

The complete structure of the small-subunit processome

Jonas Barandun^{1,4} , Malik Chaker-Margot^{1,2,4} , Mirjam Hunziker^{1,4} , Kelly R Molloy³, Brian T Chait³  & Sebastian Klinge¹ 

The small-subunit processome represents the earliest stable precursor of the eukaryotic small ribosomal subunit. Here we present the cryo-EM structure of the *Saccharomyces cerevisiae* small-subunit processome at an overall resolution of 3.8 Å, which provides an essentially complete near-atomic model of this assembly. In this nucleolar superstructure, 51 ribosome-assembly factors and two RNAs encapsulate the 18S rRNA precursor and 15 ribosomal proteins in a state that precedes pre-rRNA cleavage at site A1. Extended flexible proteins are employed to connect distant sites in this particle. Molecular mimicry and steric hindrance, as well as protein- and RNA-mediated RNA remodeling, are used in a concerted fashion to prevent the premature formation of the central pseudoknot and its surrounding elements within the small ribosomal subunit.

Eukaryotic ribosome assembly is a highly dynamic process involving in excess of 200 non-ribosomal proteins and RNAs. This process starts in the nucleolus where rRNAs for the small ribosomal subunit (18S rRNA) and the large ribosomal subunit (25S and 5.8S rRNA) are initially transcribed as part of a large 35S pre-rRNA precursor transcript. Within the 35S pre-rRNA, the 18S rRNA is flanked by the 700-nucleotide 5' external transcribed spacer (5' ETS) and internal transcribed spacer 1 (ITS1), which both have to be removed during later stages of ribosome maturation¹.

The earliest stable precursors of the small subunit (SSU), referred to here as SSU processomes, have been identified on Miller spreads by means of EM as terminal structures of pre-rRNA transcribed by RNA polymerase I (refs. 2,3). In addition to containing the 18S rRNA precursor flanked by the 5' ETS and parts of ITS1, these particles contain the U3 small nucleolar RNA (snoRNA) and a large number of ribosome-biogenesis factors, including U three proteins (Utps)⁴. A subsequent cleavage step between the 5' ETS and the 18S (at site A1) defines the mature 5' end of the 18S rRNA, and iterative processing at the 3' end within ITS1 (either at site A2 or A3) separates the large-subunit and small-subunit maturation pathways. A final endonucleolytic cleavage event at the mature 3' end of the 18S rRNA (D site) results in the final 18S rRNA in the cytoplasm⁵ (**Supplementary Fig. 1**).

The formation of the SSU processome has previously been shown to occur in a stage-specific manner^{6–8}. The 5' ETS serves as a scaffold responsible for the recruitment of a 2-MDa assembly called the 5'-ETS particle. Within this particle, the multisubunit protein complexes UtpA (Utp4, Utp5, Utp8, Utp9, Utp10, Utp15 and Utp17) and UtpB (Utp1, Utp6, Utp12, Utp13, Utp18 and Utp21) are used to chaperone both the 5' ETS and the U3 snoRNA⁹. The U3 snoRNA acts as molecular guide by base-pairing with both the 5' ETS and the sequences within the 18S rRNA^{10,11}. Together, the 5'-ETS particle components provide a platform for the ensuing steps of SSU-processome formation

during which the four rRNA domains of the 18S rRNA (5', central, 3' major and 3' minor) are bound by a set of specific ribosome-assembly factors^{7,8}. Biochemical studies indicated that these factors and the 5'-ETS particle probably contribute to the independent maturation of the domains of the SSU^{7,8}. A vital process during later stages of SSU assembly is the formation of the central pseudoknot and its surrounding elements. Here, multiple RNA and protein elements at the interface of all four domains are juxtaposed and determine the final orientations of those domains within the mature SSU.

Recently, the molecular architectures of SSU processomes from the thermophilic eukaryote *Chaetomium thermophilum* and the yeast *S. cerevisiae* were determined by means of cryo-EM at resolutions of 7.4, 5.1 and 4.5 Å^{12–14}. To obtain these related states of the SSU processome, standard growth conditions¹², nutrient starvation¹³ and depletion of the exosome-associated RNA helicase Mtr4 (ref. 14) were employed. Although these structures have elucidated the general architecture of this early ribosome-assembly intermediate, including the positions of many ribosome-assembly complexes, such as UtpA, UtpB and the U3 snoRNP, limited resolution has so far precluded the complete and unambiguous assignment of all components within the SSU processome.

Here we present the structure of the *S. cerevisiae* SSU processome at an overall resolution of 3.8 Å with local resolutions in the core near 3 Å. The combination of cryo-EM with cross-linking and MS data has enabled us to build an essentially complete near-atomic model of the SSU processome containing three RNAs (5' ETS, pre-18S and U3 snoRNA), 51 ribosome-assembly factors and 15 ribosomal proteins.

RESULTS

Overall structure

As previously reported, we used nutrient starvation to accumulate and purify the SSU processome¹³. This particular state may represent

¹Laboratory of Protein and Nucleic Acid Chemistry, The Rockefeller University, New York, New York, USA. ²Tri-Institutional Training Program in Chemical Biology, The Rockefeller University, New York, New York, USA. ³Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, New York, USA. ⁴These authors contributed equally to this work. Correspondence should be addressed to S.K. (klinge@rockefeller.edu).

Received 19 July; accepted 25 August; published online 25 September 2017; doi:10.1038/nsmb.3472

Table 1 Cryo-EM data collection, refinement and validation statistics

| | Overall map 1 (EMD-8859, PDB-5WLC) | Core map | 3' domain map | Overall map 2 | Central domain map |
|---|---------------------------------------|----------|---------------|---------------|--------------------|
| Data collection and processing | | | | | |
| Magnification | | | | | |
| Voltage (kV) | 300 | | | | |
| Pixel size (Å) | 1.3 | | | | |
| Electron exposure (e ⁻ /Å ²) | 50 | | | | |
| Defocus range (μm) | 0.6–2.6 | | | | |
| Symmetry imposed | C1 | | | | |
| Initial particle images | 772,120 | | | | |
| Final particle images | 284,213 | 284,213 | 284,213 | 123,843 | 43,415 |
| Resolution (Å) | 3.8 | 3.6 | 4.1 | 4.1 | 7.2 |
| FSC threshold | 0.143 | | | | |
| Map-sharpening <i>B</i> factor (Å ²) | -118 | -103 | -163 | -112 | -180 |
| Refinement | | | | | |
| Initial model used | 5TZS | | | | |
| Model composition | | | | | |
| Non-hydrogen atoms | 196,921 | | | | |
| Protein residues | 22098 | | | | |
| RNA bases | 1682 | | | | |
| Ligands | 1 | | | | |
| R.m.s. deviations | | | | | |
| Bond length (Å) | 0.01 | | | | |
| Angles (°) | 0.91 | | | | |
| Validation | | | | | |
| MolProbity score | 1.86 | | | | |
| Clashscore | 8.02 | | | | |
| Poor rotamers (%) | 0.79 | | | | |
| Good sugar puckers (%) | 97.62 | | | | |
| Ramachandran plot | | | | | |
| Favored (%) | 94.44 | | | | |
| Allowed (%) | 5.48 | | | | |
| Outlier (%) | 0.08 | | | | |

a storage particle or a nonproductive assembly intermediate of the SSU in response to these stress conditions^{13,15,16}. Extensive data collection and an improved 3D classification and refinement strategy enabled us to improve the resolution of the entire SSU processome to 3.8 Å (Table 1 and Supplementary Figs. 2 and 3). The core, which contains approximately 80% of the proteins present in the SSU processome, was refined to an overall resolution of 3.6 Å. Large regions in the center of the particle are well resolved near 3 Å (Supplementary Fig. 3), and clear density can be observed for side chains and bases (Supplementary Fig. 4). Focused classification and refinement resulted in considerably improved maps of more peripheral areas of the structure, such as the head region, containing the 5' domain, or the 3' region, containing parts of the 3' domain (both at 4.1-Å overall resolution) (Table 1, Fig. 1 and Supplementary Figs. 2 and 3). These improved maps also feature clear side chain density for peripheral and more exposed areas. In addition, RNA densities in different regions of the SSU processome show clear separation of nucleotides (Supplementary Fig. 4).

As previously shown, the central domain is largely flexible under the growth and purification conditions we used to obtain SSU processomes¹³. Extensive iterative 3D classification has enabled us to isolate one particular conformation of this domain, where density can be visualized at an overall resolution of 7.2 Å (Fig. 1a and Supplementary Figs. 2, 3 and 5).

The combined use of these maps with cross-linking and MS data (Supplementary Fig. 6 and Supplementary Data Set 1) allowed us to unambiguously trace proteins not only in the core but also in the periphery of the particle to obtain the first largely complete near-atomic model of the SSU processome (Figs. 1 and 2, Supplementary Table 1 and Supplementary Notes 1–20). In solvent-exposed regions, such as the central domain, where atomic resolution was not obtained, previously determined crystal structures were fitted or polyaniline models were built *de novo*. We note that in these regions, the sequence register of proteins is less reliable.

Like many eukaryotic superstructures, the SSU processome contains multiple β-propellers (WD40 domains)¹⁷ and helical-repeat structures, in addition to ribosomal components and other ribosome-assembly factors (Fig. 1 and Supplementary Video). Whereas helical-repeat elements are frequently used to encapsulate RNA and protein elements, β-propellers perform a range of different functions. The rigid scaffold provided by seven blades of a β-propeller provides a unique platform for the individual diversification of the exposed loops, which are used for protein–protein as well as RNA–protein recognition in the 20 different β-propellers of the SSU processome. In addition, N- and C-terminal extensions provide further functional regions to interact with RNA and protein elements.

Helical-repeat elements are used toward the top of the structure near the 5' and central domains to encapsulate and stabilize RNA helices

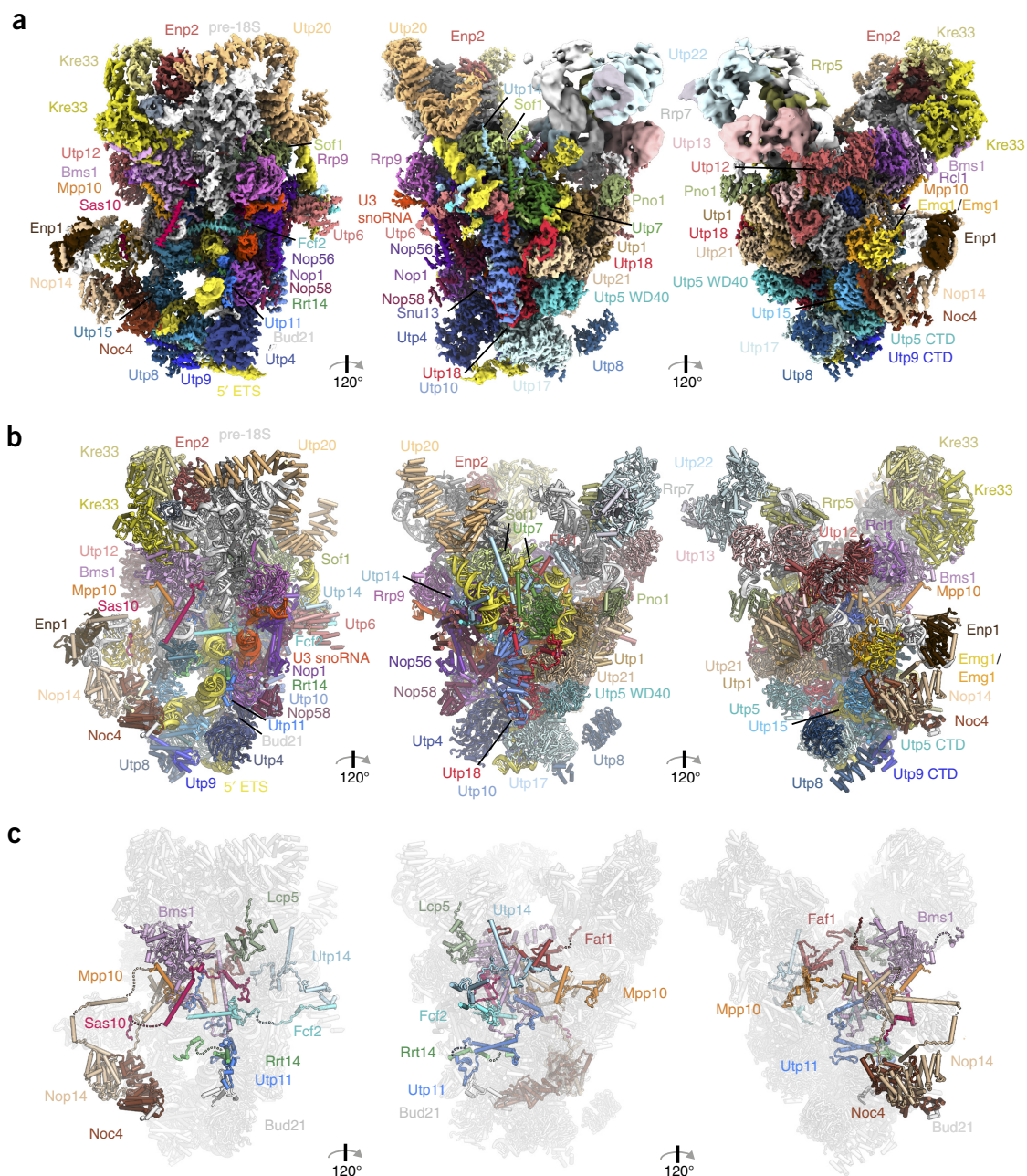


Figure 1 Cryo-EM reconstruction and complete near-atomic model of the *S. cerevisiae* SSU processome. **(a)** Three views of a composite cryo-EM map consisting of the 3.6-Å core, the 4.1-Å overall map 2, the 4.1-Å 3' domain and the 7.2-Å central domain maps. Densities for SSU processome components are color coded with analogous labels. Subunits of complexes are shown in shades of blue (UtpA), red (UtpB), purple (U3 snoRNP), brown (Nop14–Noc4) and light pink (Bms1–Rcl1). Ribosomal proteins are depicted in shades of gray. RNAs are yellow (5' ETS), red (U3 snoRNA) and white (pre-18S rRNA). **(b)** Cartoon representation of the atomic model of the SSU processome with orientations and color scheme same as in **a**. **(c)** Cartoon models of newly identified extended peptides, colored as in **a**, within the whole particle (white).

in a particular conformation. In the context of the 5' domain, Utp20 provides a structural support for RNA expansion segments ES3A and ES3B near the Kre33 heterodimer and Enp2 (**Supplementary Fig. 5**). Similarly, the tetratricopeptide repeat of Rrp5 (ref. 18), which is necessary for pre-18S processing^{19,20}, is positioned in proximity to the UtpC complex²¹, Krr1 (ref.22) and the central domain (**Supplementary Fig. 5**). It provides a cradle to stabilize helix 24 in a different conformation with respect to the mature SSU.

The SSU processome can be subdivided into subcomplexes, which include UtpA, UtpB, UtpC, the U3 snoRNP, the Mpp10 complex and

individual proteins. UtpA and UtpB provide 16 of the 20 β -propellers within the SSU processome, with Sof1, Utp7, Enp2 and Rrp9 (ref. 23) (U3 snoRNP) providing the remaining four (**Fig. 1a,b**).

Limited resolution has previously prevented the correct assignment of all β -propellers within the SSU processome, in particular of Utp4 and Utp5 (UtpA), as well as of Utp18 (UtpB)^{12–14}. It further prohibited the tracing of intricate loops, extensions and linkers that determine the subunit-specific function of these redundant folds. By using high-resolution density maps, RNA–protein⁹ and protein–protein cross-linking data, we were able to assign all subunits of UtpA and UtpB

for subsequent model building (Figs. 3 and 4 and Supplementary Notes 3–7 and 21).

We have identified and built models of ten previously unassigned ribosome-assembly factors within the SSU processome (Figs. 1c and 2 and Supplementary Notes 8–17). These include the 5'-ETS-particle proteins Utp11, Fcf2, Sas10 and Bud21 (Figs. 2 and 5), as well as

later factors such as Faf1, Lcp5, Utp14 and Rrt14 (refs. 7,8). The Nop14–Noc4 submodule²⁴ was identified as a previously unassigned helical structure in the lower half of the particle. In addition, parts of Mpp10 that extend beyond the regions interacting with Imp3 and Imp4 (refs. 25,26) have now been identified (Fig. 2 and Supplementary Note 20).

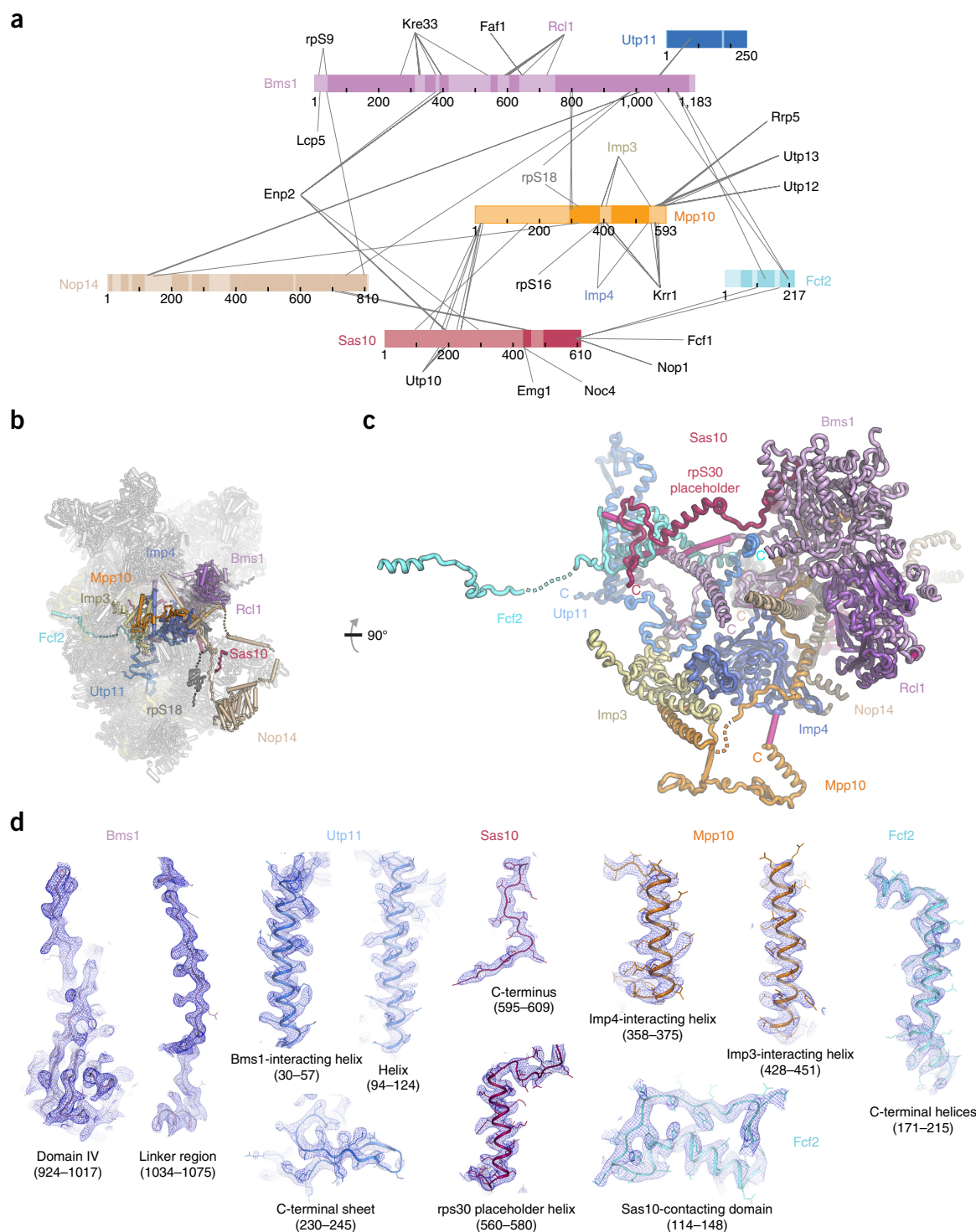


Figure 2 Identification of proteins in the SSU processome. **(a)** Schematic representation of Bms1, Utp11, Nop14, Mpp10, Fcf2 and Sas10, with modeled parts highlighted in darker shades. All DSS cross-links to Bms1, Sas10 and Mpp10 are shown as thin lines. **(b)** View of selected proteins shown in **a** in the context of the SSU processome. **(c)** Cartoon representation of proteins in **b** with cross-links between lysines shown as thick pink lines. α of identified lysines are shown as spheres. **(d)** Selected cryo-EM densities of proteins shown in **a–c**.

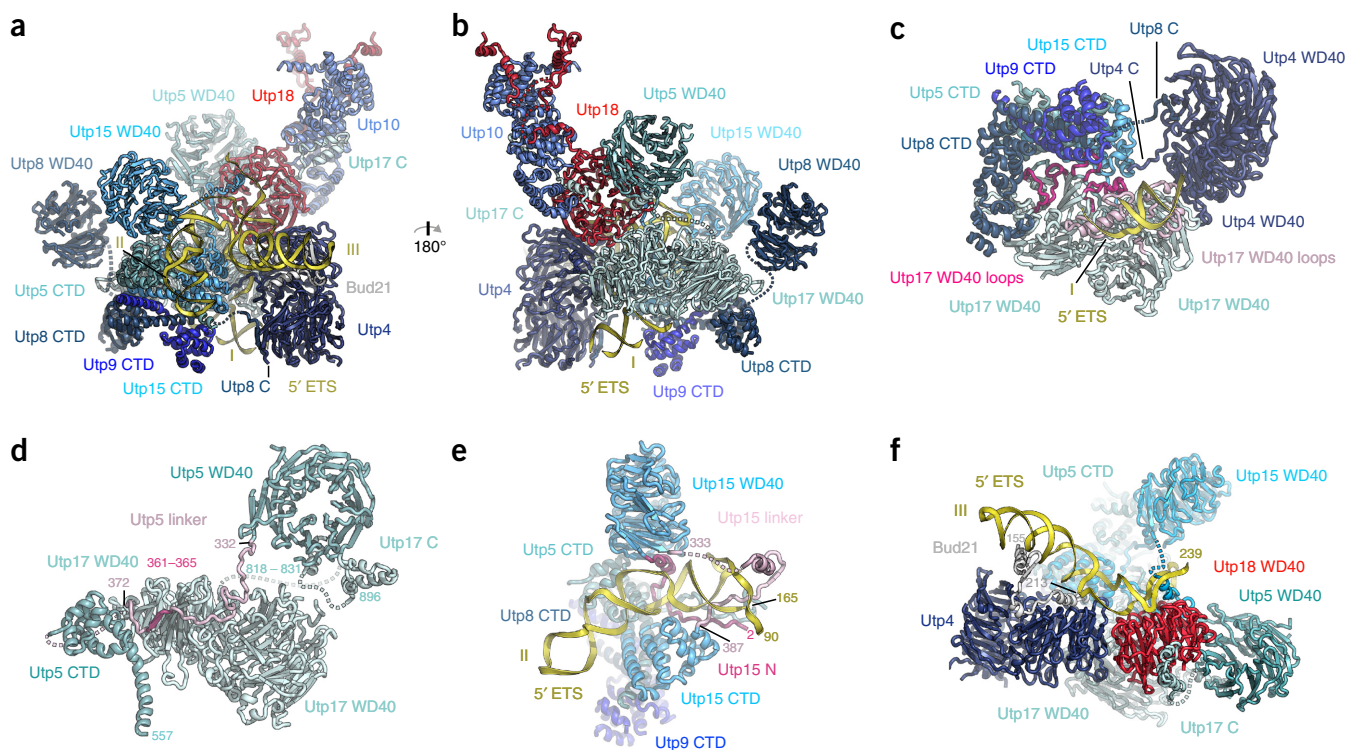


Figure 3 Architecture of the UtpA complex and its interactions with the 5' ETS and Utp18. As in **Figure 1**, UtpA subunits are shown in shades of blue. The 5' ETS is shown in yellow with its helices labeled with roman numerals I, II, and III, and the UtpB subunit Utp18 is depicted in red. All elements are shown in cartoon representation. Helical CTD, WD40 domains and N termini and C termini (N and C, respectively) of all subunits are labeled if applicable. **(a,b)** Two views of all subunits of UtpA and Utp18 bound to the first three helices of the 5' ETS. **(c)** Loops of Utp17 (light and dark pink) contact helix I of the 5' ETS and the CTD tetramer. Utp17 also interacts with Utp4 on its side surface. **(d)** A linker (light pink) between the WD40 domain and the CTD of Utp5 runs along Utp17, forming a β -strand (dark pink). **(e)** The N-terminal extension (dark pink) and a linker of Utp15 (light pink) place its WD40 domain and the CTD on opposite sides of helix II. **(f)** Helix III is coordinated by Utp4 and Bud21. The junction between helix II and III is stabilized by a linker of Utp15 and the WD40 domain of Utp18.

Coordination of the 5' ETS

UtpA forms the base of the SSU processome, where it acts as a central scaffold that recognizes the first three helices (helices I–III) of the 5' ETS (**Fig. 3**). Helix I is bound by a set of loops and helical elements on top of the tandem β -propeller of Utp17 (**Fig. 3c**). Whereas the β -propellers of Utp17 have functionalized top surfaces, Utp15 employs an N-terminal extension to its β -propeller and a long linker between its C-terminal domain (CTD) and WD40 domain to position helix II and to stabilize the junction between helices II and III (**Fig. 3a,e**). The WD40 domain of Utp15 and helix II further provide a binding platform for Noc4, which acts as the foundation of a lateral extension of the UtpA complex where the 3' domain is placed (**Figs. 3e,f** and **5b**). This extension is additionally stabilized by Bud21, which connects Noc4 with Utp4 (UtpA), Nop1 (U3 snoRNP) and helix III, which rests on top of Utp4 (**Figs. 1** and **3a,f**).

A short single-stranded RNA region between helices II and III of the 5' ETS is coordinated by two other β -propeller-containing subunits located next to Utp4 and Utp17. We unambiguously assigned these WD40 domains as Utp5 (UtpA) and Utp18 (UtpB) (**Fig. 3f** and **Supplementary Notes 5** and **6**).

Utp5 is integrated within UtpA through its CTD and a linker peptide, which runs along a conserved groove of Utp17 (**Fig. 3d**). Akin to this interaction, a C-terminal peptide expansion of Utp17 contacts the β -propeller of Utp5 before binding to Utp10, the only subunit of UtpA composed solely of helical repeats. Utp5, Utp10 (UtpA), Utp21 and Utp18 (UtpB) form the junction between the UtpA and UtpB complexes. Within this junction, Utp18 serves as a central nexus.

The placement of the WD40 domain of Utp18 between three UtpA subunits (Utp4, Utp17 and Utp5) and near two UtpA linker regions (Utp5 and Utp15) interlocks the two largest subcomplexes of the SSU processome (**Figs. 1a,b, 3** and **4**).

Like Utp17, Utp18 employs extensive peptide-like motifs to facilitate protein–protein interactions (**Figs. 4** and **5a**). Three regions within the 230-residue N terminus of Utp18 interact with the UtpB subunits Utp6, Utp21, the U3 snoRNP component Nop58 and Utp10 (UtpA) (**Figs. 3a,b, 4a–d** and **5a**). The first two segments are employed to interact with Utp6 (residues 13–28) and Utp10 (residues 29–44). The third segment (residues 123–183) forms an intricate interface with the surface of the first β -propeller of Utp21 and a conserved C-terminal peptide of Nop58 (**Figs. 4c** and **5a**). Additionally an Mtr4 arch-interacting motif (AIM)²⁷ is located in a disordered region between the second and third segment of the N terminus of Utp18 (**Fig. 4c**). The WD40 domain of Utp18 interacts with Snu13 (U3 snoRNP) and stabilizes a single-stranded region of the 5' ETS immediately upstream of the 3' hinge (nucleotides 275–280). Downstream of the 3'-hinge duplex (nucleotides 293–332), the 5' ETS is mostly single stranded with a short stem–loop (nucleotides 299–308) that is stabilized by Utp21 and Utp1 on one side and by Utp18 and Utp7–Sof1–Utp14 on the other side (**Fig. 4d,e**). The following single-stranded region between helices V and VI of the 5' ETS (nucleotides 393–396) is bound by Imp3, which interacts with an N-terminal helix (residues 2–34) of Imp4. Together with the UtpB tetramer, Imp3 and Imp4 provide a binding surface for Mpp10 (**Fig. 2a–c**).

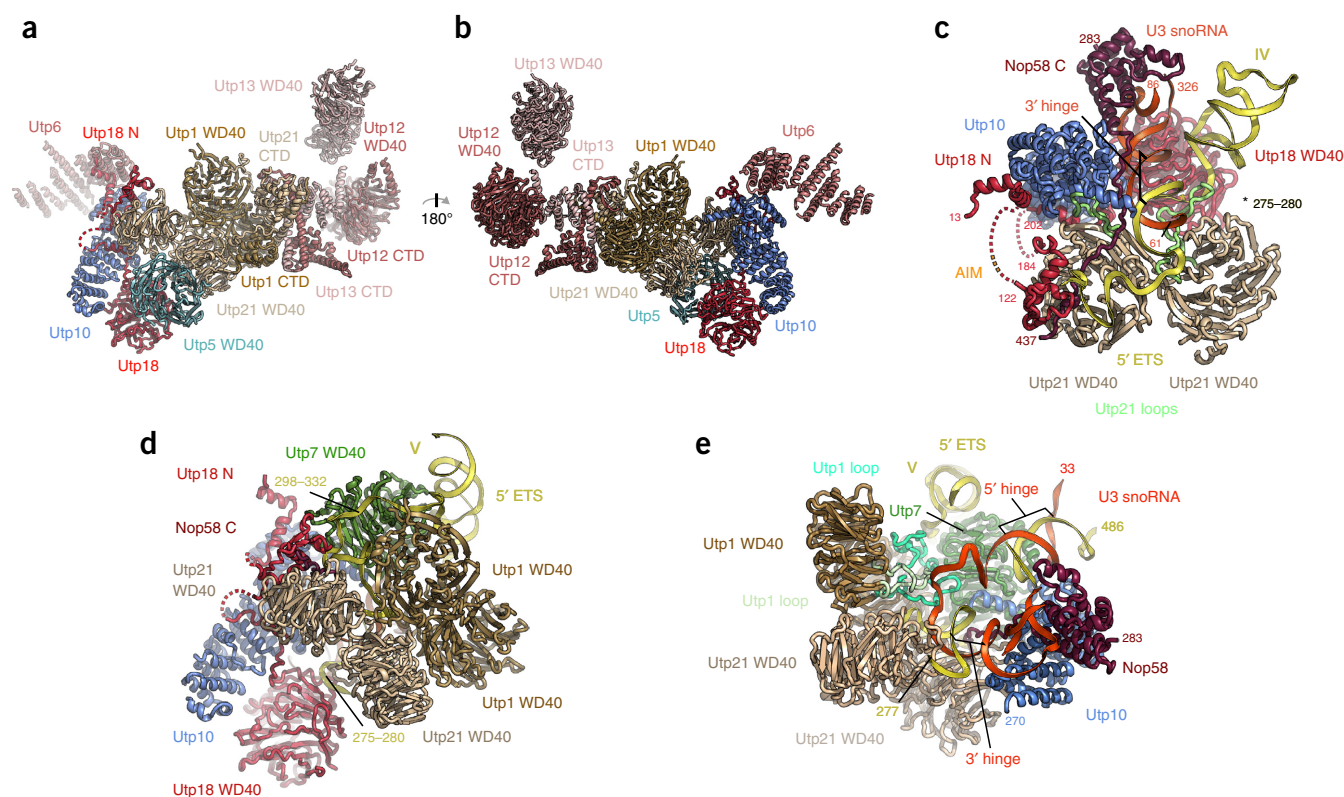


Figure 4 Architecture of the UtpB complex and its interactions with RNA, the UtpA complex, and Utp7. UtpB subunits are colored in different shades of red and brown. The 5' ETS is shown in yellow with its helices labeled with roman numerals (IV and V), and the U3 snoRNA is shown in red with important sequence elements (3' hinge, 5' hinge) indicated. The UtpA subunits Utp10 and Utp5 are depicted in blue, and Utp7 and Nop58 are illustrated in dark green and burgundy, respectively. All elements are shown in cartoon representation. Helical CTD, WD40 domains and N and C termini (N, C) of all subunits are labeled if applicable. (a,b) Two views of all subunits of UtpB, the Utp5 WD40 domain and Utp10. (c) A single-stranded region of the 5' ETS (275–280) leading into the 3'-hinge duplex is stabilized by Utp18 and a long loop (pale green) of Utp21. A second loop (pale green) of Utp21 binds Utp10. The exosome-recruiting AIM present in a linker of the N-terminal region of Utp18 is depicted in orange. (d) A predominantly single-stranded region of the 5' ETS (298–332) is chaperoned by the N-terminal linker of Utp18 and the WD40 domains of Utp21, Utp1 and Utp7. (e) Two loops of Utp1 separate the 5' ETS and the U3 snoRNA between the 3' and 5' hinges.

Roles of extended proteins in the SSU processome

Many of the newly identified proteins share striking structural properties, as they all contain long linkers that are used to weave through the structure and connect distant parts within the SSU processome (Figs. 1c and 5). At the base of the structure, Bud21 acts as a connector between Utp4 and Noc4 (Supplementary Notes 7–9), while a conserved C-terminal proline-glycine-rich sequence of Rrt14 (Supplementary Note 12) stabilizes the interaction of the L1-domain-containing (Supplementary Fig. 7) protein Utp30 with helix IV of the 5' ETS (Fig. 5b,f). In addition, Rrt14 contacts Utp11, located on the other side of helix IV where Utp11 interacts with Nop1 and Bud21, and connects this region with the core of the SSU processome (Fig. 5c). Sas10, Utp11 and Fcf2 (Supplementary Notes 11,13,14) each span at least 100 Å and are used to interconnect the U3 snoRNP with other important regions of the SSU processome (Figs. 1c and 5a–c). Strikingly, Nop1 (fibrillarin) is used as a binding platform for five proteins (Fcf2, Sas10, Utp24, Utp11 and Bud21) (Fig. 5g). The surfaces of the two Nop1 subunits—one located at the lower and the other at the upper end of U3 snoRNA—are used distinctively by these proteins. Here, peptide backbone elements are used to form shared secondary structure elements within a β -barrel (Fcf2) or an extended β -sheet within Nop1 (Utp11, Bud21). Peptides from Sas10 and Utp24 interact similarly with Nop1 (Fig. 5g).

Utp14, Faf1, Lcp5 and Mpp10 are located near the top of the core of the SSU processome (Fig. 1c). Here, Utp14 forms a highly unusual split structure in a functionally important region near the Utp7–Sof1 dimer and the A1 cleavage site (Fig. 5a,d and Supplementary Note 16). An N-terminal segment of Utp14 is used to connect Sof1 with Utp6, and a separate C-terminal segment of Utp14 links Utp7 with Sof1. Several hundred residues of Utp14 connect these two fragments. This segment contains the binding site for Dhr1, the essential DEAH box helicase that is responsible for displacing U3 snoRNA from early ribosome-assembly intermediates^{28,29}. Faf1 is positioned in proximity to Utp14 and Utp7 and directly interacts with U3 snoRNA and pre-18S rRNA near the A1 cleavage site (Fig. 1b,c and Supplementary Note 15). Lcp5 interacts extensively with rRNA of the 5' domain and rpS9 (Supplementary Note 17), whereas Mpp10 extends from Bms1 via Imp4 and Imp3 to the UtpB complex and helix 44. Several of the newly identified proteins (Fcf2, Sas10, Mpp10, Utp11 and Nop14) interact with the centrally positioned GTPase Bms1 (ref. 30) (Fig. 2, Supplementary Fig. 8 and Supplementary Note 19). In addition to the translational GTPase fold, Bms1 contains a long C-terminal helix, which is used to anchor this enzyme in the core of the SSU processome.

In contrast to most of the other newly identified proteins, which have no visible globular domains, Nop14 contains a core helical repeat in addition to its long linker peptides (Supplementary Note 10).

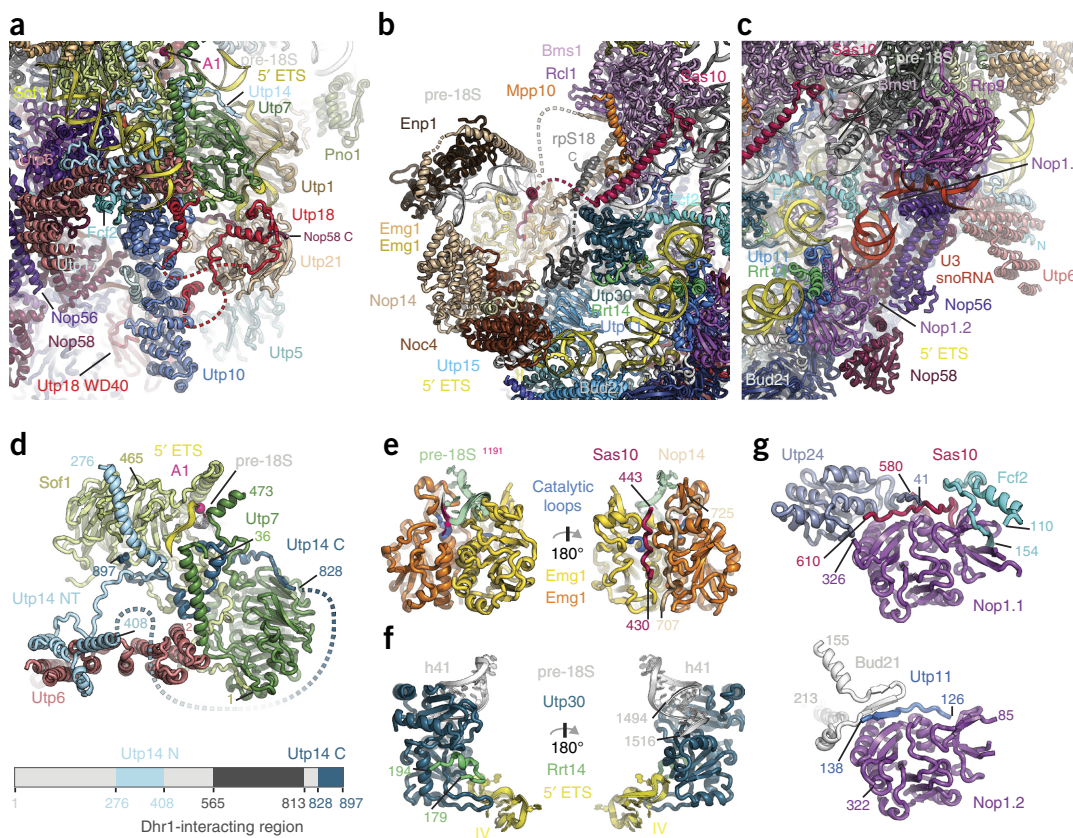


Figure 5 Diverse roles of peptides in the SSU processome. (a–c) Views of the network of peptides traversing through the SSU processome. (a) Interactions of Utp18 (UtpB) with Utp10 (UtpA), Nop58 (U3 snoRNP), Utp21 and Utp6 (UtpB). Sof1, Utp7 and Utp14 organize the A1 cleavage site (pink sphere). (b) Nop14, Noc4 and Enp1 cover the 3' domain of the pre-18S RNA and the methyltransferase Emg1. (c) Fcf2, Utp11, Bud21 and Sas10 bind to the U3 snoRNP (shades of purple) and extend through the SSU processome. (d) Accommodation of the A1 cleavage site (pink) by Utp7, Sof1, Utp14, and Utp6. N- and C-terminal parts of Utp14 are colored light blue and dark blue, respectively. A schematic representation of modeled parts of Utp14 (shades of blue) is shown below. (e) Views of the Emg1 homodimer (orange and yellow) with catalytic loops (blue). Substrate pre-18S RNA (green) with target nucleotide (1191, pink) located in one active site while peptides of Sas10 (pink) and Nop14 (beige) occupy the other. (f) Utp30 and Rrt14 recognize the pre-18S and 5'-ETS RNA. (g) Cartoon representation of the two copies of Nop1 (Nop1.1, Nop1.2) and interacting segments of Utp24, Sas10, Fcf2, Bud21 and Utp11.

By directly binding to a second repeat protein, Noc4, the core domain of Nop14 further expands into a larger repeat structure. This composite helical repeat serves as a lateral structural extension of the UtpA complex and provides a scaffold for Enp1, Emg1 and parts of the 3' domain of the 18S rRNA (Fig. 5b and Supplementary Notes 9 and 10), which later forms the beak structure in the mature SSU. The repeat of Nop14 is flanked by N- and C-terminal extensions. A C-terminal 75-amino-acid helix docks Nop14 into an opening between the Mpp10–Imp4 dimer and the Bms1–Rcl1 complex³¹ and points its C-terminal end toward the central cavity between the central and the 5' domains (Fig. 2c). The N-terminal extension of Nop14 binds Enp1, which caps the 3' domain of the 18S rRNA. Surprisingly, the N-terminal segment of Nop14 loops around this entire region and extends back to where the C-terminal long helix is located (Supplementary Note 10). Similar to Bms1, terminal extensions are used to fully integrate the Nop14–Noc4 complex within the center of the SSU processome (Figs. 1c and 5b). In addition to stabilizing Enp1 on top of the rRNA, this arrangement positions the rRNA substrate in the active site of one of the Emg1 dimers. Peptides from Nop14 and Sas10 are used to provide structural support for the dimeric Emg1 while also blocking the active site of one of its subunits. This enforces a structural asymmetry of the two Emg1 methyltransferase subunits so that only one active site is available for the methylation of base 1191 of the 3' domain (Fig. 5e).

Near the base of helix 41 of the 18S rRNA, we have identified rpS18, a ribosomal protein that is already positioned in a near-mature configuration with respect to the 18S rRNA (Fig. 5b and Supplementary Note 18). The C terminus, which binds elements of the beak structure in the mature 40S subunit, adopts a different conformation in the context of the SSU processome, where the beak structure has not yet formed. Here, the C terminus of rpS18 is stabilized by the long C-terminal helix of Nop14, as well as domain IV of Bms1 and a linker region of Mpp10 (Fig. 5b).

Protein-assisted RNA remodeling by U3 snoRNA

U3 snoRNA occupies a central position within the SSU processome and reaches from the outside into the core of the particle (Fig. 6a). By base-pairing with its 5' and 3' hinges to nucleotides within the 5' ETS, it rigidifies the structural scaffold provided by the 5' ETS, which has been described previously¹³. The 5' end of U3 snoRNA reaches further into the center of the SSU processome and base-pairs with two regions of the pre-18S rRNA. A similar base-pairing pattern was recently proposed for the SSU processome captured in a state upon Mtr4 depletion¹⁴. Whereas Box A (U3 nucleotides 16–22) is base-paired with nucleotides 9–15 of the pre-18S rRNA and organizes the pre-18S 5' end near the A1 cleavage site, Box A' (U3 nucleotides 3–13) is base-paired with nucleotides 1111–1122 and reorganizes this region that is later in proximity

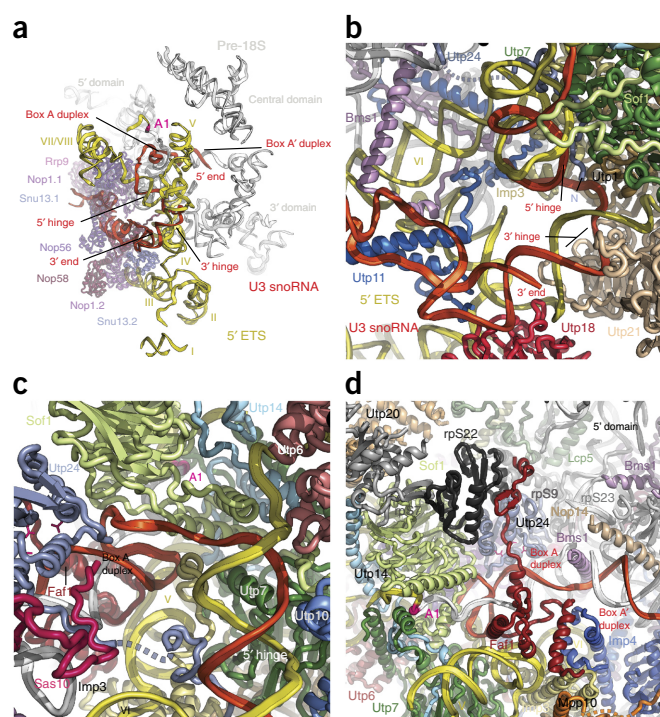


Figure 6 U3 snoRNA-mediated RNA remodeling. (a) Overview of the U3 snoRNP proteins (purple), U3 snoRNA (red) and their interactions with the 5' ETS (yellow) and pre-18S RNA (white). (b) View of protein elements stabilizing the 3' and 5' hinges. (c) View of proteins organizing the 5' hinge and the Box A duplex. (d) Overview of the Box A and Box A' duplexes of U3 snoRNA with pre-18S RNA. (a–d) Proteins and RNAs are color coded. The N terminus of Utp24 is indicated with an N. Helices of the 5' ETS are labeled with roman numerals. Important functional sequence elements of the U3 snoRNA (5' hinge, 3' hinge, Box A duplex, Box A' duplex) and visible domains of the 18S rRNA (3', 5' and central domain) are indicated. The cleavage site A1 is shown as a pink sphere. Active site residues of Utp24 are highlighted as pink sticks.

to the central pseudoknot (Fig. 6a and Supplementary Fig. 9). A range of ribosome-assembly factors is responsible for the stabilization of the four RNA duplexes that U3 snoRNA forms with the 5' ETS and the 18S precursor. Toward the outside of the particle, three proteins of the UtpB complex (Utp18, Utp21 and Utp1) stabilize the 3' hinge. While Utp18 is involved in rigidifying the junctions between helices II, III and IV of the 5' ETS, Utp21 and Utp10 (UtpA) form a clamp around the 3' hinge (Fig. 4e). Utp1 is used to bind the single-stranded regions of both U3 snoRNA and 5' ETS with two long, structured loops (residues 556–580 and 616–680). These loops act as a rudder, thereby separating the 5' ETS and the U3 snoRNA (Fig. 4e).

A short loop of U3 snoRNA and the 5' hinge are coordinated by Imp3, Utp11, Bms1 and the N terminus of Utp24 (Fig. 6b). Upstream of the 5' hinge, U3 snoRNA is stabilized by the Sof1–Utp7–Utp14 complex, which also provides a composite binding site for a single-stranded RNA that contains the A1 cleavage site (Figs. 5d and 6c,d). The Box A and Box A' duplexes are organized by the long C-terminal helices of Nop14 and Bms1, as well as Utp24, Faf1 and rpS23. Importantly, the captured state of the SSU processome contains an A0-cut precursor¹³ in a pre-A1-cleavage state in which Utp24 is positioned close to its substrate but cleavage has not yet occurred. Faf1 is positioned between helix V of the 5' ETS and the Box A duplex near Imp3 and Imp4. A linker of Faf1 locks the Box A duplex in place and interacts with rpS22 on the opposite side, thereby occluding access to the active site of the nuclease Utp24 (Fig. 6d).

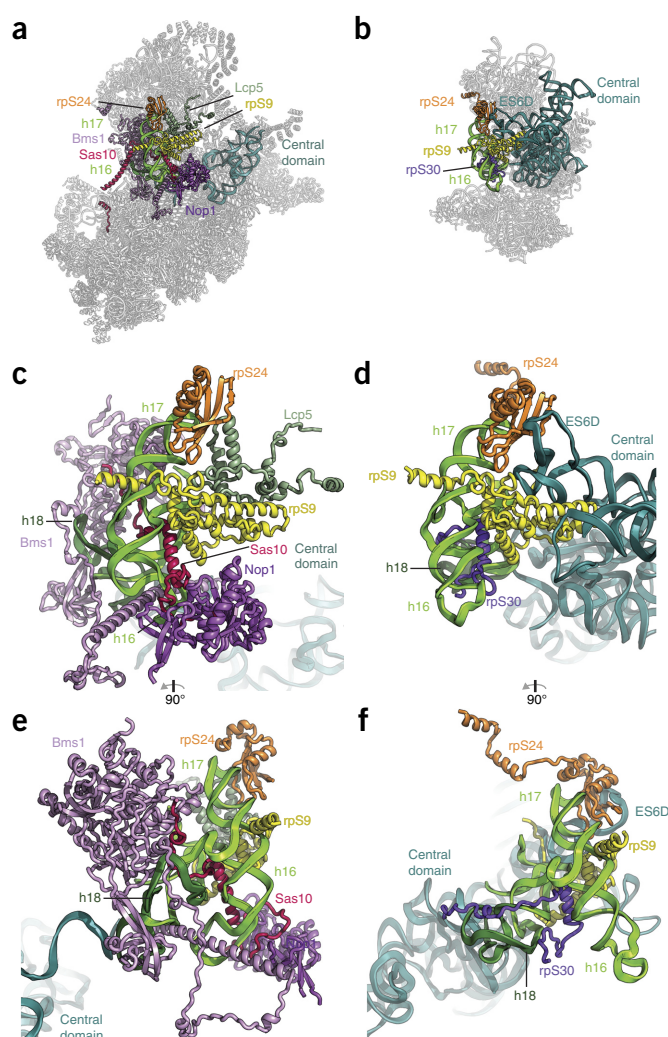


Figure 7 Steric hindrance and molecular mimicry prevent premature rRNA folding. (a) Conformations of helices h16, h17 (green) and h18 (dark green) of the 5' domain and the central domain (teal) of the pre-18S RNA within the SSU processome (gray). Interacting ribosome-assembly factors and ribosomal proteins are indicated. (b) Conformation of the same elements as in a in the context of the mature SSU (PDB 4V88) (ref. 33). (c,d) Detailed views of the conformations of h16 and the central domain in the SSU processome (c) and the SSU (d). Sas10 mimics rpS30 and occupies its binding site on h16. Lcp5 blocks the central domain from occupying its mature position by steric hindrance while Bms1 bends h16. (e,f) Orthogonal views to panels c and d.

RNA remodeling prevents central-pseudoknot formation

In the context of the SSU processome, Lcp5 and Sas10 are multifunctional proteins. In addition to Utp18, which contains a peptide that can interact with the exosome-associated helicase Mtr4 (ref. 27), Lcp5 and Sas10 contain exosome-interaction motifs (Sas10 domains)³². Lcp5 is positioned next to rpS9 and the 5' domain, a location that is occupied by expansion segment ES6D of the central domain in the mature SSU (Fig. 7). In particular, the presence of ES6D near rpS9 indicates a mature conformation of the 5' and central domains with respect to each other, which is not the case in the context of the SSU processome. In contrast to those of Lcp5, the ordered parts of Sas10 are more elongated. Sas10 contains a solvent-exposed N-terminal Sas10 domain, blocks the active site of Emg1 with a short peptide and bridges between the 5' and 3' domains with a long helix near Utp30 (Fig. 1b,c). In addition, it interacts with the U3 snoRNP through its

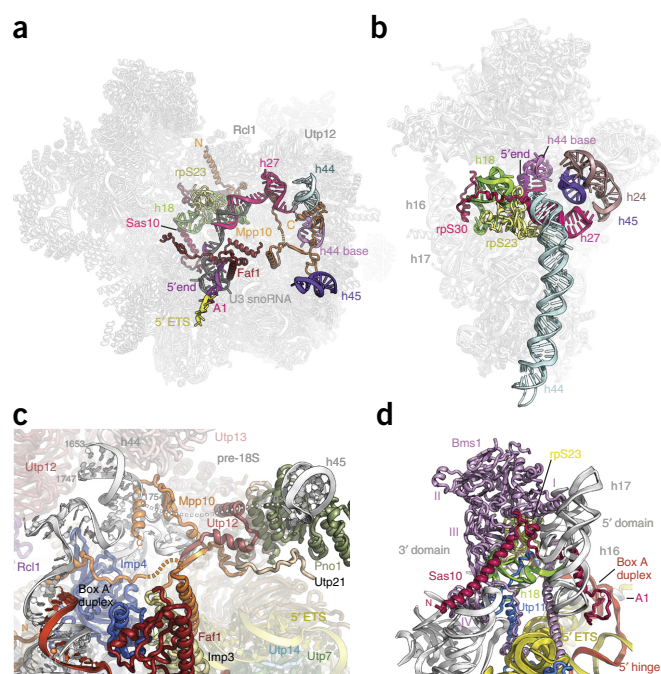


Figure 8 RNA remodeling prevents central-pseudoknot formation. (a) The central pseudoknot and 18S rRNA elements in its vicinity are shown in their immature positions in the SSU processome (gray) and labeled with their corresponding mature 18S rRNA helix numbers. Chaperoning RNA, ribosomal proteins and ribosome-assembly factors are color coded and the A1 cleavage site is highlighted in pink. Termini of Mpp10 are indicated with N and C. (b) Color coded RNA elements close to the central pseudoknot in the mature SSU (gray) labeled as in **a** (PDB 4V88) (ref. 33). (c) Mpp10 (orange) and its interactions with pre-18S RNA (white). Elements of the 18S rRNA (h44 and h45) and U3 snoRNA (Box A' duplex, red) are labeled. Nucleotide positions of the pre-18S RNA are indicated by white numbers. (d) Bms1-mediated remodeling of h18 (green) of the pre-18S RNA (white). Domains of Bms1 (purple) are numbered with roman numerals. Structural elements (h16, h17) and domains of the 18S rRNA (3' domain, 5' domain), as well as the U3 snoRNA (5' hinge, Box A duplex), are labeled. Other factors assisting in remodeling (Utp11, Sas10) and the ribosomal protein rpS23 are shown.

C terminus and is involved in RNA-protein remodeling by adopting a similar conformation and occupying the same RNA-binding site as rpS30 in the mature SSU, close to rpS9 (Figs. 2c, 5b,c and 7).

The continued presence of Sas10 or Lcp5 is therefore mutually exclusive with the mature SSU conformation in terms of protein and RNA occupancy, respectively, and may therefore signal an incomplete or faulty assembly state during later stages of ribosome assembly.

Within the SSU processome, the four structured domains of the 18S rRNA (5', central, 3' major and 3' minor domain) are segregated into different regions of the particle, thereby facilitating their separate maturation¹³. RNA elements that are positioned in the vicinity of the central pseudoknot in the mature SSU³³ are distinctly separated in the SSU processome (Fig. 8a,b and Supplementary Fig. 5). This separation is organized through multiple mechanisms. RNA-mediated chaperoning of sequences close to the central pseudoknot in the 5' and central domains is accomplished through base-pairing with U3 snoRNA boxes A and A' (Fig. 6a,c,d). As a consequence of these interactions, sequences vicinal to the central pseudoknot, such as helix 27, adopt different conformations in the SSU processome, where a new stem-loop forms between Rcl1 and Utp12 (Fig. 8a).

Mpp10 plays a central role in the remodeling of nucleotides close to helices 44 and 45. A partial unwinding of the region upstream of

helix 44 results in an RNA loop (nucleotides 1628–1639) that is stabilized by Mpp10 (Fig. 8c). Owing to this partial unwinding, 16 nucleotides of the opposite strand (nucleotides 1755–1769) are available to serve as a linker to helix 45, which is positioned 60 Å away on top of Pno1, which is held in place by Utp1 and Utp21 of the UtpB complex (Fig. 8c).

A second important location for protein-mediated RNA remodeling is the binding site of ribosomal protein rpS23, which is positioned in the mature SSU close to all other remodeled RNA elements next to helix 18 (Fig. 8b). In the SSU processome, conserved elements of Bms1, Utp11 and Sas10 are employed in a concerted fashion to dramatically remodel helix 18 (h18, nucleotides 558–590) of the 18S rRNA (Figs. 7c–f and 8d). The C-terminal-linker region and domain IV of Bms1 together with the conserved N-terminal segment of Utp11 and a conserved linker region of Sas10 stabilize the remodeled RNA as well as rpS23, which is located in proximity to domains I–III of Bms1. Combined with the chaperone functions of Utp20 and Rrp5 in the 5' and central domains (Supplementary Fig. 5), these examples highlight the precise and elaborate mechanisms that are employed to control tertiary interactions between rRNA domains.

DISCUSSION

Emerging eukaryotic regulatory mechanisms

The evolution of eukaryotes has been accompanied by the emergence of dedicated multiprotein complexes that are uniquely suited to fulfill specific tasks. Eukaryotic ribosome biogenesis is catalyzed by a complex assembly of specialized protein factors, many of which have no counterparts in prokaryotes. The near-atomic structure of the SSU processome provides a molecular snapshot of approximately one-quarter of these 200 factors. These proteins share an essential role in providing an additional level of control to ribosome biogenesis. This is achieved through concerted RNA remodeling that prevents the premature formation of the junction between all four 18S rRNA domains.

Defects in ribosome assembly are associated with a set of human diseases called ribosomopathies^{34,35}, of which some can now be mapped in a structural context (Supplementary Fig. 10). In particular, two residues near the Bms1–rpS23 interface have been implicated in human diseases. While one residue is positioned in a dynamic loop of rpS23 (ref. 36) (Supplementary Fig. 10d,e,g), the other is located within Bms1 (ref. 37) and interacts with rpS23 (Supplementary Fig. 10f). It is tempting to speculate that mutations in either of these residues may affect the structural transitions required for the resulting incorporation of rpS23 into the mature SSU.

A second function of the eukaryotic ribosome biogenesis machinery is the encapsulation and guided stabilization of a series of early ribosome-assembly intermediates, which ultimately result in the mature SSU^{9,38}. The formation of the SSU processome may be directed by an initially flexible set of peptide interactions. As SSU processome maturation progresses, the number of these peptide interactions likely increases and results in this stable nucleolar superstructure.

The current state of the SSU processome highlights the need for extensive RNA and protein remodeling by specialized enzymes. This is exemplified by the numerous base-pairing interactions between U3 snoRNA and 18S rRNA, which need to be unwound by enzymes such as Dhr1 to facilitate further maturation of the SSU. In addition, the order of the catalytic reactions performed by enzymes within the SSU processome, such as the acetyltransferase and helicase Kre33, the methyltransferase Emg1, the GTPase Bms1 and the nuclease Utp24, is still unknown.

Ultimately, the separation of the 5' ETS, U3 snoRNA and 18S rRNA precursor represents a major hurdle. To accomplish this separation, degradation of the 5' ETS and release of the 18S rRNA precursor may be coupled via the three proteins (Utp18, Lcp5 and Sas10) that can interact with the RNA surveillance machinery. This approach would provide an elegant mechanism to combine the recycling of ribosome-assembly factors, release of a pre-40S particle and quality control.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the [online version of the paper](#).

Note: Any Supplementary Information and Source Data files are available in the [online version of the paper](#).

ACKNOWLEDGMENTS

We thank M. Ebrahim and J. Sotiris for outstanding support with data collection at the Evelyn Gruss Lipper Cryo-EM resource center at The Rockefeller University. We further thank T. Walz, G. Alushin and Y. Shi for helpful discussions. J.B. is supported by an EMBO long-term fellowship (ALTF 51-2014) and a Swiss National Science Foundation fellowship (155515). M.C.-M. is supported by a postgraduate scholarship from the Natural Sciences and Engineering Research Council of Canada (NSERC) and S.K. is supported by the Robertson Foundation, the Alfred P. Sloan Foundation, the Irma T. Hirschl Trust, the Alexandrine and Alexander L. Sinsheimer Fund, the Human Frontier Science Program Career Development Award and the NIH New Innovator Award (1DP2GM123459). B.T.C. is supported by National Institute of Health grant nos. P41GM103314 and P41GM109824.

AUTHOR CONTRIBUTIONS

S.K. established purification conditions, and M.C.-M. and J.B. acquired cryo-EM data. K.R.M. performed MS experiments and analyzed the resulting data with B.T.C. J.B., M.C.-M., M.H. and S.K. determined the cryo-EM structure of the yeast SSU processome, built the atomic model, interpreted the results and wrote the manuscript. All authors edited the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Woolford, J.L. Jr. & Baserga, S.J. Ribosome biogenesis in the yeast *Saccharomyces cerevisiae*. *Genetics* **195**, 643–681 (2013).
2. Mougey, E.B. *et al.* The terminal balls characteristic of eukaryotic rRNA transcription units in chromatin spreads are rRNA processing complexes. *Genes Dev.* **7**, 1609–1619 (1993).
3. Osheim, Y.N. *et al.* Pre-18S ribosomal RNA is structurally compacted into the SSU processome prior to being cleaved from nascent transcripts in *Saccharomyces cerevisiae*. *Mol. Cell* **16**, 943–954 (2004).
4. Dragon, F. *et al.* A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature* **417**, 967–970 (2002).
5. Fatica, A., Tollervey, D. & Dlakić, M. PIN domain of Nob1p is required for D-site cleavage in 20S pre-rRNA. *RNA* **10**, 1698–1701 (2004).
6. Pérez-Fernández, J., Román, A., De Las Rivas, J., Bustelo, X.R. & Dosil, M. The 90S preribosome is a multimodular structure that is assembled through a hierarchical mechanism. *Mol. Cell. Biol.* **27**, 5414–5429 (2007).
7. Chaker-Margot, M., Hunziker, M., Barandun, J., Dill, B.D. & Klinge, S. Stage-specific assembly events of the 6-MDa small-subunit processome initiate eukaryotic ribosome biogenesis. *Nat. Struct. Mol. Biol.* **22**, 920–923 (2015).
8. Zhang, L., Wu, C., Cai, G., Chen, S. & Ye, K. Stepwise and dynamic assembly of the earliest precursors of small ribosomal subunits in yeast. *Genes Dev.* **30**, 718–732 (2016).
9. Hunziker, M. *et al.* UtpA and UtpB chaperone nascent pre-ribosomal RNA and U3 snoRNA to initiate eukaryotic ribosome assembly. *Nat. Commun.* **7**, 12090 (2016).
10. Marmier-Gourrier, N., Cléry, A., Schlotter, F., Senty-Ségault, V. & Branlant, C. A second base pair interaction between U3 small nucleolar RNA and the 5'-ETS region is required for early cleavage of the yeast pre-ribosomal RNA. *Nucleic Acids Res.* **39**, 9731–9745 (2011).
11. Dutca, L.M., Gallagher, J.E.G. & Baserga, S.J. The initial U3 snoRNA:pre-rRNA base pairing interaction required for pre-18S rRNA folding revealed by in vivo chemical probing. *Nucleic Acids Res.* **39**, 5164–5180 (2011).
12. Kornprobst, M. *et al.* Architecture of the 90S pre-ribosome: a structural view on the birth of the eukaryotic ribosome. *Cell* **166**, 380–393 (2016).
13. Chaker-Margot, M., Barandun, J., Hunziker, M. & Klinge, S. Architecture of the yeast small subunit processome. *Science* **355**, eaal1880 (2017).
14. Sun, Q. *et al.* Molecular architecture of the 90S small subunit pre-ribosome. *eLife* **6**, e22086 (2017).
15. Talkish, J. *et al.* Disruption of ribosome assembly in yeast blocks cotranscriptional pre-rRNA processing and affects the global hierarchy of ribosome biogenesis. *RNA* **22**, 852–866 (2016).
16. Kos-Braun, I.C., Jung, I. & Koš, M. Tor1 and CK2 kinases control a switch between alternative ribosome biogenesis pathways in a growth-dependent manner. *PLoS Biol.* **15**, e2000245 (2017).
17. Makarova, K.S., Wolf, Y.I., Mekhedov, S.L., Mirkin, B.G. & Koonin, E.V. Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucleic Acids Res.* **33**, 4626–4638 (2005).
18. Khoshnevis, S. *et al.* The DEAD-box protein Rok1 orchestrates 40S and 60S ribosome assembly by promoting the release of Rrp5 from pre-40S ribosomes to allow for 60S maturation. *PLoS Biol.* **14**, e1002480 (2016).
19. Torchet, C. & Jacq, C. Hermann-Le Denmat, S. Two mutant forms of the S1/TPR-containing protein Rrp5p affect the 18S rRNA synthesis in *Saccharomyces cerevisiae*. *RNA* **4**, 1636–1652 (1998).
20. Eppens, N.A., Rensen, S., Granneman, S., Raué, H.A. & Venema, J. The roles of Rrp5p in the synthesis of yeast 18S and 5.8S rRNA can be functionally and physically separated. *RNA* **5**, 779–793 (1999).
21. Lin, J., Lu, J., Feng, Y., Sun, M. & Ye, K. An RNA-binding complex involved in ribosome biogenesis contains a protein with homology to tRNA CCA-adding enzyme. *PLoS Biol.* **11**, e1001669 (2013).
22. Zheng, S., Lan, P., Liu, X. & Ye, K. Interaction between ribosome assembly factors Krr1 and Faf1 is essential for formation of small ribosomal subunit in yeast. *J. Biol. Chem.* **289**, 22692–22703 (2014).
23. Zhang, L., Lin, J. & Ye, K. Structural and functional analysis of the U3 snoRNA binding protein Rrp9. *RNA* **19**, 701–711 (2013).
24. Kühn, H. *et al.* The Noc-domain containing C-terminus of Noc4p mediates both formation of the Noc4p-Nop14p submodule and its incorporation into the SSU processome. *PLoS One* **4**, e8370 (2009).
25. Lee, S.J. & Baserga, S.J. Imp3p and Imp4p, two specific components of the U3 small nucleolar ribonucleoprotein that are essential for pre-18S rRNA processing. *Mol. Cell. Biol.* **19**, 5441–5452 (1999).
26. Zheng, S. & Ye, K. Purification, crystallization and preliminary X-ray diffraction analysis of Imp3 in complex with an Mpp10 peptide involved in yeast ribosome biogenesis. *Acta Crystallogr. F Struct. Biol. Commun.* **70**, 918–921 (2014).
27. Thoms, M. *et al.* The exosome is recruited to RNA substrates through specific adaptor proteins. *Cell* **162**, 1029–1038 (2015).
28. Zhu, J., Liu, X., Anjos, M., Correll, C.C. & Johnson, A.W. Utp14 recruits and activates the RNA helicase Dhr1 to unlock U3 snoRNA from the pre-ribosome. *Mol. Cell Biol.* **36**, 965–978 (2016).
29. Sardana, R. *et al.* The DEAH-box helicase Dhr1 dissociates U3 from the pre-rRNA to promote formation of the central pseudoknot. *PLoS Biol.* **13**, e1002083 (2015).
30. Gelperin, D., Horton, L., Beckman, J., Hensold, J. & Lemmon, S.K. Bms1p, a novel GTP-binding protein, and the related Tsr1p are required for distinct steps of 40S ribosome biogenesis in yeast. *RNA* **7**, 1268–1283 (2001).
31. Delprato, A. *et al.* Crucial role of the Rcl1p-Bms1p interaction for yeast pre-ribosomal RNA processing. *Nucleic Acids Res.* **42**, 10161–10172 (2014).
32. Mitchell, P. Rrp47 and the function of the Sas10/C1D domain. *Biochem. Soc. Trans.* **38**, 1088–1092 (2010).
33. Ben-Shem, A. *et al.* The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science* **334**, 1524–1529 (2011).
34. Sondalle, S.B. & Baserga, S.J. Human diseases of the SSU processome. *Biochim. Biophys. Acta* **1842**, 758–764 (2014).
35. Robson, A., Owens, N.D.L., Baserga, S.J., Khokha, M.K. & Griffin, J.N. Expression of ribosomopathy genes during *Xenopus tropicalis* embryogenesis. *BMC Dev. Biol.* **16**, 38 (2016).
36. Paolini, N.A. *et al.* A ribosomopathy reveals decoding defective ribosomes driving human dysmorphism. *Am. J. Hum. Genet.* **100**, 506–522 (2017).
37. Marneros, A.G. BMS1 is mutated in aplasia cutis congenita. *PLoS Genet.* **9**, e1003573 (2013).
38. Lebaron, S. *et al.* Proofreading of pre-40S ribosome maturation by a translation initiation factor and 60S subunits. *Nat. Struct. Mol. Biol.* **19**, 744–753 (2012).

ONLINE METHODS

Purification of the SSU processome. The SSU processome was purified as previously described¹³ from a *S. cerevisiae* BY4741 strain harboring a TEV-protease-cleavable C-terminal GFP-tag on Utp1 (Pwp2) (Utp1-3myc-TEV-GFP-3FLAG) and a second streptavidin-binding peptide tag on Kre33 (Kre33-sbp). Yeast cultures were grown to an optical density of 0.6–1 in full synthetic media containing 2% raffinose (w/v) at 30 °C before the addition of 2% galactose (w/v) and subsequently grown to saturation. Cells were harvested by centrifugation at 4,000g for 5 min at 4 °C. The pellet was washed with ice-cold ddH₂O, first without then supplemented with protease inhibitors (E-64, Pepstatin, PMSF). Washed cells were flash frozen in liquid nitrogen and lysed by cryogenic grinding using a Retsch Planetary Ball Mill PM100.

The obtained yeast powder was resuspended in buffer A (50 mM Tris-HCl, pH 7.7 (20 °C), 150 mM NaCl, 1 mM EDTA, 0.1% Triton-X100, PMSF, Pepstatin, E-64), cleared by centrifugation at 4 °C, 40,000g for 10 min, and incubated with anti-GFP nanobody beads (Chromotek) for 2 h at 4 °C. Beads were washed three times in buffer A before bound protein complexes were eluted through TEV-protease cleavage (1 h, 4 °C). The eluted supernatant was subjected to a second affinity purification step by incubation with streptavidin beads (Sigma) in buffer B (50 mM Tris-HCl, pH 7.7 (20 °C), 150 mM NaCl, 1 mM EDTA) for 1 h at 4 °C.

For cryo-EM grid preparation, the streptavidin beads were subsequently washed four times in buffer B and the SSU processome was eluted in the same buffer, supplemented with 5 mM D-biotin. For protein–protein cross-linking analysis, the beads were washed in buffer C (50 mM HEPES-NaOH, pH 7.7 (4 °C), 150 mM NaCl, 1 mM EDTA) and eluted in buffer C supplemented with 5 mM D-biotin.

Cryo-EM grid preparation. Grids were prepared from separate SSU processome purifications to collect four separate data sets (ds1–ds4). First, the sample (in 50 mM Tris-HCl, pH 7.7 (20 °C), 150 mM NaCl, 1 mM EDTA, 5 mM D-biotin) at an absorbance of 1.2–2.4 mAU at 260 nm (Nanodrop 2000, Thermo Scientific) was supplemented with 0.03% Triton-X100 (ds1) or 0.1% Triton-X100 (ds2, ds3, ds4). Subsequently, 3.5–4 µl of sample was applied onto glow-discharged grids (30 s at 50 mA) and flash frozen in liquid ethane using a Vitrobot Mark IV (FEI Company) (100% humidity, blot force of 0 and blot time 2 s). The grids for ds1, ds3 and ds4 were prepared using lacey-carbon grids (TED PELLA, Inc, Prod. No. 01824), and Quantifoil R 1.2/1.3 Cu 400 mesh grids (Agar Scientific) were used for ds2. Both grid types contained an ultra-thin carbon film.

Cryo-EM data collection and image processing. 10,029 micrographs were collected in four sessions (ds1–ds4) on a Titan Krios (FEI Company) operated at 300 kV, mounted with a K2 Summit detector (Gatan, Inc.). The micrographs from ds1 have been obtained previously¹³ and have been included in this larger data set and reprocessed together with ds2, ds3 and ds4 (Supplementary Fig. 2). SerialEM³⁹ was used to automatically acquire micrographs with a defocus range of 0.6–2.6 µm at a pixel size of 1.3 Å. Movies with 32 frames were collected at a dose of 10.5 electrons per pixel per second over an exposure time of 8 s, resulting in a total dose of 50 e⁻/Å². Data collection parameters can be found in Table 1.

All 32 movie frames were gain corrected, aligned and dose weighted using MotionCor2 (ref. 40). CTFFIND 4.1.5 was used for estimating the contrast transfer function (CTF)⁴¹. Manual inspection and the elimination of micrographs with bad CTF fits or drift reduced the number of micrographs from 10,029 to 8,406. Particles were first picked automatically using the RELION-2.0 (ref. 42) Autopick feature and then subjected to manual curation, which yielded a total of 772,120 particles. These particles were extracted with a box size of 400 pixels (520 Å) for 3D classification. 2D classification was skipped to retain rare views of the particles. 3D classification was performed with five classes using EMD-8473 (ref. 13), low-pass filtered to 60 Å, as an input model (Supplementary Fig. 2). This 3D classification produced two good classes, both combined containing a total of 284,213 particles. Autorefinement and post processing in RELION-2.0 yielded a map with an overall resolution of 3.8 Å with large areas in the center of the particle near 3-Å local resolution (Supplementary Fig. 3a). A focused refinement using a mask encompassing the 'core' region further improved the quality of the map in the best-resolved areas of the particle (Supplementary Figs. 2 and 3b). By using a mask encompassing UtpA, the Nop14–Noc4 heterodimer and the 3' domain, we were able to obtain continuous

density for the RNA and the Enp1 repeat protein and also considerably improved the density for UtpA (Supplementary Figs. 2 and 3d).

To improve the peripheral areas near the top of the particle, iterative 3D classification (first without, later with image alignment) was done using a mask around the head region, including the Utp20 helical-repeat protein. The best class from these classification steps was used for subsequent 3D refinement without mask. This strategy yielded a reconstruction at an overall resolution of 4.1 Å, with good density throughout the particle and an improvement in the head region (Supplementary Figs. 2 and 3c).

Similarly, iterative focused classifications (with and without image alignment) were used for the central domain, where one specific conformation was isolated. This conformation is present in 15% of the particles used to generate overall map 1. Focused 3D refinement led to an improvement of the resolution of this domain to 7.2 Å, allowing the docking of crystal structures (Supplementary Figs. 2, 3e and 5). Local resolution was estimated using Resmap⁴³. All computation was performed on a single Thinkmate SuperWorkstation 7048GR-TR equipped with four NVIDIA QUADRO P6000 video cards, two Twenty-two Core Intel Xeon 2.40 GHz Processors and 512 GB RAM.

Model building and refinement. The polyaniline model of the SSU processome provided by PDB 5TZS (ref. 13) served as a starting scaffold for the building of the current model. SSU processome proteins and RNA were either *de novo* modeled, or if applicable, available crystal structures were docked and manually adjusted. Phyre models were used as an initial template for the model building of some proteins⁴⁴. All model building was done in Coot⁴⁵. A complete list of templates, crystal structures and maps used to build the model can be found in Supplementary Table 1.

The model was refined against overall map 1 (3.8 Å) in PHENIX⁴⁶ with phenix.real_space_refine and secondary structure restraints for proteins and RNAs. Model statistics can be found in Table 1.

Model validation. The final refined model was resampled and scaled to the final cryo-EM map using Chimera⁴⁷. Model and maps were masked and model vs. map FSC curves were calculated using EMAN2 (ref. 48). To test for overfitting, atoms of the final model were randomly displaced by 0.5 Å using PHENIX⁴⁶ and refined against half-map 2 using the refinement strategy outlined above. The refined model was resampled, scaled and low-pass filtered to the nominal resolution of the final map using RELION-2.0 (ref. 42). FSC curves were calculated as described above with half-map 1 or half-map 2.

Map and model visualization. Structure analysis and figure preparation were performed using PyMOL Molecular Graphics System, Version 1.8 (Schrödinger, LLC) and Chimera⁴⁷. Molecular graphs and analyses were also performed with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics and the University of California, San Francisco (supported by NIGMS P41-GM103311).

DSS cross-linking mass spectrometry sample preparation and analysis. Final elution fractions of tandem-affinity purified SSU processome samples (in 50 mM HEPES-NaOH, pH 7.7 (4 °C), 150 mM NaCl, 1 mM EDTA, 5 mM D-biotin) with an absorbance of 0.5 mAU at 260 nm (Nanodrop 2000, Thermo Scientific) were pooled (total volume 3 ml) and split into twenty 150-µl cross-linking reaction aliquots.

To each aliquot, 1.5 µl of disuccinimidylsuberate (DSS; 50 mM in DMSO, Creative Molecules Inc.) was added to yield a final DSS concentration of 0.5 mM, and samples were cross-linked for 30 min at 25 °C with 450-r.p.m. constant mixing. The reactions were quenched with 50 mM ammonium bicarbonate (final concentration) and precipitated by adding methanol (Alfa Aesar, LC–MS grade) to a final concentration of 90%, then incubated overnight at –80 °C. Precipitated cross-linked SSU processomes were collected in one tube by repeated centrifugation at 21,000g, 4 °C for 30 min. The resulting pellet was washed three times with 1 ml cold 90% methanol, air dried and resuspended in 50 µl of 1× NuPAGE LDS buffer (Thermo Fisher Scientific).

DSS cross-linked SSU processomes in LDS buffer were reduced with 25 mM DTT, alkylated with 100 mM 2-chloroacetamide, separated by SDS–PAGE in

three lanes of a 3–8% Tris-Acetate gel (NuPAGE, Thermo Fisher Scientific) and stained with Coomassie-blue. The gel region corresponding to cross-linked complexes was sliced and digested overnight with trypsin to generate cross-linked peptides. After digestion, the peptide mixture was acidified and extracted from the gel as previously described^{49,50}. Peptides were fractionated offline by high-pH reverse-phase chromatography, loaded onto an EASY-Spray column (Thermo Fisher Scientific ES800: 15 cm × 75 µm ID, PepMap C18, 3 µm) via an EASY-nLC 1000 and gradient-eluted for online ESI-MS and MS/MS analyses with a Q Exactive Plus mass spectrometer (Thermo Fisher Scientific). MS/MS analyses of the top eight precursors in each full scan used the following parameters: resolution: 17,500 (at 200 Th); AGC target: 2×10^5 ; maximum injection time: 800ms; isolation width: 1.4 m/z; normalized collision energy: 24%; charge: 3–7; intensity threshold: 2.5×10^3 ; peptide match: off; dynamic exclusion tolerance: 1,500 mmu. Cross-linked peptides were identified from mass spectra by pLink⁵¹. All spectra reported here were manually verified as previously⁴⁹ and all cross-links are listed in **Supplementary Data Set 1**. Cross-links were visualized using xiNET⁵².

A **Life Sciences Reporting Summary** for this article is available.

Data availability. The cryo-EM density maps for the yeast SSU processome have been deposited in the EM Data Bank with the accession code EMD-8859. Corresponding atomic coordinates have been deposited in the Protein Data Bank under PDB 5WLC. A PyMOL session for the analysis of the structure is available (**Supplementary Data Set 2**). All other data and materials are available from the corresponding authors upon reasonable request.

39. Mastrorade, D.N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
40. Zheng, S.Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
41. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
42. Kimanius, D., Forsberg, B.O., Scheres, S.H. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *eLife* **5**, e.18722 (2016).
43. Kucukelbir, A., Sigworth, F.J. & Tagare, H.D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
44. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. & Sternberg, M.J.E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
45. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
46. Adams, P.D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
47. Pettersen, E.F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
48. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
49. Shi, Y. *et al.* Structural characterization by cross-linking reveals the detailed architecture of a coatamer-related heptameric module from the nuclear pore complex. *Mol. Cell. Proteomics* **13**, 2927–2943 (2014).
50. Shi, Y. *et al.* A strategy for dissecting the architectures of native macromolecular assemblies. *Nat. Methods* **12**, 1135–1138 (2015).
51. Yang, B. *et al.* Identification of cross-linked peptides from complex samples. *Nat. Methods* **9**, 904–906 (2012).
52. Combe, C.W., Fischer, L. & Rappsilber, J. xiNET: cross-link network maps with residue resolution. *Mol. Cell. Proteomics* **14**, 1137–1147 (2015).

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

▶ Experimental design

1. Sample size

Describe how sample size was determined.

N/A

2. Data exclusions

Describe any data exclusions.

N/A

3. Replication

Describe whether the experimental findings were reliably reproduced.

N/A

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

N/A

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

N/A

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- A statement indicating how many times each experiment was replicated
- The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
- A description of any assumptions or corrections, such as an adjustment for multiple comparisons
- The test results (e.g. P values) given as exact values whenever possible and with confidence intervals noted
- A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- Clearly defined error bars

See the web collection on [statistics for biologists](#) for further resources and guidance.

► Software

Policy information about [availability of computer code](#)

7. Software

Describe the software used to analyze the data in this study.

SerialEM, pLINK, Relion, Coot, Chimera, Resmap, Pymol, xiNET

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* [guidance for providing algorithms and software for publication](#) provides further information on this topic.

► Materials and reagents

Policy information about [availability of materials](#)

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

There are no restrictions on availability.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

N/A

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

N/A

b. Describe the method of cell line authentication used.

N/A

c. Report whether the cell lines were tested for mycoplasma contamination.

N/A

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

N/A

► Animals and human research participants

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

N/A

Policy information about [studies involving human research participants](#)

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

N/A