons shown at 9.5 Å resolution) and *S. cerevisiae* (green) tri-snRNP, which are superpositioned in (B). The small diagrams on the left are the corresponding side views. The 3D structures are aligned such that the 3D structures of the U5 proteins Prp8 and Snu114 and the U5 Sm core structure are brought into alignment. Superpositioning reveals the large difference in the positions of Brr2 and the U4 Sm core in the two models. At the same time, relative to the “long axis” of Prp8 (i.e. going from “RT” to “En”), which is the same in all three representations, the U4/U6 proteins and Prp6 are closer to the RT domain in the yeast versus human tri-snRNP.
Figure S13. Brr2 and Sad1 can be docked at the corresponding positions in the yeast tri-snRNP 3D structure where they are found in the human tri-snRNP. (A) The *S. cerevisiae* cryo-EM 3D structure (9) is constricted in the central part. (B) The Brr2 tandem helicase domain (bound to Jab1) and the Sad1 protein, which is connected by Brr2’s PWI domain in the human tri-snRNP (Fig. 5), fit together as a rigid body into matching cavities of the yeast tri-snRNP structure. The fit includes the same Brr2-Sad1 positioning relative to Prp8 and Snu114 as found in the human tri-snRNP structure. In the different views it can be seen that there are only minor clashes upon docking of these proteins, mainly at the N-terminal TPRs of Prp6. (C) Space filling presentation of Sad1 and Brr2 illustrating the good spatial fit.
**Figure S14. Proposed sequential tri-snRNP remodeling steps during spliceosome assembly and activation.** (A) Left: Initial docking of the A complex (grey oval) at the top of the tri-snRNP, involving base pairing of U2 RNA (green dashed line) and U6 snRNA (blue dashed line), and association of Prp8's RH domain with U1 snRNA (yellow dashed line) based paired to the pre-mRNA 5’ splice site (5’SS). Green and yellow circles: the U1 and U2 snRNP, respectively. Middle: Movement of Prp28 from its outward position through the cleft to the Prp8 RH domain where it mediates the handover of the 5'SS from U1 to U6 snRNA, leading to spliceosomal B complex formation. Right: Spliceosome activation requires Brr2 movement from its pre-activation position across the cleft to engage its U4/U6 snRNA substrate, which may be triggered by dissociation of Sad1. Snu114 might transiently undergo a structural change from a compact to a more extended form, which would require Brr2 movement away from its pre-activation position to prevent steric clashes with Snu114 domains D4. (B) Prp8 undergoes a substantial conformational change during catalytic activation of the spliceosome. While the overall-structure of Prp8’s large N-terminal NTD1 domain is similar, the RT/En domain adopts a significantly different position in the human tri-snRNP (left) and *S. pombe* spliceosome (right).
Figure S15. Prp8 undergoes a large structural change after tri-snRNP incorporation into the spliceosome. (A) Prp8 undergoes a substantial conformational change during catalytic activation of the spliceosome. While the overall-structure of Prp8’s large N-terminal NTD1 domain is similar, the RT/En domain adopts a significantly different position in the human tri-snRNP (left) and S. pombe spliceosome (right). In the tri-snRNP Prp8 is in an open conformation with the En-end of the elongated RT/En domain well-separated from the upper region (UR, helical region) of the NTD1 domain. The lower right region (LL, long loop, comprising amino acids V283 to R309) of NTD1, as well as the Snu114 D3 domain, are connected to the RT end of the RT/En domain and represent the pivot point around which the RT/En domain likely rotates (left). To achieve the spatial orientation observed in the S. pombe spliceosome (right) (10) where Prp8 is in a more closed conformation, the RT/En domain must move toward the NTD1 domain, such that the En domain now tightly contacts the NTD1 at the UR site. The RT/En domain must not only tilt around the pivot point (the RT-LL interface) along its long axis but also along its short axis. As a result of these movements, the RT-LL interface is greatly reduced and the RT contact to Snu114-D3 present in the tri-snRNP is not maintained in the S. pombe spliceosome. Surface charge representation of Prp8 in the human tri-snRNP (B) or S. pombe spliceosome (C), where blue represents positively-charged and red negatively-charged residues. The open Prp8 conformation in the human tri-snRNP (B) provides binding sites for the Dim1 protein and the Prp28-RecA2 domain. The Dim1-binding region of Prp8 is highly positively-charged and encompasses the RT/En-NTD1-NTD2 interface region. Binding of these proteins likely prevents Prp8 from adopting the closed conformation observed in the S. pombe spliceosome (C). The closed conformation generates the binding pocket for the catalytic RNA network that is formed during catalytic activation (C, right) (32).
Figure S16. Distribution of the Ca-Ca distances between BS3-crosslinked residues.
The Euclidian distances were measured in the 7 Å model of the human tri-snRNP using PyMOL 1.7. More than 95% of all crosslink-assigned spectra correspond to crosslink distances of 30 Å or less.
Supplementary Tables

**Table S1: Protein and model building information for all modeled tri-snRNP proteins.** Protein names, their molecular weight and detailed information about the model building process are provided. The correlation coefficient was calculated as the correlation between the simulated 7 Å 3D map of each protein domain and the experimental EM 3D density map using the commands "molmap" and "measure correlation" of UCSF Chimera (43).

**Table S2: BS3-crosslinks between and within known structured regions of human tri-snRNP proteins**
Statistics (Spectral Counts and Score$_{\text{max}}$) of the CX-MS data for the proteins of the purified human tri-snRNP. “Inter” and “Intra” indicate inter-protein and intra-protein crosslinks, respectively. Numbers in the Residue 1 and 2 columns indicate the position of the crosslinked lysine or N-terminal methionine residue.
Table S1. Positioning of proteins in the human tri-snRNP cryo-EM 3D structure

<table>
<thead>
<tr>
<th>Protein Names</th>
<th>MW</th>
<th>Length</th>
<th>Domain</th>
<th>Amino Acids</th>
<th>Structure Source</th>
<th>Reference</th>
<th>Note</th>
<th>Correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>hPrp8 U5-220K</td>
<td>273.6kDa</td>
<td>2335</td>
<td>NTD1</td>
<td>S.p. 34-685 (Spp42) PDB 3JB9</td>
<td>(10) orthologue model (Swiss Model)</td>
<td>0.874</td>
<td>The structure of the N-terminal domain (NTD) of Prp8 in the <em>S. pombe</em> spliceosome was recently solved by Cryo-EM (10). Based on this structure, we have divided the NTD into a large domain (NTD1) at the very N terminus, a downstream linker (NTD2) and remaining C-terminal region (NTD2). In the <em>S. pombe</em> spliceosome (NTD1) is largely α-helical and compact, but on one side a long, mostly unstructured region (between M271 and N307 in the human orthologue) protrudes, with its two loops oriented in different directions. An unstructured linker also connects the helix at the very beginning of NTD1 to the globular main NTD1 body. The homology-modelled NTD1 structure fits very well into a density of the tri-snRNP 3D structure located in the lower part close to U5 stem-loop 1, which is largely covered by the helical region. The lasso-shaped loop of the large unstructured region fits into a circular density surrounding Smu1·4-D1 (fig. S8B) and the other loop (named long loop, LL; V283-R309) runs above Smu1·4-D4·D5 (fig. S15A). When this NTD1 is fitted as a rigid body into the density, the N-terminal helix is positioned close to the U5-40K WD40 protein. The overall position of NTD1 in the tri-snRNP is extremely similar to that found in the <em>S. pombe</em> spliceosome (10). Only the N-terminal helix has to be moved somewhat – into a density element above the U5-40K WD40 domain – to attain the corresponding <em>S. pombe</em> position (fig. S8C). Several crosslinks support this location (K29, K36, K43, K48, K50 of U5-40K·WD40 to K216 and/or K349 of NTD1).</td>
<td></td>
</tr>
<tr>
<td>NTD2</td>
<td>663-677</td>
<td>678-798</td>
<td>NTD2</td>
<td>S.p. 686-821 (Spp42) PDB 3JB9</td>
<td>(10) orthologue model (Swiss Model)</td>
<td>0.813</td>
<td>In the <em>S. pombe</em> spliceosome structure, the NTDL, which is about 15 amino acids long, spans the gap between NTD1 and NTD2. The latter is small and contains four helices that are attached to the RT region of Prp8. In the tri-snRNP the two C-terminal helices of the homology-modelled NTD2 fit very well into density elements positioned relative to the RT domain in the same way as found in the <em>S. pombe</em> spliceosome. The two N-terminal helices of NTD2 are arranged in a V-like manner. Close by there are two similarly arranged densities, into which the two helices can be fitted perfectly by twisting them around the connection between the second N-terminal helix and the first C-terminal helix by about 30 degrees. As in the <em>S. pombe</em> spliceosome, the Prp8-RT/En domain is positioned relative to the NTD1 domain quite differently compared to the situation in the human tri-snRNP (figs. S14 and S15); the RT-attached NTD2 is also differently positioned with respect to NTD1. Therefore, the NTDL linking NTD1 and NTD2 has to be tilted to reach the NTD1-C-terminus to which it connects. A density for the stretched NTDL sequence is barely visible, but a number of crosslinks determine its likely path (fig. S8A). Several crosslinks also confirm the NTD2 positioning close to the RT domain (fig. S8A).</td>
<td></td>
</tr>
<tr>
<td>Prp8-RT/En</td>
<td>790-1755</td>
<td>(Spp42)</td>
<td>(Spp42)</td>
<td>PDB 3E9L</td>
<td>(10) orthologue model (Swiss Model)</td>
<td>0.844</td>
<td>In the ca. 12 nm long RT/En domain, the En and linker domains form, together with the thumb helices of the RT domain, a slightly twisted tail that originates from the ca. 7 nm broad RT ‘head’ structure. In the cryo-EM reconstruction of the human tri-snRNP, the tapered wedge-like RT/En structure fits into a correspondingly-shaped density element at the center of the particle (Fig. 3A). The broad RT side is positioned in a region surrounded by large density elements, while the pointed En end protrudes slightly up and outwards at the periphery of the tri-snRNP, ending just below the U4 Sm complex. The model based on the <em>S. pombe</em> orthologue (Spp42) structure in the late spliceosome (10) gives a better fit than the model based on the <em>S. cerevisiae</em> Prp8 crystal structure (44). The spliceosome catalytic core site of the RT/En structure, where the rearranged catalytic U2/U6 snRNA network and U5 snRNA loop 1 dock, faces almost directly the U4/U6 three-way junction, but is well separated from it. The U5 stem-loop 1 is located just below the RT-linker region, with the loop closest to the RT-thumb helices.</td>
<td></td>
</tr>
<tr>
<td>Prp8-RH</td>
<td>1760-2015</td>
<td>PDB 3E9L</td>
<td>(45) crystal structure</td>
<td>0.910</td>
<td>The Prp8-RH domain was placed close to Prp8’s RT/En domain into a well-fitting density element by initially keeping the arrangement found in the Prp8-Aar2 co-crystal (44). Compared with the crystal structure, only a small shift and a rotation around the axis along the RH β-finger are necessary to achieve a very good fit (see also Fig.3A and fig. S7). In this fit, the β-finger is well-positioned in a small, elongated, protruding density region. Thus, the RH domain is situated centrally at the top of the tri-snRNP with the presumed RNA-binding surface oriented towards the space between the U4 Sm complex and the RT domain. The RH domain is connected to the RT domain by a density band that extends up to the arc that bridges the RH domain and the U4/U6 site. For intermolecular crosslinks of the RH domain with other tri-snRNP proteins see fig. S9.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prp8-Jab1</td>
<td>2067-2335</td>
<td>PDB 4K1T</td>
<td>(20) co-crystal with Brr2</td>
<td>0.902</td>
<td>The Prp8-Jab1 domain was inserted into the tri-snRNP reconstruction by straightforward positioning of the Brr2-Jab1 co-crystal structure (20). As described in more detail in fig. S9, taking our intra- and intermolecular crosslinks into account we have also positioned part of the linker region connecting the Jab1 and Prp8 RH domains, close to the Jab1 domain.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hBrr2 U5-200K SNRNP200</td>
<td>244.5kDa</td>
<td>2136</td>
<td>Br12-helicase</td>
<td>PDB 4K1T</td>
<td>(20) N-terminal (NC) and C-Terminal (CC)</td>
<td>0.919</td>
<td>For the large co-crystal structure of Br12’s entire helicase region in complex with Prp8’s Jab1 domain (20), a perfectly fitting peripheral density region positioned at the RT end of Prp8’s RT/En structure is clearly apparent. In this fit, the C-terminal (inactive) helicase cassette takes up a peripheral position, whereby its Ig domain is the</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Mass (kDa)</td>
<td>Orthologue Model</td>
<td>Crystal Structure</td>
<td>Remarks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
<td>------------------</td>
<td>-------------------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prp8-Jab1</td>
<td>22</td>
<td>Comparative (SpltProt3D)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prp8-Jab1 complex in the tri-snRNP is well reflected in its compact crystal structure. Only the position of a small helix (T1967-G1978) close to the C-terminus was slightly adjusted to fit into a close-by density element.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brr2-Jab1</td>
<td>22</td>
<td>Orthologue model (Swiss Model)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The crystal structure shows a well-defined structure of the Brr2-Jab1 complex in the tri-snRNP, with D2 representing a relatively rigid building block. In the tri-snRNP reconstruction, these two domains were placed into a well-fitting density region at the bottom of the particle just above the U5 Sm ring, with D1 proximal to U5 snRNA and D2 distal to it. The nucleotide-binding site is then oriented inwardly, in the direction of the U5 snRNA stem-loop 1 (5'SL).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Snu114</td>
<td>39.3</td>
<td>Orthologue model (Swiss Model)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The position of Snu114 in the tri-snRNP is consistent with the crystal structure of the Snu144 orthologue Cwf10 observed in the S. pombe spliceosome (10). According to the numerous ribosomal translocase crystal structures, domains D1 and D2 represent a relatively rigid building block. In the tri-snRNP reconstruction, these two domains were placed into a well-fitting density region at the bottom of the particle just above the U5 Sm ring, with D1 proximal to U5 snRNA and D2 distal to it. The nucleotide-binding site is then oriented inwardly, in the direction of the U5 snRNA stem-loop 1 (5'SL).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dm1</td>
<td>16.8</td>
<td>Crystal structure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For the small Dm1 protein, 8 crosslinks to the Prp8 domains RT/En or NTD2 were detected (fig. S8A). There is only one density, a node-like multiply interconnected element just between the three-way junction and the U5 snRNA and very close to the site where the catalytic center will form, into which the Dm1 protein fits well while allowing the formation of all the crosslinks (Fig. 3B).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prp28-RecA1</td>
<td>352-629</td>
<td>Crystal structure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Prp28-RecA1 domain must be located in close proximity to the C-terminal RecA2 domain, because the linker connecting them is short. The RecA1 domain fits into a knob-like protrubance linked to the density element containing Prp28-RecA2 (Fig. 3C). This arrangement indicates that the Prp28 helicase adopts an inactive, open conformation in the tri-snRNP, which is consistent with our finding that in the isolated tri-snRNP Prp28 cannot bind ATP (17). All available crystal structures of the isolated Prp28 helicase also show an open conformation, in which the two RecA-like domains are turned away from one another.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prp6</td>
<td>106.9</td>
<td>Comparative (SpltProt3D)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Our analysis revealed a deep cleft between the U5 and U4/U6 domains in the tri-snRNP, only a few defined bridges seem to connect the proteins of these two snRNPs. The most remarkable feature is an arc-like bridge at
<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
<th>Molecular Weight (kDa)</th>
<th>Accession or Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>hPrp3</td>
<td>U4/U6-90K protein Prp6</td>
<td>683</td>
<td>PDB 4YHW</td>
</tr>
<tr>
<td>hPrp31</td>
<td>U4/U6-61K protein Prp31</td>
<td>499</td>
<td>PDB 2OZB</td>
</tr>
<tr>
<td>hSnu13</td>
<td>U4/U6-15.5K protein NHP2L1</td>
<td>128</td>
<td>PDB 3-128</td>
</tr>
<tr>
<td>hPrp4</td>
<td>U4/U6-60K protein Prp4</td>
<td>521</td>
<td>PFAM WD40, PDB 3-521</td>
</tr>
<tr>
<td>CypH</td>
<td>U4/U6-20K protein PPLaseH</td>
<td>177</td>
<td>PDB 1MZW</td>
</tr>
<tr>
<td>hSad1</td>
<td>U5/U4/U6-65K protein SNU4</td>
<td>565</td>
<td>S.c. 29-124, PDB 3-565</td>
</tr>
<tr>
<td>hSad6</td>
<td>U5,U4/U6-110K protein SART1, SNU1, SNU2</td>
<td>800</td>
<td>S.c. 29-124, PDB 3-800</td>
</tr>
<tr>
<td>27K</td>
<td>U5,U4/U6-27K protein SNRP2/27</td>
<td>155</td>
<td>S.c. 29-124, PDB 3-155</td>
</tr>
<tr>
<td>RBM42</td>
<td>SNRP2/27, USP39, SNUT3</td>
<td>480</td>
<td>S.c. 29-124, PDB 3-480</td>
</tr>
</tbody>
</table>

These proteins are largely unstructured and could not be assigned to a particular density in the 3D cryo EM structure of the tri-snRNP. This also applies to various low complexity regions of other tri-snRNP proteins, including the N-terminal regions of Prp3, Prp6 and Prp28 and the C-terminal region of Prp31.

The top that structures the clef is a ribbon-like structure. The size of the ribbon has striking similarities to chains of α-helical TPRs. Prp6 is the only tri-snRNP protein in which the sequence indicates extended TPR repeats. Most of its 19 repeats are also confirmed by intramolecular crosslinks, and the intermolecular crosslinks establish that Prp6 bridges U5 and U4/U6 proteins and define the chain polarity, as illustrated in Fig. 2B and fig. S4. The N-terminal of the TPR ribbon interacts first with the Prp8 RT and R1H domains, and two TPRs later it also contacts the N-terminal cassette of Brr2. It then forms a bridge that connects to the U4/U6 site, connecting all of the known structural modules of Prp3, Prp4, Prp31 and 15.5K by forming a circle that involves the C-terminal most seven of its 19 TPRs.
| Smn | SNR PB | 24.6 kDa | 240 | SmnB | 6-95 | PDB 4WZJ | (52) | co-crystal with SmnD1,3,5,6,7,8, U6RNA orthologue model (Swiss Model) | 0.897 (U5) | 0.856 (U4) | U5 Smn
The 5' stem-loop (5'SL) is connected to the U5 Smn binding site by eight unpaired nucleotides that can easily bridge the 3' end of the 5'SL to the center of the large circular density at the very bottom of the tri-snRNP (Fig. 1B). In this density, the typical heptameric ring structure fits perfectly, in a tilted orientation, with the side containing the final 3' stem-loop facing diagonally downwards. The 5' m3G cap of the U5 snRNA is found directly at the lower end of the 5' stem. This position fits nicely with the result of an earlier localization by immuno-EM, in which it was found at the bottom of the particle (12).

U4 Smn
At the top of the particle, a second Smn-shaped circular density element is present at some distance from the U4/U6 stem I site. Twenty one nucleotides are situated between the end of stem I and the stem-loop directly preceding the Smn binding site. This stretch is long enough to bridge the distance between the U4-Smn and the U4/U6 helix connection sites. The crystal structure of the U4 Smn core RNP fits well into the circular density (Fig. 1B). Protruding centrally on both sides, are smaller regions of density that can accommodate the two stem-loops of U4 snRNA that flank the Smn site. The U4 Smn ring is tilted just above the Prp8-En end of the RT/En domain, with its flat side, containing the penultimate stem-loop, pointing outwards (like the RT/En domain).

| SmnD | SNR PD1 | 13.3 kDa | 119 | Smn D1 | 2-95 | PDB 4WZJ | (52) | co-crystal with SmnB,D2,D3,E,F,G, U4RNA | 0.967 (U5) | 10.4 (U4) | U5 Smn
The 5' stem-loop (5'SL) is connected to the U5 Smn binding site by eight unpaired nucleotides that can easily bridge the 3' end of the 5'SL to the center of the large circular density at the very bottom of the tri-snRNP (Fig. 1B). In this density, the typical heptameric ring structure fits perfectly, in a tilted orientation, with the side containing the final 3' stem-loop facing diagonally downwards. The 5' m3G cap of the U5 snRNA is found directly at the lower end of the 5' stem. This position fits nicely with the result of an earlier localization by immuno-EM, in which it was found at the bottom of the particle (12).

U4 Smn
At the top of the particle, a second Smn-shaped circular density element is present at some distance from the U4/U6 stem I site. Twenty one nucleotides are situated between the end of stem I and the stem-loop directly preceding the Smn binding site. This stretch is long enough to bridge the distance between the U4-Smn and the U4/U6 helix connection sites. The crystal structure of the U4 Smn core RNP fits well into the circular density (Fig. 1B). Protruding centrally on both sides, are smaller regions of density that can accommodate the two stem-loops of U4 snRNA that flank the Smn site. The U4 Smn ring is tilted just above the Prp8-En end of the RT/En domain, with its flat side, containing the penultimate stem-loop, pointing outwards (like the RT/En domain).
**Movie S1.** The movie shows the tri-snRNP 3D structure rotating along its long axis. The fitted proteins and RNA parts of the tri-snRNP are shown in the semi-transparent envelope of the EM density.

**References**


51. H. Hadjivassiliou, O. S. Rosenberg, C. Guthrie, The crystal structure of *S. cerevisiae* Sad1, a catalytically inactive deubiquitinase that is broadly required for pre-mRNA splicing. *RNA* **20**, 656–669 (2014). Medline doi:10.1261/rna.042838.113


doi:10.1371/journal.pone.0036768