2019

Elucidation of the Functional Architecture of the Early Pre-Ribosomal Processing Machinery in Yeast

Mirjam Hunziker

ELUCIDATION OF THE FUNCTIONAL ARCHITECTURE OF THE EARLY PRE-RIBOSOMAL PROCESSING MACHINERY IN YEAST

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Mirjam Hunziker

June 2019

ELUCIDATION OF THE FUNCTIONAL ARCHITECTURE OF THE EARLY PRE-RIBOSOMAL PROCESSING MACHINERY IN YEAST

Mirjam Hunziker, Ph.D.

The Rockefeller University 2019

Ribosomes carry out one of the most fundamental functions of life - the translation of genetic information into functional proteins. The pivotal role of the ribosome in the cell is reflected in its immensely complicated and energy-consuming assembly pathway.

The maturation of a eukaryotic ribosome involves more than 200 non-ribosomal factors and the activity of all three RNA polymerases. In yeast, ribosome biogenesis starts with the transcription of the 35S pre-ribosomal RNA in the nucleolus. This large RNA molecule contains three of the four ribosomal RNAs separated by several internal and external transcribed spacer regions. The 5' external transcribed spacer (5'ETS) is the first RNA domain of the 35S pre-rRNA being transcribed. As it emerges from the RNA polymerase it is bound by UtpA, a 660 kDa complex consisting of 7 essential subunits in yeast.

By binding to the nascent pre-rRNA, UtpA triggers the association of multiple other proteins and complexes, which leads to the formation of the ~2 MDa 5' ETS particle. As transcription continues through the ensuing small subunit rRNA gene more ribosome biogenesis factors as well as ribosomal proteins are recruited and the 5' ETS particle evolves into the small subunit processome. The small subunit processome, a giant particle, unique and essential to eukaryotes, coordinates the cleavage of the 35S pre-rRNA to separate the maturation of the small and large ribosomal subunit.

So far, a functional understanding of the initial events in ribosome biogenesis has been impeded by a lack of structural and biochemical data about the protein complexes facilitating this process and the pre-ribosomal particles they form.

To gain mechanistic insights into these earliest steps we set out to delineate the role of UtpA as first building block, vital structural component and organizer of the 5' ETS particle and the small subunit processome. By using protein-protein and RNA-protein cross-linking techniques combined with negative stain electron microscopy and biochemical assays we were able to define the composite RNA binding site of UtpA and characterize its molecular architecture in the absence of high-resolution structural data (**Chapter II**).

Subsequent structure determination of the small subunit processome by cryo-electron microscopy has not only provided the first fully assigned atomic model of UtpA but visualized how ribosome biogenesis factors keep the ribosomal RNA domains in spatially separated compartments of this large particle (**Chapter III**). In the small subunit processome, the 5' ETS particle forms the base onto which the segregated ribosomal RNA domains are folded. To investigate whether the 5' ETS particle serves as a structural mold for the maturing rRNA domains during earlier assembly stages, we solved the cryo-EM structures of the 5' ETS particle in intermediates preceding the formation of the small subunit processome (**Chapter IV**). Combined with the *in vivo* analysis of artificial pre-rRNA fragments, the architecture of the 5' ETS particle shows that the initial steps of ribosome assembly are governed by the functional independence of all rRNA domains and the 5' ETS particle. Completion of ribosomal gene transcription then leads to a conformational change in the 5' ETS particle and small subunit processome formation.

In summary, our work provides structural snapshots and biochemical information on more than 50 ribosome assembly factors during different stages of the initiating steps in eukaryotic ribosome biogenesis. These data form the basis for a three-dimensional model of these essential events in the eukaryotic cell.

Acknowledgement

I am sincerely grateful to Sebastian Klinge, Jonas Barandun and Malik Chaker-Margot for their intellectual camaraderie during my graduate thesis studies. Working alongside them on our shared projects has allowed me to grow more than I would ever have working on my own and was one of the greatest experiences during my PhD. Sebastian Klinge has been a dedicated mentor, whose enthusiasm and optimism, scientific creativity and openness for discussion has created an encouraging and collaborative working environment.

I thank Jonas Barandun not only for being an inspiring colleague, but for being a true partner in all aspects of my life. Malik Chaker-Margot has been an exceptional fellow since the first day of my PhD - his incredible memory and generous kindness impressed me equally. Sharing my workplace with Zahra Assur-Sanghai, Linamarie Miller and Camille Cheng, whose scientific prowess and organizational skills are only matched by their coolness and likability, has enriched my life at Rockefeller daily. Sameer Singh, who recently joined our lab, has contributed greatly to its humorous spirit and I am excited to see where his scientific endeavors take him. For his cheerfulness and help during the first years of my PhD I thank Antoine Gruet.

I would like to thank Thomas Walz for his support and advice as well as being on my thesis committee. To Seth Darst I am grateful for serving as the head of my committee and I thank John L. Woolford for agreeing to be my external thesis examiner.

Rajendra Boggavarapu, Gabriella Angiulli, Hiroshi Suzuki, Kelly H. Kim, Yixiao Zhang and Martin Bush – thanks for all your help with electron microscopy matters and your friendship. Richard Hite and members of his laboratory, Greg Alushin, Ayala Carl and Deena Oren have contributed to this work through helpful discussions. David Tollervey, Elisabeth Petfalski, Clementine Delan-Forino and Hywel Dunn-Davis from the University of Edinburgh welcomed me to their lab and taught me how to perform and analyze RNA-protein cross-linking studies – thank you very much.

I am indebted to many scientists here at the Rockefeller University who contributed to this work in a collaborative effort: Kelly Molloy, Yi Shi and Junjie Wang from Brian Chait's laboratory, Stefanie Gerstberger, Devrim Acehan and Kunihiro Uryu from the electron microscopy research center, Caitlin Steckler, Milica Tesic Mark, Brian D. Dill and Henrik Molina from the proteomics facility, Mark Ebrahim and Johanna Sotiris for their support at the cryo-electron microscopy center.

Many thanks to Liz Campbell for contributing to my inner well-being by teaching spin classes at Rockefeller. I also thank Elizabeth Wasmuth and Zac Hann for their company.

Lastly, I would like to acknowledge the continuous support and encouragement from my family, Jonas' family and our friends back home and here.

Table of contents

List of figures

List of tables

Chapter 1 | Introduction

1.1 The Ribosome

In a biological system the flow of genetic sequence information is enabled by three different classes of biopolymers – DNA, RNA and protein. The central dogma, postulated by Francis Crick in 1958 (Crick 1958), states that genetic sequence information can flow from nucleic acid to nucleic acid, and from nucleic acid to protein, but not from protein to nucleic acid or from protein to protein. In living cells three types of general information transfers occur (**Figure 1.1**). DNA replication, a process in which DNA is copied to DNA by DNA polymerases, is one of them. DNA-encoded information can be transcribed into messenger RNAs (mRNAs) by RNA polymerases. This constitutes the second general type of sequence information transfer. The ribosome then translates the information stored in mRNAs into proteins, thus representing the third form of general genetic sequence information transfer.

By translating the genetic information carried by mRNAs into proteins, the ribosome accomplishes one of the most fundamental tasks of life. All living cells rely on accurate and efficient protein synthesis by the ribosome to maintain their structure, catalyze biochemical reactions and integrate signals from the environment.

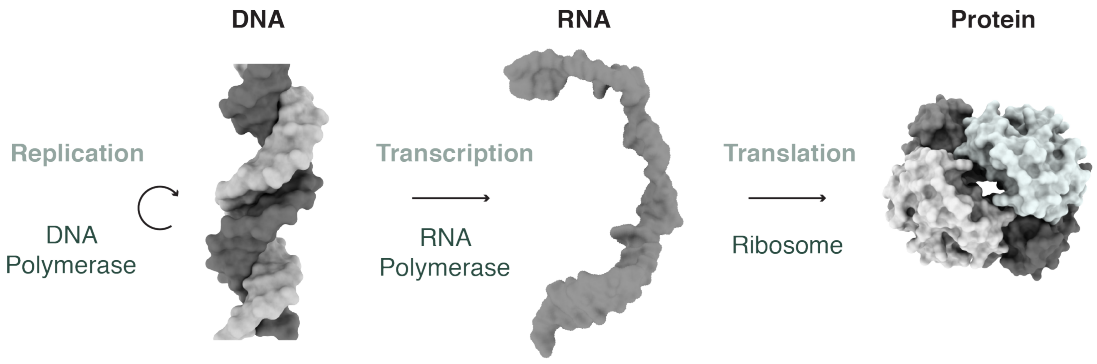


Figure 1.1 | General flow of genetic sequence information in living cells.
The three general genetic sequence information transfers are enabled by three biopolymers: DNA (from PDB 1D66), RNA (from PDB 4V6G) and protein (PDB 1FDH). The processes (replication, transcription and translation) and enzymes (DNA- and RNA polymerases, and the ribosome) responsible for these transfers are indicated.

The ribosome itself depends on protein synthesis as it consists of 79 proteins and 4 ribosomal RNAs (rRNA) in *S. cerevisiae* (**Figure 1.2**). Ribosomal proteins and rRNA molecules form two asymmetrical subunits – the small (40S) and the large (60S) ribosomal subunit.

The small ribosomal subunit, containing the 18S rRNA and 33 ribosomal proteins in yeast, harbors the decoding center and monitors the complementarity between incoming transfer RNA (tRNA) anticodons and the mRNA (**Figure 1.2 a,f**) (Yusupova & Yusupov 2014). The three anticodon residues of tRNAs correlate with the amino acids covalently loaded onto their 3' end. Thus, tRNAs play a vital role in translation as physical adaptors between the genetic code stored in mRNAs and the

amino acid sequence of proteins. The large ribosomal subunit, consisting of the 25S, 5.8S and 5S rRNA and 46 ribosomal proteins, catalyzes peptide bond formation between the amino acids carried by the tRNAs (**Figure 1.2 a,f**). All enzymatic activity of the ribosome is carried out by ribosomal RNA whereas ribosomal proteins have a structural role (Noller et al. 1992; Nissen et al. 2000; Ban et al. 2000). Hence, the ribosome is a ribozyme.

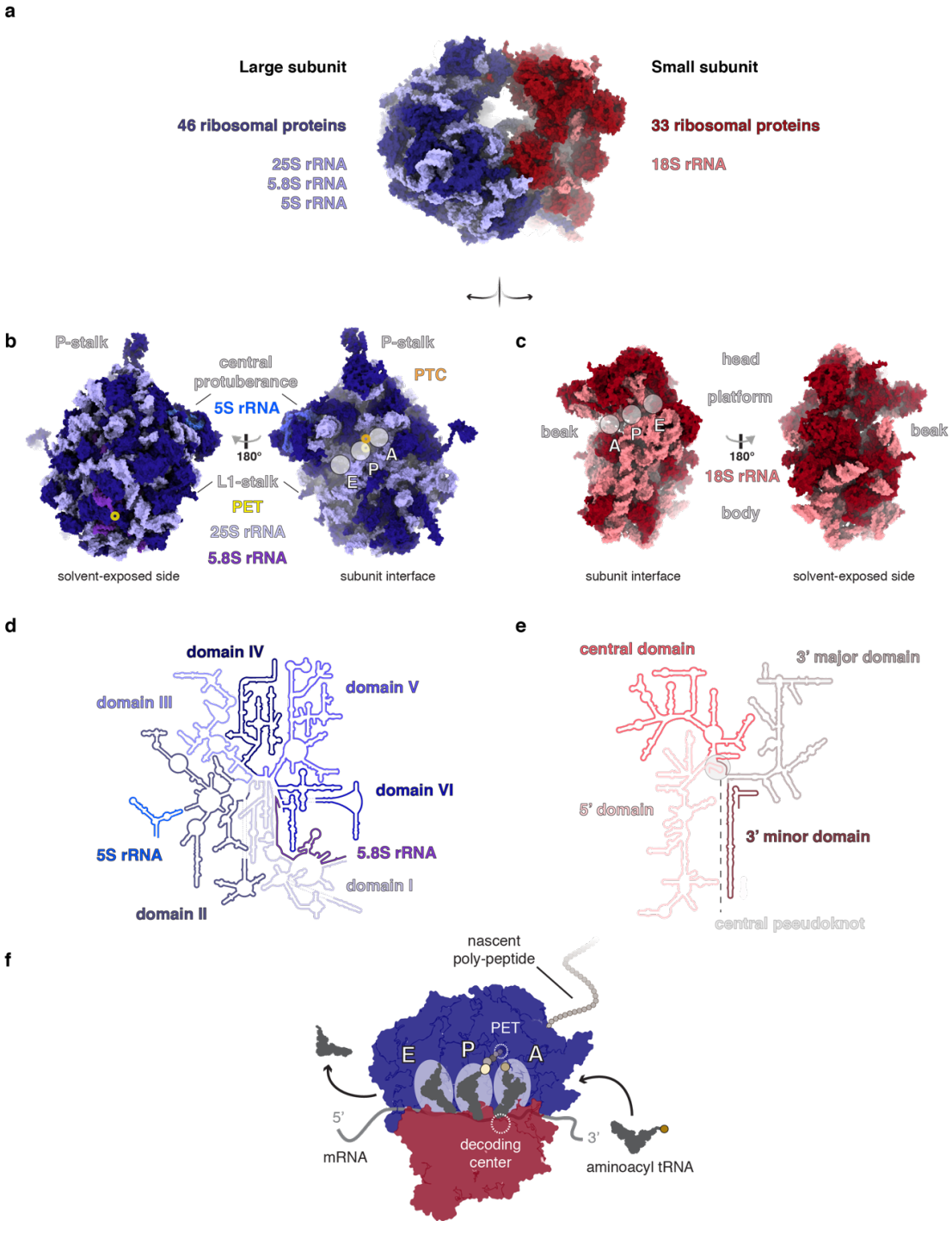
Since the ribosome was discovered (Palade 1955), great efforts have been directed at understanding the cellular context, function and structure of this macromolecular machinery. In a scientific breakthrough, the crystal structures of the prokaryotic (Ban et al. 2000; Schluenzen et al. 2000; Yusupov et al. 2001; Harms et al. 2001; Clemons et al. 1999) and later the eukaryotic ribosome (Rabl et al. 2011; Klinge et al. 2011; Ben-Shem et al. 2011) were solved. These studies revealed for the first time the structural basis of protein translation at high resolution.

Structures of the ribosome have shown that the small subunit has two distinct architectural features - the body and the head (**Figure 1.2c**). In between these features, located at the subunit interface, runs a groove containing the mRNA binding sites. Close to the mRNA binding path are the tRNA binding sites (A, P and E) (**Figure 1.2c**) and the decoding center, where codon-anticodon complementarity is monitored to ensure translation fidelity (**Figure 1.2f**). The 18S rRNA, which constitutes these functionally relevant sites in the small subunit, can be divided into

four domains: the 5'-, central -, 3'-major and 3'-minor domains (**Figure 1.2e**). With the 3' major domain forming the head of the small subunit, the other three domains form the body. Their structural organization with respect to each other is determined by the central pseudoknot. The central pseudoknot is a tertiary structural element formed by base-pairing RNA sequences from the beginning of the 5' domain and the middle of the 18S rRNA (**Figure 1.2e**).

Architectural landmarks of the large ribosomal subunit are the central protuberance, the L1-stalk and the P-stalk (**Figure 1.2b**). Functionally important structures – the tRNA binding sites (A, P and E) which are formed between the large and small subunit and the peptidyl transfer center (PTC) – are located on the subunit interface (**Figure 1.2b**). Adjacent to the PTC is the entrance to the polypeptide exit tunnel (PET), a channel through which nascent polypeptides proceed to exit the ribosome on the solvent-exposed side of the large subunit (**Figure 1.2 b,f**). Similar to the small subunit, the large subunit rRNA can be subdivided into six structured domains (I-VI) (**Figure 1.2d**). The PET is located at the interface of these rRNA domains in the mature 60S.

Figure 1.2 | Structure and functional sites in the *S. cerevisiae* ribosome.
[a] Surface representation of the 80S ribosome from *S. cerevisiae* (PDB 4V88) (Ben-Shem et al. 2011). The components of the large- and small-ribosomal subunit are listed and colored as in the structure. **[b]** The subunit interface and solvent-exposed side of the large subunit are shown with architectural features (P-stalk, L1-stalk, central protuberance) labeled in grey, the three rRNAs in shades of blue and purple with color-coded labels and the approximate locations of the functionally important peptidyl-transferase center (PTC, orange) and peptide exit tunnel (PET, yellow) indicated. The tRNA binding sites (A, P, E) are marked with white-transparent circles on the subunit interface. **[c]** The small ribosomal subunit viewed from its solvent-exposed as well as subunit interface side. Architectural landmarks (body, platform, head, beak) are labeled in grey and the approximate location of the tRNA binding sites (A, P, E) are indicated. **[d,e]** Secondary structure diagrams of the large [d] and small [e] subunit rRNAs. Subdomains and are labeled in a color-coded manner. The structurally important central-pseudoknot is indicated in grey in [e]. **[f]** Simplified schematic representation of ribosomal translation. Aminoacyl-tRNAs (dark-grey) charged with amino acids (brown circle) are bound in the A-site cavity (white) formed between the two ribosomal subunits (blue, red). After base-pairing with the mRNA (light-grey) in the decoding center (dashed white circle), a peptide-bond is formed between the amino acid of the tRNA in the A- and P-site. The nascent poly-peptide chain (shades of brown) exits through the PET (white). Uncharged tRNAs dissociate from the ribosome.


a
Large subunit
46 ribosomal proteins
25S rRNA
5.8S rRNA
5S rRNA
Small subunit
33 ribosomal proteins
18S rRNA
b
P-stalk
central protuberance
5S rRNA
180°
L1-stalk
PET
25S rRNA
5.8S rRNA
solvent-exposed side
P-stalk
PTC
A
P
E
subunit interface
c
head
platform
beak
A
P
E
180°
18S rRNA
body
subunit interface
beak
solvent-exposed side
d
domain IV
domain III
domain V
domain VI
5S rRNA
5.8S rRNA
domain I
domain II
e
central domain
3' major domain
5' domain
3' minor domain
central pseudoknot
f
nascent poly-peptide
PET
E
P
A
5'
mRNA
decoding center
3'
aminoacyl tRNA

Eukaryotic ribosomal translation starts with the formation of an initiation complex between the small subunit, the mRNA and several initiation factors (reviewed in Hashem and Frank 2018). This initiation complex scans the mRNA for the start codon which is then bound by a methionine-charged tRNA. The initiation factors dissociate from the complex and eventually the large subunit is recruited. The ribosome subsequently commences translation elongation. During translation elongation tRNAs charged with amino acids (aminoacyl-tRNAs) are loaded into the A-site cavity between the two subunits (**Figure 1.2f**). When the delivered tRNA anti-codon base-pairs stably with the mRNA-codon in the decoding site, peptide bond formation between the amino acids of the tRNAs in the A-site and P-site is catalyzed in the PTC (not shown). The nascent poly-peptide chain formed by this reaction exits the ribosome through the proximally located PET (**Figure 1.2f**). Through a rotation of the two ribosomal subunits with respect to each other, the amino-acid-carrying ends of the tRNAs are moved to the E- and P- sites respectively, while the anti-codon ends stay in the A- and P-sites. A eukaryotic elongation factor binds the ribosome and hydrolyzes GTP to induce translocation of the mRNA, which also shifts the anti-codon ends of the tRNAs to the E- and P-sites. Uncharged tRNAs in the E-site dissociate from the ribosome (**Figure 1.2f**). The ribosomal translation cycle has been reviewed in detail (Dever et al. 2018; Dever et al. 2016).

While biochemical, structural and cell biological studies have elucidated the mature eukaryotic ribosome, its interactions with other proteins and the translation cycle at great depth and resolution (Graille & Séraphin 2012; Yusupova & Yusupov 2014; Dever et al. 2016; Dever et al. 2018; Ling & Ermolenko 2016), the process by which these essential particles are made is less well understood.

1.2 Ribosome Biogenesis in *Saccharomyces cerevisiae*

1.2.1 Transcription of ribosomal RNA precursors

Ribosome biogenesis is a complex, multidimensional pathway involving all three RNA polymerases and more than 200 non-ribosomal assembly factors in yeast (Woolford & Baserga 2013). It starts in the nucleolus, a compartment of the nucleus defined by and dedicated to this essential process. Rather than being determined by a skeletal framework or membrane-enclosure, the nucleolus is a direct product of active ribosome biogenesis (Oakes et al. 1993; Oakes et al. 1998; Prieto & McStay 2007). In yeast, ribosomal RNA is encoded on chromosome XII which harbors tandem arrays of 100-200 polycistronic ribosomal DNA (rDNA) repeats (**Figure 1.3a**). Each of these repeats contains the sequence of all four rRNAs interspersed by external- and internal transcribed spacers (ETS and ITS) and non-transcribed spacers (NTS) (**Figure 1.3b**). While the precursor of the 5S rRNA, which is part of the large subunit, is transcribed separately by RNA polymerase III, the other three rRNAs (18S, 5.8S and 25S) together with the ETS and ITS regions are transcribed as a single precursor RNA (35S pre-rRNA) by RNA polymerase I (**Figure 1.3b**).

While the transcribed spacer regions (5' ETS, 3' ETS, ITS1, ITS2) are part of the ribosomal precursor RNA, the non-transcribed spacers harbor cis-acting DNA elements important for transcription and DNA replication (Nomura et al. 2004) (**Figure 1.3 c,d**). NTS1 contains two transcription terminators downstream of the 3' ETS

(**Figure 1.3d**) as well as an enhancer element (not shown) for the 35S rRNA promoter, which is located in NTS2. While the enhancer sequence in NTS1 is dispensable for RNA polymerase I transcription of rDNA genes (Wai et al. 2001), the two elements of the 35S rRNA gene promoter, the upstream element and the core, are required for high levels of transcription and accurate transcription initiation, respectively (Kulkens et al. 1991; Keys et al. 1996; Choe et al. 1992).

The recruitment of a transcription initiation complex at the 35S rDNA promoter requires two basal transcription factor assemblies: the upstream activating factor (UAF) associated with the TATA-box binding protein (TBP) and the 3-subunit core factor (CF) (**Figure 1.3c**). Additionally, RNA Polymerase I has to bind the Pol I-specific initiation factor Rrn3 to render the enzyme initiation-competent before associating with the CF at the promoter (Peyroche 2000). After initiation RNA polymerase I separates from Rrn3 and leaves the promoter. With an elongation rate of 40–60 nucleotides per second, transcription by RNA polymerase I is highly processive (French et al. 2003; Koš & Tollervey 2010).

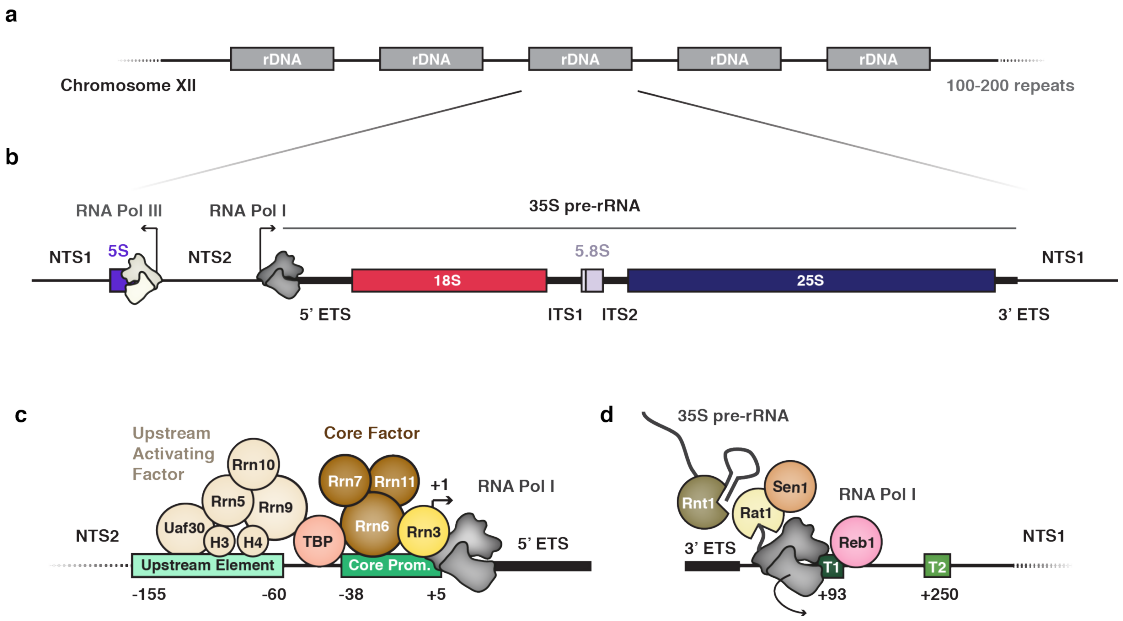


Figure 1.3 | Functional organization of rDNA repeats in yeast.

[a] *S. cerevisiae* chromosome XII contains 100-200 rDNA repeats (grey). **[b]** One repeat contains all four rRNAs (5S, 18S, 5.8S and 25S, as colored boxes) and external- and internal transcribed spacers (5' ETS, 3' ETS, ITS1 and ITS2, as thick black lines) as well as non-transcribed spacers (NTS1 and NTS2, as thin black lines). RNA polymerase III (light-grey) and RNA polymerase I (dark grey) are schematically shown at their transcription start sites. The DNA elements coding for the 35S pre-rRNA are grouped with a black line. **[c]** Transcription initiation factors and DNA sequences needed for RNA polymerase I initiation. The 6-subunit Upstream Activating Factor (subunits labeled, beige circles) binds to the upstream element (light-green) and the TATA-Box binding protein (TBP, pink), whereas the 3-subunit Core Factor (subunits labeled, brown) is associated with the Core promoter (green). Rrn3 (yellow) binding to RNA polymerase I (grey) renders the enzyme initiation-competent. The nucleotide positions of the transcription start site (+1) and the functional promoter elements are indicated with black numbers. **[d]** Transcription termination is facilitated by Reb1 (pink), termination 1 sequence (T1, dark green), the nucleases Rnt1 (green) and Rat1 (yellow) as well as Sen1 (brown). A fail-safe terminator sequence (T2, light green) is located downstream.

Downstream of the 3' ETS, 35S pre-rRNA transcription is terminated through a multistep process involving DNA elements of NTS1, regulatory proteins and rRNA processing factors (**Figure 1.3d**). Reb1, the yeast homolog of the mammalian transcription termination factor I, binds to elements of the first NTS1 terminator sequence (T1) which causes the polymerase to pause (Jansa & Grummt 1999) (**Figure 1.3d**). The endonuclease Rnt1 recognizes and cleaves a stem loop in the transcribed 3' ETS (Kufel et al. 1999), which instigates the nuclease Rat1 and the RNA-helicase Sen1 to digest the polymerase associated RNA cleavage product (El Hage et al. 2008; Kawauchi et al. 2008; Braglia et al. 2011; Braglia et al. 2010). RNA polymerase I dissociates from the DNA when Rat1 reaches the enzyme. While 90% of all transcripts are terminated at the T1 site, a second terminator sequence further downstream (T2) exists as a fail-safe mechanism (Nomura et al. 2004) (**Figure 1.3d**).

1.2.2 Processing of the 35S pre-rRNA

To yield a mature ribosome the external- and internal transcribed spacer regions have to be excised from the 35S precursor RNA. The 35S pre-rRNA can be processed both co- and post- transcriptionally, with most of the precursors undergoing co- rather than post-transcriptional processing (Osheim et al. 2004; Koš & Tollervey 2010). Cleavage at A2 in the co-transcriptional, or at A3 in the post-transcriptional pathway, separate small and large subunit rRNA maturation (**Figure 1.4**). During co-transcriptional processing A0 cleavage in the 5' ETS precedes the

coordinated A1- and A2 processing steps in ITS1 (Hughes & Ares 1991). In contrast, A3 cleavage in ITS1 and thus the separation of the two subunits occurs before processing of the 5' ETS in the post-transcriptional pathway.

The site-specific cleavages of the 35S pre-rRNA are catalyzed by different nucleases. A multi-subunit complex, RNAse MRP, cleaves the A3-site in the post-transcriptional pathway (Lygerou et al. 1994). Utp24, a PIN-domain protein, is responsible for A1 site processing and suspected to cut the A2 site as well during co-transcriptional processing (Bleichert et al. 2006; Tomecki et al. 2015; Wells et al. 2016). It remains unknown which factor catalyzes A0 cleavage.

After the initial ITS1 cleavage, the small and large subunit rRNAs mature independently. Small subunit maturation in the nucleolus leads to the removal of the 5' ETS after A0 cleavage. This is mediated by the 3' to 5' exonuclease activity of the exosome and its associated helicase Mtr4 (Allmang et al. 2000). Subsequent processing of the A1-site not only separates the remaining parts of the 5' ETS from the pre-18S but also defines the mature 5' end of the small subunit rRNA. The resulting 18S precursor, the 20S rRNA, is exported from the nucleolus to the cytoplasm. Here the mature 18S rRNA emerges after Nob1 cuts the D-site (Fatica et al. 2004).

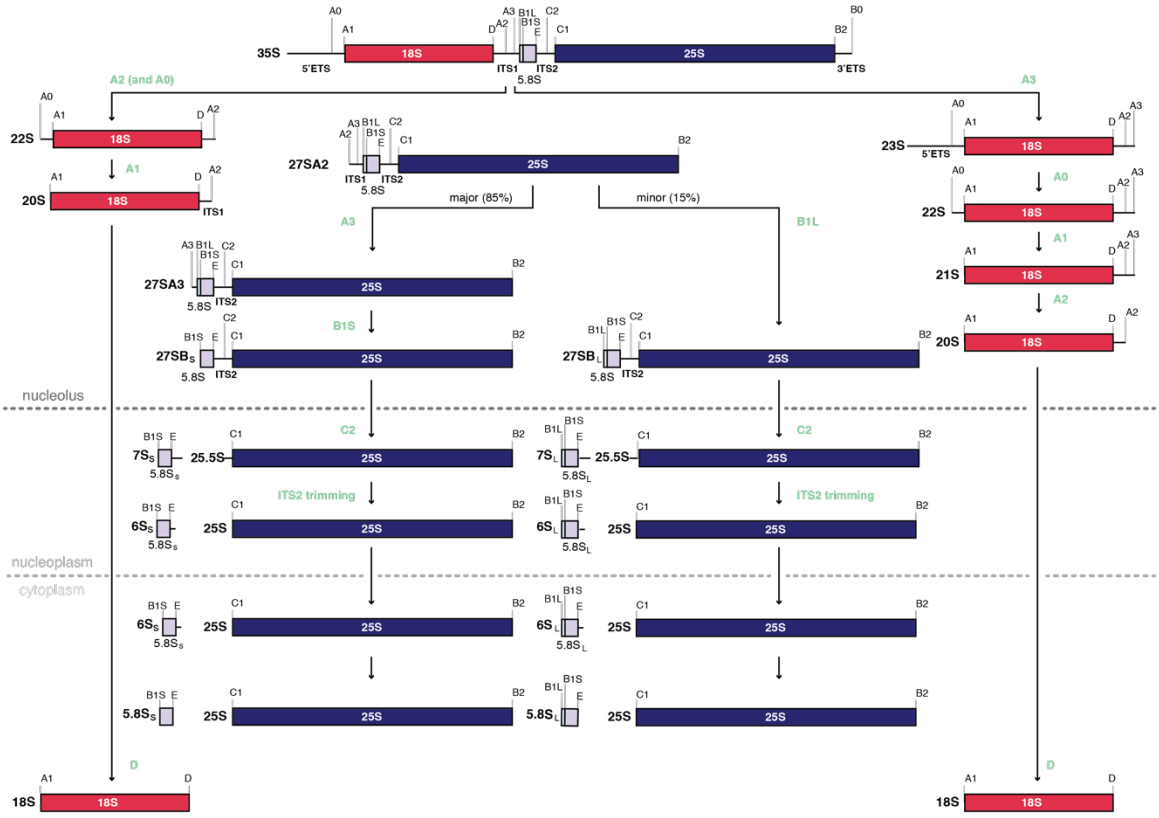


Figure 1.4 | Processing pathways of the 35S pre-rRNA in yeast.

The RNA polymerase I transcribed 35S precursor contains the 18S (red), 5.8S (light-blue) and 25S (dark-blue) rRNA flanked by the 5'- and 3' external spacers (5' ETS, 3' ETS, black line) and separated by two internal spacers (ITS1, ITS2, black line). Locations of processing sites are indicated above the different pre-rRNA species in black. The different site-specific cleavages and pre-rRNA processing events taking place in the nucleolus, nucleoplasm or cytoplasm (dotted lines) are labeled in green.

The large subunit precursor RNA is processed at its 3' end first. After cleavage of an RNA stem-loop at site B0 by Rnt1 (Kufel et al. 1999), a step which contributes to RNA polymerase I transcription termination (**Figure 1.3d**), the remaining 3' ETS sequence is trimmed to the mature 3' end of the 25S rRNA. This could be mediated by the exonuclease Rex1 (Kempers-Veenstra et al. 1986).

The 5' end of the vast majority of the 27SA2 precursor molecules is cut at the A3 site by RNAse MRP (Shuai & Warner 1991; Lindahl et al. 1991; Schmitt & Clayton 1993; Chu et al. 1994; Lygerou et al. 1996; Woolford & Baserga 2013) and then trimmed to the B1S site by Rat1 and Rrp17 yielding the 27SBS (Henry et al. 1994; Oeffinger et al. 2009) (**Figure 1.4**). A minority of the 27SA2 pre-rRNA is directly processed at the B1L site by an unknown nuclease which results in the 27SBL precursor. Both of the 27SB variants are cut at the C2 site in ITS2 by the Las1-complex giving rise to the 25.5S and the 7SS/L (Gasse et al. 2015; Fromm et al. 2017). Trimming of the 5' end of the 25.5S by the exonuclease Rat1 removes all remaining spacer region sequences from the 25S rRNA. The ITS2 parts of the 7SS/L precursors are degraded by the exosome (Mitchell et al. 1997). A short unprocessed overhang (6S pre-rRNA) is removed by Ngl2 after export of the large subunit pre-rRNAs into the cytoplasm (Thomson & Tollervey 2010). The RNA polymerase III transcribed pre-5S rRNA is processed separately by Rex1 (van Hoof et al. 2000) (not shown).

1.2.3 Chemical modifications of the ribosomal RNA

In addition to being extensively processed, the pre-rRNA is also chemically modified (reviewed in Sloan et al. 2017). Chemical modifications of the rRNA occur during all stages of ribosome maturation. In the mature ribosome these modifications contribute to the stability of the rRNA scaffold (Polikanov et al. 2015) and improve the efficiency and accuracy of translation. Nucleotide modifications are enriched around functionally important elements of the rRNA such as the decoding center, the tRNA binding sites (A, P and E), the PTC and the subunit interface (Ben-Shem et al. 2011; Decatur & Fournier 2002; Sloan et al. 2017). Despite this clustering close to functional sites, only very few of the individual chemical modifications have an impact on ribosome function and cell survival when deleted. Instead, these modifications seem to act in a cumulative manner since the loss of grouped modifications leads to significant cell growth phenotypes (King et al. 2003; Liang et al. 2007; Liang et al. 2009; Baxter-Roshek et al. 2007; Baudin-Baillieu et al. 2009).

In yeast, the most abundant rRNA modifications are 2'-O-methylation of the ribose moiety and isomerization of uridine to pseudo-uridine. A total of 55 2'-O-methylation and 45 pseudo-uridylation sites have been identified in *S. cerevisiae* (Birkedal et al. 2015; Taoka et al. 2016; Sloan et al. 2017). Most of them are introduced by two classes of small nucleolar ribonucleoparticles (snoRNPs) termed box C/D and box H/ACA respectively (Ganot et al. 1997; Kiss-Laszlo et al. 1996; Ni et al. 1997) (**Figure 1.5**).

In eukaryotic snoRNPs, a defined set of common core-proteins binds to conserved sequence motifs of the snoRNA, which base-pairs with its target pre-rRNA and thereby positions the catalytic subunits of its core proteins for target site modification.

In box H/ACA snoRNPs the snoRNA forms two hairpins, which are bound by the proteins Nhp2, Nop10, Gar1 and Cbf5. While Cbf5 is the catalytically active subunit and facilitates the pseudo-uridylation of the target nucleotide (Lafontaine et al. 1998), the other core proteins have structural roles in the RNP. The substrate RNA and snoRNA are base-paired within a hairpin and form a pseudo-uridylation pocket. In this RNA-RNA duplex the target nucleotide is accessible for isomerization by Cbf5 as it is left non-base-paired. The conserved Box H and ACA-sequence are the name-giving elements of this class of snoRNAs.

Through base-pairing of the conserved box C and C' motifs with the box D and D' motifs respectively, the box C/D snoRNAs adopt a single hairpin structure. The box C/D snoRNA scaffold is bound by two copies of Snu13 and the methyltransferase Nop1 as well as one copy of Nop56 and Nop58. Substrate RNA is recruited downstream of the box D and D' elements, stabilized by the structural core proteins and methylated by Nop1.

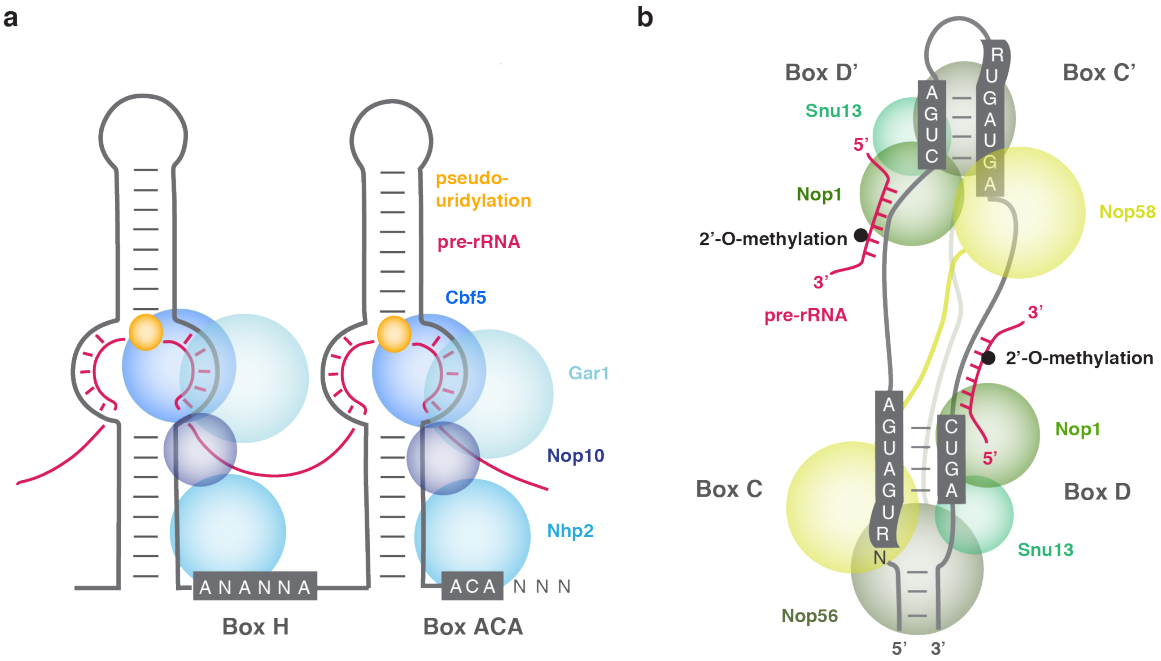


Figure 1.5 | General architecture of eukaryotic snoRNPs mediating chemical rRNA modifications.

[a] Schematics representation of H/ACA snoRNA (grey) architecture and the bound core proteins (in shades of blue). The sequences of the conserved box H motif and the ACA sequence are highlighted in grey boxes. Substrate pre-rRNA (pink) is bound within one hairpin and pseudo-uridylation (orange) of the target nucleotide is catalyzed by Cbf5. **[b]** Base-pairing between the conserved box C and box D as well as box C' and box D' (grey boxes) results in the single hairpin structure of box C/D snoRNAs (grey). Two copies of Snu13 and Nop1 and one copy of Nop56 and Nop58 constitute the general core proteins (shades of green). Nop1 mediated methylation (black circle) of the base-paired RNA substrate (pink) occurs upstream of the box D/D' motifs.

Despite their capacity for chemical modifications several H/ACA and box C/D snoRNPs seem to have a solely structural role in regulating pre-rRNA folding rather than guiding chemical alterations. U3 snoRNA, the most abundant box C/D snoRNA in the cell, base-pairs with the 5' ETS and 18S rRNA but does not mediate nucleotide methylation of either RNA species (Beltrame & Tollervey 1995; Sharma & Tollervey 1999; Dutca et al. 2011; Marmier-Gourrier et al. 2011). Another example is snR30, an essential box HACA snoRNA which affects pre-18S rRNA processing but has no known modification target (Fayet-Lebaron et al. 2009; Lemay et al. 2011; Atzorn et al. 2004).

Enzymes which are not part of a snoRNP framework can also mediate rRNA modifications. The methyltransferases Emg1 (Leulliot et al. 2008) and Dim1 (Lafontaine et al. 1994) as well as the acetyltransferase and helicase Kre33 (Sharma et al. 2015) are examples for stand-alone enzymes modifying the 18S rRNA.

1.2.4 Early steps in eukaryotic ribosome biogenesis lead to the assembly of the small subunit processome

Pre-rRNA processing and the introduction of chemical rRNA modifications occur in the context of pre-ribosomal particles. These particles initially form co-transcriptionally through the assembly of ribosome biogenesis factors and ribosomal proteins on the nascent 35S pre-rRNA (**Figure 1.6a**). As early as 1969 pre-ribosomal

particles could be visualized by negative stain electron microscopy (EM) as terminal knob structures on actively transcribed rDNA (Miller & Beatty 1969) and were identified as rRNA processing complexes in 1993 (Mougey et al. 1993). With the advent of tandem affinity purification and mass spectrometry methods the composition of specific pre-ribosomal particles could be characterized in unprecedented depth. This led to the discovery of the small subunit (SSU) processome (Dragon et al. 2002; Grandi et al. 2002).

The SSU processome contains more than 70 proteins, the 23S pre-rRNA and the U3 snoRNA (Dragon et al. 2002). Many of the SSU processome components are comprised of protein-protein interaction domains such as β-propellers and α-helical repeat structures, but some also contain known RNA binding motifs, RNA helicase domains, GTPase folds and RNA modification enzyme structures. Several proteins associated with the SSU processome have not been characterized before and since they all co-precipitated with the U3 snoRNA they were named U-three-proteins (Utp). Later on, the large Utp-complexes UtpA, UtpB and UtpC were discovered to exist as pre-assembled submodules in the cell (Krogan et al. 2004).

While the 660kDa-complex UtpA is composed of seven subunits (Utp4, Utp5, Utp8, Utp9, Utp10, Utp15, Utp17) and the 550kDa-complex UtpB of six (Utp1, Utp6, Utp12, Utp13, Utp18, Utp21) (**Figure 1.6b**), UtpC has two members (Utp22 and Rrp7) (Baudin-Baillieu et al. 1997) which can associate with the casein kinase 2

complex (Cka1, Cka2, Ckb1 and Ckb2) (Krogan et al. 2004). UtpC and casein kinase 2 have also been shown to form the CURI-complex which includes the transcription factor Ifh1. The CURI-complex represents a link between pre-rRNA processing and the transcription of mRNAs of ribosomal proteins (Rudra et al. 2007; Albert et al. 2016).

U3 snoRNP is another large pre-assembled subcomplex of the SSU processome (**Figure 1.6b**). In addition to the 333-nucleotide long U3 snoRNA and the core box C/D proteins (**Figure 1.5b**), it also contains the U3 specific factor Rrp9 (Venema et al. 2000). In its 5' sequence, upstream of the architecture defining box C/D and box C'/D' duplexes, the U3 snoRNA harbors four elements that base-pair with the 5' part of the 18S rRNA (Box A and A') and the 5' ETS (3' and 5' hinge) (Beltrame & Tollervey 1995; Sharma & Tollervey 1999; Marmier-Gourrier et al. 2011; Dutca et al. 2011).

UtpA, UtpB and U3 snoRNP were proposed to be early binding factors in the co-transcriptional assembly process of the SSU processome (**Figure 1.6b**). The recruitment of UtpB and U3 snoRNP was shown to depend on the presence of UtpA subunits, but the integration of UtpA in pre-ribosomal particles was independent of UtpB or U3 snoRNP (Pérez-Fernández et al. 2007). Hence it was suggested that UtpA initiates the hierarchical assembly of the SSU processome and that its binding

is a pre-requisite for subsequent ribosome biogenesis factor recruitment (Pérez-Fernández et al. 2007; Pérez-Fernández et al. 2011).

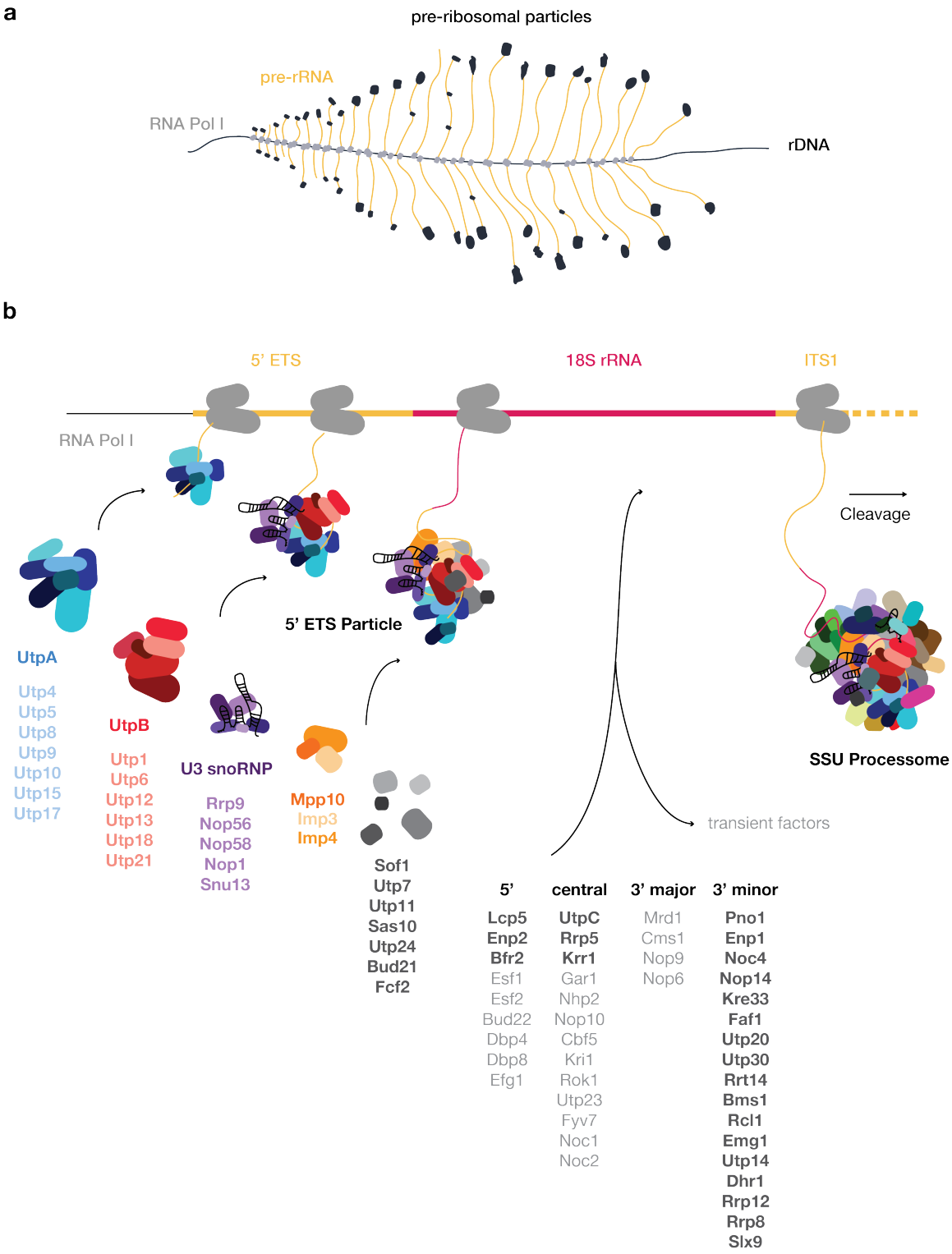
Decades of genetic and proteomic studies have revealed a plethora of proteins and snoRNAs involved in the nucleolar steps of small subunit ribosome biogenesis (Woolford & Baserga 2013). However, the order of their assembly on the nascent pre-rRNA, their precise roles and interaction networks have remained unclear. While the work described herein was in progress, two biochemical studies using truncated rRNA fragments mimicking different transcriptional stages of the pre-18S rRNA have shed light on the temporal assembly order of small subunit ribosome biogenesis factors (Chaker-Margot et al. 2015; Zhang et al. 2016). Since this order provides a structured framework for the introduction of the large number of early ribosome biogenesis factors, the results of these studies are described in the subsequent section.

Transcription of the first domain of the 35S pre-rRNA, the complete 5' ETS, leads to the formation of a 2-MDa particle containing UtpA, UtpB, U3 snoRNP and several additional factors such as Mpp10, Imp3 and Imp4 (Mpp10 complex) (**Figure 1.6b**). The three proteins of the Mpp10 complex were originally identified by yeast-two-hybrid screens as U3 snoRNP interacting proteins (Lee & Baserga 1999) and shown to influence not only A2 cleavage but also U3 snoRNA stability in the SSU processome (Wehner et al. 2002). Utp7 and Sof1 (Jansen et al. 1993) are two

individual β-propellers recruited to the 5' ETS, along with Utp11, Sas10, Bud21 and Utp24. While Sas10 contains an exosome interaction motif (Mitchell 2010), Utp24 harbors endonuclease activity and is responsible for cleavage of the A1 site, which separates the 5' ETS and the 18S rRNA (Bleichert et al. 2006; Tomecki et al. 2015; Wells et al. 2016).

Figure 1.6 | Early steps in eukaryotic ribosome biogenesis.

[a] Illustration of actively transcribed rDNA (black) as seen in Miller chromatin spreads (Miller and Beatty 1969). Pre-ribosomal particles (black) form on the nascent rRNA (yellow) which is transcribed by RNA polymerase I (grey). **[b]** UtpA (shades of blue), UtpB (shades of red), U3 snoRNP (purple, black), the Mpp10 complex (shades of orange) and individual factors (grey) assemble co-transcriptionally on the pre-rRNA and form the 5' ETS particle. Transcription of the 18S rRNA domains (5' domain, central domain, 3' major domain, 3' minor domain) leads to the recruitment of multiple factors. These factors are listed below each domain label. Protein names in bold and dark-grey indicate recruited subunits also present in the SSU processome whereas fine light-grey names show transient factors, which dissociate before SSU processome formation. The SSU processome catalyzes cleavage of the pre-rRNA and the particle is matured further. Assembly order of the listed proteins is based on (Chaker-Margot et al. 2015) and (Zhang et al. 2016).


a
pre-ribosomal particles
pre-rRNA
RNA Pol I
rDNA
b
5' ETS
18S rRNA
ITS1
RNA Pol I
Cleavage
5' ETS Particle
SSU Processome
transient factors
UtpA
Utp4
Utp5
Utp8
Utp9
Utp10
Utp15
Utp17
UtpB
Utp1
Utp6
Utp12
Utp13
Utp18
Utp21
U3 snoRNP
Rrp9
Nop56
Nop58
Nop1
Snu13
Mpp10
Imp3
Imp4
Sof1
Utp7
Utp11
Sas10
Utp24
Bud21
Fcf2
5'
Lcp5
Enp2
Bfr2
Esf1
Esf2
Bud22
Dbp4
Dbp8
Efg1
central
UtpC
Rrp5
Krr1
Gar1
Nhp2
Nop10
Cbf5
Kri1
Rok1
Utp23
Fyv7
Noc1
Noc2
3' major
Mrd1
Cms1
Nop9
Nop6
3' minor
Pno1
Enp1
Noc4
Nop14
Kre33
Faf1
Utp20
Utp30
Rrt14
Bms1
Rcl1
Emg1
Utp14
Dhr1
Rrp12
Rrp8
Slx9

While all the components bound at the 5' ETS stage are part of the SSU processome (Dragon et al. 2002; Grandi et al. 2002), the consecutive expansion of the rRNA transcripts from the 5' ETS to 18S rRNA domains also triggers the association of ribosome biogenesis factors not included in the SSU processome (**Figure 1.6b**) (Chaker-Margot et al. 2015; Zhang et al. 2016).

These transient factors include the 5' domain-associated RNA helicases Dbp4 and Dbp8. Dbp4 is presumably involved in the removal of the equally transient U14 snoRNA and was shown to bind to the 5' domain previously (Koš & Tollervey 2005; Granneman et al. 2006; Soltanieh et al. 2014; Soltanieh et al. 2015). In contrast to the helicase itself, the binding partners of Dbp4, Bfr2 and Enp2 (Liang & Fournier 1995; Soltanieh et al. 2014), stay throughout SSU processome assembly. Lcp5, Esf1, Bud22 and Efg1 are further factors recruited at the 5' domain stage. Lcp5, like Sas10, contains an exosome interacting motif (Mitchell 2010) and is later on also part of the SSU processome, while Efg1, Bud22 and Esf1 are transient factors.

Expansion of the construct to the central domain leads to the binding of UtpC, the box H/ACA snoRNA snR30, the RNA-helicase Rok1 and six other ribosome assembly factors (Rrp5, Noc1/Mak21, Noc2, Utp23, Fyv7, Krr1, Kri1) (**Figure 1.6b**). Akin to U14 in the 5' domain, snR30 is a transient snoRNA (Zhang et al. 2016), which depends on the presence of Utp23 and Rok1 for its removal (Hoareau-Aveilla et al. 2012; Martin et al. 2014; Khoshnevis et al. 2016).

While the addition of the 3' major domain to the growing pre-rRNA construct only adds transient factors (Mrd1, Nop6, Nop9, Cms1), the 18S completion and continuation into ITS1 mark the formation of the SSU processome. This step leads to the association of more than 18 factors (**Figure 1.6b**) (Chaker-Margot et al. 2015; Zhang et al. 2016). Amongst them are the GTPase Bms1 and its co-factor Rcl1. GTP- and Rcl1- binding by Bms1 are essential for A2 cleavage (Wegierski et al. 2001; Gelperin et al. 2001; Karbstein et al. 2005; Delprato et al. 2014). Other factors recruited are the acetyltransferase Kre33, the methyltransferase Emg1, the large helical-repeat protein Utp20, the RNA-helicase Dhr1 with Utp14, the Nop14-Noc4 complex, Pno1 and the D-site nuclease Nob1 (Turowski et al. 2014).

1.2.5 Nuclear export and final maturation steps of small subunit precursors

For the SSU processome to transition into an export competent pre-40S particle multiple steps have to occur. The 5' ETS particle has to be removed and degraded and new ribosome assembly factors and ribosomal proteins have to associate. The details and order of these events, as well as their regulation and the potential prerequisite of additional folding steps are yet to be defined.

5' ETS degradation is thought to be mediated by the nuclear exosome as mutations in its core and associated factors lead to the accumulation of the 23S pre-rRNA (Allmang et al. 2000). 5' ETS removal leads to the dissociation of many SSU

processome factors but Rrp12, Enp1, Nob1 and Pno1 stay on the nuclear pre-40S particle. Additional assembly factors that bind in the nucleus are Rio2, Dim1, Ltv1 and Tsr1 (Schäfer et al. 2003). Tsr1 is an inactive GTPase and a homolog to Bms1 (McCaughan et al. 2016), while Dim1 is a methyltransferase shown to modify the 3' loop of the 18S rRNA in the cytoplasm (Lafontaine et al. 1994) and Rio2 an essential serine kinase (Geerlings et al. 2003).

Cryo-EM structures of pre-40S particles (Strunk et al. 2011; Heuer et al. 2017; Scaiola et al. 2018) and RNA-protein cross-linking studies (Granneman et al. 2010) have determined the position of some of these factors on these precursor particles, which adopt an overall architecture similar to the mature 40S (**Figure 1.7**). Tsr1, Rio2 and Dim1 are located on the subunit interface and Pno1 and Nob1 on the platform, thereby blocking the binding sites of translation initiation factors on the still maturing subunit (Strunk et al. 2011; Heuer et al. 2017; Scaiola et al. 2018). Enp1 and Ltv1 prevent mRNA binding by prohibiting interactions of the ribosomal protein S3 needed to open the mRNA binding channel.

Pno1, Ltv1 and Rio2 have predicted nuclear export signals in their sequences and can recruit nuclear export machinery components. The pre-40S particles have been shown to rely on multiple, overlapping pathways to achieve transport through the nuclear pore (reviewed in Woolford & Baserga 2013; Chaker-Margot 2018). In the cytoplasm the pre-40S particles undergo final maturation steps. These include

removal of ribosome assembly factors, integration of ribosomal proteins, functional quality control and processing of the 20S pre-rRNA into the 18S rRNA by Nob1-catalyzed D-site cleavage.

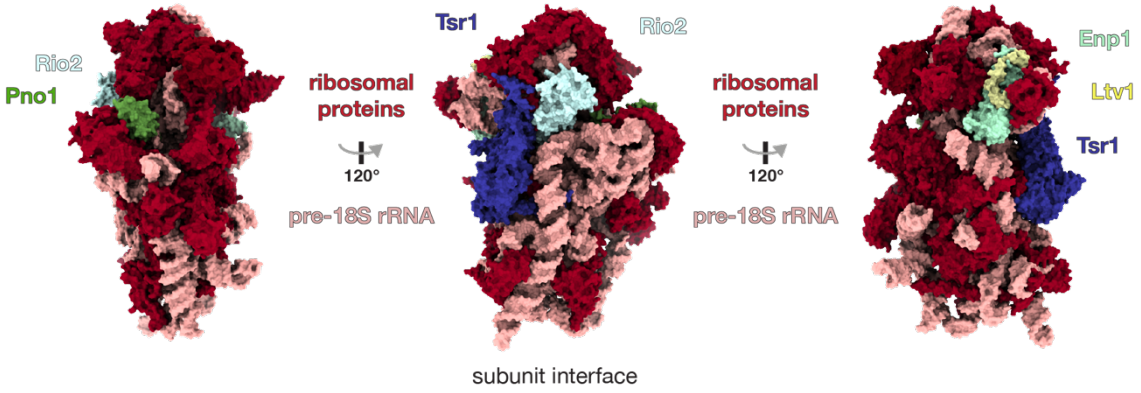


Figure 1.7 | Cytoplasmic precursors of the small ribosomal subunit are bound by specific assembly factors.

Three views of the late pre-40S particle (PDB 6FAI, (Scaiola et al. 2018)). Ribosomal proteins are colored dark-red, the 20S rRNA pink and the ribosome assembly factors in green (Pno1), light-blue (Rio2), purple (Tsr1), cyan (Enp1) and yellow (Ltv1).

In addition to blocking translation initiation, some ribosome assembly factors bound to the newly exported pre-40S subunit also prevent the premature association of the small and large subunit in the cytoplasm. Enp1 and Ltv1 leave the pre-40S particle after phosphorylation by the conserved kinase Hrr25 (Schäfer et al. 2006; Ghalei et al. 2015).

The still maturing pre-40S particles are then subjected to a functional quality control step involving a translation-like process. To this end, the pre-40S and a mature 60S subunit form an 80S-like complex (Granneman et al. 2005; Lebaron et al. 2012; Strunk et al. 2012). Tsr1 and Rio2, which are located on the subunit interface (**Figure 1.7**) of the pre-40S, dissociate during or after the formation of these 80S-like particles. The 80S-like ribosome cannot synthesize proteins during this test cycle as the binding sites for the mRNA and initiation factors were blocked prior to subunit joining by assembly factors and are continued to be occupied by the 60S subunit in the 80S-like particle (Strunk et al. 2012). 80S-like complex formation is promoted by the ATPase Rio1 (Turowski et al. 2014) and the translational GTPase Fun12 (eIF5B) (Lebaron et al. 2012). The ATPase Fap7 then induces a test-translocation, which releases Dim1 (Ghalei et al. 2017). Functional 80S-like particles are dissociated by the ATPase activity of Rio1 (Ferreira-Cerca et al. 2014), the termination factor Rli1 and Dom34 (Strunk et al. 2012). In the translation cycle of the mature ribosome, Rli1 and Dom34 also facilitate dissociation of the 80S ribosome and Fun12 mediates methionine-tRNA binding during initiation complex formation and aids the recruitment of the large subunit. Therefore, this process not only assesses the ability of the pre-40S to associate with and position the large subunit, but also its capabilities of binding Fun12 and stimulating its GTPase activity, recruiting Dom34 and binding the termination factor Rli1 (Karbstein 2013).

The released pre-40S particle undergoes D-site cleavage by Nob1 which produces the mature 18S rRNA. Since Pno1 is placed in close proximity to the D-site while Nob1 is flexibly attached (Heuer et al. 2017; Scaiola et al. 2018), it may protect the cleavage site during early steps of ribosome assembly and either presents the substrate to Nob1 during the final maturation stages or its dissociation from the particle renders the site accessible for nuclease.

1.2.6 Maturation of the large ribosomal subunit in yeast

The fully assembled SSU processome coordinates the cleavage of the 35S pre-rRNA at site A2 in ITS1 to separate small and large subunit maturation pathways. Thereafter the large subunit is assembled by a multitude of dedicated ribosome biogenesis factors and proceeds through numerous distinct assembly intermediates in the nucleolus, nucleus and cytoplasm (**Figure 1.8**) (reviewed in Konikkat & Woolford, 2017; Woolford & Baserga 2013).

As for the small subunit, the spatiotemporal assembly order of nucleolar large subunit biogenesis factors has recently been determined using truncated pre-rRNA fragments (**Figure 1.8**) (Chen et al. 2017; Chaker-Margot unpublished). Transcription of a construct encompassing the 5.8S, ITS2 and domains I to III of the 25S rRNA recruits more than 20 assembly factors (**Figure 1.8**). Some of these factors are known ITS2 binding proteins and others were shown to interact with domains I-III by

RNA-protein cross-linking analysis and structural work (**Figure 1.8**) (Wu et al. 2016). Most of the assembly proteins recruited at this stage are also called A3-factors (Ebp2, Brx1, Pwp1, Nop12, Nop7, Ytm1, Erb1, Rlp7, Nop15, Cic1, Drs1, Rrp1 and Has1) as they were found to associate hierarchically with the 27SA precursor and were required for ITS1 removal (Sahasranaman et al. 2011; Woolford & Baserga 2013). Rrp5, Noc1 and Noc2 are ribosome biogenesis proteins involved in both the early stages of SSU processome maturation and binding of large subunit precursors (Chaker-Margot et al. 2015; Zhang et al. 2016; Chen et al. 2017).

While domains III, IV and V are bound by only a small number of transient factors, transcription of domain VI leads to the assembly of a multitude of proteins usually associated with the 27SB precursor RNA. Several of them belong to the group of B-factors, which are required for C2 cleavage in ITS2 during nuclear maturation stages. The pre-5S RNP is incorporated into the maturing pre-60S towards the end of pre-rRNA transcription in the nucleolus. A characteristic of nucleolar pre-60S particles containing the 27SB rRNA is the presence of Nsa1 (Kressler et al. 2008), which is recruited early during domain I-II transcription.

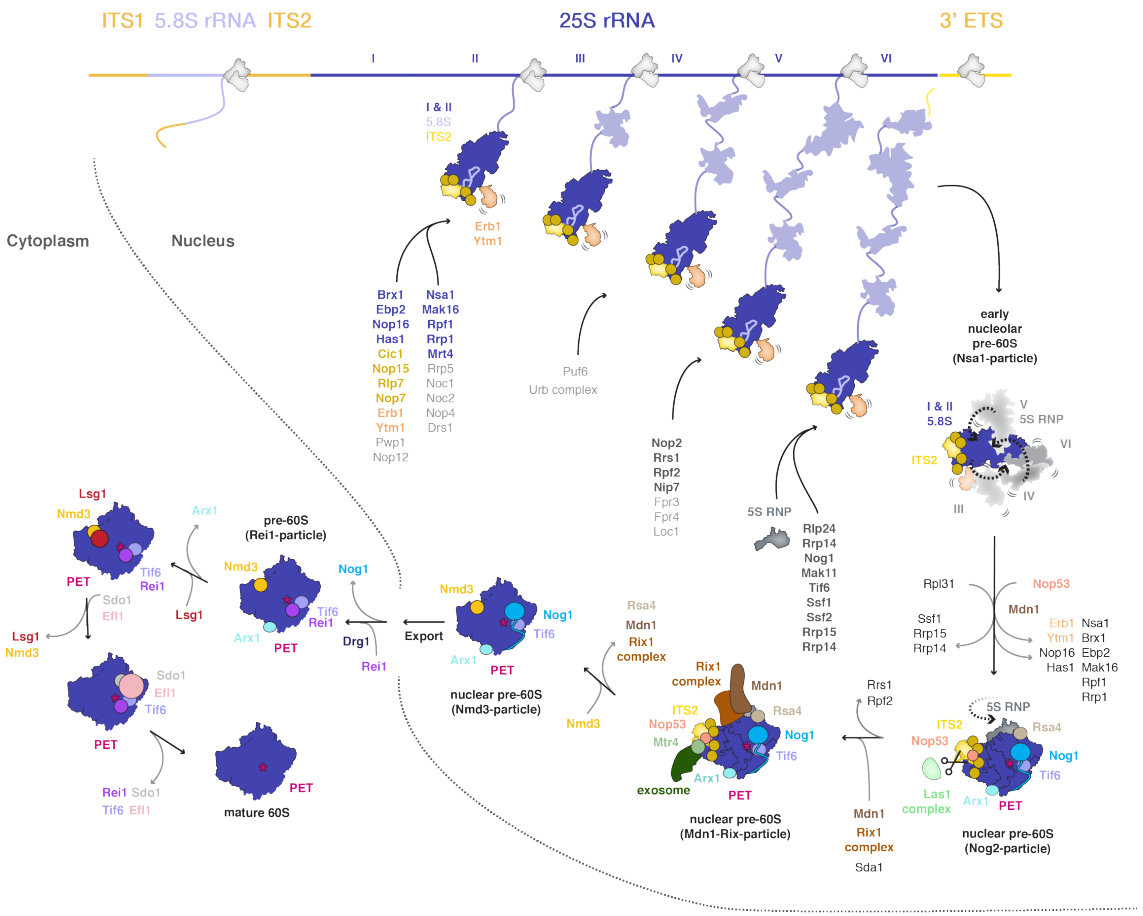


Figure 1.8 | Maturation of the large ribosomal subunit in yeast.

Schematic overview of the nucleolar, nuclear and cytosolic maturation steps in 60S maturation. Transcription of the second half of the rDNA locus occurs in the nucleolus and is mediated by RNA polymerase I (grey). The precursor RNA is bound by ribosome assembly factors in a co-transcriptional manner. Factors associating at different transcription stages are indicated as lists below. Factors in blue bind domain I or II of the stable core of the maturing pre-60S, while yellow factors bind ITS2 and the orange Ytm1-Erb1 complex plays a role in stabilizing domain I and III. Thin light-grey names indicate transient factors and bold grey names stable factors that are not resolved in the structure of the nucleolar pre-60S structure. Subsequent removal, exchange and addition of assembly factors is indicated with curved arrows on top of stage-transition arrows.

Recent cryo-EM structures of Nsa1 and 27SB containing pre-60S particles have visualized the different maturation intermediates present at this early nucleolar stage and have shown a controlled modular assembly mechanism of the 25S rRNA domains (I-VI) which leads to PET formation (Kater et al. 2017; Sanghai et al. 2018; Zhou et al. 2018). Domains I, II and parts of domain VI together with ITS2 and the 5.8S rRNA form a stable core, which is bound by several ribosome assembly factors (**Figure 1.8**). The other domains are still flexible in the earliest state of this nucleolar precursor (**Figure 1.8**). Stabilization of either domain III through Ytm1-Erb1, or VI via Ssf1-Rrp14-Rrp15, on the core of the particle is followed by the Mdn1 dependent removal or exchange of assembly factors and folding of domain V and then IV. Mdn1 (also known as Rea1) is a large dynein related AAA+ protein that can bind and remove proteins containing ubiquitin-like domains, such as Ytm1 or Rsa4, using its ATPase activity (Ulbrich et al. 2009; Baßler et al. 2010; Raman et al. 2016). The PET formed by the pre-25S folding events is blocked by an extension of Nog1 and further stabilized by Arx1. While the nucleolar stages of large subunit assembly guide pre-rRNA domain folding and PET formation, the nuclear stages organize the remodeling of the central protuberance and the removal of ITS2.

The pre-60S particle transitions from the nucleolus to the nucleus and is bound by new assembly factors while others dissociate. Nog2, a putative GTPase, is the defining factor for the early nuclear maturation stages (Saveanu et al. 2001). A cryo-EM structure of the Nog2-particle (Wu et al. 2016), has revealed that the pre-25S and

5.8S adopt near-mature structures but are still bound by ribosome biogenesis proteins. Especially the subunit interface, which harbors the functionally important PTC, PET, A-P-E sites, is covered by assembly factors. ITS2 can be intact or processed in the Nog2 particle and is bound by Nop53 which contains an exosome recruitment motif (Thoms et al. 2015). The 5S rRNA is placed in the particle but has to undergo a 180-degree rotation in order to adopt its mature position (Leidig et al. 2014).

Remodeling of the central protuberance, of which the 5S rRNA is the major component, is an important maturation step in the nucleus. The 5S rRNA is chaperoned by Rpf2 and Rrs1 in the Nog2-particle. Rotation of the 5S rRNA likely results in Rpf2 and Rrs1 dissociation and subsequent binding of Sda1, the Rix1 complex and Mdn1. Structural characterization of a nuclear pre-60S particle with the Rix1 complex, Sda1 and Mdn1 bound (Mdn1-Rix1 particle) (Barrio-Garcia et al. 2016) showed that the binding site of Rpf2 and Sda1 overlap and that the 5S rRNA adopts a near-mature conformation at this stage. Thus, 5S rotation and Rpf2 and Rrs1 dissociation probably precede Mdn1, Sda1 and Rix1 complex binding. Mdn1 and the Rix1 complex remove Rsa4 and complete the central protuberance remodeling (Chen et al. 2018).

ITS2 is cleaved by the Las1 complex at site C2 in the nucleus (Gasse et al. 2015). This results in the 7S, a precursor of the 5.8S rRNA, and the 25.5S rRNA. The 3' end of the 7S precursor is engaged by the exosome-associated helicase Mtr4 which is bound via the AIM motif of Nop53 (Thoms et al. 2015). Processing by the exosome leads to the 6S precursor which is matured into the 5.8S rRNA in the cytosol (Fromm et al. 2017; Schuller et al. 2018).

Export of the pre-60S from the nucleus relies on binding of Nmd3, whose binding site overlaps with Nog1 and Nog2 (Wu et al. 2016; Ma et al. 2017; Malyutin et al. 2017). While Nog2 is released from the particle through GTP hydrolysis, Nog1 is still bound in Nmd3 particles. This suggests alternative binding sites or conformational flexibility of Nmd3 and Nog1 (Kallstrom et al. 2003; Matsuo et al. 2014). Transportation through the nuclear pore is facilitated by the nuclear export signal in the sequence of Nmd3.

In the cytosol the AAA-ATPase Drg1 removes the ribosomal protein homolog Rlp24 from the pre-60S and rpL24 can bind. rpL24 then recruits Rei1 which in turn initiates the Jjj1 and Hsp70 dependent dissociation of Arx1. As a consequence of Rlp24 removal Nog1 leaves the particle (Pertschy et al. 2007; Kappel et al. 2012). Upon Nog1 dissociation Rei1 inserts an extension into the PET (Greber et al. 2016).

Nmd3 and Tif6 block subunit joining in the cytoplasm and are removed as one of the last steps in 60S maturation. Their removal occurs through a proof-reading mechanism for functional sites of the large subunit. Post Nog1-release the conformational space of the PET is probed by Rei1 and the adjacent P-site bound by Sdo1 (Weis et al. 2015; Wu et al. 2016; Greber et al. 2016; Ma et al. 2017). Through a complicated mechanism the GTPase Efl1 and its co-factor Sdo1, Tif6 and the sarcin-ricin loop of the 25S rRNA test drive the P-site and translational GTPase activating structure of the large subunit and thereby trigger the Lsg1-dependent release of Nmd3 from the PTC as well as the dissociation of Tif6 from GTPase activating center (Konikkat & Woolford, 2017).

1.2.7 Regulation of ribosome biogenesis

Ribosome biogenesis is a highly efficient and energy consuming process. In actively growing yeast cells 2000 ribosomes are produced each minute (Warner 1999), which uses more than 60% of the available cellular ATP (Zhou et al. 2015). Hence, this process has to be tightly controlled in response to shifting nutrient availability and growth. The TOR (Target of Rapamycin) kinase complex occupies a central role in the multiple mechanisms used to tune ribosome production. When energy supply is high TOR is active and phosphorylates its downstream effector proteins (Albert & Hall 2015). The transcription of pre-rRNA, the activity of pre-rRNA processing factors and the expression of ribosomal protein genes constitute three key processes regulated in a nutrient-dependent manner.

Synthesis of the 35S precursor is controlled on the level of activation or inactivation of rDNA repeats (Sandmeier et al. 2002), transcription initiation (French et al. 2003) and elongation (Zhang et al. 2010). RNA polymerase I is an indirect target of TOR regulation (Philippi et al. 2010). Under nutrient-depleted conditions Rrn3 binding to RNA Polymerase I is down regulated by phosphorylation of Rrn3 (Torreira et al. 2017). The polymerase forms inactive homodimers which are stored in the cell (Torreira et al. 2017). Regulation of RNA polymerase I activity also occurs at the elongation step through the Paf1-complex, which is presumed to stop stimulating transcription elongation under nutrient-deprived conditions or following TOR complex inhibition by rapamycin (Zhang et al. 2010).

Processing of the pre-rRNA is influenced by environmental conditions (Talkish et al. 2016; Kos-Braun et al. 2017). Starvation, heat-shock, nitrogen-limitation and rapamycin lead to the accumulation of the 23S and 27S species, representing a pause in processing activity. The casein kinase 2 complex, is involved in a nutrient-dependent switch from A2 to A3 cleavage in the pre-rRNA processing pathway that could give rise to the 23S and 27S precursors (Kos-Braun et al. 2017). However, the exact mechanism responsible for the accumulation of these precursors and halting of ribosome maturation are still elusive.

The CURI-complex, comprised of UtpC, the casein kinase 2 complex and Ifh1, regulates transcription of ribosomal protein genes. The transcription factor Ifh1 can associate with UtpC and casein kinase 2 when TOR is inhibited, which titrates it off ribosomal gene promoters and leads to a reduction in RNA polymerase II dependent transcription of these genes (Albert et al. 2016).

Chapter 2 | Biochemical and Structural Characterization of UtpA in Isolation

UtpA is an essential protein complex thought to initiate the early steps of eukaryotic ribosome biogenesis by binding to pre-ribosomal RNA and recruiting downstream-binding multi-subunit complexes such as UtpB and U3 snoRNP (Pérez-Fernández et al. 2007; Pérez-Fernández et al. 2011).

Despite the fundamental importance of UtpA to all eukaryotes little was known aside from its molecular composition. Its function, structure and RNA binding site were elusive. To delineate the role of UtpA in ribosome biogenesis I set out to characterize the molecular architecture of this complex in isolation by using size-exclusion chromatography, biochemical assays of subcomplexes, negative-stain electron-microscopy and DSS cross-linking and mass spectrometry analysis.

Furthermore, in collaboration with the laboratory of David Tollervey at the Wellcome Center for Cell Biology in Edinburgh, Scotland, we determined the RNA binding site of all UtpA subunits, and hence defined the composite binding site of the whole complex, through UV-induced cross-linking and cDNA analysis (CRAC) (Granneman et al. 2009). The results of this study, combined with findings about the architecture and RNA binding sites of UtpB obtained by Jonas Barandun, were published in (Hunziker et al. 2016).

Elisabeth Petfalski and I performed CRAC experiments on UtpA subunits. UtpB subunit CRAC data was generated by Elisabeth Petfalski. Clémentine Delan-Forino, Hywel Dunn-Davies and I analyzed the CRAC data. Kelly R. Molloy, Yi Shi and Brian T. Chait collected mass spectrometry data of the DSS-cross-linked UtpA sample and manually curated and analyzed the cross-links.

2.1 UtpA is highly flexible in isolation

UtpA from yeast is composed of seven protein subunits (Utp4, Utp5, Utp8, Utp9, Utp10, Utp15, Utp17), all of which consist of well described structural motifs such as WD40 β-propellers and α-helical repeats (**Figure 2.1**). Rather than enzymatic activity, β-propellers and α-helical repeat structures predominantly have scaffolding functions in cellular processes (Makarova et al. 2005). Therefore, the function of UtpA in ribosome biogenesis is likely of a structural nature and the elucidation of its architecture important for defining its role in this essential pathway.

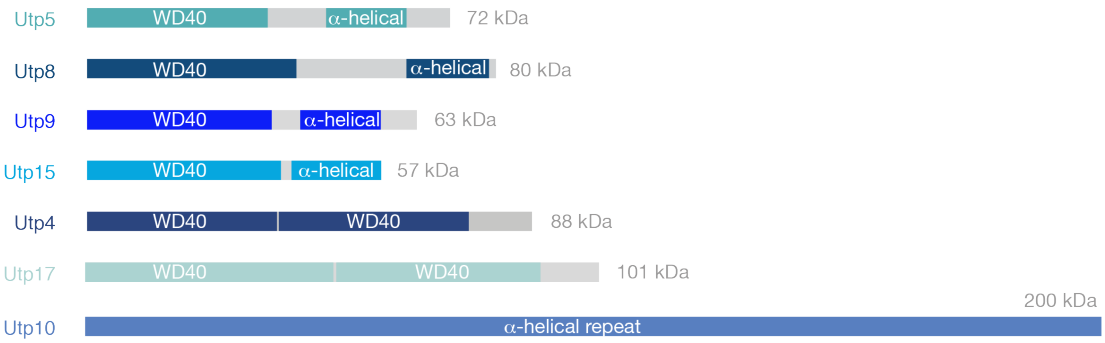


Figure 2.1 | Schematic domain architecture of UtpA subunits from *S. cerevisiae*.
Individual subunit names are indicated on the left with their approximate molecular weight labeled on the right. Folded domains are shown as colored and labeled blocks while predicted unstructured regions in the protein sequence are depicted in grey.

To biochemically and structurally characterize UtpA in isolation, a purification protocol from endogenous sources was established first. Utp10, the largest subunit of UtpA, was C-terminally tagged with protease-cleavable GFP. Affinity purification using anti-GFP-nanobody coupled resins followed by size exclusion

chromatography, yielded pure and stoichiometric UtpA (**Figure 2.2a**). As the amounts of pure protein obtained by endogenous purification were limiting for biochemical assays and certain structural characterization techniques, the seven subunits of UtpA were cloned into plasmids suitable for stable integration in the yeast genome and galactose driven over-expression (**see 6.3**). Purification of over-expressed UtpA using the same strategy as for the isolation from endogenous sources, improved the protein amounts while yielding the same purity and stoichiometry of the complex (**Figure 2.2b**).

To visualize the molecular architecture of UtpA, purified protein samples were analyzed by negative stain electron microscopy (EM) (**Figure 2.3a**). This revealed that UtpA is a highly flexible complex composed of a body and an elongated feature (**Figure 2.3b**). The elongated feature of UtpA likely consists of Utp10, a 200 kDa α-helical repeat protein and the largest subunit of UtpA. The high degree of flexibility observed under negative stain EM is also reflected in the broad elution profile of UtpA on size exclusion chromatography (**Figure 2.2 a,b**).

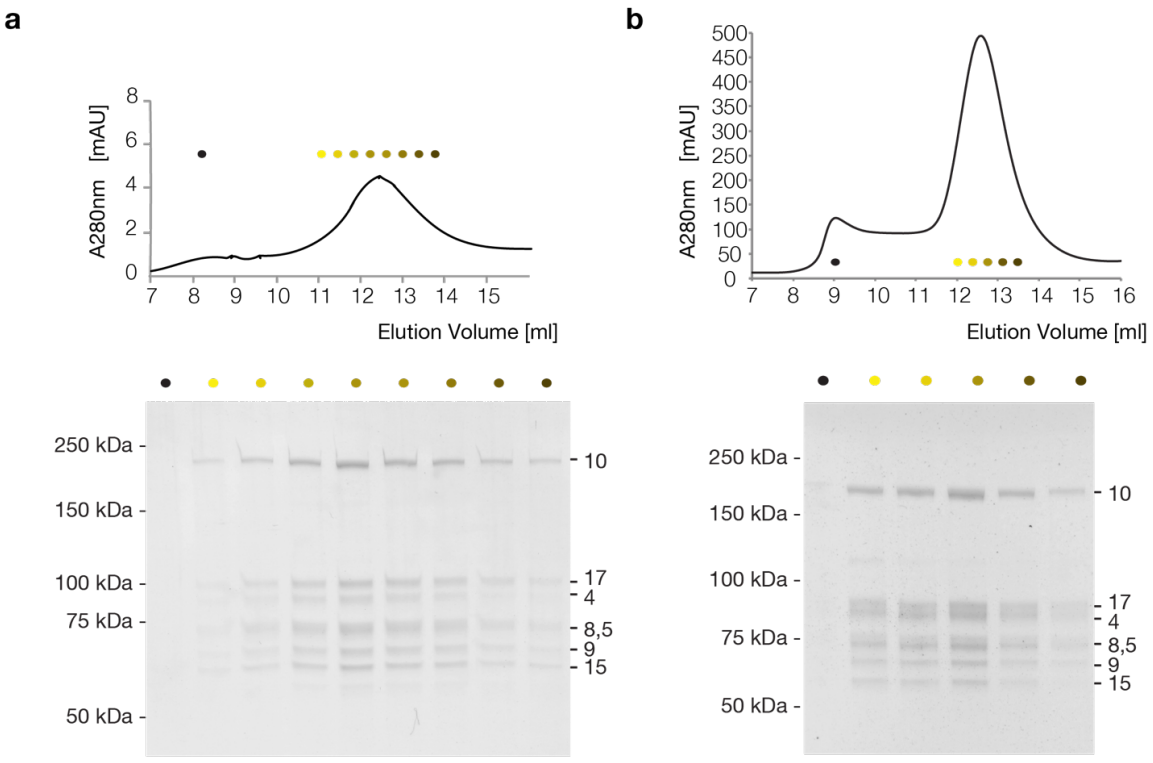


Figure 2.2 | Purification of endogenous and overexpressed UtpA from *S. cerevisiae.*

[a] Size-exclusion chromatogram and SDS-PAGE analysis of UtpA purified from endogenous sources. Each lane in the Coomassie-blue stained SDS-PAGE corresponds to a fraction in the size-exclusion elution profile. Corresponding lanes and fractions are marked with a dot of the same color. Co-eluting U-three-proteins (Utp) of UtpA are marked with their number on the right side of the denaturing protein gel. Molecular marker positions are shown on the left side. **[b]** Purification of UtpA overexpressed in *S. cerevisiae*. Elements and labeling are as described in [a].

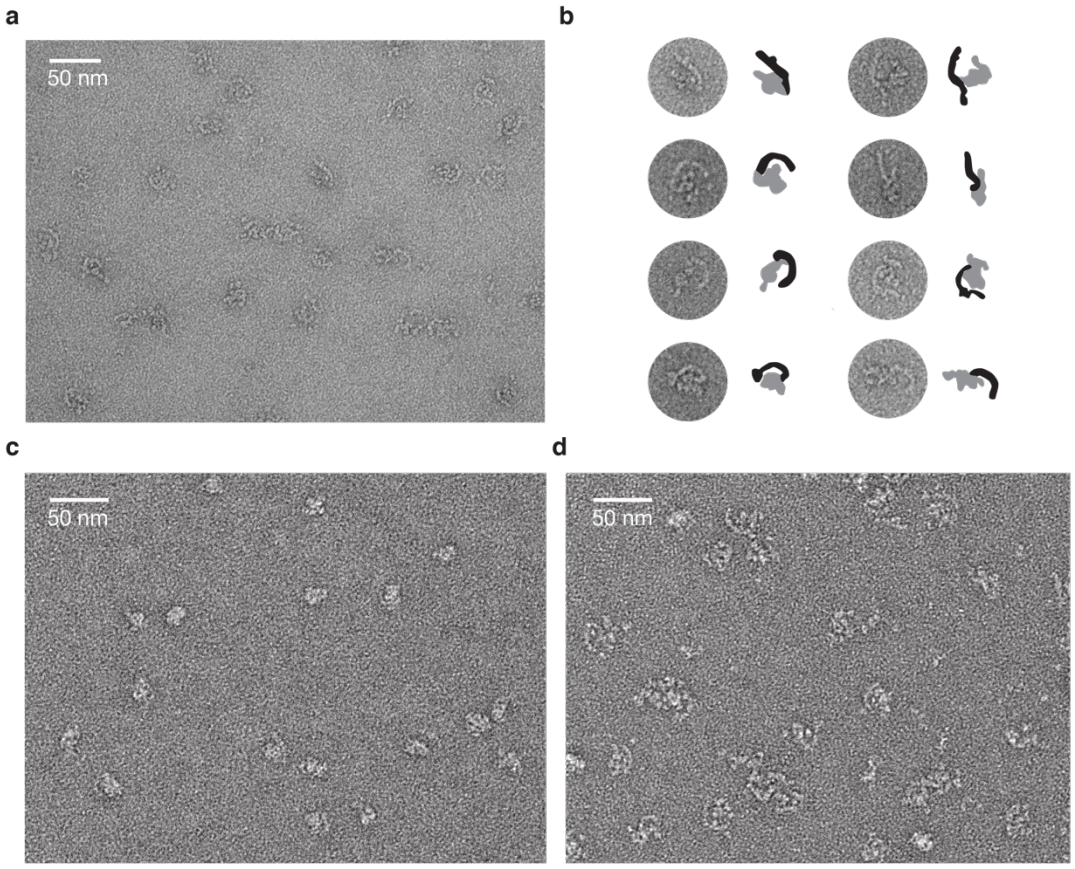


Figure 2.3 | Visualization of purified UtpA from yeast by negative-stain electron-microscopy.

[a] A section of a representative micrograph with a size-bar in the top left corner. **[b]** Selection of particles cropped from the micrograph in [a] showing the conformational flexibility of the complex. Outlines of the particles were traced and are displayed as schematics on the right of each particle. An elongated feature is colored in black, while the body is colored in grey. The diameter of the circle is 56 nm. **[c]** Section of a micrograph of on-column cross-linked UtpA. A size bar is indicated on the top. **[d]** The same sample as in [c] before cross-linking.

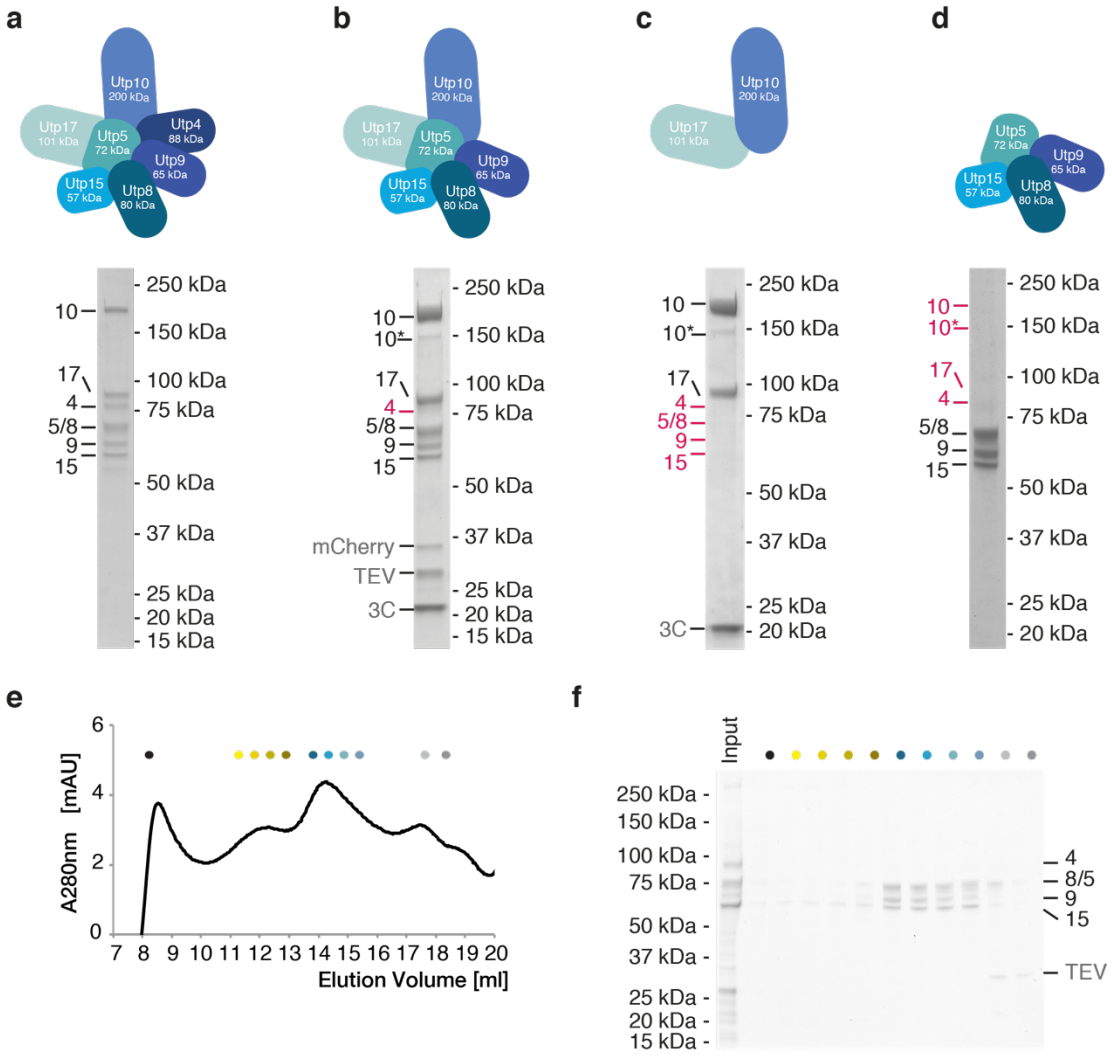
This inherent flexibility of UtpA and the resulting conformational heterogeneity in the sample prohibited further structural analysis of the complex in isolation by EM or X-ray crystallography. Several trials to obtain 2D averages from negative stain EM data sets failed (data not shown). To increase the structural homogeneity, UtpA was cross-linked with glutaraldehyde during size exclusion chromatography as described previously (Shukla et al. 2014). However, this resulted in a distorted complex and did not improve the amenability of the sample for EM analysis (**Figure 2.3 c,d**).

2.2 UtpA is composed of two stable sub-complexes

Visualization of UtpA by negative-stain EM has highlighted the dynamic nature of UtpA. To gain further insights into the molecular organization of UtpA in the absence of three-dimensional structural information, the complex was purified under low salt conditions (200mM NaCl) and immobilized on anti-GFP-nanobody coupled resin via the GFP-tagged Utp10 subunit. Subsequently, the heteroheptameric complex (Figure 2.4**a**) was washed and incubated with increasing salt concentrations. The first subunit to disassociate from UtpA with increasing salt concentration was Utp4 (**Figure 2.4b**). A tetrameric subcomplex consisting of Utp5, Utp8, Utp9 and Utp15 separated at intermediate salt concentration from the stable Utp10-Utp17 dimer (Figure 2.4**c**). This suggests that UtpA is composed of two main subcomplexes with the largest two subunits (Utp10, Utp17) forming a stable dimer interacting with a pentameric subcomplex (Utp4, Utp5, Utp8, Utp9, Utp15). Further, this result indicates that within the pentamer Utp4 is the most salt sensitive subunit.

Figure 2.4 | UtpA is composed of two sub-complexes.

[a] The heteroheptameric UtpA complex can be purified under low salt conditions (200 mM NaCl). A schematic illustration of the complex with its subunits labeled and their molecular weights indicated is shown above the SDS-PAGE gel lane of the purified sample. Protein bands corresponding to U-three-proteins (Utps) are labeled with their respective numbers on the left of the gel lane**. [b,c]** Schematic depictions and SDS-PAGE analysis of UtpA subcomplexes resulting from purifications with buffers of different ionic strengths. Tagged UtpA (Utp10-3C-GFP and Utp15-TEV-mCherry) was purified at 200mM NaCl on anti-GFP sepharose and incubated with buffers containing either 400mM NaCl, yielding UtpA without Utp4 [c], or 800mM NaCl, yielding the Utp10-Utp17 dimer [d]. The positions of dissociated Utp proteins are indicated by a pink label of their number. 10* labels a degradation product of Utp10. TEV and 3C proteases (grey) were used for the elution and the removal of mCherry (grey). **[d]** SDS-PAGE analysis and schematic depiction of the main peak fraction of co-eluting Utp5, Utp8, Utp9 and Utp15 on size-exclusion chromatography. The pentameric complex comprised of Utp4, Utp5, Utp8, Utp9 and Utp15-TEV-mCherry was overexpressed in yeast and affinity purified. Utp4 dissociated from the complex during size-exclusion chromatography resulting in the elution of the heterotetramer shown. **[e]** Chromatogram of the size-exclusion run described in [d]. **[f]** Coomassie-blue stained SDS-PAGE analysis of the overexpressed pentameric complex of Utp4, Utp5, Utp8, Utp9 and Utp15 injected on size-exclusion (Input) and of color-coded peak fractions.


a
Utp10
200 kDa
Utp17
101 kDa
Utp5
72 kDa
Utp4
88 kDa
Utp9
65 kDa
Utp15
57 kDa
Utp8
80 kDa
10
17
4
5/8
9
15
250 kDa
150 kDa
100 kDa
75 kDa
50 kDa
37 kDa
25 kDa
20 kDa
15 kDa
b
10*
mCherry
TEV
3C
c
d
e
A280nm [mAU]
Elution Volume [ml]
f
Input
4
8/5
9
15
TEV

To elucidate if the five smaller subunits can form a subcomplex in the absence of the Utp10-Utp17 heterodimer, Utp4, Utp5, Utp8, Utp9 and a protease-cleavable Utp15-mCherry fusion protein were over-expressed in *S. cerevisiae*. After anti-mCherry-nanobody based affinity purification, all five subunits were present (**Figure 2.4f**), but Utp4 dissociated from Utp5, Utp8, Utp9 and Utp15 during the subsequent size-exclusion step (**Figure 2.4e,f**). The loss of Utp4 in low-salt buffer conditions during size-exclusion chromatography suggests either a weak association of Utp4 with the other subunits or the necessity for Utp10 and/or Utp17 for its stable integration within UtpA.

To investigate the protein–protein interactions within UtpA further, we cross-linked purified UtpA with DSS and analyzed the resulting cross-links by mass spectrometry (**Figure 2.5**). The DSS concentration used to cross-link UtpA was determined by titrating increasing concentrations of DSS while keeping the concentration of UtpA constant. The extent of cross-linking at each DSS-concentration was assessed by the mobility shift of protein bands on SDS-PAGE (data not shown). At a concentration of 0.2 mM DSS the majority of individual UtpA subunit bands has shifted upwards to an area below the well of the gel. Hence this concentration was used to cross-link a large-scale purification of UtpA. Cross-linked complexes were precipitated and subjected to mass spectrometry analysis.

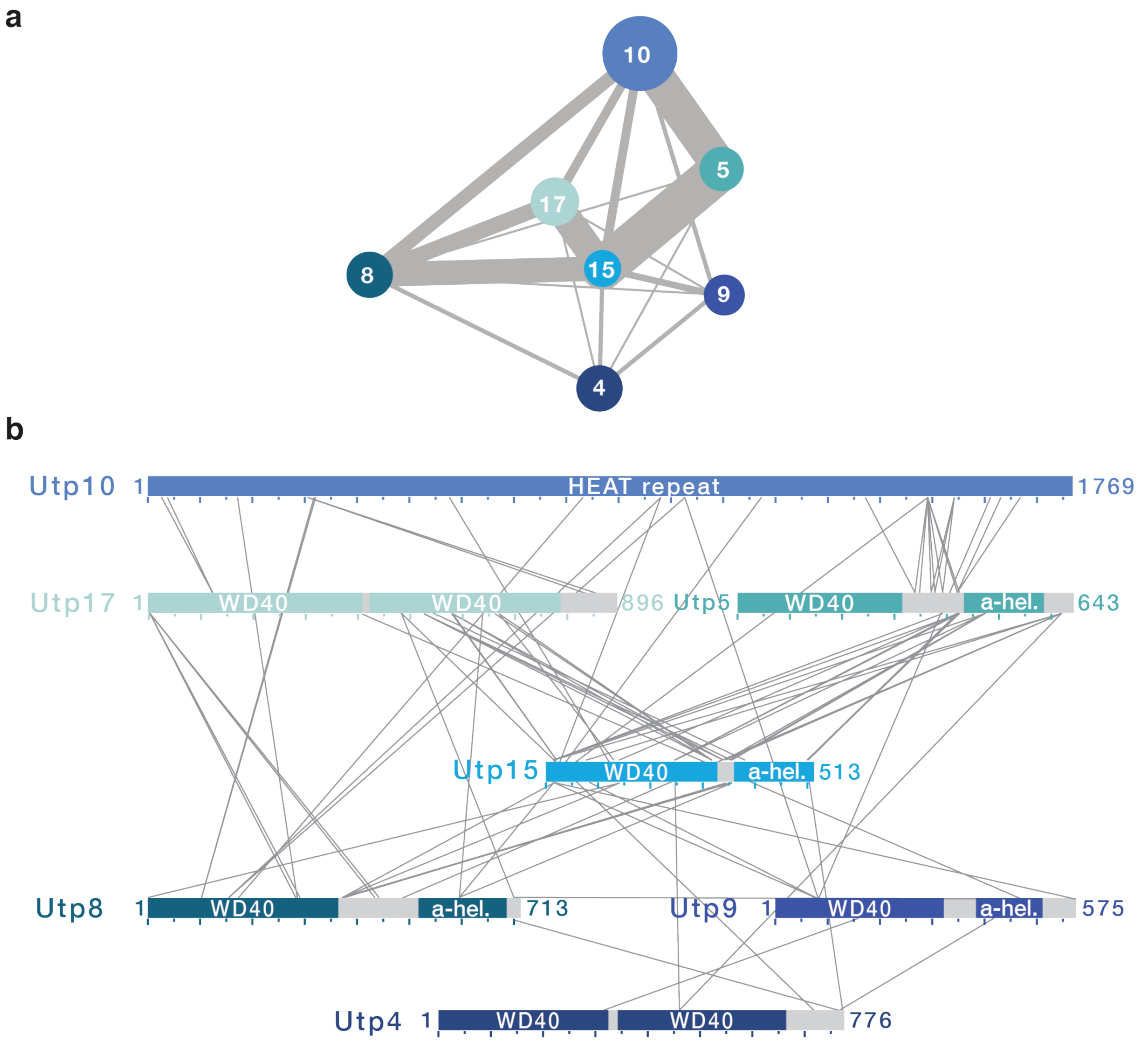


Figure 2.5 | Visualization of DSS cross-linking and mass spectrometry analysis of UtpA.

[a] Schematic representation of inter-subunit DSS cross-links (grey lines) of UtpA subunits (circles colored in different shades of blue). The thickness of the line represents the number of cross-links shared between the subunits. **[b]** Detailed depiction of inter-subunit DSS cross-links of UtpA. Cross-links are shown as grey lines and protein subunits as grey bars with their domain architectures in color. Residue numbers for all UtpA subunits are indicated on the left and right of the respective schematic.

A dense network of cross-links was identified between Utp10, Utp5, Utp15, Utp8 and Utp17, suggesting that these are located in close proximity within UtpA (**Figure 2.5a**). As expected, Utp10 shares cross-links with Utp17 but also shares a large number of cross-links with Utp5 and Utp8 (**Figure 2.5b**). Utp4 was not strongly cross-linked to other subunits of UtpA and only shared few cross-links with Utp8, Utp9 and Utp15 (**Figure 2.5b**). Taken altogether, these observations suggest a molecular organization of UtpA in which Utp10 and Utp17 form a stable dimer that has spatial proximity to Utp5, Utp8 and Utp15 and to a lesser extent to Utp9 and the salt-labile Utp4.

2.3 UtpA and UtpB bind on distinct sites of the pre-rRNA

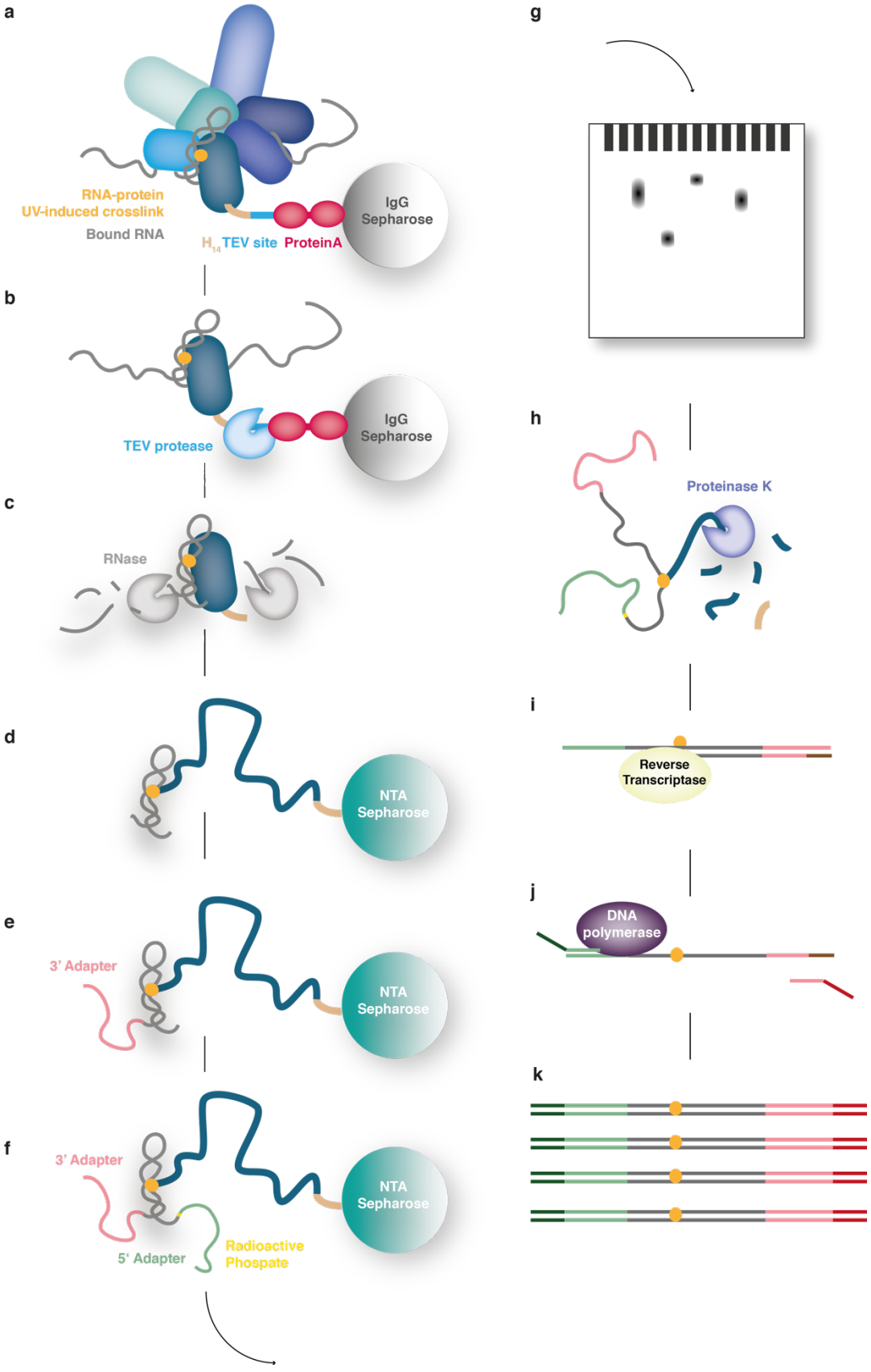
The β-propeller and α-helical repeat structures present in all UtpA subunits and the absence of evident enzymatic activity suggests that UtpA is providing a structural framework for other ribosome assembly factors and the nascent pre-rRNA. To understand how UtpA binds pre-rRNA we set out to determine which UtpA subunits are involved in RNA recognition and what their binding sites are. Cross-linking and analysis of cDNA (CRAC) experiments were performed on all individual subunits of the complex.

CRAC is a method that combines *in vivo* UV cross-linking with deep sequencing of cDNA (Granneman et al. 2009). It allows for the unbiased identification of RNA binding sites of a protein with single nucleotide resolution and high specificity. This is achieved by using a bipartite affinity tag (HTP tag), which is fused to the target subunit. The first part of the affinity tag is used for an initial purification step under native conditions. The other part is a His-tag, which allows for a second purification step under denaturing conditions.

To obtain the ensemble of CRAC datasets needed to describe the composite binding site of UtpA, seven yeast strains each containing one HTP-tagged UtpA subunit were subjected to the following protocol.

Figure 2.6 | Overview of steps in the cross-linking and analysis of cDNA (CRAC) protocol.

[a] A yeast strain harboring a C-terminal H_{14}-TEV-2xProteinA tag on one of the UtpA subunits (in shades of blue) is irradiated with UV-light to induce protein-RNA cross-links (orange). The tagged protein:RNA complex is purified from lysate via IgG beads. **[b]** Associated proteins are washed away with high-salt buffer before the protein:RNA complex is eluted by TEV-protease cleavage (light-blue). **[c]** The cross-linked and purified RNA is subjected to RNAse-foot printing by RNAseA and RNAse T1 (in grey). **[d]** Cross-linked protein and RNA are further purified under denaturing conditions using the H_{14}-moiety of the protein tag and NTA sepharose. **[e]** The purified RNA is dephosphorylated (not shown) and an adenylated RNA adapter (pink) is ligated to the 3' end. The excess of 3' adapter is washed away (not shown). **[f]** Prior to the ligation of a second RNA adapter (green) to the 5' end, the target RNA is labeled with P32 (yellow). **[g]** The radioactively labelled protein:RNA complex is eluted through imidazole, separated on a 4-20% SDS-PAGE and then transferred onto a nitrocellulose membrane. To specifically isolate the labelled complex the membrane is exposed to a phosphor screen and the radioactive area around the molecular weight of the tagged UtpA subunit is excised. **[h]** Proteinase K digest (purple) of the protein-RNA hybrid elutes the RNA from the excised membrane and removes the protein moiety. **[i]** A DNA strand complementary to the eluted RNA is synthesized by RNA reverse transcriptase (yellow) after binding of a DNA primer (brown) to the 3' end of the RNA adapter. **[j,k]** The single stranded DNA templates from the reverse transcription step are amplified by DNA Polymerase (purple) in a PCR reaction. Primers used in this reaction contain sites complementary to the RNA adapter sequences (light-green, light-pink) and to the Illumina sequencing platform (dark-green, dark-red). The protein binding site can be identified by micro-deletions caused by leftover protein residues at the cross-linking site (orange).


a
RNA-protein
UV-induced crosslink
Bound RNA
H14 TEV site ProteinA
IgG
Sepharose
b
TEV protease
IgG
Sepharose
c
RNase
d
NTA
Sepharose
e
3' Adapter
NTA
Sepharose
f
3' Adapter
5' Adapter
Radioactive
Phospate
NTA
Sepharose
g
h
Proteinase K
i
Reverse
Transcriptase
j
DNA
polymerase
k

In the first step of this procedure, yeast cultures harboring an endogenously tagged subunit are UV-irradiated. Crosslinks between proteins directly interacting with RNA are created (**Figure 2.6a**). The cross-linked RNA-protein complex is isolated from cell lysate by affinity purification through the Protein A moiety of the tag (**Figure 2.6a**).

Associated, but uncross-linked, proteins are washed away by multiple incubations in high-salt buffers. The tagged subunit is eluted by protease cleavage (**Figure 2.6b**) and subjected to a partial RNA digest to shorten the bound RNA fragments to the minimal length protected by the protein (**Figure 2.6c**). Subsequently, protein contaminants still bound to the target RNA are removed by a second purification step under denaturing conditions (**Figure 2.6d**). While still bound to the resin, the cross-linked RNA is dephosphorylated and ligated at the 3' end to an adenylated RNA adapter (**Figure 2.6e**). Excess of the 3' RNA adapter is washed away and the 5' end of the RNA is labelled with radioactive phosphate prior to the ligation of a second RNA adapter (**Figure 2.6f**).

The covalently cross-linked protein:RNA complex is eluted and separated on SDS-PAGE. Radioactively labelled complexes are identified after transfer onto a nitrocellulose membrane and exposure to a phosphor screen (**Figure 2.6g**). The radioactive area around the approximate molecular weight of the tagged protein subunit is excised and the RNA eluted by proteinase digestion (**Figure 2.6h**).

Eluted RNA is reverse-transcribed into DNA (**Figure 2.6i**) and the resulting DNA fragments are amplified by PCR (**Figure 2.6j**) prior to deep sequencing. Cross-linked sites are identified by micro deletions and mutations in the sequences.

As negative control, CRAC experiments were carried out with untagged wildtype yeast. For all subunits of UtpA a radioactive signal was observed after separation on SDS PAGE (data not shown). This indicates, that all subunits bind RNA. From the bound RNA cDNA libraries were prepared and deep sequenced. Sequenced cDNA libraries were mapped on to the whole yeast genome with Novoalign and analyzed with the pyCRAC software package (Webb et al. 2014).

Mapping of UtpA subunit datasets showed an enrichment of reads for all seven subunits on the ribosomal RNA locus (RDN37) relative to the non-tagged wildtype control data set (**Figure 2.7a**). All subunits of UtpA showed predominant cross-linking in the 5' proximal region of the 5' ETS (**Figure 2.7a**), consistent with a key role in initiating the ribosome assembly process. The wild-type control showed only a prominent peak in the 25S rRNA, which has been seen in many experiments and represents a common contaminant (Bradatsch et al. 2012; Granneman et al. 2010; Schneider et al. 2012).

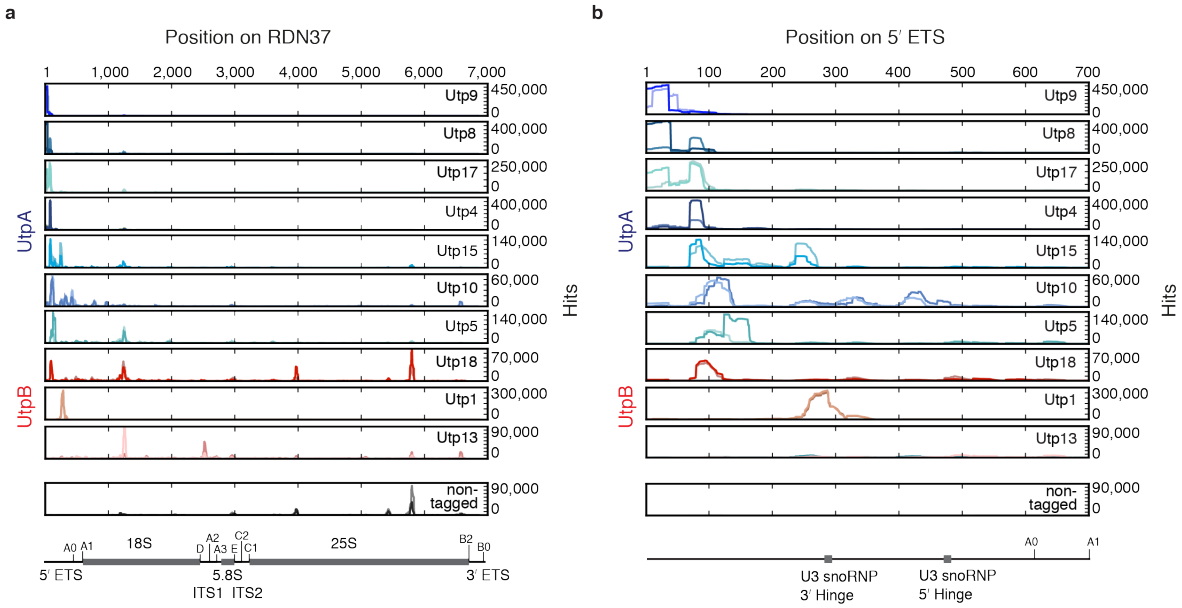


Figure 2.7 | Binding sites of UtpA and UtpB within the 35S pre-rRNA.

[a] Sequences obtained from CRAC experiments with all UtpA subunits (in shades of blue, duplicates in light blue), selected UtpB subunits (in shades of red, duplicates in light red) and the non-tagged control (in grey, duplicate in light grey) were aligned to the rDNA locus (RDN37) and plotted as frequency of recovery (hits per million mapped reads) at each nucleotide position (indicated above all panels). Individual scales for the frequency of recovery are shown on the right of each subunit panel. The 35S pre-rRNA encoded by RDN37 is schematically depicted below the traces with cleavage sites (A0, A1, D, A2, A3, E, C2, C1, B2, B0), internal and external spacer regions (ITS1, ITS2, 5' ETS, 3' ETS) and ribosomal RNA (18S, 5.8S, 25S) indicated. **[b]** Expanded view of CRAC library hits on the 5' ETS (nucleotides 1–700 of RDN37). Positions of U3 snoRNA base-pairing sites (3' Hinge, 5' Hinge) and pre-rRNA cleavage sites (A0, A1) are shown on a schematic representation of the 5' ETS below.

Expansion of the 5' ETS region (**Figure 2.7b**) shows differences in the peak cross-linking sites for different components, suggesting the pathway of the pre-rRNA through the complex. Utp9 had the most 5' proximal position with strong cross-linking only within the first 40 nucleotides of the 5' ETS. Utp8 and Utp17 also bind within these first 40 nucleotides but showed additional cross-linking around nucleotide 90. Utp4 showed only the peak at 90 nucleotides, while Utp15 was cross-linked at this site and further 3' around nucleotide 250, close to the binding site for the U3 snoRNA 3'-hinge region. The large Utp10 protein showed peak cross-linking around nucleotide 110, with weaker binding at sites from the 5' end to around residue 500, suggesting that it interacts with the pre-rRNA in an extended conformation. Finally, the peak of Utp5 cross-linking was seen around nucleotide 130. Altogether these data revealed that the UtpA complex incorporates the 5' end of the nascent pre-rRNA, with extensive interactions up to around nucleotide 150.

CRAC experiments were also performed with UtpB complex members, as the RNA binding sites of this large protein complex were unknown. Furthermore, UtpB recruitment to the nascent pre-rRNA has been shown to depend on the presence of UtpA *in vivo* (Pérez-Fernández et al. 2007; Pérez-Fernández et al. 2011), which suggests protein-protein or RNA mediated interactions between the two largest sub-complexes of the small subunit processome. As C-terminal tagging was only successful for four out of the six UtpB subunits and one subunit, Utp21, did not significantly cross-link to RNA, reads were obtained only for Utp18, Utp1 and Utp13.

The recovered reads from these three subunits aligned to the ribosomal DNA locus (**Figure 2.7a**). Surprisingly, Utp18 bound in the center of the UtpA binding region, around nucleotide 90 of the 5' ETS, suggesting a role for Utp18 in mediating contacts between the two complexes (**Figure 2.7b**). Utp1 was predominantly cross-linked around the U3 snoRNA binding site at nucleotide 280 of the 5' ETS (**Figure 2.7b**) whereas Utp13 showed peak cross-linking around cleavage site D at the 3' end of the 18S rRNA (**Figure 2.7a**). These binding sites indicate that UtpB brings together functionally important sites that are dispersed in the pre-rRNA sequence.

In addition, several proteins, notably Utp5, Utp15 and Utp18, showed peaks of cross-linking around +1200 nucleotide in the 35S precursor RNA (+500 within the 18S rRNA) (**Figure 2.7a**). In the 18S rRNA secondary structure this site lies close to the 'central pseudoknot', a key structural feature of the small ribosomal subunit, and may also be closely located in different stages of early pre-ribosomes.

The pre-rRNA cross-linking data place the UtpA and UtpB complexes in close proximity to binding sites for U3 snoRNA and we therefore also analyzed reads corresponding to this RNA (**Figure 2.8a**). Utp1 has particularly high reads close to the 3' hinge region of the 5' ETS. If Utp1 binds to a formed U3:5' ETS heteroduplex, sequencing reads for this subunit should be peak on both – the U3 snoRNA 3' hinge and the corresponding 5' ETS sequence. Notably, all UtpA and UtpB components showed U3 snoRNA cross-linking that was substantially higher than the negative

control. However, the cross-linking of Utp10 and Utp1 to U3 snoRNA was more than 10-fold higher than that of other UtpA or UtpB subunits. Since the read numbers in **Figure 2.8** are expressed as hits per million mapped reads, this reflects relatively strong cross-linking of these proteins to U3 snoRNA compared with the pre-rRNA (peak heights in **Figure 2.7** and **Figure 2.8**).

The U3 snoRNP has a pronounced domain structure, with a large, highly structured 3' domain that binds the core snoRNA proteins including Nop56, Nop58, Nop1 (fibrillarin) and Rrp9 (Granneman et al. 2009). The 5' domain is relatively unstructured and contains pre-rRNA base-paring regions, including the 5' and 3' hinge regions and box A and A' (**Figure 2.8b**) (Beltrame & Tollervey 1995; Sharma & Tollervey 1999; Marmier-Gourrier et al. 2011; Dutca et al. 2011). Utp10 predominately cross-linked to the 3' domain of U3 snoRNA, adjacent to major binding sites for the snoRNP proteins (**Figure 2.8b**). In contrast, Utp1 cross-linked only over the 3' hinge region in the 5' domain of U3 snoRNA (**Figure 2.8b**). This is consistent with the specific binding of Utp1 to a formed heteroduplex of the 5' ETS and the 3' hinge of U3 snoRNA (**Figure 2.7b**).

A second UtpB subunit, Utp18, showed a low level of cross-linking to the 3' hinge in U3 snoRNA (**Figure 2.7a**). Other Utp proteins all had low levels of reads within the large terminal stem of the U3 snoRNA 3' domain; either on the 5' side (Utp17) or 3' side (Utp4, Utp5, Utp8, Utp13, Utp15, Utp18). Notably, no significant cross-linking was seen to the other experimentally confirmed pre-rRNA-binding sites in the 5' region of U3 snoRNA; the 5' hinge and box A and A'.

Altogether these data suggest that Utp10 within the UtpA complex is proximal to the U3 snoRNP, while Utp1 within the UtpB complex subsequently interacts with both RNA strands in the base-paired U3 snoRNA - pre-rRNA interaction at nucleotide 280 in the 5' ETS.

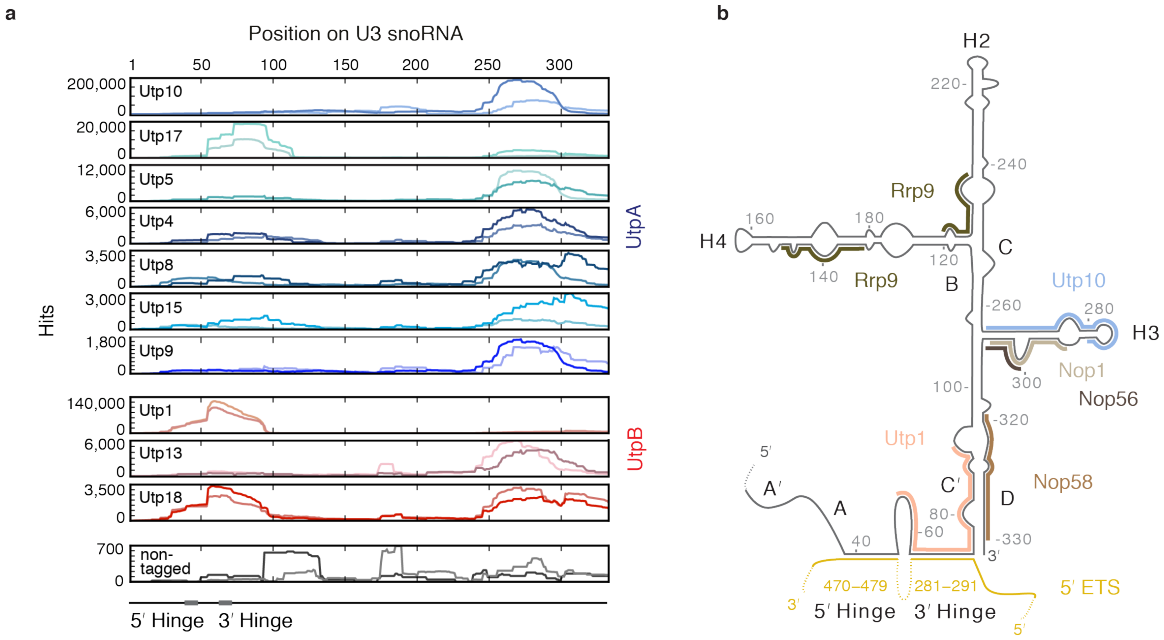


Figure 2.8 | UtpA and UtpB contact the U3 snoRNA.

[a] CRAC library sequences of UtpA (in shades of blue, duplicates in light blue), UtpB (in shades of red, duplicates in light red) and the non-tagged control (in grey, duplicate in light grey) mapped to the spliced SNR17A (encoding the U3 snoRNA) are plotted as frequency of recovery (hits per million mapped reads) at each nucleotide position (indicated above all panels). Individual scales for the frequency of recovery are shown on the left of each subunit panel. Subunit panels are ordered by protein complex and further by their respective number of hits per million mapped reads. The positions of the 3' and 5' hinges that base-pair with the 5' ETS are indicated below the traces. **[b]** Schematic secondary structure of U3 snoRNA (black) with base-paired 5' ETS (yellow) and CRAC-based binding sites for the U3 snoRNP proteins Rrp9, Nop1, Nop56 and Nop58 in different shades of brown, as determined previously (Granneman et al. 2009), binding sites for Utp10 (light blue) from UtpA and Utp1 (salmon) from UtpB. Helices are labelled with H and the conserved Box A'/A/B/C/C'/D sequence elements are marked with their single letter respectively.

2.4 Conclusions

By combining biochemical and structural biology approaches with ensembles of RNA-protein cross-linking data, we were able to provide the first detailed insights into the molecular organization and the key RNA-protein contacts of the essential multi-subunit complex UtpA. These results, together with findings about the RNA-binding sites and architecture of UtpB obtained by Jonas Barandun, allowed us to suggest a model of potential early co-transcriptional events in eukaryotic ribosome biogenesis (**Figure 2.9**).

In isolation, UtpA is a highly dynamic complex composed of an elongated feature, most likely Utp10, and a core (**Figure 2.3**). Utp10 forms a salt-stable subcomplex with Utp17. However, Utp10 shares most protein-protein cross-links with Utp5 and not Utp17. This indicates that the strength of this protein-protein interaction is not exclusively based on proximity. The second subcomplex of UtpA is the tetramer formed by Utp5, Utp8, Utp9 and Utp15 (**Figure 2.4**). The proximity between Utp5 and Utp10 might bridge these two subcomplexes within UtpA.

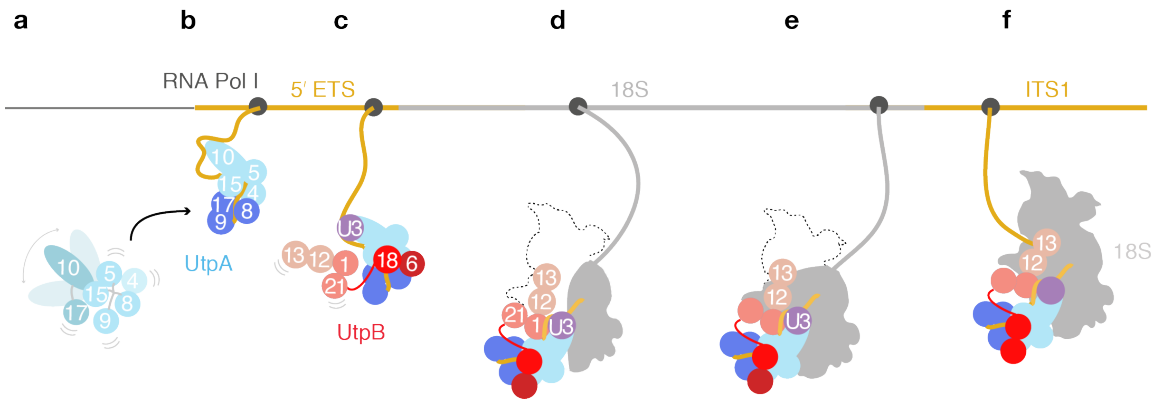


Figure 2.9 | Schematic model of early co-transcriptional events in eukaryotic ribosome biogenesis.

[a] The apo-complex of UtpA (shades of blue) is highly flexible and can adopt different conformations. It consists of two stable subcomplexes, Utp10-Utp17 and Utp5-Utp15-Utp9-Utp8, and Utp4, a less strongly integrated subunit. **[b]** UtpA co-transcriptionally captures the RNA polymerase I (dark grey) synthesized 5' ETS (yellow). Utp17, Utp8 and Utp9 bind to the first ~50 nucleotides of the 5' ETS. **[c]** More subunits of UtpA and Utp18 of UtpB (shades of red) bind to the 5' ETS. The U3 snoRNP (purple) is recruited to UtpA and UtpB. **[d]** The U3 snoRNP (purple) is base-paired at the 3' hinge with the 5' ETS and in close proximity to Utp1. **[e]** Transcription continues and more parts of the 18S rRNA (light grey) are generated. **[f]** The completion of the SSU processome with a largely folded 18S enables Utp13 to bind in the vicinity of the 3' end of the 18S rRNA.

As RNA polymerase I synthesizes the 5' ETS, UtpA binds to the nascent transcript. Even though the RNA-protein interactions identified by CRAC correspond to a population of different ribosome assembly stages *in vivo,* and therefore it is not possible to distinguish sequential RNA-binding events of individual UtpA or UtpB subunits, it seems likely that the 5' to 3' location of cross-linking sites at least partially reflect the order of binding. We therefore suggest that during the early stages of transcription UtpA engages with the 5' end of the 5' ETS through Utp9, Utp8 and Utp17 (**Figure 2.9b**). The other four subunits bind the pre-rRNA further downstream (**Figure 2.9b**). UtpB is potentially recruited via Utp18, which has an overlapping RNA binding site with UtpA subunits (**Figure 2.9c**). Structural analysis of UtpB has revealed its elongated and flexible structure (data obtained by J. Barandun, Hunziker et al. 2016), making it ideally suited to bridge distinct RNA binding sites. After an initial 5' ETS binding through Utp18, the core of UtpB interacts with both RNA strands of the 5' ETS-U3 snoRNA 3' hinge duplex via Utp1 (**Figure 2.9d**). During later stages of small subunit processome assembly, the completion of the 18S rRNA and resulting structural changes enable Utp13 to interact with the 3' end of the 18S rRNA (**Figure 2.9 e,f**).

The analysis of RNA-protein cross-linking highlights that overlapping binding sites exist for different subunits of UtpA and UtpB both within the 5' ETS as well as on the U3 snoRNA. These may either reflect close spatial proximity within pre-ribosomal particles, or dynamic structural changes within these assemblies.

The interactions of UtpA and UtpB with U3 snoRNA seem to be with specific sub-structures of this RNA, since no interactions were observed with the Box A or A' motifs, the 5' hinge or regions previously implicated in Rrp9 binding (Granneman et al. 2009). This observation supports a temporal order of U3 snoRNA binding with distinct sites in the pre-rRNA, in which the most 5' interaction site, at nucleotide 280 in the 5' ETS, is bound by U3 snoRNA before the sites at nucleotide 470 and within the 18S rRNA. The U3 snoRNA-5' ETS interaction at position 280 is required for subsequent pre-rRNA processing, but involves only a relatively short, 11 nucleotide long, region of complementarity (Dutca et al. 2011). We therefore speculated that U3 snoRNP recruitment is stimulated by UtpA via the Utp10-U3 snoRNA interaction, while specific U3 snoRNA-5' ETS base-pairing may be facilitated by UtpB, via Utp1 bridging the 3' hinge interaction site.

The inherent flexibility of UtpA and UtpB in isolation might be reduced upon interaction with their target pre-rRNA structure and protein binding partners. To determine their role in assembling a large structural framework for the maturing small ribosomal subunit further, high-resolution structures of pre-ribosomal particles containing these protein complexes are needed.

Chapter 3 | Structural studies of the small subunit processome – a nucleolar precursor of the small ribosomal subunit

UtpA binds the 5' end of nascent pre-ribosomal RNA thereby initiating the stage-wise formation of the small subunit (SSU) processome (Pérez-Fernández et al. 2007; Pérez-Fernández et al. 2011; Chaker-Margot et al. 2015; Zhang et al. 2016). The SSU processome, a large particle unique and essential to eukaryotes, is composed of over 75 components (Dragon et al. 2002) and coordinates the nucleolar maturation events of the small ribosomal subunit. A functional and mechanistic understanding of the SSU processome and its individual components has long been impeded by a lack of structural data. Recent improvements in cryo-EM made the characterization of large, flexible and low-abundance complexes like the SSU processome more feasible.

A cryo-EM structure of the SSU processome from the thermophilic filamentous fungus *Chaetomium thermophilum* was solved at 7.3 Å resolution (Kornprobst et al. 2016). This reconstruction allowed for the visualization of RNA helices, β-propeller and α-helical repeat structures as well as the approximate assignment of cryo-EM densities to large sub-complexes, such as UtpA, UtpB and U3snoRNP. Several factors previously characterized by crystallography could be placed in the density. This study provided the first architectural description of the

SSU processome but the unambiguous assignment of the majority of individual subunits, *de novo* atomic model building and the interpretation of the functional core of the particle were not possible at this resolution.

The primary model organism to study eukaryotic ribosome biogenesis is *Saccharomyces cerevisiae.* Since a complete structural description of ribosome assembly factors in this particular model organism would help to integrate the wealth of existing biochemical and genetic data and serve as a basis for the design of new studies, we set out to provide a high resolution cryo-EM reconstruction of the SSU processome from *S. cerevisiae*.

Jonas Barandun, Malik Chaker-Margot, Sebastian Klinge and I were able to obtain a cryo-EM reconstruction and build an atomic model of the yeast SSU processome at an overall resolution of 3.8 Å. We further determined the processing state of the pre-rRNA in the purified SSU processome sample by northern blotting and used DSS cross-linking and mass spectrometry analysis to elucidate the protein-protein interaction network within this large particle.

Sebastian Klinge established the purification protocol for the yeast SSU processome. Malik Chaker-Margot carried out the initial characterization of the particle by negative stain and cryo-EM, which lead to a density map at 5.1 Å resolution.

Jonas Barandun, Sebastian Klinge and Malik Chaker-Margot built the model in this medium resolution density map and I carried out the RNA analysis of the sample. The obtained findings were published in (Chaker-Margot et al. 2017).

Subsequently, Jonas Barandun and Malik Chaker-Margot improved the resolution of the cryo-EM map to 3.8 Å. Sebastian Klinge, Malik Chaker-Margot, Jonas Barandun and I built the near-complete atomic model of the SSU processome. To aid the correct assignment and *de novo* model building of all proteins in the SSU processome I performed DSS cross-linking on the purified sample. Cross-linked residues were identified by mass spectrometry analysis carried out by Kelly H. Molloy in the laboratory of Brian T. Chait at the Rockefeller University. The atomic model of the SSU processome was published in (Barandun et al. 2017).

In a joint effort Jonas Barandun, Sebastian Klinge and I summarized the structural and biochemical insights gained from the most recent small subunit ribosome biogenesis studies in a review (Barandun et al. 2018). The published papers (Barandun et al. 2017; Chaker-Margot et al. 2017) and the review (Barandun et al. 2018) form the basis of the following chapter.

3.1 Purification and biochemical characterization of the SSU processome from *S. cerevisiae*

Ribosome synthesis and cell growth are coupled processes (Mayer & Grummt 2006; Woolford & Baserga 2013). Cells regulate their energy consumption to adapt to changing environmental conditions. Since ribosome biogenesis is an energy-intense process its control is fundamental for survival. In response to nutrient-depletion in yeast cultures grown to high optical density, pre-ribosomal particles were shown to accumulate (Talkish et al. 2016). Northern blotting analysis of pre-rRNA compositions in cells at different stages of the growth, indicated that a specific pre-rRNA species, the 23S, accumulated after the post-diauxic shift (Talkish et al. 2016). The 23S pre-rRNA is a precursor of the small subunit rRNA defined by A3 site cleavage in ITS1 (**Figure 1.4**).

When purifying the SSU processome from yeast for structural studies, cells were grown in nutrient-limited conditions to take advantage of the pre-ribosomal particle accumulation observed under these settings. Two protein factors previously shown to be part of the SSU processome (Dragon et al. 2002; Chaker-Margot et al. 2015; Zhang et al. 2016), Utp1 and Kre33, were endogenously tagged with a protease-cleavable GFP and streptavidin-binding-peptide (sbp) respectively (**Figure 3.1**). The two affinity tags were used in tandem, with the GFP-tag as an initial capture step followed by the sbp-based immobilization (**Figure 3.1a**).

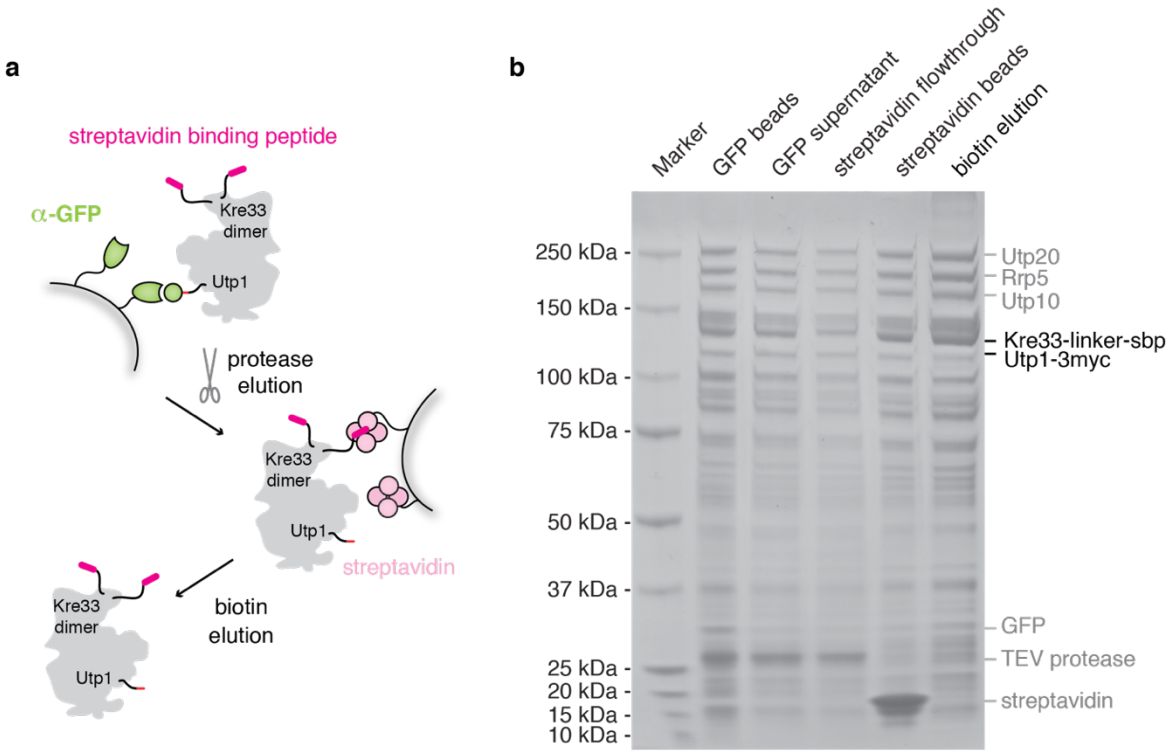


Figure 3.1 | Purification of the small subunit processome from *S. cerevisiae*. **[a]** Schematic overview of the tandem affinity purification procedure. SSU processomes are isolated from a yeast strain harboring a protease cleavable GFP tag (green) on Utp1 and a streptavidin binding peptide tag (pink) on Kre33. Utp1 was chosen as a first bait because it binds the 5' end of the 23S pre-rRNA and Kre33 was selected as a second bait because it binds the 3' end. This ensures the purification of intact particles. Particles containing Utp1 are immobilized on anti-GFP nanobody beads (green) and eluted by protease cleavage. The eluted particles are subsequently further purified by incubation with streptavidin beads (light-pink). SSU processomes are eluted from streptavidin beads with biotin. **[b]** Coomassie blue stained 4-12% SDS-PAGE analysis of individual purification steps with the most prominent protein bands labeled on the right side and the molecular weight markers indicated on the left.

Even though trials to purify the SSU processome from the same yeast strain grown in rich media yielded particles with SSU processome protein compositions, the stoichiometry of the components and amounts were suboptimal (data not shown). Growing cells to saturation in full synthetic drop-out media containing galactose instead of glucose as a carbon source, not only increased the overall yields of the purification but also improved the stoichiometry of the protein components. Since SSU processome can be purified under both, rich and nutrient-starved conditions, this particle is likely not only an early small subunit intermediate but also a storage particle in the cell. However, it is currently unclear if the SSU processome accumulated under stress conditions is a non-productive assembly intermediate targeted for degradation or if these particles are further matured when nutrient-availability improves again (Talkish et al. 2016; Chaker-Margot et al. 2017; Kos-Braun et al. 2017).

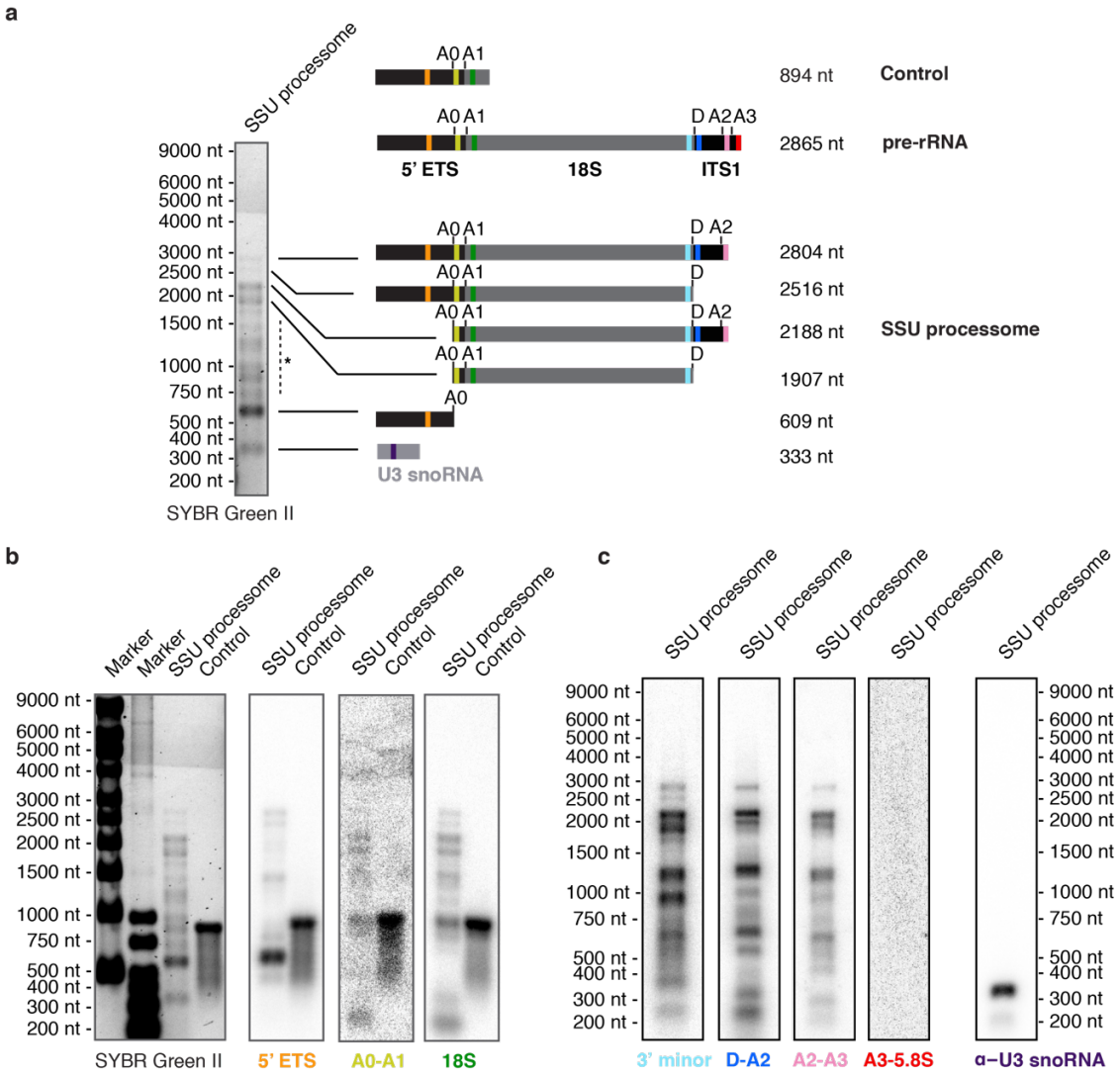
The protein composition and stoichiometry of the tandem-affinity purified SSU processome was assessed by mass spectrometry (data not shown) and SDS-PAGE analysis (**Figure 3.1b**). To characterize the state and composition of the pre-rRNA in the SSU processome, RNA was extracted from the sample and separated on a denaturing agarose gel. SYBR-Green II staining revealed the presence of multiple RNA species in the sample (**Figure 3.2a**). Subsequent comparative northern blotting analysis (**Figure 3.2 b,c**) allowed us to determine the processing state and identity of the most prominent bands observed under SYBR-Green II staining (**Figure 3.2a**).

To define the state of the 5' end of the pre-rRNA present in the SSU processome sample, we blotted with a probe complementary to the middle of the 700-nucleotide long 5' ETS. Strong signal was observed at a band of approximately 600 nucleotides in length and weaker signal was seen for three larger bands (**Figure 3.2b**). The strong signal around 600 nucleotides suggests that the majority of the pre-rRNA in the SSU processome sample has been cleaved either within the 5' ETS (A0) or between the 5' ETS and 18S rRNA (A1). The A0 cleavage site is located at nucleotide 610 within the 5' ETS, which matches the approximate length of the observed band.

To confirm A0 cleavage within the 5' ETS, probes complementary to the sequences between A0-A1 and the first ~50 residues of the 18S rRNA were used (**Figure 3.2b**). Since no radioactive signal was expected in the SSU processome sample around the 5' ETS position with these two probes if it were indeed A0-cleaved, an *in vitro* transcribed RNA spanning the 5' ETS and the first 194 nucleotides of the 18S rRNA served as a positive control for the blot (**Figure 3.2 a,b**). The absence of signal around 600 nucleotides in the SSU processome sample for both the A0-A1 and the 18S probe indicates that the 5' ETS in the isolated SSU processome is primarily A0 but not A1 cut (**Figure 3.2b**).

Figure 3.2 | Composition and processing-state analysis of pre-rRNA in the SSU processome.

[a] SYBR Green II stained denaturing agarose gel of RNA extracted from purified SSU processomes. RNA species identified by northern blotting in [b] and [c] are indicated schematically on the right (SSU processome) and unassigned pre-rRNA degradation products are indicated with an Asterix. An imaging artifact in the SYBR Green II stained gel ([a] and [b]) results in a color change above 4000 nt. A schematic pre-rRNA and the *in vitro* transcribed control RNA used in [b] are shown on the top right with the positions of blotting probes used in [c] and [d] indicated. **[b]** Comparative analysis of SSU processome RNA and control pre-rRNA fragment (5' ETS to nucleotide (nt) 194 of 18S rRNA). RNAs are visualized by SYBR Green II and northern blotting with 5' ETS, 18S and A0-A1 probes. **[c]** Northern blot analysis of SSU processome RNA components using 3'minor, D-A2, A2-A3, A3-5.8S and U3 snoRNA probes.


a
SSU processome
9000 nt
6000 nt
5000 nt
4000 nt
3000 nt
2500 nt
2000 nt
1500 nt
1000 nt
750 nt
500 nt
400 nt
300 nt
200 nt
SYBR Green II
A0 A1
894 nt
Control
D A2 A3
2865 nt
pre-rRNA
5' ETS
18S
ITS1
D A2
2804 nt
D
2516 nt
2188 nt
SSU processome
1907 nt
A0
609 nt
333 nt
U3 snoRNA
b
Marker
Marker
SSU processome Control
SSU processome Control
SSU processome Control
SSU processome Control
SYBR Green II
5' ETS
A0-A1
18S
c
SSU processome
3' minor
D-A2
A2-A3
A3-5.8S
α–U3 snoRNA

To determine the processing state of the 3' end of the SSU processome pre-rRNA, comparative northern blotting analysis with probes binding to the 18S rRNA (3' minor) and ITS1 cleavage sites (D-A2, A2-A3, A3-5.8S) was performed (**Figure 3.2c**). The northern blots showed that the state of the pre-rRNA 3' end in the SSU processome is heterogeneous (**Figure 3.2c**). However, all pre-rRNA in the sample is at least A3 cut (**Figure 3.2c**). Surprisingly, a portion of the pre-rRNA is cleaved at or close to the D-site. *In vivo,* D-site cleavage occurs in pre-40S particles in the cytoplasm and is catalyzed by the endonuclease Nob1 (Fatica et al. 2004). This could either indicate RNA degradation during the purification procedure or inadvertent cleavage of the D-site by the already associated nuclease (Chaker-Margot et al. 2016). Several degradation products of the pre-rRNA were observed in the SSU processome sample (**Figure 3.2a**). Finally, northern blotting also revealed the expected presence of the U3 snoRNA (**Figure 3.2c**).

3.2 Initial analysis of the SSU processome by negative stain and cryo-electron microscopy

To characterize the SSU processome from yeast structurally, the purified sample was analyzed by negative stain EM (**Figure 3.3a**). Most of the visualized particles were homogenous in size and showed a compact structure (**Figure 3.3a**). 2D class averaging of manually selected 36,000 particles revealed a stable core and flexible regions on top (**Figure 3.3b**). In cryo-EM similar characteristics were observed, with the 2D class averages obtained from 79,414 manually picked

particles showing two moving domains in the top region of the particle and a well resolved core (**Figure 3.3 c,d**). Subsequent 3D classification and refinement of this first cryo-EM dataset led to an density map of the SSU processome at 5.1 Å resolution (Chaker-Margot et al. 2017). This reconstruction contained 33,813 particles(Chaker-Margot et al. 2017).

The good quality of most parts of the reconstruction allowed for the placement of available crystal structures or homology models of ribosome biogenesis factors, large parts of the 18*S* rRNA, and several ribosomal proteins (**Table 3-1**). It further permitted the tracing of the 5′ ETS and U3 snoRNA. The large subcomplexes UtpA, UtpB and U3 snoRNP were shown to encapsulate the lower part of the structure containing the 5' ETS, whereas the 18S rRNA domains were kept in an open conformation at the top of the structure(Chaker-Margot et al. 2017). One of the flexible parts seen in the 2D class averages corresponds to the central domain of the 18S rRNA that could not be resolved in the reconstruction.

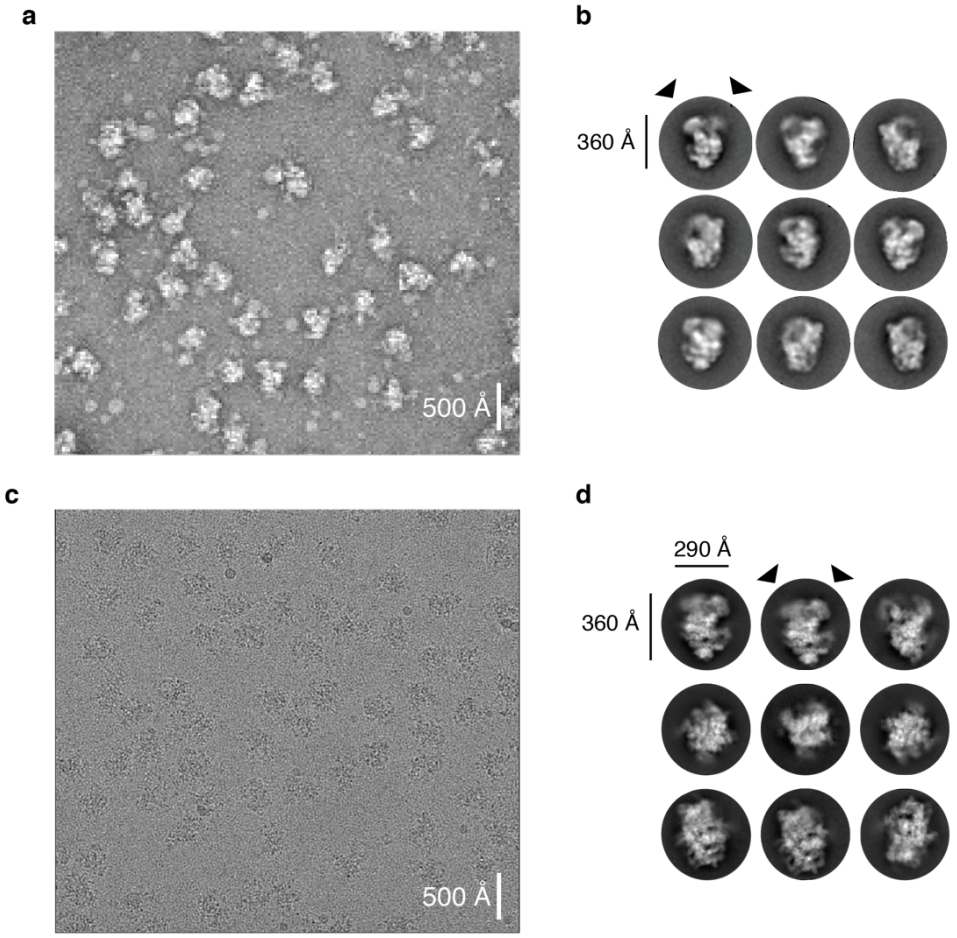


Figure 3.3 | Negative stain and cryo-electron microscopy analysis of SSU processome particles.

[a] Section of a sample micrograph of SSU processome particles stained with 2% uranyl acetate collected at a magnification of 29,000x. A size bar is indicated in the lower right corner. **[b]** Selection of 2D class averages generated from 36,000 manually picked particles using EMAN2 (Tang et al. 2007). A size bar is shown on the left. Flexible regions are marked with black triangles. **[c]** Section of a representative cryo-EM micrograph of SSU processome particles at a magnification of 22,500x. A size bar is drawn in the lower right corner. **[d]** Selection of RELION-2 (Kimanius et al. 2016) 2D class averages generated from 79,414 manually picked particles. Dimensions of the particle are shown with size bars outside the top left class. Black triangles point towards flexible regions.

Even though the density map at 5.1 Å allowed for the generation of an architectural model of the SSU processome which explained aspects of the role of the 5' ETS and its large subcomplexes as an organizing platform for the maturing 18S rRNA, the limited resolution has prohibited the identification of several components (**Table 3-1**). The presence of 20 β-propellers and numerous α-helical repeats in combination with the medium resolution made it difficult to unambiguously assign these repetitive elements of density to specific subunits (**Table 3-1**). Furthermore, the current resolution did not allow for *de novo* model building, which hampered the full interpretation of the particle.

3.3 Improvement of SSU processome cryo-EM density maps and atomic model building

To improve the resolution of the SSU processome density and build a molecular model, additional data were collected and subjected to an extensive 3D classification and refinement strategy (**Figure 3.4**). This resulted in a reconstruction of the SSU processome at 3.8 Å overall resolution (**Figure 3.4, Figure 3.5a**). Through focused refinement with a mask around the core of the particle, the resolution in this region improved further to 3.6 Å (**Figure 3.4, Figure 3.5b**). Additional focused classifications and refinements with masks around the central domain, the UtpA - 3' domain area and the head-region resulted in higher-resolved individual maps of these peripheral regions (**Figure 3.4, Figure 3.5 c,d,e**).

Table 3-1 | Overview of the assignment status and model quality of SSU processome subunits in reconstructions at 5.1 Å and 3.8 Å.

Sub-complex	Molecule name	Chaker-Margot et al. (5.1 Å)	Barandun et al. (3.8 Å)
RNA	5' ETS	idealized helices	atomic
	18S rRNA	crystal structure	atomic
	ITS1	disordered	disordered
	U3 snoRNA	homology model	atomic
UtpA	Utp4	misassigned	atomic
	Utp5	unassigned	atomic
	Utp8	unassigned	atomic/poly-Alanine
	Utp9	unassigned	atomic/poly-Alanine
	Utp10 N-term	unassigned	atomic
	Utp15	unassigned	atomic
	Utp17	misassigned	atomic
UtpB	Utp1	homology model	atomic
	Utp6	poly-Alanine	atomic/poly-Alanine
	Utp12	homology model	atomic
	Utp13	homology model	atomic/poly-Alanine
	Utp18	misassigned	atomic
	Utp21	crystal structure	atomic
U3 snoRNP	Nop56	homology model	atomic
	Nop58	homology model	atomic
	Nop1	homology model	atomic
	Nop1	homology model	atomic
	Snu13	crystal structure	crystal structure
	Snu13	crystal structure	crystal structure
	Rrp9	crystal structure	crystal structure
other ribosome assembly factors	Fcf2	unassigned	atomic
	Utp14	unassigned	atomic
	Sas10	unassigned	atomic
	Lcp5	unassigned	atomic
	Bud21	unassigned	atomic
	Faf1	unassigned	atomic
	Nop14	unassigned	atomic/poly-Alanine
	Noc4	unassigned	atomic/poly-Alanine
	Rrt14	unassigned	atomic/poly-Alanine
	Mpp10	major parts unassigned	atomic
	Rrp5	unassigned	crystal structure

Sub-complex	Molecule name	Chaker-Margot et al. (5.1 Å)	Barandun et al. (3.8 Å)
other ribosome assembly factors	Utp11	unassigned	atomic
	Sof1	unassigned	atomic
	Utp7	unassigned	atomic
	Bms1	homology model	atomic
	Enp2	homology model	atomic
	Utp24	homology model	atomic
	Imp4	homology model	atomic
	Utp30	homology model	atomic
	Imp3	homology model	atomic
	Enp1	unassigned	crystal structure
	Utp22	unassigned	crystal structure
	Rrp7	unassigned	crystal structure
	Krr1	unassigned	crystal structure
	Rcl1	crystal structure	crystal structure
	Emg1	crystal structure	crystal structure
	Emg1	crystal structure	crystal structure
	Pno1	unassigned	homology model
	Kre33	homology model	homology model
	Kre33	homology model	homology model
	Utp20	poly-Alanine	poly-Alanine
ribosomal proteins	rpS23_US12	crystal structure	crystal structure
	rpS13_US15	unassigned	crystal structure
	rpS14_US11	unassigned	crystal structure
	rpS4_ES4	crystal structure	crystal structure
	rpS5_US7	crystal structure	crystal structure
	rpS6_ES6	crystal structure	crystal structure
	rpS7_ES7	crystal structure	crystal structure
	rpS8_ES8	crystal structure	crystal structure
	rpS9_US4	crystal structure	crystal structure
	rpS16_US9	crystal structure	crystal structure
	rpS11_US17	crystal structure	crystal structure
	rpS22_US8	crystal structure	crystal structure
	rpS24_ES24	crystal structure	crystal structure
	rpS28_ES28	crystal structure	crystal structure
	rpS18_US13	unassigned	crystal structure

Due to the small size of the central domain relative to the overall particle and a resulting alignment problem, central domain focused classification and refinement was performed without image alignment using the refined angles of the overall reconstruction. Using the individually improved maps and the architectural model of the initial cryo-EM reconstruction (Chaker-Margot et al. 2017) as starting coordinates, atomic models were built for most SSU processome components (**Table 3-1**).

In the core region of the particle, which encompasses ~ 80 % of all SSU processome proteins, density for side-chains and bases could be observed in the new cryo-EM maps. When attempting to build atomic models in the densities previously assigned to the β-propeller subunits Utp4, Utp17 and Utp18, the protein sequences did not fit into the experimentally determined density map. These subunits have been wrongly assigned in the 5.1 Å structure (**Table 3-1**) (Chaker-Margot et al. 2017) and have either been unassigned or mis-assigned in other published SSU processome structures (Kornprobst et al. 2016; Sun et al. 2017).

.

Figure 3.4 | Cryo-EM data processing strategy.

10,029 micrographs were collected in 4 independent sessions (Datasets 1-4) and aligned using Motioncor2 (Zheng et al. 2017) with dose weighting. Manual inspection and elimination of low quality micrographs reduced this number to 8,406 used for particle picking in RELION 2.0 (Kimanius et al. 2016) (Autopicking and extensive manual cleanup). 3D classification with five classes yielded 2 good classes containing 284,213 particles. Overall 3D refinement yielded a reconstruction at a resolution of 3.8 Å. Focused refinement was performed for the core (3.6 Å) and the 3' domain (4.1 Å). Focused and iterative 3D classification using a head mask (pink dashed line) or a central domain mask yielded improved maps for these regions.

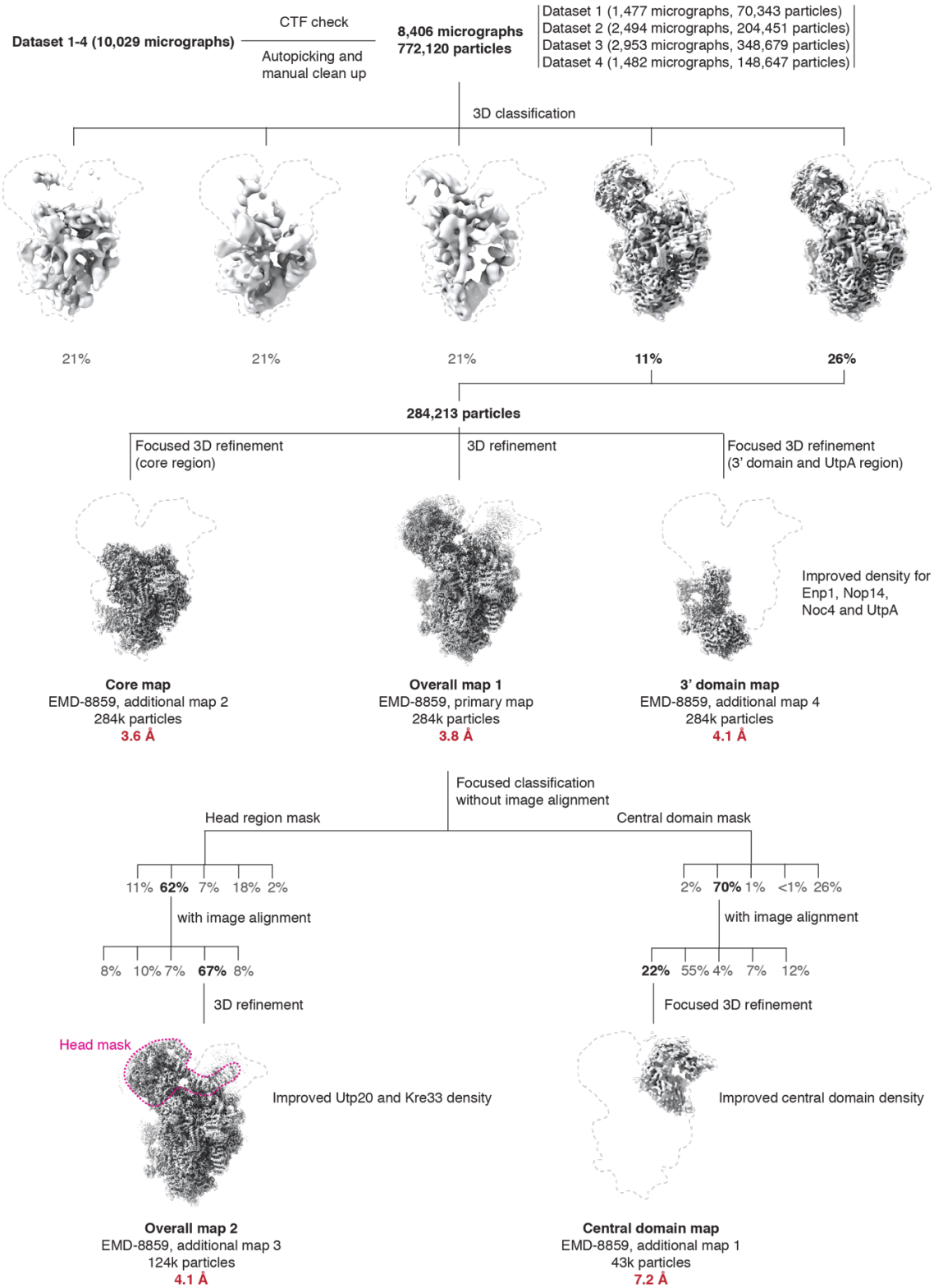
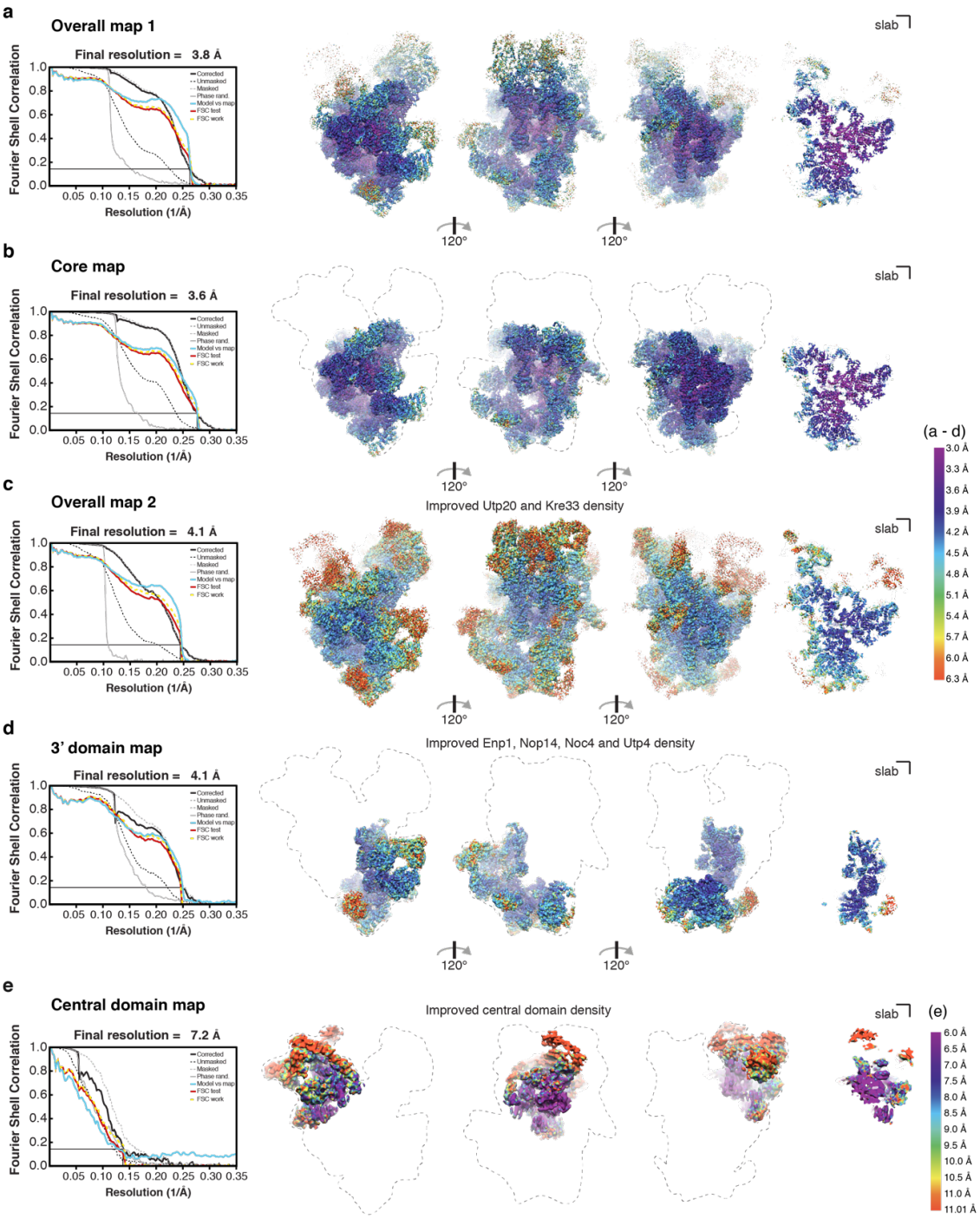

Dataset 1-4 (10,029 micrographs)
CTF check
Autopicking and manual clean up
8,406 micrographs
772,120 particles
Dataset 1 (1,477 micrographs, 70,343 particles)
Dataset 2 (2,494 micrographs, 204,451 particles)
Dataset 3 (2,953 micrographs, 348,679 particles)
Dataset 4 (1,482 micrographs, 148,647 particles)
3D classification
21%
21%
21%
11%
26%
284,213 particles
Focused 3D refinement (core region)
3D refinement
Focused 3D refinement (3' domain and UtpA region)
Improved density for Enp1, Nop14, Noc4 and UtpA
Core map
EMD-8859, additional map 2
284k particles
3.6 Å
Overall map 1
EMD-8859, primary map
284k particles
3.8 Å
3' domain map
EMD-8859, additional map 4
284k particles
4.1 Å
Focused classification without image alignment
Head region mask
Central domain mask
11% 62% 7% 18% 2%
2% 70% 1% <1% 26%
with image alignment
with image alignment
8% 10% 7% 67% 8%
22% 55% 4% 7% 12%
3D refinement
Focused 3D refinement
Head mask
Improved Utp20 and Kre33 density
Improved central domain density
Overall map 2
EMD-8859, additional map 3
124k particles
4.1 Å
Central domain map
EMD-8859, additional map 1
43k particles
7.2 Å

Figure 3.5 | Overall and local resolution estimation of all obtained cryo-EM maps.

Overall and local resolution estimation of **[a]**, the overall map at 3.8 Å (overall map 1), **[b]**, the core focused map at 3.6 Å (core map), **[c]**, the overall map with focus on the head region at 4.1 Å (overall map 2), **[d]**, 3' domain and UtpA focused map at 4.1 Å (3' domain map) and **[e]**, the central domain focused map at 7.2 Å (central domain map). [a-e] The left panel shows Fourier Shell Correlation (FSC) curves for the unmasked (dashed black line), phase randomized (solid grey line), the masked (dashed grey line) and the corrected map (solid black line). An FSC value of 0.143 is indicated by a thin black line. Three views related by a 120-degree rotation of the obtained cryo-EM map are shown colored according to local resolution. The fourth density panel shows a slab view visualizing the resolution in the center. Local resolution was calculated using Resmap (Kucukelbir et al. 2014).


a
Overall map 1
Final resolution = 3.8 Å
b
Core map
Final resolution = 3.6 Å
c
Overall map 2
Final resolution = 4.1 Å
d
3' domain map
Final resolution = 4.1 Å
e
Central domain map
Final resolution = 7.2 Å
Fourier Shell Correlation
Resolution (1/Å)
Corrected
Unmasked
Masked
Phase rand.
Model vs map
FSC test
FSC work
slab
120°
Improved Utp20 and Kre33 density
Improved Enp1, Nop14, Noc4 and Utp4 density
Improved central domain density
(a - d)
(e)

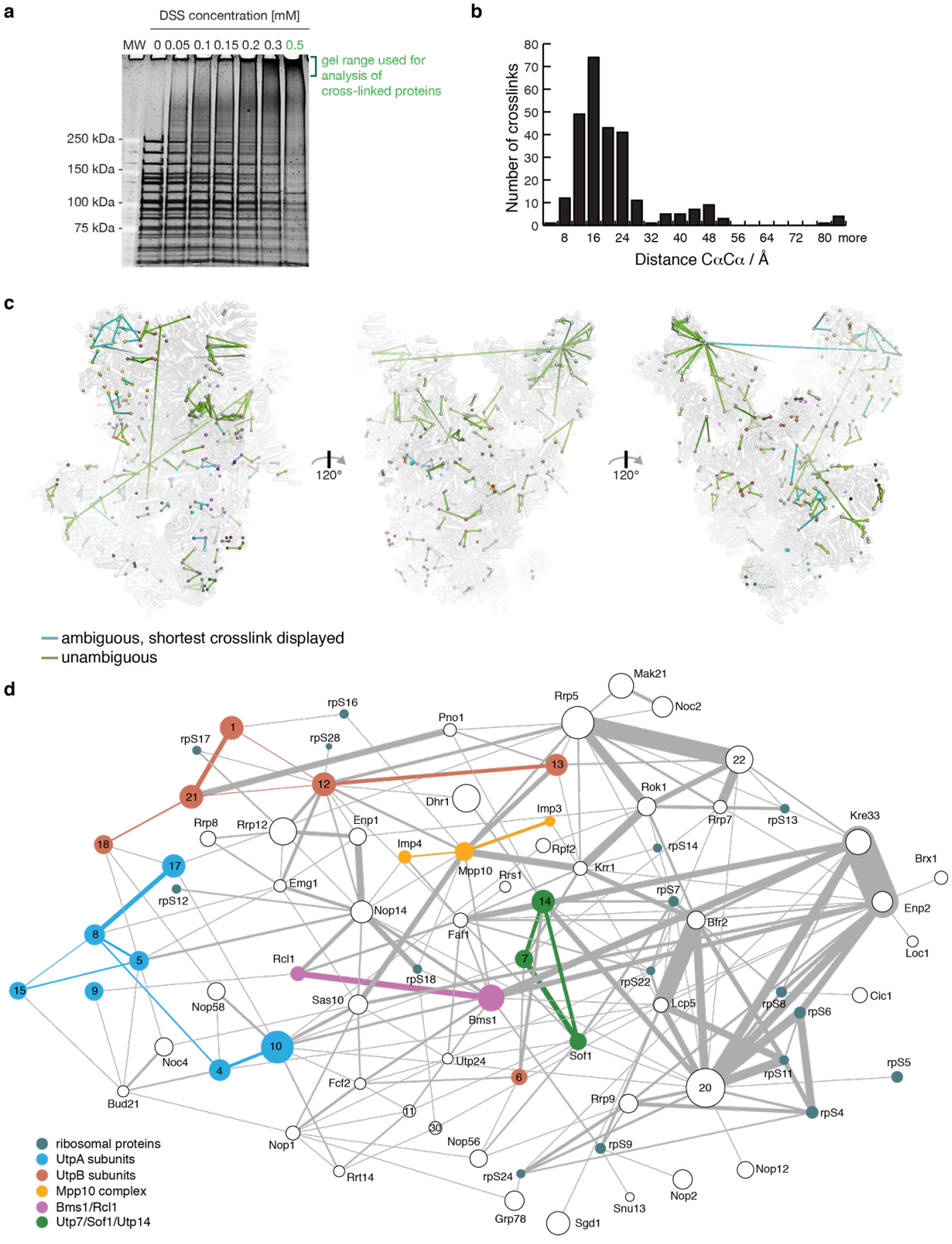
3.4 Protein-protein cross-links provide distance restraints for atomic model building of the SSU processome

While the quality of the cryo-EM density map in the core of the SSU processome allowed for the identification of protein sequences and their register in globular folds, unambiguous density-based assignment of peptide-like proteins weaving through the particle was not always possible. In the peripheral areas of the particle where the resolution of the cryo-EM maps is limited, proteins could not be identified solely based on density fits. To aid the identification and model building of these subunits and to confirm the density-based assignments, DSS cross-linking and mass spectrometry analysis of the SSU processome was performed (**Figure 3.6**).

Akin to the approach used for cross-linking analysis of UtpA in isolation, the purified SSU processome sample was titrated with increasing concentrations of DSS (**Figure 3.6a**). The extent of cross-linking at different concentrations resulted in various mobility shifts of protein bands on SDS-PAGE (**Figure 3.6a**). At a DSS concentration of 0.5 mM most individual bands have shifted upwards to a region slightly below the well. This region was cut out and analyzed by mass spectrometry.

Figure 3.6 | DSS cross-linking and mass spectrometry analysis reveals the protein interaction network within the SSU processome.

[a] SDS-PAGE analysis of a purified SSU processome sample cross-linked with increasing concentrations of DSS. The gel region and DSS concentration used for mass spectrometry analysis experiments are highlighted in green. **[b]** Histogram of all Cα cross-link distances in Å. 87.2 % of all cross-link distances are within 32 Å. **[c]** Cross-links plotted onto the structure of the SSU processome shown as direct connection between the Cα of individual lysine residues. All Cα atoms found in the cross-linking analysis are shown as spheres. In cases where two copies of a protein are present (Kre33, Emg1, and Nop1), the shorter cross-link is displayed. Conformational flexibility of the central domain and a reconstruction of this domain based on only a small subpopulation of the data (15%) may explain the high abundance of cross-links with longer distances in this region. These cross-links may result from other conformational states of the central domain. **[d]** Two-dimensional visualization of all inter-protein DSS cross-links obtained for the SSU processome sample generated with xiNET (Combe et al. 2015). Protein subunits are represented as spheres. The size of each sphere is proportional to the molecular weight of the corresponding protein. Subunits belonging to complexes or those forming a structural unit are highlighted with the same color. The thickness of the line connecting two subunits is proportional to the number of shared cross-links. All Utps (U three proteins) are labeled with their respective number


a
DSS concentration [mM]
MW 0 0.05 0.1 0.15 0.2 0.3 0.5
gel range used for analysis of cross-linked proteins
250 kDa
150 kDa
100 kDa
75 kDa
b
Number of crosslinks
Distance CαCα / Å
c
120°
120°
ambiguous, shortest crosslink displayed
unambiguous
d
ribosomal proteins
UtpA subunits
UtpB subunits
Mpp10 complex
Bms1/Rcl1
Utp7/Sof1/Utp14

The distance restraints provided by the cross-links were used to guide protein identification and build the model of the SSU processome. Therefore, they cannot serve as an independent validation of the built coordinates. Most C_α-C_α distances of cross-linked lysine residues are below the expected 32 Å threshold (Merkley et al. 2014) in the final model (**Figure 3.6b**). This illustrates that the restraints posed by the experimentally determined cross-links and cryo-EM density are represented well in the built atomic model.

Cross-linking distance outliers can be observed in the top regions of the particle (**Figure 3.6c**). In these solvent-exposed areas atomic resolution was not achieved, therefore previously determined crystal structures were fitted or polyalanine models were built *de novo*. The sequence register of proteins in these regions is thus less reliable. The high flexibility of these domains may also lead to various cross-links which are the product of multiple conformations and do not represent the one conformation resolved in the structure. Several distance outliers are originating from a cross-linking hot spot at the top of the particle. It remains unclear why this residue cross-linked so frequently to spatially distant areas.

Taken together, the protein-protein cross-linking and mass spectrometry analysis in combination with high resolution cryo-EM maps have permitted the significant improvement of the previous architectural models of the SSU processome (Kornprobst et al. 2016; Chaker-Margot et al. 2017; Sun et al. 2017). The following

section describing the architecture and structure of the SSU processome and its components will use the complete atomic model of this particle (Barandun et al. 2017) to discuss new insights gained from the high-resolution analysis as well as principles already discovered in previous studies (Kornprobst et al. 2016; Chaker-Margot et al. 2017; Sun et al. 2017). After a general architectural overview, important concepts discovered through the analysis of the SSU processome structure are introduced before the structure is described in detail. The detailed structural description will follow the path of the pre-rRNA starting with the 5' ETS ensued by the four structured domains of the 18S rRNA.

3.5 Overview of the architecture of the small subunit processome

The center of the SSU processome is formed by RNA. The 5' ETS composes the base of the structure whereas ribosomal RNA is located on top (**Figure 3.7a**). U3 snoRNA runs on the side of the particle, but reaches into the center, where all interactions with the 5' ETS and 18S rRNA occur (**Figure 3.7a**). Through base pairing with the 5' ETS at the 3' and 5' hinges, U3 snoRNA rigidifies the structure of the 5' ETS. The RNA duplexes formed by U3 snoRNA and the 18S rRNA, the Box A and A' structures, outline the positions of the 5', central- and 3' major domain of the pre-18S rRNA in the SSU processome (**Figure 3.7a**). While the 5' and central domains contribute to the two top lobes of the particle, the 3' major domain is placed laterally to the 5' ETS (**Figure 3.7a**).

The 18S rRNA is bound by ribosomal proteins (**Figure 3.7b**), which adopt predominantly the same conformation as in the mature ribosome. Out of the fifteen modelled ribosomal proteins only rpS6, rpS18 and rpS23 assume slightly different, yet near mature conformations.

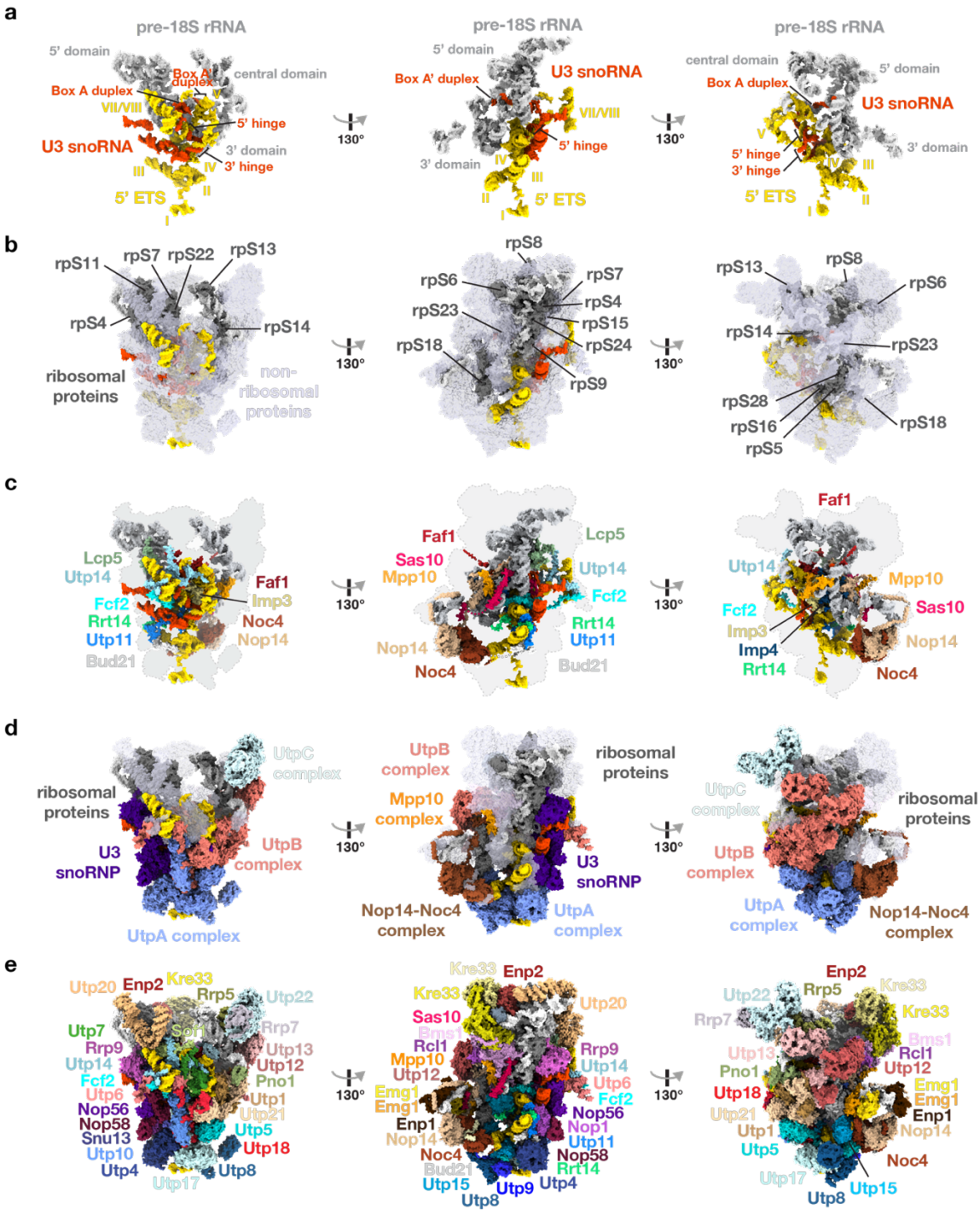
A large shell of more than 51 ribosome assembly factors encapsulates the pre-ribosomal RNA and ribosomal proteins. The innermost layer of this shell is formed by extended peptide-like proteins, which weave through the entire particle (**Figure 3.7c**). Members of this group include the multi-modular proteins Faf1, Lcp5, Mpp10, Sas10, Fcf2, Rrt14, Utp11 and Utp14, which are characterized by their unusual folds and many interaction partners as described later.

Several large multi-subunit complexes (**Figure 3.7d**) as well as individual ribosome assembly factors (**Figure 3.7e**) provide the outer shell of the SSU processome. In agreement with the determined RNA-protein cross-linking profiles, UtpA stabilizes the first half of the 5' ETS and is located at the bottom of the particle. UtpB wraps around the side and back of the SSU processome connecting distant sites through its elongated architecture. UtpC, another pre-assembled subcomplex of the SSU processome is bound to the central domain of the pre-18S rRNA on top of the particle.

The outermost shell of the SSU processome is formed by many additional ribosome assembly factors (**Figure 3.7e**). These include the acetyltransferase and helicase Kre33, which rests on the Bms1-Rcl1 GTPase complex at the top of the structure, and the methyltransferase Emg1 (**Figure 3.7e**), which is positioned on a lateral extension formed by the Nop14-Noc4 complex (**Figure 3.7d**). Lastly, Utp20, Utp10, Rrp5 and the Nop14-Noc4 complex provide large helical repeat structures to support and bridge distant regions of the particle.

Figure 3.7 | Structural organization of the yeast small subunit processome.

[a] RNA molecules of the SSU processome are shown as surfaces with 5' ETS (yellow), U3 snoRNA (red) and pre-18S (white). Structural elements of RNAs and helices of the 5' ETS are indicated. **[b]** Ribosomal proteins are represented in dark-grey, non-ribosomal assembly factors in transparent light-blue, and RNA species as in [a]. [c] Surface representation of centrally located ribosome assembly factors. **[d]** Visualization of the complexes UtpA (blue), UtpB (red), U3 snoRNP (purple), UtpC (light-blue), the Nop14-Noc4 complex (brown) and the Mpp10 complex (orange). **[e]** Surface representation of all individual components of the small subunit processome.


a
pre-18S rRNA
5' domain
Box A duplex
central domain
Box A duplex
VII/VIII
5' hinge
U3 snoRNA
3' domain
3' hinge
III
IV
5' ETS
II
I
130°
pre-18S rRNA
5' domain
Box A' duplex
U3 snoRNA
VII/VIII
5' hinge
3' domain
IV
III
II
5' ETS
I
130°
pre-18S rRNA
central domain
5' domain
Box A duplex
U3 snoRNA
V
5' hinge
3' hinge
IV
III
3' domain
5' ETS
II
I
b
rpS11
rpS7
rpS22
rpS13
rpS4
rpS14
ribosomal proteins
non-ribosomal proteins
130°
rpS8
rpS6
rpS7
rpS23
rpS4
rpS15
rpS18
rpS24
rpS9
130°
rpS13
rpS8
rpS6
rpS14
rpS23
rpS28
rpS16
rpS18
rpS5
c
Lcp5
Utp14
Faf1
Fcf2
Imp3
Rrt14
Noc4
Utp11
Nop14
Bud21
130°
Faf1
Lcp5
Sas10
Mpp10
Utp14
Fcf2
Rrt14
Utp11
Nop14
Noc4
Bud21
130°
Faf1
Utp14
Mpp10
Fcf2
Sas10
Imp3
Nop14
Imp4
Rrt14
Noc4
d
UtpC complex
ribosomal proteins
U3 snoRNP
UtpB complex
UtpA complex
130°
UtpB complex
ribosomal proteins
Mpp10 complex
U3 snoRNP
Nop14-Noc4 complex
UtpA complex
130°
UtpC complex
ribosomal proteins
UtpB complex
UtpA complex
Nop14-Noc4 complex
e
Utp20
Enp2
Kre33
Rrp5
Utp22
Utp7
Sof1
Rrp7
Rrp9
Utp13
Utp14
Utp12
Fcf2
Pno1
Utp6
Utp1
Nop56
Utp21
Nop58
Utp5
Snu13
Utp18
Utp10
Utp4
Utp17
Utp8
130°
Kre33
Enp2
Kre33
Utp20
Sas10
Bms1
Rcl1
Rrp9
Mpp10
Utp14
Utp12
Utp6
Emg1
Fcf2
Emg1
Nop56
Enp1
Nop1
Nop14
Utp11
Noc4
Nop58
Bud21
Rrt14
Utp15
Utp9
Utp4
Utp8
130°
Enp2
Utp22
Rrp5
Kre33
Rrp7
Kre33
Bms1
Utp13
Rcl1
Pno1
Utp12
Utp18
Emg1
Utp21
Emg1
Enp1
Utp1
Nop14
Utp5
Noc4
Utp17
Utp15
Utp8

3.6 General concepts in structural features of small subunit processome assembly factors

The atomic model of the SSU processome provided the first structural information on many ribosome assembly factors involved in the maturation of the small subunit. The structural analysis of these protein and RNA components revealed repeated architectural concepts in ribosome assembly factors which fulfill distinct functions.

A common characteristic of ribosome assembly factors is a high degree of flexibility in isolation. Many proteins in the SSU processome either lack canonical protein folds completely or contain long flexible linkers and extensions. This allows for an initially dynamic co-transcriptional assembly process. Through the formation of multiple, chronologically ordered binding interactions the conformational freedom of individual factors is reduced as the assembly process progresses. In the rigidified SSU processome, the initially flexible ribosome biogenesis factors adopt an ordered state.

Similar to the highly flexible protein components of the SSU processome, sequential base-pairing of the U3 snoRNA with the pre-18S rRNA drives the reduction in flexibility and defines the chronology of SSU processome formation. U3 snoRNA base-pairs with sequences in the 5' ETS and the 18S rRNA and thereby acts as a vital RNA chaperone during the assembly steps. Furthermore, the U3

snoRNA base-pairing with pre-18S rRNA sequences prevents the premature formation of the central pseudoknot and thus enables the spatial separation of the individual pre-18S rRNA domains into sub-compartments of the SSU processome. In addition to U3 snoRNA, several ribosome biogenesis proteins remodel the pre-rRNA.

The temporal binding order of ribosome biogenesis factors and ribosomal proteins is often regulated by assembly factors that employ molecular mimicry. Earlier ribosome assembly factors occupy the binding sites of late-binding factors and ribosomal proteins on the pre-rRNA or on other assembly factors. Thereby the premature binding of ribosomal proteins and factors is sterically hindered.

3.7 UtpA coordinates the first three helices of the 5' ETS

The revised SSU processome model, with the correct assignment of all β-propellers and α-helical repeats, provides the first high-resolution structure of UtpA (**Figure 3.8 a,b**). The general architecture of the complex is characterized by a large number of WD40-domains, long inter-domain linkers as well as a tetramerization module formed by the α-helical C-terminal domains (CTDs) of Utp5, Utp8, Utp9 and Utp15 (**Figure 3.8a**). The four subunits contributing to the tetrameric α-helical bundle have been identified as a stable sub-complex of UtpA in biochemical assays (**Figure 2.4**). The second biochemically characterized subcomplex of UtpA is the Utp10-Utp17 dimer (**Figure 2.4**). In contrast to the tetramer, the Utp10-Utp17 dimer contact is conveyed by only a small interaction interface formed by a C-terminal peptide-like extension of Utp17 and the α-helical repeat of Utp10 (**Figure 3.8a**). This interface is located distant from the tandem β-propeller of Utp17 (**Figure 3.8 a,b**).

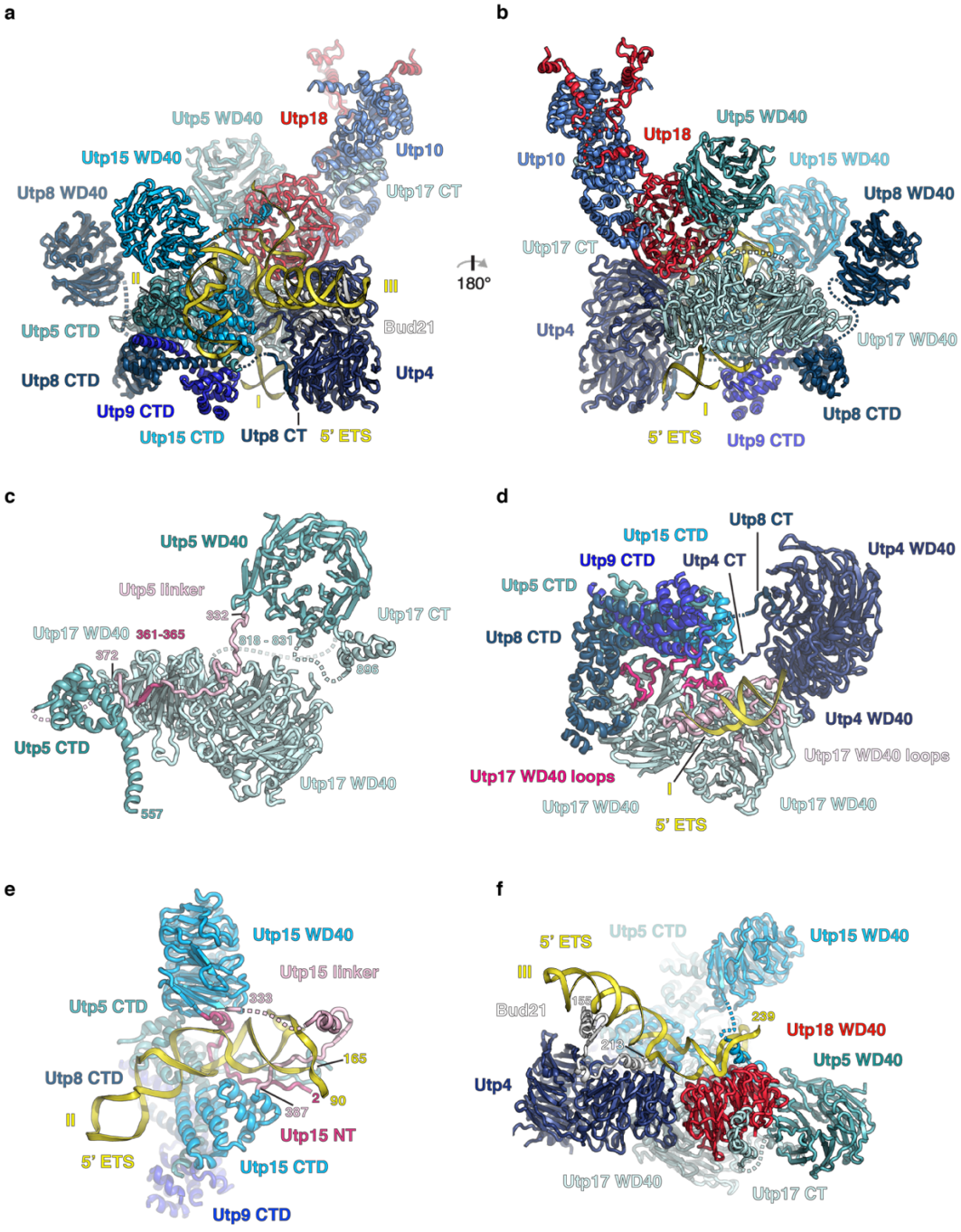
Long linker-peptides mediating subunit-subunit interactions without necessitating spatial proximity of globular domains have contributed to the mis-assignment of several UtpA subunits (**Table 3-1**) and explain the conformational heterogeneity of the complex in isolation (**Figure 2.3**). Utp5, a previously unassigned subunit of UtpA, is another example for the long distances covered by the inter-domain linkers in UtpA. Utp5 is integrated within UtpA mostly through its CTD but a long linker peptide, which runs along a conserved groove of Utp17 (**Figure 3.8 a,c**), places its β-propeller moiety in a spatially distant region of the particle.

The reduced flexibility of UtpA in context of the SSU processome permitted the building of near-complete models for most subunits of the complex. The exceptions being the solvent exposed β-propeller of Utp9 and C-terminal part of Utp10. In the previously published SSU processome structures (Kornprobst et al. 2016; Chaker-Margot et al. 2017; Sun et al. 2017) density for the C-terminal part of Utp10 was observed in low-pass filtered or low-resolution maps. While its N-terminal part is bound at the back of the particle (**Figure 3.7e, Figure 3.8 a,b**) the C-terminal repeats span from the SSU processome base, close to Utp4, up to the U3 snoRNP component Rrp9. Hence, Utp10 wraps around the particle connecting UtpA with UtpB (N-terminal parts) and U3 snoRNP (C-terminal part). RNA binding analysis of Utp10 supports this positioning as it showed cross-linking to helix 3 of U3 snoRNA (**Figure 2.8**). Helix 3 of U3 snoRNA is solvent exposed and not resolved in this cryo-EM reconstruction.

RNA binding site analysis of UtpA has further shown that three of its subunits, Utp8, Utp9 and Utp17, bind the first ~90 nucleotides of the 5' ETS whereas the other four subunits have binding sites up to nucleotide ~250 (**Figure 2.7**). The structure of the SSU processome reveals how UtpA coordinates this sequence range of the 5' ETS, which forms the first three helices of the pre-rRNA.

Figure 3.8 | Architecture of the UtpA complex and its interactions with the 5′-ETS and Utp18.

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]** A linker (light pink) between the WD40 domain and the CTD of Utp5 runs along Utp17, forming a β-strand (dark pink). [d] 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. **[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.


a
Utp5 WD40
Utp18
Utp10
Utp15 WD40
Utp17 CT
Utp8 WD40
II
III
Utp5 CTD
Bud21
Utp4
Utp8 CTD
Utp9 CTD
I
Utp15 CTD
Utp8 CT
5' ETS
180°
b
Utp18
Utp5 WD40
Utp10
Utp15 WD40
Utp8 WD40
Utp17 CT
Utp4
Utp17 WD40
Utp8 CTD
I
5' ETS
Utp9 CTD
c
Utp5 WD40
Utp5 linker
332
Utp17 CT
Utp17 WD40
361-365
372
818 - 831
886
Utp5 CTD
Utp17 WD40
557
d
Utp8 CT
Utp15 CTD
Utp4 WD40
Utp9 CTD
Utp4 CT
Utp5 CTD
Utp8 CTD
Utp4 WD40
Utp17 WD40 loops
Utp17 WD40 loops
I
Utp17 WD40
Utp17 WD40
5' ETS
e
Utp15 WD40
Utp15 linker
Utp5 CTD
333
165
Utp8 CTD
2
90
387
II
Utp15 NT
5' ETS
Utp15 CTD
Utp9 CTD
f
5' ETS
Utp5 CTD
Utp15 WD40
III
Bud21
155
239
Utp18 WD40
213
Utp5 WD40
Utp4
Utp17 WD40
Utp17 CT

Helix I is recognized by a set of loops and helical elements on top of the tandem β-propeller of Utp17 (**Figure 3.8d**). Parts of helix I are not resolved in the density and additional binding of Utp9 and Utp8 could occur in the solvent exposed section of it. A single stranded region of the 5' ETS connecting helix I and II is enclosed by Utp4 and the CTD-tetramer.

While 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 CTD and WD40 domains to position helix II and stabilize the junction between helix II and III (**Figure 3.8e**). 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 of the 18S rRNA is placed (**Figure 3.7 d,e**). This extension is additionally stabilized by Bud21, which connects Noc4, UtpA and U3 snoRNP by binding to Utp4, Nop1 and helix III, which rests on top of Utp4 (**Figure 3.7e, Figure 3.8f**).

A short single stranded RNA region between helix II and III of the 5' ETS, is coordinated by two β-propellers located next to Utp4 and Utp17. A UtpB subunit, Utp18, and the WD40-domain of Utp5 stabilize this RNA region leading into helix IV of the 5' ETS (**Figure 3.8f**).

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) is in good agreement with the determined RNA-protein cross-linking data sets, as binding peaks of UtpA and Utp18 overlap (**Figure 2.7**).

3.8 UtpA and UtpB share an intricate binding interface and architectural similarities

The two largest subcomplexes of the SSU processome, UtpA and UtpB, share an intertwined binding interface formed by Utp18, Utp5 and Utp10 (**Figure 3.9**). Utp18 acts as a central nexus in this junction (**Figure 3.9 a,b,c**). Like Utp17, Utp18 employs extensive peptide-like motifs to mediate protein-protein interactions (**Figure 3.9 b,c**). Three regions within the 230-residue N-terminus of Utp18 mediate its interactions with the UtpB subunits Utp6, Utp21, the U3 snoRNP component Nop58 and Utp10 (UtpA) (**Figure 3.9b**). The first segment (residues 13-28) is employed to interact with both Utp6 and Utp10, while the second (residues 123-183) forms an intricate interface with the surface of the first β-propeller of Utp21 and a conserved C-terminal peptide sequence of Nop58 (**Figure 3.9 b,c, Figure 3.10c**). Additionally, it features an exosome-associated helicase (Mtr4) recruitment peptide (AIM motif) which is located in a disordered region between the first and second N-terminal segment of Utp18 (Thoms et al. 2015) (**Figure 3.10c**).

UtpA and UtpB share not only an interaction interface, but also architectural similarities. Like UtpA, the hexameric UtpB complex is characterized by a high number of β-propellers, the presence of one subunit composed solely of α-helical repeats (Utp6), long peptide-like linkers and a tetramerization module formed by the CTDs of four of its subunits (Utp12, Utp13, Utp21, Utp1) (**Figure 3.9 c,d**). The likeness of the tetramerization element in UtpA and UtpB is evident on the sequence and structural level. As noted previously (Chaker-Margot et al. 2017) its occurrence in UtpA and UtpB subunits suggests a common evolutionary origin of their CTDs. The proteins involved in forming the tetramerization module differ in UtpA and UtpB by the use of single- (UtpA) or tandem (UtpB) β-propeller (Chaker-Margot et al. 2017).

3.9 UtpB stabilizes the 3' hinge of U3 snoRNA and bridges distant sites in the SSU processome

Within the SSU processome UtpB is coordinating sequences of the pre-rRNA located in distant regions of the particle (**Figure 3.9a**). It connects the first third of the 5' ETS (helices III, IV) with the last third (helices VII,VIII) through Utp18 and Utp6 and further links the 5' ETS with the end of the 18S rRNA through Utp12 and Utp13 (**Figure 3.10**).

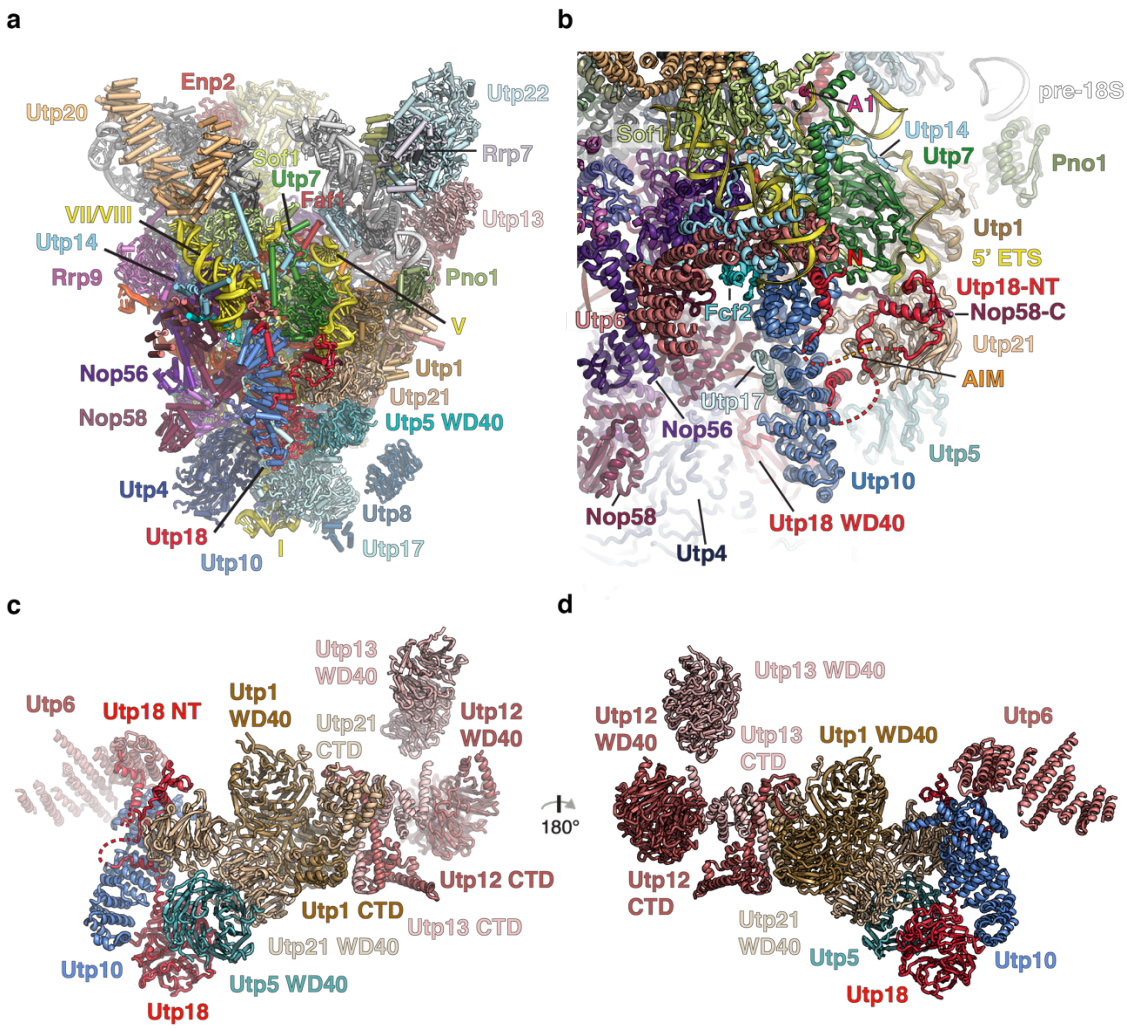


Figure 3.9 | Utp18 acts as a central nexus in the UtpA – UtpB junction.

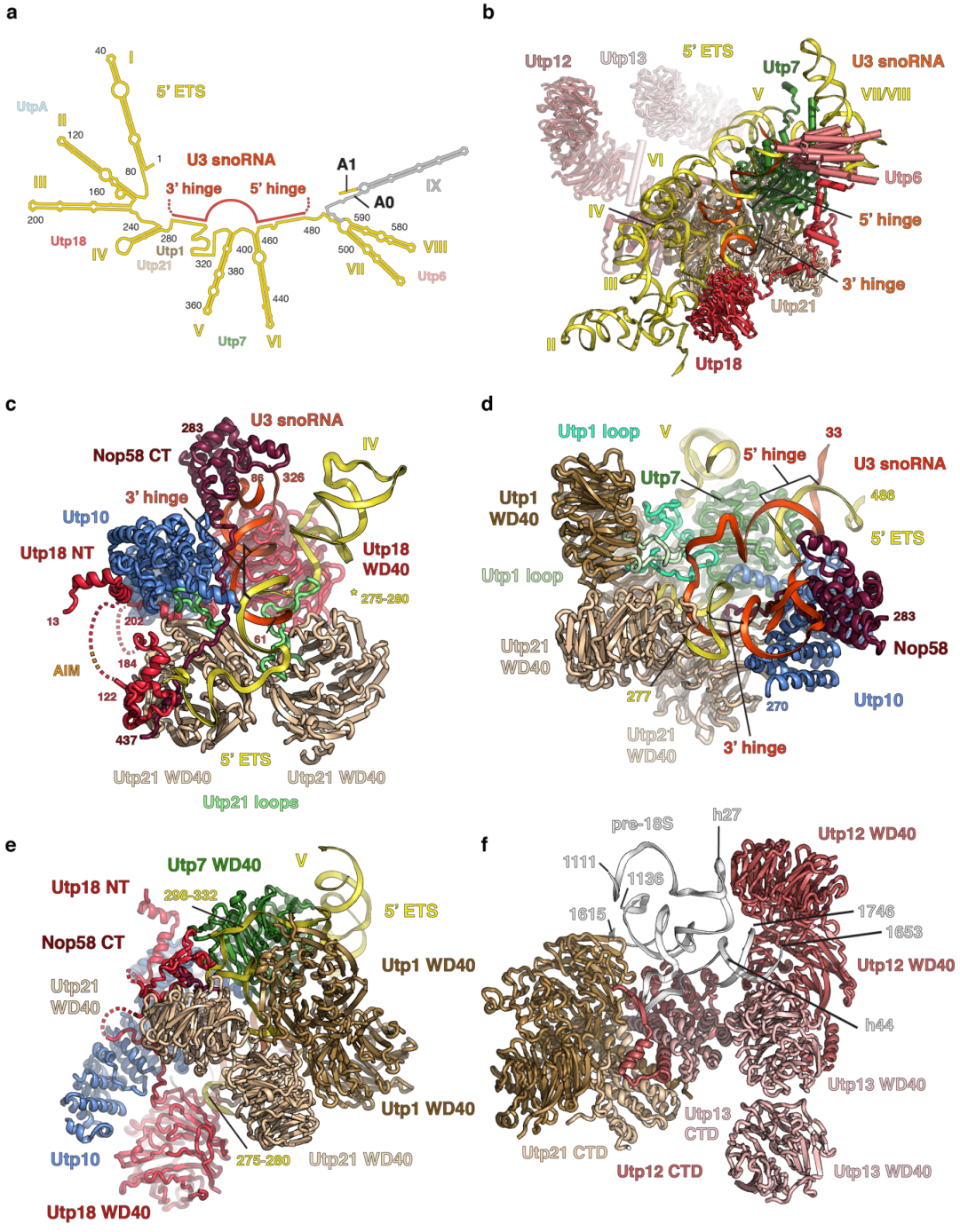
[a] Cartoon representation of the complete atomic model of the SSU processome with individual subunits and their labels color-coded. Selected helices of the 5' ETS are labeled with roman numbers. **[b]** Zoomed in, and slightly turned, view of [a] showing the UtpA-UtpB interface with focus on the N-terminal extension of Utp18 (UtpB). Interactions of the N-terminal peptide of Utp18 with Utp10 (UtpA), Nop58 (U3 snoRNP), Utp21 and Utp6 (UtpB). Sof1, Utp7 and Utp14 organize the A1 cleavage site (pink sphere) and interact with multiple UtpB subunits. **[c,d]** Overviews of the UtpB complex and its interactions with Utp5 and Utp10.

Its most 5' RNA binding site is recognized by the WD40-domain of Utp18. Single stranded RNA of the 5' ETS leading in and out of helix IV is running on top of this β-propeller (**Figure 3.10 a,b,c**). Downstream of helix IV, from nucleotide 280 to 293, the 5' ETS forms an RNA duplex with the U3 snoRNA – the 3' hinge. Loops extending from Utp21 and the helical repeat of Utp10 form a clamp around this duplex (**Figure 3.10c**). The 5' ETS and U3 snoRNA are single stranded again downstream of the 3' hinge. Utp1 is used to bind these single-stranded regions of both U3 snoRNA and 5' ETS with two long, structured loops (residues 556–580 and 616–680) (**Figure 3.10d**). These loops act as a rudder, separating the 5' ETS and the U3 snoRNA after the 3' hinge.

While the 5' ETS (nucleotides 293 – 332) is guided towards the periphery of the SSU processome by Utp21 and Utp1 on one side, and Utp18 and Utp7/Sof1/Utp14 on the other side (**Figure 3.10e**), U3 snoRNA is directed towards the 5' hinge on the inside of the particle (**Figure 3.10d**). Before base-pairing with U3 snoRNA in the 5' hinge (**Figure 3.10d**), the 5' ETS forms helix V, which rests on top of Utp7 and Utp1 (**Figure 3.10 b,e**), and helix VI, which points downwards and contacts helix IV (**Figure 3.10b**). Downstream of the 5' hinge, the 5' ETS helices VII and VIII are bound by Utp6 (**Figure 3.10b**).

Figure 3.10 | RNA interactions of the UtpB complex, Utp7 and Utp10.

[a] Schematic secondary structure diagram of the 5' ETS (yellow) and its interactions with the U3 snoRNA (red) 3'- and 5' hinges. 5' ETS helices are labeled with roman numerals. As helix IX is not resolved in the SSU processome structure it is colored grey. The A0 and A1 cleavage sites are indicated. Approximate nucleotide positions of the 5' ETS are labeled in black. RNA structures coordinated by the UtpA complex (blue), UtpB subunits (shades of red and brown) and Utp7 are marked. **[b]** Cartoon representation of UtpB (shades of red and brown) and Utp7 (green) bound to the 5' ETS (yellow) and the U3 snoRNA hinges (red). Utp12 and Utp13 stabilize the 3' end of the pre-18S rRNA (not shown) in a distant region. **[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]** Two loops of Utp1 (cyan and pale green) separate the 5′ ETS and the U3 snoRNA between the 3′ and 5′ hinges. **[e]** 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. **[f]** Helix 27 and helix 44 of the 3' domain of the pre-18S rRNA (white) are coordinated by the WD40 domains of Utp12. The CTD-tetramer of UtpB binds a looped-out region of the pre-18S rRNA 3' domain.


a
5' ETS
U3 snoRNA
3' hinge
5' hinge
A1
A0
UtpA
Utp18
Utp1
Utp21
Utp7
Utp6
b
Utp12
Utp13
5' ETS
Utp7
U3 snoRNA
VII/VIII
Utp6
5' hinge
3' hinge
Utp21
Utp18
c
U3 snoRNA
Nop58 CT
3' hinge
Utp10
Utp18 NT
Utp18 WD40
275-280
AIM
5' ETS
Utp21 WD40
Utp21 WD40
Utp21 loops
d
Utp1 loop
5' hinge
U3 snoRNA
Utp7
Utp1 WD40
5' ETS
Utp1 loop
Nop58
Utp21 WD40
Utp10
3' hinge
e
Utp7 WD40
Utp18 NT
5' ETS
Nop58 CT
Utp1 WD40
Utp21 WD40
Utp1 WD40
Utp10
Utp21 WD40
Utp18 WD40
f
pre-18S
h27
Utp12 WD40
Utp12 WD40
h44
Utp13 WD40
Utp13 CTD
Utp21 CTD
Utp12 CTD
Utp13 WD40

In addition to the coordination of different 5' ETS structures UtpB recognizes the 3' end of the 18S pre-rRNA (3' major domain - helix 27 and 3' minor domain - helix 44) through its CTD tetramer and the tandem β-propeller of Utp12 (**Figure 3.10f**). Helix 27 is stabilized by the side surface of one WD40-domain of Utp12, whereas the two top surfaces of the tandem β-propeller are used to form a clamp around helix 44 (**Figure 3.10f**). Only the base of the ~100 nucleotide-long helix 44 could be modeled, but density representing the non-modeled parts is clearly visible in the central domain maps (**not shown**). The continuation of helix 44 is positioned between Utp12 and Utp13 and reaches Utp22, a subunit of the central domain bound UtpC complex, with its tip.

3.10 The A1 cleavage site is organized by Utp7, Sof1 and Utp14

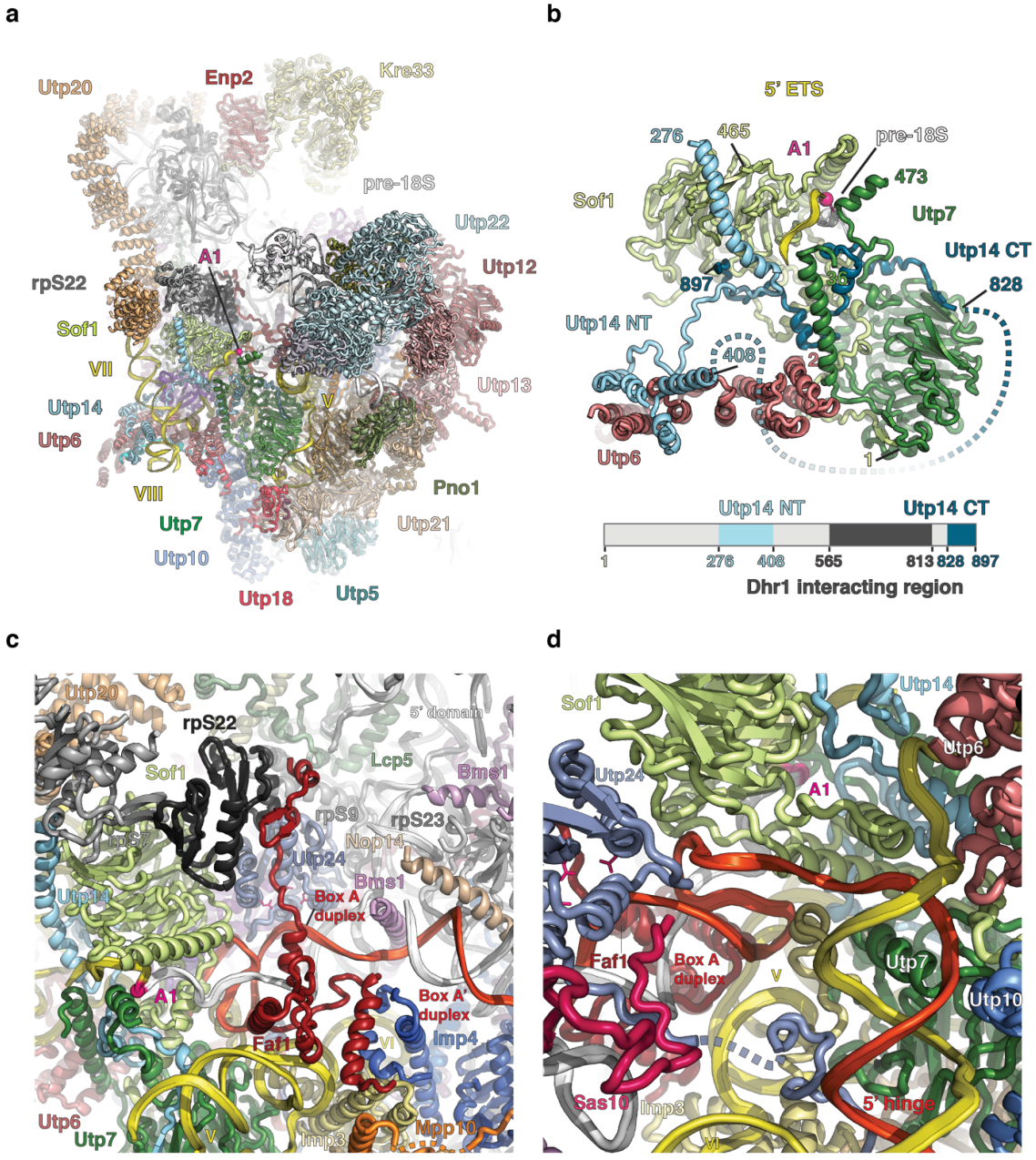
The 5' ETS helices VII and VIII, located above the UtpB-subunit Utp6, are the most 3' helices of the 5' ETS resolved in the SSU processome structure (**Figure 3.10a**). Despite density for helix IX not being discernible, northern blotting of the SSU processome sample has shown that the pre-rRNA is A0 but not A1 cut (**Figure 3.2**), indicating that helix IX is still attached through the uncut A1 site. The A1 site, and 6 nucleotides of the 5' ETS preceding it, are coordinated by N- and C-terminal extensions of the WD40-domains of Utp7 and Sof1 as well as Utp14 (**Figure 3.11**).

Similar to the β-propellers of UtpA and UtpB, Sof1 and Utp7 have functionalized peptide extensions of their WD40 domains. These extensions, together with Utp14, protect the A1 site on the solvent exposed side of the particle (**Figure 3.11b**). Here, Utp14 forms an unusual split structure. An N-terminal segment of Utp14 is used to connect Sof1 with Utp6 while a separate C-terminal segment of Utp14 links Utp7 with Sof1 (**Figure 3.11b**). Several hundred disordered residues of Utp14 bridge these two fragments. Close to the modeled C-terminal segment, this bridging sequence contains the binding site for Dhr1, the essential DEAH-box helicase that is responsible for displacing U3 snoRNA from early ribosome assembly intermediates (Sardana et al. 2015; Zhu et al. 2016).

The nuclease responsible for cleaving the Sof1-Utp7-Utp14 bound A1 site, Utp24, is already integrated in the SSU processome (Bleichert et al. 2006; Tomecki et al. 2015; Wells et al. 2016). The PIN-domain nuclease is positioned near its substrate inside the particle, but the active site of the nuclease is occluded by the Box A U3:18S RNA duplex held in place by Faf1 (**Figure 3.11c**). U3 snoRNA engages in the Box A interaction with the 18S rRNA shortly upstream of the 5' hinge formed with the 5' ETS (**Figure 3.11d**).

Figure 3.11 | Sof1, Utp7 and Utp14 shelter the A1 cleavage site.

All panels are cartoon representations with proteins and RNAs color coded. Helices of the 5′ ETS are labeled with roman numerals. The cleavage site A1 is shown as a pink sphere. **[a]** Top-back view of the SSU processome showing the location of the A1 cleavage site in the context of the particle. **[b]** 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. **[c,d]** The Box A duplex of U3 snoRNA (red) and pre-18S rRNA (white) is held in place by a Faf1 and occludes the active site (highlighted as pink sticks) of the A1-nuclease Utp24.


a
Utp20
Enp2
Kre33
pre-18S
Utp22
Utp12
A1
rpS22
Sof1
VII
Utp13
Utp14
V
Utp6
VIII
Pno1
Utp7
Utp21
Utp10
Utp18
Utp5
b
5' ETS
276
465
A1
pre-18S
473
Sof1
Utp7
Utp14 CT
897
828
Utp14 NT
408
2
Utp6
1
Utp14 NT
Utp14 CT
1
276
408
565
813
828
897
Dhr1 interacting region
c
Utp20
rpS22
5' domain
Lcp5
Sof1
Bms1
rpS9
rpS23
rpS7
Nop14
Utp24
Utp14
Bms1
Box A duplex
A1
Box A' duplex
Imp4
Faf1
VI
Utp6
Utp7
V
Imp3
Mpp10
d
Sof1
Utp14
Utp6
Utp24
A1
Faf1
Box A duplex
V
Utp7
Utp10
Sas10
Imp3
5' hinge
VI

3.11 U3 snoRNP and bound adaptor proteins reach into the center of the SSU processome

U3 snoRNA occupies a functionally central position within the SSU processome. 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. Additionally, the 5' end of U3 snoRNA base pairs in two regions with the pre-18S rRNA. Its Box A (U3 nucleotides 16-22) motif base pairs with the pre-18S rRNA (nucleotides 9-15) near the A1 cleavage site while the Box A' motif (U3 nucleotides 3-13) is base paired with nucleotides 1111-1122 of the pre-18S rRNA (**Figure 3.12a**). A range of ribosome assembly factors is responsible for the stabilization of the four RNA duplexes that U3 snoRNA forms with the 5' ETS (**Figure 3.10**) and the 18S precursor (**Figure 3.11 c,d**).

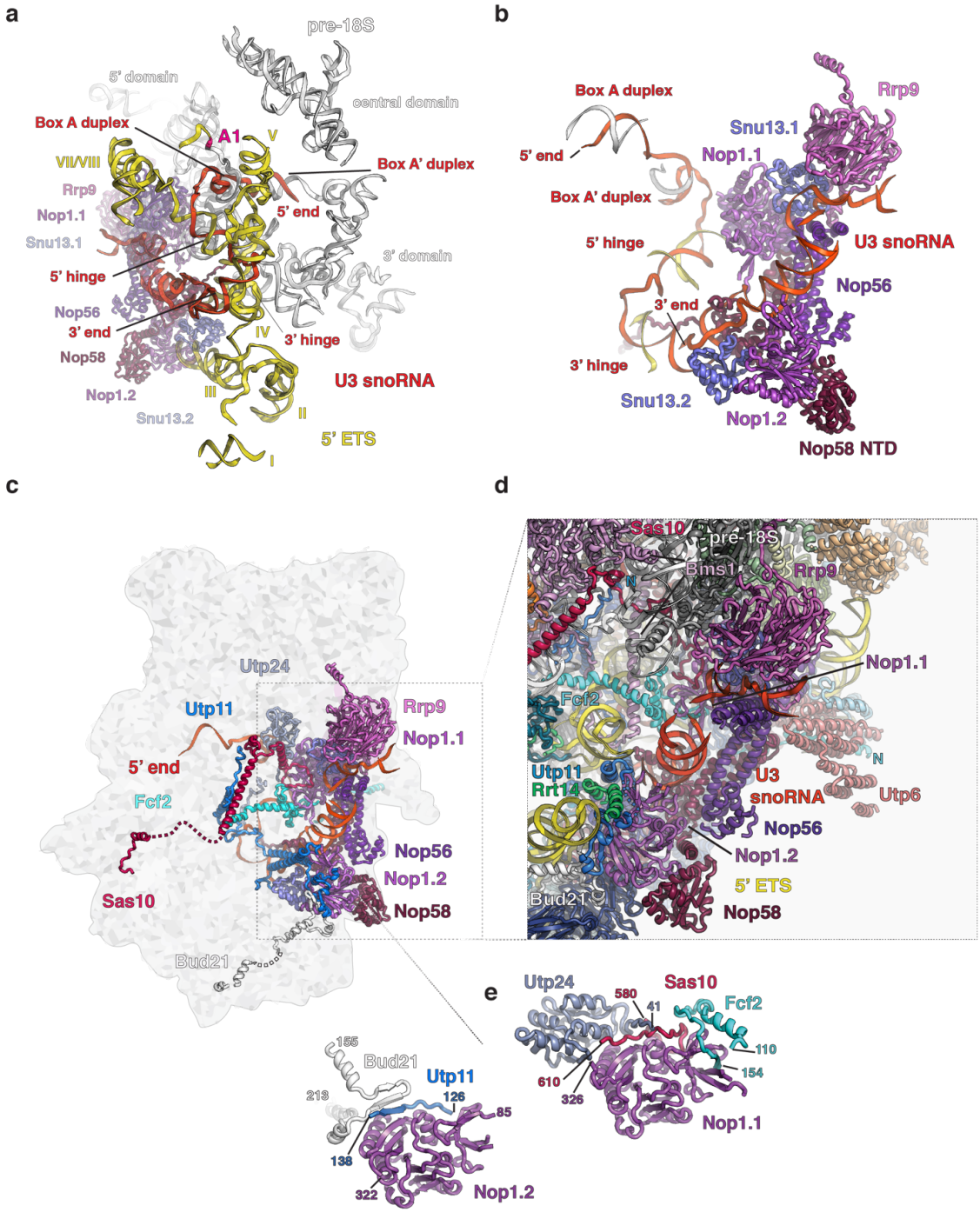
While the 5' part of U3 snoRNA is used to base-pair with the pre-rRNA in the center of the SSU processome, the 3' part which contains the conserved Box B/C and Box C'/D motifs is bound by the core box C/D snoRNA proteins Nop1, Nop56, Nop58 and Snu13 as well as the U3 specific factor Rrp9 (**Figure 3.12b**). All of the U3 snoRNP protein subunits are located on the periphery of the SSU processome (**Figure 3.12c**). The core proteins form an almost-symmetrical arrangement on the U3 snoRNA scaffold. Two copies of Snu13 and Nop1 are placed on opposite sides of the U3 snoRNA with the heterodimer of Nop56 and Nop58 in between (**Figure 3.12b**). The CTD and NTD of Nop56 contact the Nop1-Snu13 subunits proximal to

the β-propeller of Rrp9, while the CTD and NTD of Nop58 bind the second Nop1-Snu13 copies in a similar manner. Despite harboring methyltransferase activity, the two copies of Nop1 in U3 snoRNP are not in an active state and substrate methylation is prohibited by steric hindrance (Sun et al. 2017).

In the SSU processome the Nop1 subunits have a function besides RNA base methylation. Surprisingly, these common box C/D snoRNA proteins serve as binding platforms for five SSU processome components (Fcf2, Sas10, Utp24, Utp11 and Bud21) (**Figure 3.12 c,d,e**). The surfaces of the two Nop1 subunits are used distinctively by these proteins (**Figure 3.12e**). Peptide backbone elements of these subunits 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. Four of the Nop1-bound proteins (Sas10, Utp11, Fcf2, Bud21) traverse through and around the core of the SSU processome, thereby connecting the U3 snoRNP to distant pre-rRNA domains (**Figure 3.12d**). Sas10, Utp11 and Fcf2 each span at least 100 Å and use conserved motifs to bind multiple interaction partners within the particle and fulfill distinct roles at each site. A conserved sequence in Sas10 is also capable of recruiting binding partners outside the particle, as it features an exosome recruitment domain (Mitchell 2010).

Figure 3.12 | Structure of the U3 snoRNP base-paired to the 5' ETS and pre-18S rRNA.

[a] Overview of the U3 snoRNP proteins (purple), U3 snoRNA (red) and their interactions with the 5' ETS (yellow) and pre-18S RNA (white). Proteins and RNAs in this cartoon representation are color-coded and the 5'-, central – and 3'- domain of the pre-18S rRNA, the helices of the 5' ETS and all functionally relevant sequence elements of U3 snoRNA indicated. **[b]** View of the U3 snoRNP with focus on the U3 snoRNA associated core proteins (purple). **[c]** The Nop1 bound proteins Sas10 (pink), Utp24 (grey), Utp11 (blue), Fcf2 (cyan) and the 5' end of the U3 snoRNA connect the peripheral part of U3 snoRNP with distant regions in the SSU processome (shown as schematic with outline). **[d]** Detailed view of the protein and RNA environment surrounding the core U3 snoRNP proteins in the SSU processome. **[e]** The two copies of Nop1 (Nop1.1, Nop1.2) are bound by a distinct set of SSU processome components. Residue ranges of the proteins mediating the interactions are labeled in a color-coded manner.


a
pre-18S
5' domain
central domain
Box A duplex
A1
V
VII/VIII
Box A' duplex
Rrp9
Nop1.1
5' end
Snu13.1
3' domain
5' hinge
Nop56
3' end
IV
3' hinge
Nop58
U3 snoRNA
Nop1.2
III
Snu13.2
II
5' ETS
I
b
Box A duplex
Rrp9
Snu13.1
5' end
Nop1.1
Box A' duplex
5' hinge
U3 snoRNA
Nop56
3' end
3' hinge
Snu13.2
Nop1.2
Nop58 NTD
c
Utp24
Utp11
Rrp9
Nop1.1
5' end
Fcf2
Nop56
Nop1.2
Sas10
Nop58
Bud21
d
Sas10
pre-18S
Bms1
Rrp9
N
Nop1.1
Fcf2
N
Utp11
Rrt14
U3
snoRNA
Utp6
Nop56
Nop1.2
5' ETS
Bud21
Nop58
e
Utp24
Sas10
580
41
Fcf2
110
154
610
326
Nop1.1
155
Bud21
Utp11
213
126
85
138
322
Nop1.2

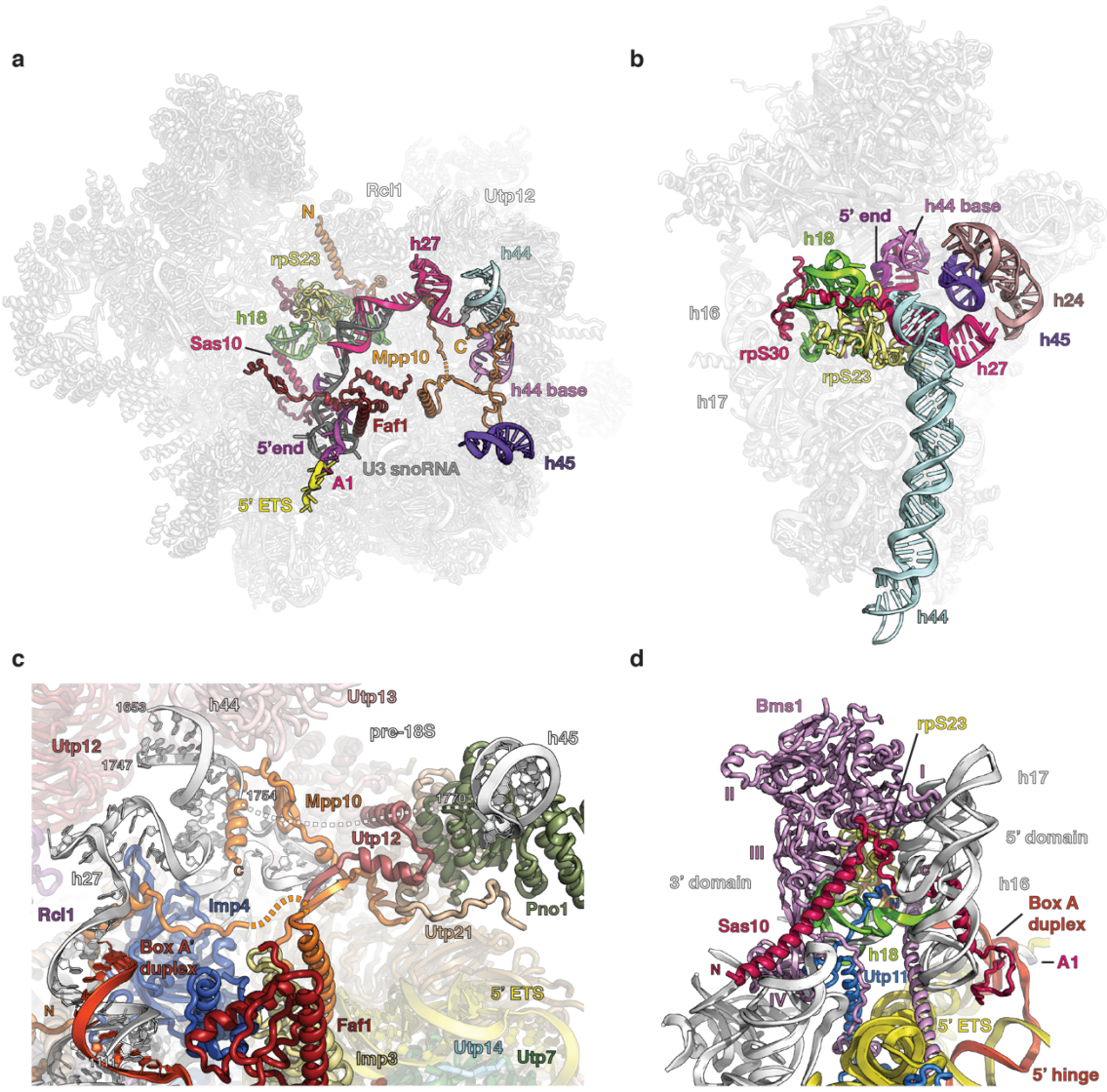
3.12 U3 snoRNA- and protein-mediated remodeling of central pseudoknot elements

The 18S rRNA sequence elements engaging with the Box A and A' motifs of U3 snoRNA in the SSU processome are part of the central pseudoknot in the mature small subunit (**Figure 3.13 a,b**). In the ribosome the central pseudoknot determines the position of the four structured 18S rRNA domains (5'-, central-, 3;' major and 3' minor) relative to each other. In their mature positions, the four domains adopt the compact conformation of the small ribosomal subunit (**Figure 3.13b**). As a consequence of U3 snoRNA base-pairing, central pseudoknot formation is prohibited in the SSU processome, which leads to the spatial separation of the 18S domains in the particle. In addition to U3 snoRNA several other ribosome assembly factors are involved in stabilizing this separation and further RNA remodeling events (**Figure 3.13a, Figure 3.14**), thus facilitating the independent maturation of the 18S domains in the SSU processome (Chaker-Margot et al. 2017).

The Box A/A' interactions lead to the remodeling of helix 27, which is located downstream of these RNA duplexes (**Figure 3.13 a,c and Figure 3.14**). As the RNA sequence forming the base of helix 27 in the mature ribosome is bound by U3 snoRNA, helix 27 has to adopt a different conformation in the SSU processome (**Figure 3.13 a,b**). Consequentially, a new RNA stem loop forms, which is supported by Mpp10, Utp12 and Rcl1 (**Figure 3.13 a,c**).

Figure 3.13 | RNA remodeling prevents central pseudoknot formation.

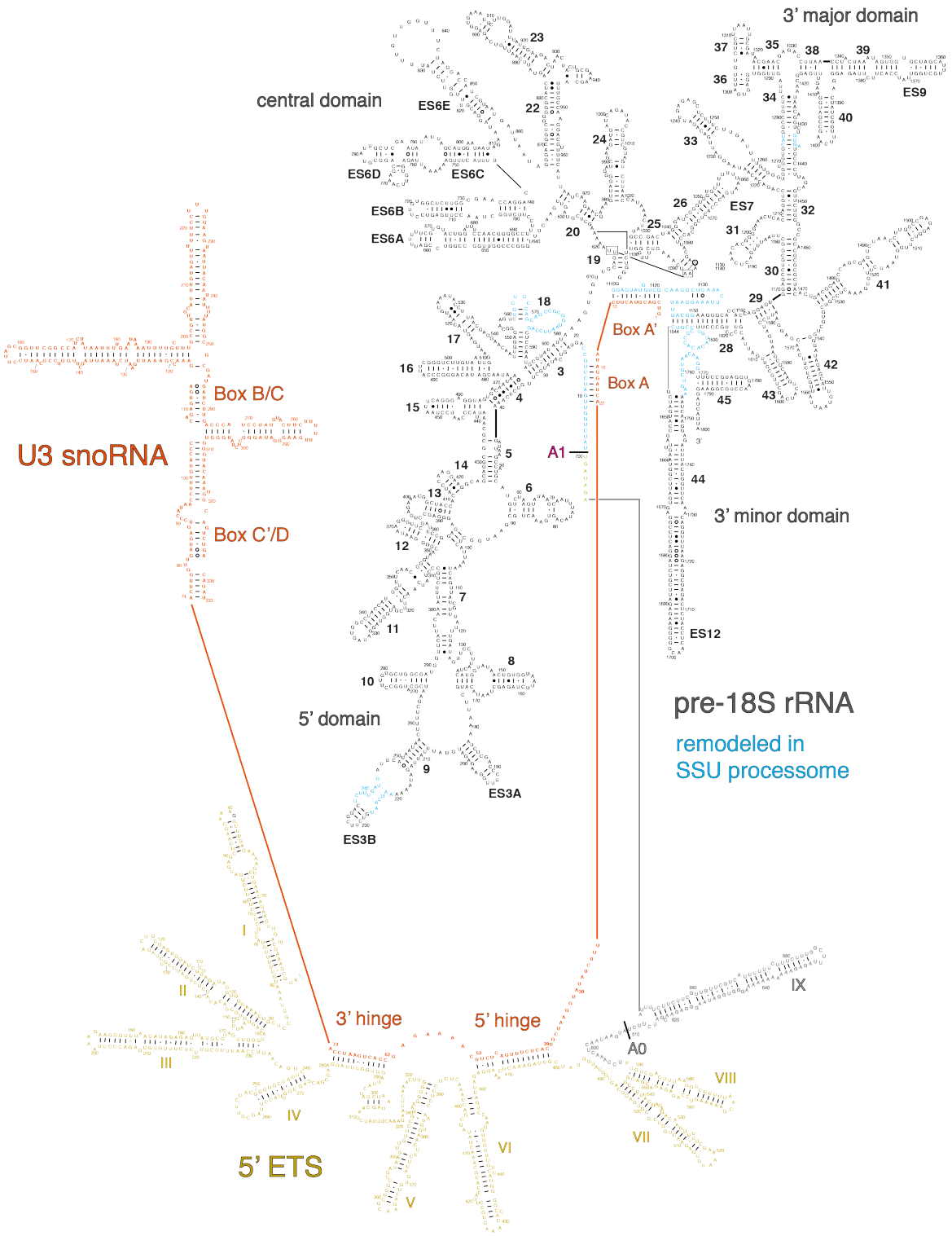
[a] The central pseudoknot and 18S rRNA elements in its vicinity are shown in color in their immature positions in the SSU processome (grey) and labeled with their corresponding mature 18S rRNA helix (h) number. Chaperoning RNA, ribosomal proteins and ribosome assembly factors are color-coded, and the A1 cleavage site is highlighted in pink. Termini of Mpp10 are indicated with N and C respectively. **[b]** Color-coded RNA elements close to the central pseudoknot in the mature small ribosomal subunit (grey) labeled as in a (PDB 4V88) (Ben-Shem et al. 2011). **[c]** Mpp10 (orange) and its interactions with pre-18S RNA (white). Elements of the 18S rRNA (helices 44 and 45) 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 helix 18 (h18, green) of the pre-18S RNA (white). Domains of Bms1 (purple) are numbered with roman letters. 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 the remodeling (Utp11, Sas10) and the ribosomal protein rpS23 are shown.


a
Rcl1
Utp12
N
h27
rpS23
h44
h18
Sas10
Mpp10
C
h44 base
Faf1
5'end
U3 snoRNA
h45
A1
5' ETS
b
5' end
h44 base
h18
h24
h16
h45
rpS30
rpS23
h27
h17
h44
c
1653
h44
Utp13
pre-18S
h45
Utp12
1747
1754
Mpp10
1770
Utp12
h27
C
Rcl1
Imp4
Pno1
Utp21
Box A
duplex
5' ETS
N
Faf1
Imp3
Utp14
Utp7
d
Bms1
rpS23
I
h17
II
5' domain
III
3' domain
h16
Box A
duplex
Sas10
A1
N
h18
Utp11
IV
5' ETS
5' hinge

Mpp10 also plays a central role in the remodeling of nucleotides close to helices 44 and 45 (**Figure 3.13 a,c**). These two helices are proximal to the central pseudoknot in the mature small subunit (**Figure 3.13b**). A partial unwinding of the region upstream of helix 44 results in an RNA loop (nucleotides 1628-1639) that is stabilized by Mpp10 (**Figure 3.13 a,c**). Due to this partial unwinding, 16 nucleotides of the opposite strand (nucleotides 1755-1769) are available to serve as a long linker to helix 45, which is positioned 60 Ångstroms away on top of Pno1 (**Figure 3.13c**). Pno1 is held in place by linker-peptides of Utp12 and Utp21 of the UtpB complex (**Figure 3.13c**).

Another important location for protein-mediated RNA remodeling is the binding site of ribosomal protein rpS23, which is positioned in the mature small subunit close to all other remodeled RNA elements next to helix 18 (**Figure 3.13 a,b,d**). In the SSU processome, conserved elements of the U3 snoRNP-interacting Utp11 and Sas10, and the GTPase Bms1 are employed in a concerted fashion to remodel helix 18 (nucleotides 558-590) (**Figure 3.13d**). 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.

Figure 3.14 | Secondary structure diagram of RNAs in the SSU processome. Individual nucleotides of RNAs are indicated with their base pairing interactions. 5' ETS (yellow), 18S (black) and U3 snoRNA (red) are shown. Regions of the 18S rRNA that have been remodeled in the SSU processome are highlighted in light-blue. The intact A1 site is highlighted in pink, whereas the cut A0 site is shown in grey. Helix IX of the 5' ETS is shown in grey.


3' major domain
central domain
ES6E
ES6D
ES6C
ES6B
ES6A
ES7
ES9
Box A'
Box A
Box B/C
Box C'/D
U3 snoRNA
A1
3' minor domain
ES12
pre-18S rRNA
remodeled in
SSU processome
5' domain
ES3A
ES3B
3' hinge
5' hinge
A0
5' ETS
I
II
III
IV
V
VI
VII
VIII
IX

3.13 Ribosome assembly factors stabilize the separated pre-18S rRNA domains

U3 snoRNA- and protein-dependent remodeling of pre-18S sequences in the vicinity of the central pseudoknot (**Figure 3.13 and Figure 3.14**), affects the spatial arrangement of the individual 18S rRNA domains (**Figure 3.15**). In the SSU processome the four 18S rRNA domains are set in an open conformation around an axis represented by the Box A/A' interactions of U3 snoRNA (**Figure 3.15b**). Specific ribosome assembly factors are associated with each 18S rRNA domain, fulfilling diverse functions (**Figure 3.15 a,b**).

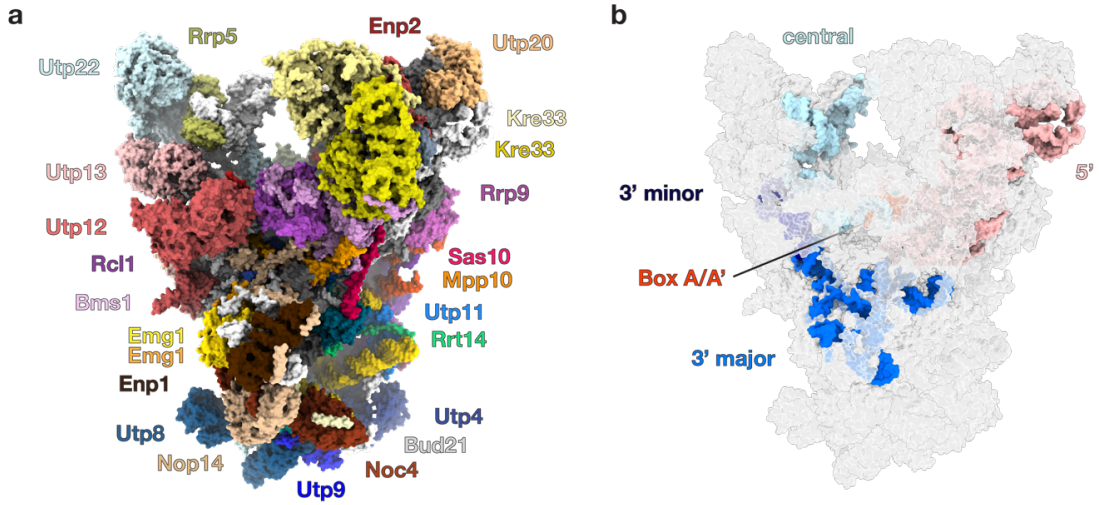


Figure 3.15 | Distinct ribosome assembly factors bind the separated 18S domains.

[a] Surface representation of the SSU processome with all visible ribosome assembly factors labeled in a color-coded manner. **[b]** Same view of the SSU processome as in [a] but with all ribosome assembly factors shown in transparent white. The 5' domain (salmon), central domain (light-blue), 3' major domain (blue) and the 3' minor domain (navy) of the pre-18S are depicted in full color. The Box A/A' base-pairing regions of U3 snoRNA (red) are indicated.

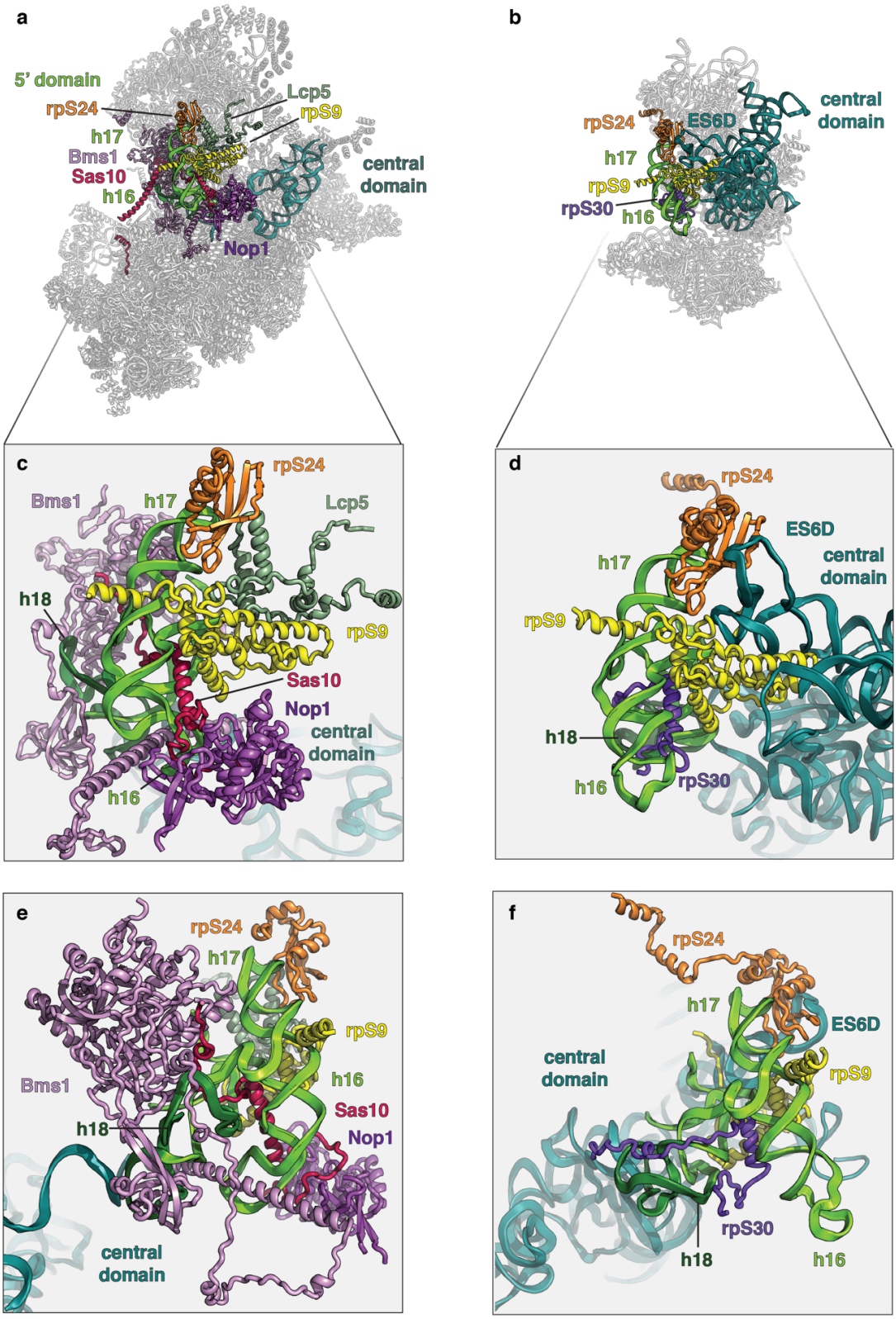
The 5’ domain is placed at the top of the particle, above the U3 snoRNP core proteins (**Figure 3.15**). While the beginning of the 5’ domain (nucleotides 9-15) is remodeled by the U3 snoRNA Box A motif, the majority of the ensuing sequence adopts a near-mature conformation with most secondary structure elements formed (**Figure 3.14**) and seven (rpS4, rpS6, rpS8, rpS9, rpS11, rpS23, rpS24) of the eight 5’-domain ribosomal proteins bound. rpS30 is the only 5’ domain-associated ribosomal protein not yet recruited. Its binding is blocked by Sas10, which employs molecular mimicry to occupy the binding site of rpS30 on helix 16 (**Figure 3.16**).

Enp2, Utp20, Lcp5, Utp11, Bms1 and Kre33 are further ribosome assembly factors interacting with the 5’ domain (**Figure 3.15**). While Sas10, Bms1 and Utp11 remodel helix 18 (**Figure 3.13d and Figure 3.16 a,c,e**), a 5’ domain rRNA structure close to the central pseudoknot, the other associated ribosome assembly factors fulfill diverse roles in quality control, rRNA base modifications, bridging of pre-rRNA domains and stabilization of premature rRNA conformations.

Lcp5, like Sas10, harbors an exosome recruitment motif in its sequence and could act as another quality control factor (Mitchell 2010). Lcp5 is positioned next to rpS9 on the 5’ domain (**Figure 3.16 a,c,e**), a location occupied by expansion segment ES6D of the central domain in the mature small subunit (**Figure 3.16 b,d,f**).

Figure 3.16 | Ribosome biogenesis factors prevent premature r-RNA folding in the 5' domain.

[a] Conformations of helices 16, 17 (green) and 18 (dark green) of the 5' domain, and the central domain (teal) of the pre-18S RNA within the SSU processome (grey). Interacting ribosome assembly factors and ribosomal proteins are shown and color-coded. **[b]** Conformation of the same elements as in a in the context of the mature small ribosomal subunit (PDB 4V88) (Ben-Shem et al. 2011). **[c,d]** Detailed views of the conformations of helix 16 and the central domain in the SSU processome [c] and the small ribosomal subunit [d]. Sas10 mimics rpS30 and occupies its binding site on helix 16. Lcp5 blocks the central domain from occupying its mature position by steric hindrance while Bms1 bends helix 16. **[e,f]** Orthogonal views to panel [c] and [d].


a
5' domain
rpS24
Lcp5
rpS9
h17
Bms1
Sas10
h16
central
domain
Nop1
b
central
domain
rpS24
ES6D
h17
rpS9
rpS30
h16
c
rpS24
Bms1
h17
Lcp5
h18
rpS9
Sas10
Nop1
central
domain
h16
d
rpS24
ES6D
h17
central
domain
rpS9
h18
h16
rpS30
e
rpS24
h17
rpS9
Bms1
h16
Sas10
h18
Nop1
central
domain
f
rpS24
h17
ES6D
central
domain
rpS9
rpS30
h18
h16

The presence of expansion segment 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 (**Figure 3.16 a,c,e**). The continued presence of Lcp5 during later stages of ribosome assembly may therefore indicate an incomplete or faulty state that could be targeted for degradation.

Kre33 is a homodimer that possesses helicase as well as acetyltransferase activity (Sharma et al. 2017). The two 18S rRNA bases modified by Kre33 (nucleotides 1280,1773) (Ito et al. 2014; Sharma et al. 2015) are located in helix 34 (3' major) and 45 (3' minor) respectively. Both of these helices are distant to the current position of Kre33. Since the pre-rRNA in the SSU processome sample was only analyzed for its composition and not its modification state it remains unclear whether Kre33 has already acted upon its substrates or if it is positioned to do so downstream in the maturation process.

The large α-helical repeat of Utp20 provides structural support for RNA expansion segments ES3A and ES3B and bridges the back of the structure, close to Sof1, with the top near Kre33 and Enp2 (**Figure 3.17a**). The β-propeller of Enp2 occupies the same position on the 5' domain as parts of helix 44 in the mature ribosome.

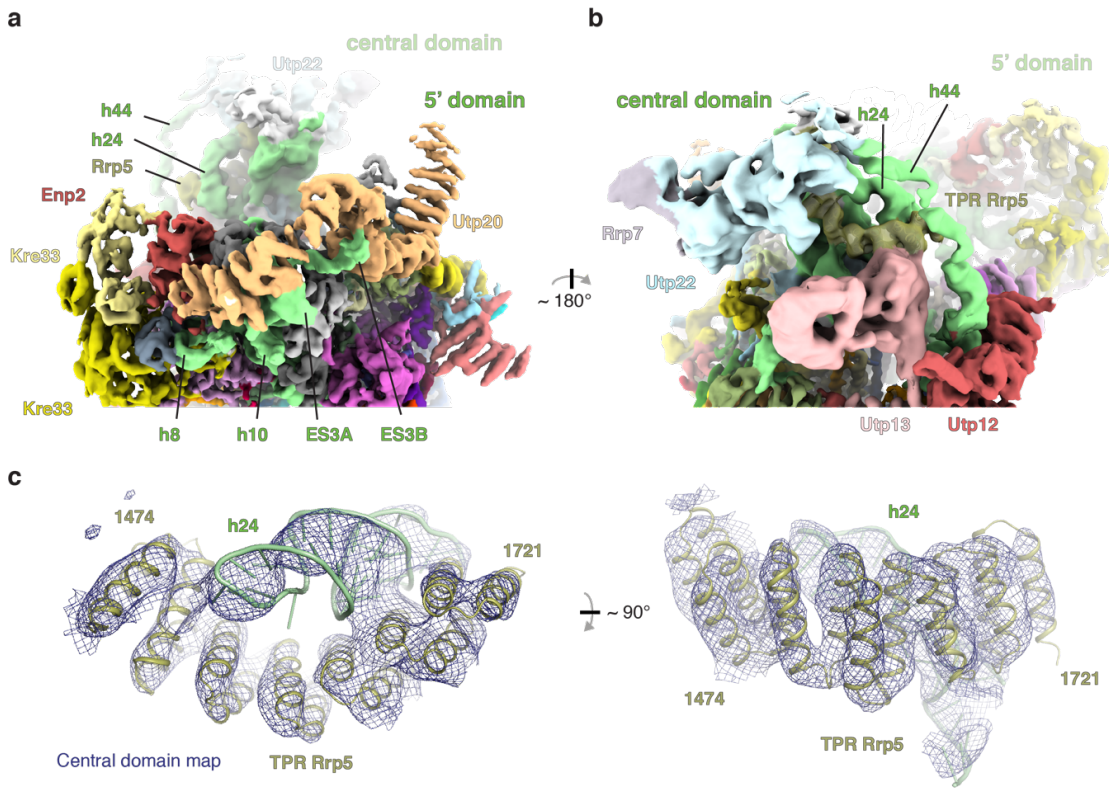


Figure 3.17 | The helical repeat proteins Utp20 and Rrp5 stabilize rRNA helices in the 5' - and central domain.

[a,b] Two views of a composite cryo-EM density map consisting of the 6 Å low-pass filtered overall map 2 and the 7.2 Å central domain map. The density is colored as in (Figure 3.7) but with the pre-18S RNA colored in pale-green. Helices (h8, h10, h24, h44) and expansion segments (ES3A, ES3B) of the 18S rRNA are labeled next to the corresponding density. In the right view, the density for the tetratricopeptide repeat (TPR) of Rrp5 is shown transparent with the docked crystal structure (PDB 5C9S). **[c]** Cryo-EM density from the central domain map with molecular fit of the TPR repeat crystal structure of Rrp5 (PDB 5C9S), shown as cartoon. The concave interface serves as binding platform for 18S rRNA helix 24 (h24, in green).

The 5' domain transitions into the central domain downstream of the Bms1-Utp11-Sas10 remodeled helix 18 (**Figure 3.14**). The central domain is located opposite of the 5' domain at the top of the particle (**Figure 3.15 and Figure 3.17 a,b**) and is highly flexible under the growth and purification conditions used to obtain the SSU processome. This is reflected in the limited resolution of the central domain density map (**Figure 3.17**).

UtpC, a pre-assembled protein complex comprised of Utp22 and Rrp7 (Krogan et al. 2004; Lin et al. 2013), as well as Rrp5 are coordinating parts of this mobile domain (**Figure 3.17 b,c**). The tetratricopeptide repeat (TPR) of Rrp5 (Khoshnevis et al. 2016), which is necessary for pre-18S processing (Torchet et al. 1998; Eppens et al. 1999), provides a cradle to stabilize helix 24 (**Figure 3.17 b,c**) in a different conformation with respect to the mature small subunit.

In contrast, the other resolved sequence sections of the central domain rRNA adopt a near-mature conformation. Two ribosomal proteins (rpS13, rpS14) could be fitted into the cryo-EM density surrounding the central domain rRNA in the SSU processome. rpS22 and rpS7, in the mature ribosome bound to the central domain, are positioned closer to the 5' domain in the SSU processome and held in place by protein-protein interactions with Faf1 and Sof1.

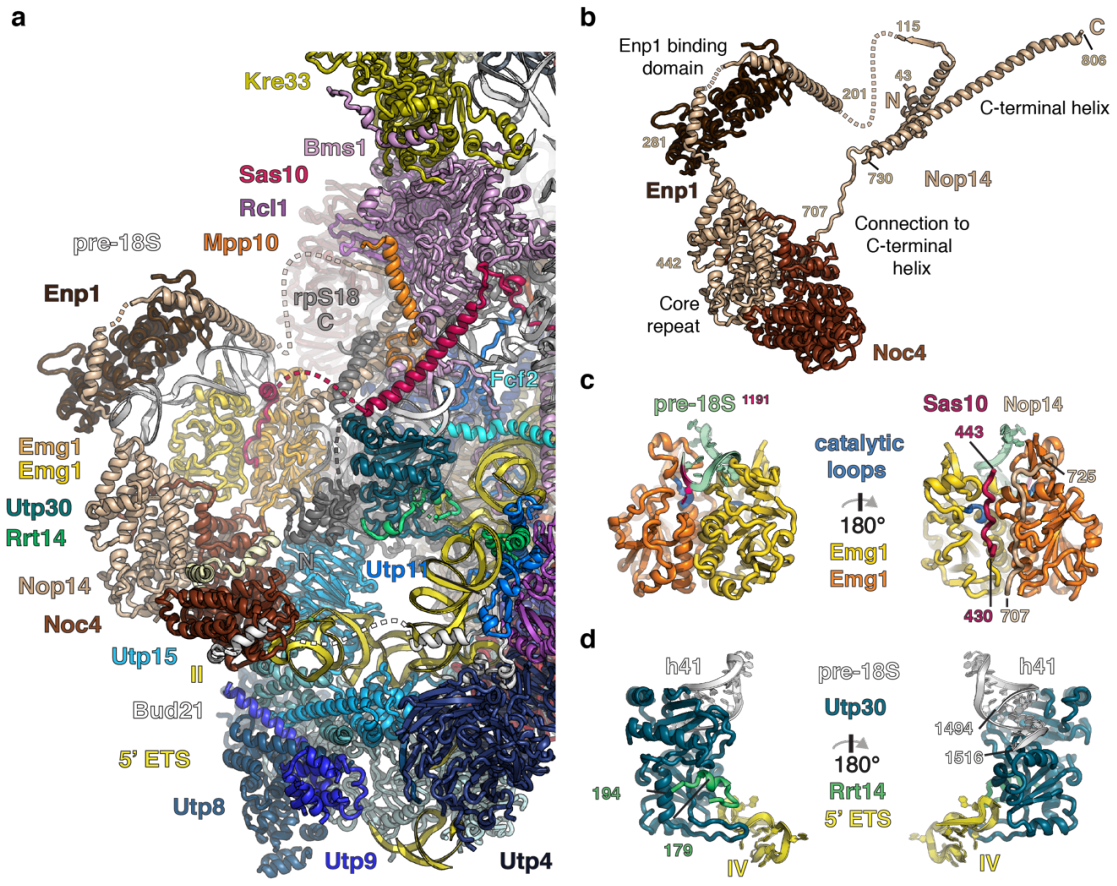


Figure 3.18 | The 3' major domain is positioned on a lateral extension of UtpA. **[a]** Overview of the protein and RNA environment of the 3' major domain in the SSU processome shown as cartoon representation. Components and their label are color coded. **[b]** Nop14 is docked into the SSU processome through its long C- and N-terminal extensions. Structural elements of Nop14 are labeled in black while its residue numbers as well as N- and C-terminus are labeled in beige. The Nop14-interacting Enp1 and Noc4 have color-coded labels. **[c]** Views of the Emg1 homodimer (orange, 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. **[d]** Utp30 and Rrt14 recognize the 3' major domain pre-18S (h41, white) and 5' ETS RNA (helix IV, yellow).

The 3' end of the central domain and U3 snoRNA form the Box A' RNA duplex. Downstream of this base-paired region is the Utp12/Mpp10/Rcl1-bound remodeled helix 27 (**Figure 3.13 a,c**), which leads into the 3' major domain. The 3' major domain, in contrast to the 5' and central domain, is largely unfolded in the SSU processome and positioned near the base of the particle on a lateral extension close to UtpA (**Figure 3.18a**). There, the β-propeller of Utp15 and helix II of the 5' ETS provide a binding platform for the α-helical repeat of Noc4, which is further connected to UtpA through Bud21 (**Figure 3.18a**). By directly binding to a second repeat protein, Nop14, the α-helical repeat structure of Noc4 is further expanded. This composite helical repeat provides a scaffold for Enp1, Emg1 and the parts of the 3' major domain forming the beak structure in the mature ribosome.

The repeat of Nop14 is flanked by N- and C-terminal extensions (**Figure 3.18b**). A 75 amino-acid-long C-terminal helix docks Nop14 into an opening between the Mpp10-Imp4 dimer and the Bms1-Rcl1 complex and points its C-terminal end towards the central cavity between the central and the 5' domains. The N-terminal extension of Nop14 is positioned close to the long C-terminal helix and binds Enp1, which caps the 3' major domain, before leading into the core repeat of the protein (**Figure 3.18b**). In addition to stabilizing Enp1 on top of the rRNA, this arrangement also positions the 3' major domain residue 1191 in the active site of the methyltransferase Emg1 (**Figure 3.18c**). 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 the target base.

In proximity to the long C-terminal helix of Nop14, Bms1 and helix 28, the C-terminus of rpS18 was identified (**Figure 3.18a**). The C-terminus of rpS18, which binds elements of the beak structure in the mature small subunit, adopts a different conformation in the context of the SSU processome, where this structure has not yet formed. The core domain of rpS18 however, is positioned in a near-mature configuration with respect to the 18S rRNA close to the base of helix 41.

Helix 41 of the 3' major domain is connected to helix IV of the 5' ETS via the L1-domain containing ribosome assembly factor Utp30 (**Figure 3.18d**). This ribosome biogenesis factor is a homolog of the bacterial ribosomal protein rpL1, which binds the 23S rRNA in the large subunit of the bacterial ribosome (Tishchenko et al. 2012). Instead of functioning as a ribosomal protein, Utp30 has adopted a role in ribosome assembly in eukaryotes. While binding helix 41 on one side using a β-sheet and short loops, helix IV is recognized through a longer loop on the other side of Utp30 (**Figure 3.18d**). A C-terminal proline-glycine rich peptide of Rrt14 binds on the β-sheet of Utp30 below helix IV, while the N-terminal part of this protein interacts with Utp11 close to helix III of the 5' ETS (**Figure 3.18d**).

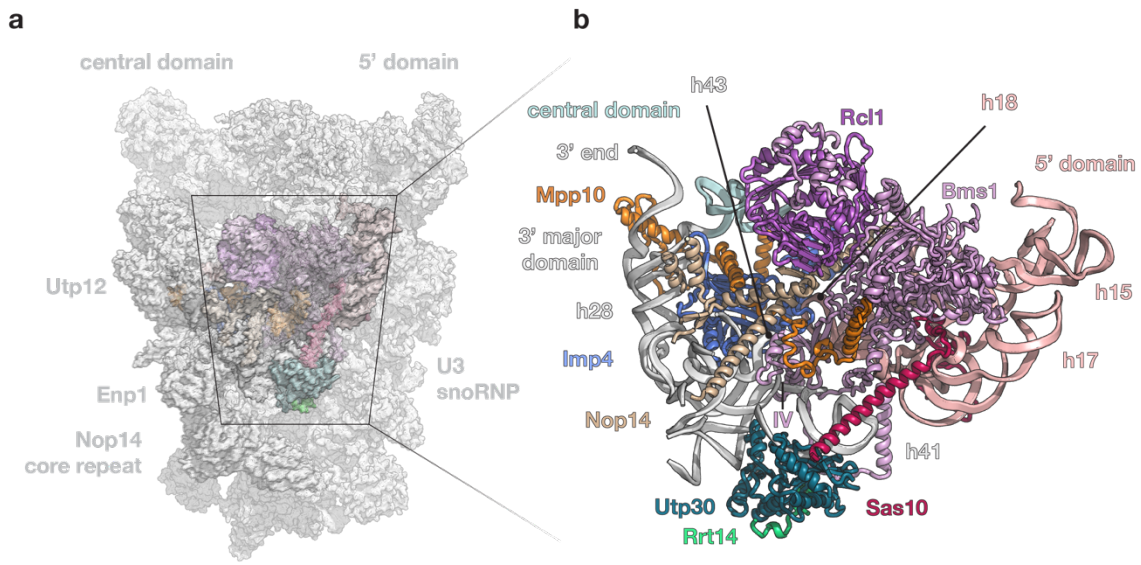


Figure 3.19 | Bms1 and the Mpp10 complex connect the 3' major- with the 5' domain.

[a] Surface representation of the SSU processome (white) with the approximate location of landmark components (Utp12, Enp1, Nop14 core repeat, central domain, 5' domain, U3 snoRNP) indicated in grey as references. **[b]** Bms1 (purple), Imp4 (blue) and Mpp10 (orange) mediate contacts between the 3' major- (white) and 5' domain (light-pink). Helices of the pre-18S rRNA are labeled in the respective color of their domains. Domain IV of Bms1 is shown in purple.

While the peripherally located beak-forming elements of the 3' major domain are engulfed by Nop14, Noc4, Enp1 and Emg1, the rest of the domain is stabilized by Bms1-Rcl1, Mpp10 and Imp4 near the core of the SSU processome (**Figure 3.19a**). Bms1 bridges the 3' major and 5' domain of the pre-18S rRNA by bending and remodeling the 5'-domain helices 16 and 18 through its domain IV, and binding helix 43 with the same domain (**Figure 3.19b**).

Imp4, which is stabilizing several 3' major helices, contributes to this inter-rRNA-domain connection as well by supporting helix18 with a C-terminal peptide loop. The core domain of Imp4 is wedged between helix 43 and helix 28. It is bound by Mpp10, whose extended structure not only couples Bms1 and Imp4, but also reaches the 3' end of helix 28 (**Figure 3.19b**). Here, the pre-rRNA has been remodeled through a partial undoing of the ensuing helix 44 base, which leads to the formation of an SSU processome specific rRNA-loop (nucleotides 1628-1639) chaperoned by Mpp10 and the UtpB complex (**Figure 3.13c and Figure 3.14**).

Downstream of this loop, the tandem β-propeller of Utp12 coordinates the base of helix 44 (**Figure 3.10f and Figure 3.17b**). Helix 45 is positioned 60 Å away at the back of the structure on top of Pno1 (**Figure 3.11a, Figure 3.13c**). Helix 44 and 45 constitute the 3' minor domain of the 18S rRNA, which ends at the D-site. The D-site separates the pre-18S rRNA from ITS1. Even though the D-site and ITS1 were not resolved in the SSU processome structure, RNA-protein cross-linking of UtpB has implicated Utp13 in binding structures within ITS1 (**Figure 2.7**).

3.14 The GTPase Bms1 is placed in a central position and mediates inter-18S rRNA domain contacts

Similar to U3 snoRNA, the GTPase Bms1 reaches deep into the SSU processome and is connected to distant pre-18S rRNA domains via long peptide-like assembly factors that traverse through the core of the particle and share binding sites with the GTPase (**Figure 3.20a**). A long C-terminal linker and helix (CTD) wrap around helix VI of the 5' ETS and thereby anchor Bms1 in the core of the particle (**Figure 3.20 b,c**). The U3 snoRNP bound adaptor proteins Sas10 and Utp11 bind helix VI too (**Figure 3.20c**) and engage in Bms1-mediated remodeling of the 5' domain helices (**Figure 3.13d**). Through its domain IV, Bms1 bridges the remodeled 5' domain helices with parts of the 3' major domain (**Figure 3.19b**). The C- and N-terminal helices of Nop14 and Sas10 link Bms1 to the beak-forming 3' major sequences on the periphery of the SSU processome (**Figure 3.18 a,b and Figure 3.20a**). Via Mpp10, Bms1 is also connected to the 3' minor domain (**Figure 3.13c and Figure 3.20 a,c**). Bms1 binds its co-factor Rcl1 through a dedicated peptide and interacts with the acetyltransferase and helicase Kre33 through a newly identified binding motif (**Figure 3.20b**). Its central placement in the particle, long range connections to pre-18S domains and GTPase activity, put Bms1 in an optimal position to modulate conformational changes during SSU processome formation or future maturation steps.

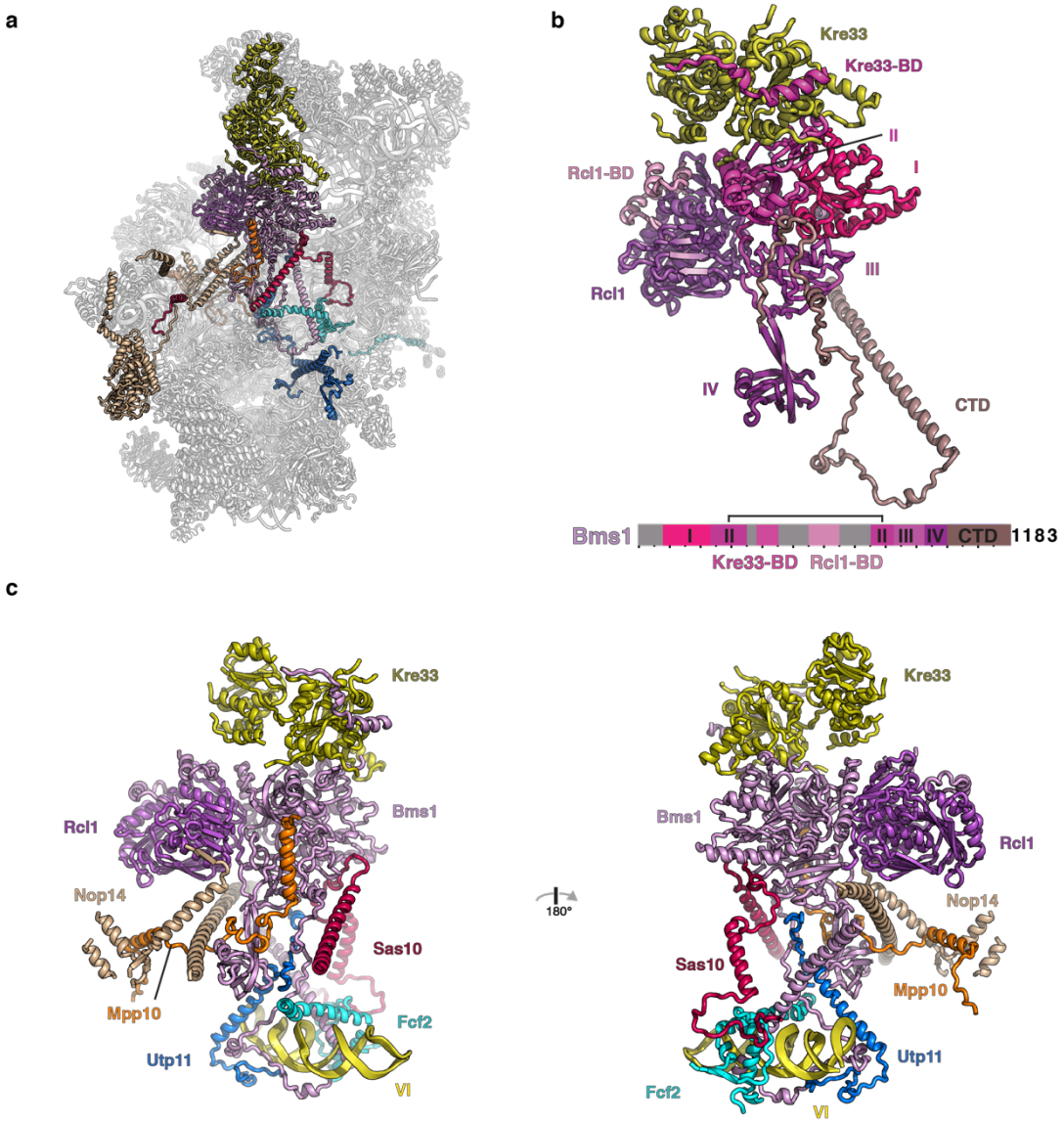


Figure 3.20 | Structural analysis of Bms1 and its interaction partners.

[a] Overview of Bms1 and its binding partners within the SSU processome. Proteins are colored as in (Figure 3.7) with a transparent outline of the SSU processome in white. **[b]** Architecture of the Bms1-Rcl1 complex with Bms1 domains I-IV, the Kre33-binding domain (Kre33-BD) and the Rcl1-binding domain (Rcl1-BD) color-coded in shades of violet. The Bms1 C-terminal domain (CTD) is highlighted in light-brown. **[c]** Two views of the interactions of Bms1 with other SSU processome subunits. Only the most N-terminal domains of one Kre33 monomer (yellow) are shown. Helix VI of the 5' ETS is labeled in yellow.

3.15 Conclusions

The cryo-EM reconstruction and atomic model of the yeast SSU processome has provided unique insights into the function of more than 50 ribosome assembly factors, the 5' ETS and U3 snoRNA. By solving the structure of this early nucleolar precursor of the small ribosomal subunit, contextualized high-resolution structural information for several of the inherently flexible ribosome biogenesis proteins was obtained for the first time.

In isolation UtpA, the largest subcomplex of the SSU processome, is a highly dynamic assembly. The structural basis for this high degree of flexibility was revealed in the SSU processome structure, where long inter-domain linkers organize the architecture of the complex at the base of the particle. UtpA subunits coordinate the first three helices of the 5' ETS, form an intertwined interaction interface with UtpB and provide a platform for the 3' major domain of the 18S rRNA. Together with UtpB, U3 snoRNP, the Mpp10 complex and other 5' ETS associated subunits, UtpA forms the structural scaffold for the spatially separated 18S rRNA domains in the SSU processome.

U3 snoRNA and protein-mediated remodeling of pre-18S rRNA sequences in the vicinity of the central pseudoknot lead to the segregation of the four structured 18S domains. The U3 snoRNA Box A and A' base-pairings are essential for the prevention of premature central pseudoknot formation. To stabilize the open

architecture of the 18S domains, ribosome assembly factors bind and link the pre-18S structures to the 5' ETS particle and each other. Many of these factors share similar folds and yet fulfill distinct functions. While 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 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. Extended proteins, which weave through the particle and employ conserved motifs to interact with multiple assembly factors and RNA segments, connect distant regions in the particle. Two of these extended proteins, Sas10 and Lcp5, may act as quality control factors as they harbor exosome recruitment motifs. With the nuclease Utp24, acetyltransferase/helicase Kre33, GTPase Bms1 and methyltransferase Emg1 several enzymes are integrated in the SSU processome. While it is clear that the A1-site nuclease Utp24 has not yet acted upon its substrate, it remains unclear if the other identified enzymes have already been active or are positioned to act at later stages of the maturation process.

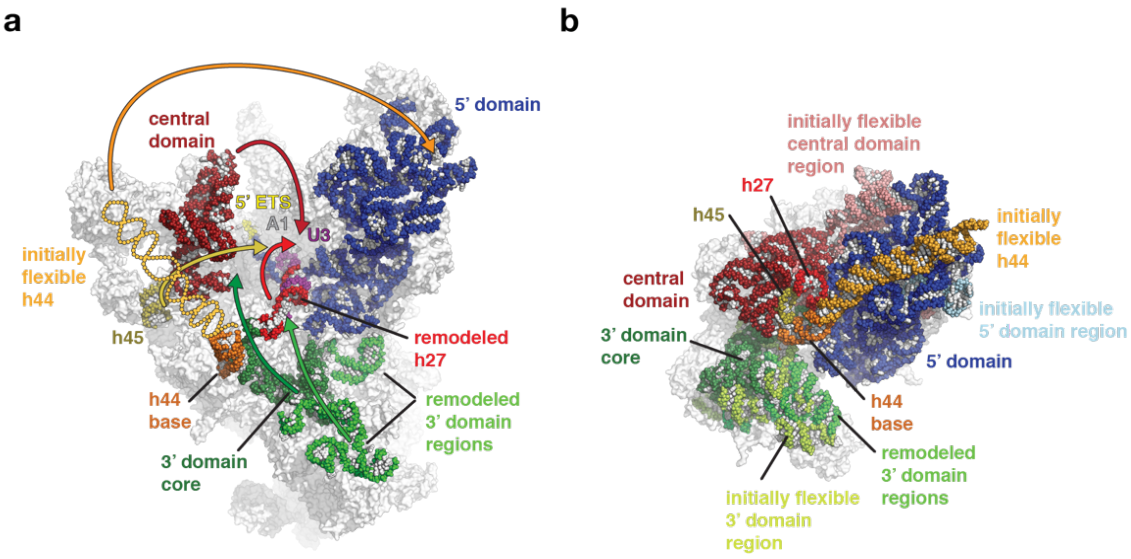


Figure 3.21 | Model of pre-rRNA domain rearrangements during small subunit maturation.

[a,b] Comparative locations of rRNA domains within the SSU processome [a] and the mature small ribosomal subunit (PDB 4v88) (Ben-Shem et al. 2011) [b]. Individual rRNA domains are colored identically with 5' domain (blue), central domain (red), 3' domain (green) and shown as spheres superimposed onto transparent outlines of the particles. In the SSU processome, the flexible helix 44 is indicated as schematic outline. Rearrangements of rRNA domains from the SSU processome [a] that are necessary to obtain the positions within the mature small ribosomal subunit [b] are indicated with arrows. The central U3 snoRNA Box A and Box A' are colored in purple. RNA elements disordered in the SSU processome are indicated in lighter shades in the mature small subunit.

Major enzymatic and structural changes are needed to transition from the SSU processome into the mature small ribosomal subunit (**Figure 3.21**). Enzymatic reactions include the unwinding of the Box A and Box A' duplexes by RNA helicases such as the Utp14-associated Dhr1 (Sardana et al. 2015), cleavage of the A1-site by

Utp24 (Bleichert et al. 2006; Tomecki et al. 2015; Wells et al. 2016), and the exchange of the GTPase Bms1 with the structurally related factor Tsr1. Additionally, structural changes such as rotational and translational movements of the central and 3' domains with respect to the 5' domain are required (**Figure 3.21**). The formation of the central pseudoknot and its surrounding elements, the formation of inter-domain base-pairing interactions between the central and 5' domains and the incorporation of additional ribosomal proteins, such as rpS30, are further maturation steps.

So far, the SSU processome is the only structurally characterized nucleolar precursor of the small ribosomal subunit. Hence, our mechanistic understanding of the maturation steps occurring before or after the SSU processome stage is very limited. Furthermore, the SSU processome has been identified as a storage particle and yet little is known about the specific signaling and mechanistic targets thereof that result in the arrest or the potential resumption of ribosome biogenesis.

Chapter 4 | Biochemical and structural studies of the earliest steps in eukaryotic ribosome biogenesis reveal mechanistic principles in SSU processome formation

Eukaryotic ribosome biogenesis starts with the transcription of the 5' ETS. The 700-nucleotide long spacer region that precedes the 18S rRNA in the primary transcript of the rDNA locus, is bound by eukaryote-specific protein complexes in a co-transcriptional manner. UtpA, UtpB and U3 snoRNP together with multiple other ribosome biogenesis factors initiate the assembly process of the SSU processome. Although studies of the SSU processome have revealed its structure (Kornprobst et al. 2016; Barandun et al. 2017; Sun et al. 2017; Cheng et al. 2017; Chaker-Margot et al. 2017), the highly controlled assembly of the SSU processome is still poorly understood.

Biochemical studies using affinity-tagged, truncated pre-rRNA transcripts mimicking sequential transcriptional stages of the 5' ETS and the ensuing 18S rRNA domains provided an *in vivo* map of the temporal association order of ribosome biogenesis factors during SSU processome formation (Chaker-Margot et al. 2015; Zhang et al. 2016). While the 5' ETS sequence alone recruits more than 27 distinct protein components, the addition of individual 18S rRNA domains leads to the binding of approximately 40 more ribosome assembly factors. Some of these factors associate only transiently with the pre-rRNA transcripts and leave the growing pre-

ribosomal particles at later stages whereas others are incorporated into the SSU processome. Transcription of the 167-nucleotide long 3' minor domain completes the 18S rRNA and not only triggers the dissociation of multiple transient factors but also recruits the highest number of ribosome assembly factors of all 18S rRNA domains. The protein compositions of the solved SSU processome structures (Kornprobst et al. 2016; Barandun et al. 2017; Sun et al. 2017; Cheng et al. 2017; Chaker-Margot et al. 2017) and the pre-ribosomal particles obtained by affinity purification of pre-rRNA transcripts covering the 5' ETS and complete 18S rRNA (Chaker-Margot et al. 2015; Zhang et al. 2016) overlap strongly, suggesting that they represent a similar assembly state. Since the cryo EM reconstructions of the SSU processome, which all represent the same pre-ribosomal particle, are the only available structural snapshots of a highly dynamic process, they do not allow for a mechanistic understanding of the earlier steps in eukaryotic ribosome biogenesis.

To embed the biochemical assembly data (Chaker-Margot et al. 2015; Zhang et al. 2016) in a structural framework and deepen our understanding of the mechanistic principles governing SSU processome formation, we determined the cryo-EM structures of three small subunit assembly stages preceding the SSU processome. We combined these structural studies with tandem affinity purifications and mass spectrometry analysis of isolated 18S rRNA domains to show that *in vivo* the pre-ribosomal context provided by the 5' ETS sequence is not needed for the recruitment of ribosome assembly factors to individual rRNA domains, but that the

presence of the 5' ETS particle is required for SSU processome formation upon completion of the 18S rRNA. The cryo-electron microscopy reconstructions of the early pre-ribosomal particles highlight that the initial independence of rRNA domains from the 5' ETS particle is structurally maintained, which prevents a premature compaction of the rRNA domains into the SSU processome. Furthermore, the compositional characterization of ribonucleoparticles containing the 5' ETS and 18S rRNA with a deleted domain suggests, that SSU processome formation represents a checkpoint for the presence and arrangement of the individual 18S rRNA domains, which facilitates the recruitment of a large number of factors including the GTPase Bms1, RNA-helicase Dhr1, the methyltransferase Emg1 and the acetyltransferase and helicase Kre33.

This unpublished study was a collaborative effort of Jonas Barandun and me. I have established the purifications of the pre-rRNA fragments with their associated proteins and performed all northern blotting experiments. Both of us have performed ribonucleoparticle purifications, prepared negative stain and cryo-EM grids and acquired electron microscopy datasets. Jonas Barandun has processed the cryo-EM data of the highest resolved 5' ETS particle structure and also built the corresponding architectural model. Henrik Molina and Caitlin Stecker performed the mass spectrometry analyses.

Malik Chaker-Margot has kindly provided a yeast strain with a tagged version of the MS2-protein integrated in the genome, which served as a basis for all the strains subsequently made for this study. Albert Antar and Sebastian Klinge have cloned the construct containing the 5' ETS and 18S rRNA with the central domain deleted.

4.1 18S rRNA domains recruit dedicated ribosome biogenesis factors in a 5' ETS-independent manner

Almost all ribosome biogenesis proteins recruited by the 5' ETS are essential. In the SSU processome they function as a structural scaffold for the segregated 18S rRNA domains with which they share multiple interaction sites. While the role of the 5' ETS particle in relation to the individual 18S rRNA domains in the SSU processome has been characterized by structural studies of this particle (Kornprobst et al. 2016; Barandun et al. 2017; Sun et al. 2017; Cheng et al. 2017; Chaker-Margot et al. 2017), its function during SSU processome assembly remains elusive. It is unclear whether the 5' ETS particle is required for the recruitment of rRNA-domain specific biogenesis factors and if its role as a structural mold for the 18S rRNA domains is essential during the earliest steps in eukaryotic ribosome biogenesis.

To determine the dependencies of the 18S rRNA domains during their initial maturation process, 18S-rRNA sequences encompassing isolated and combined domains were expressed without the pre-ribosomal context provided by the 5' ETS (**Figure 4.1**). Individual 18S rRNA-domains and combinations thereof containing a 3'-MS2-aptamer tag were overexpressed in yeast strains harboring a genomically integrated MS2-3C-GFP, which associates with the tagged rRNA *in vivo* (**Figure 4.1 a,b,c**). 18S-rRNA fragment expression levels *in vivo* were assessed by northern blotting analysis of total cellular RNA (**Figure 4.1d**). A degradation product, likely composed of the MS2-loops and additional 3' sequences, is consistently observed

in all rRNA domain samples. However, the majority of the 18S-rRNA fragments are full length including the MS2 loops and, based on their size, the CYC terminator sequence used to control their RNA Polymerase II driven expression. (**Figure 4.1d**).

In addition to the integrated MS2-3C-GFP, these yeast strains harbor a C-terminal streptavidin binding peptide (sbp) on a ribosome biogenesis factor expected to bind the respective expressed rRNA segment (**Figure 4.1a**) based on previous data (Chaker-Margot et al. 2015; Zhang et al. 2016; Kornprobst et al. 2016; Chaker-Margot et al. 2017; Sun et al. 2017; Barandun et al. 2017; Cheng et al. 2017). Esf1 was used as a bait for the 5' domain and the full 18S constructs, Kri1 for the central and 5'- to central domain pulldowns and Mrd1 for the 5'-central-3'-major segment. Tandem affinity purifications via the MS2-aptamers and the sbp-tagged assembly factors permitted the isolation of the truncated 18S-rRNAs and their associated proteins (**Figure 4.1b**). As a control for the purification and expression conditions used, pulldowns with a construct encompassing the 5' ETS and all 18S rRNA domains with Utp10 as a bait were performed. This should yield a particle composed of SSU processome subunits.

Figure 4.1 | Expression and purification strategy of isolated 18S rRNA domains. **[a]** Schematic overview of the genetic components present in yeast strains used to purify 18S rRNA domains and their bound proteins. rRNA fragments (light-blue) tagged at their 3' ends with five MS2-aptamer stem-loops (beige) are expressed from a plasmid. A galactose inducible promoter (Gal1) and a CYC terminator are used to control the RNA polymerase II driven expression of the rRNA sequences. A single galactose inducible copy of the GFP tagged MS2 protein (green, pink) is integrated in the yeast genome. Endogenous ribosome assembly factors (dark-blue) are tagged with a streptavidin binding peptide (sbp) (orange). **[b]** *In vivo* the galactose induced rRNA fragments recruit endogenous ribosome assembly proteins and the MS2-3C-GFP dimers. The RNP is immobilized on anti-GFP nanobody (light-green) covered beads through its MS2-3C-GFP moiety. 3C-protease cleavage elutes the particle which is subjected to a second affinity purification step on streptavidin (light orange) coated beads. Biotin competes with the sbp-tag for binding to streptavidin and thereby elutes the RNP. **[c]** A schematic of the 5' ETS (black) and 18S rRNA (grey) is shown with its cleavage sites indicated. The MS2 (beige) tagged constructs used in pulldowns of individual 18S rRNA domains (5' domain in pink, central domain in cyan, 3' major domain in blue and 3' minor domain in dark blue) are indicated below with nucleotide positions labeled. **[d]** Northern blotting analysis of total RNA extracted from yeast strains expressing the constructs in [d] using a MS2-aptamer binding probe. An approximate size ladder is indicated on the left and an Asterix on the right labels a common degradation product.

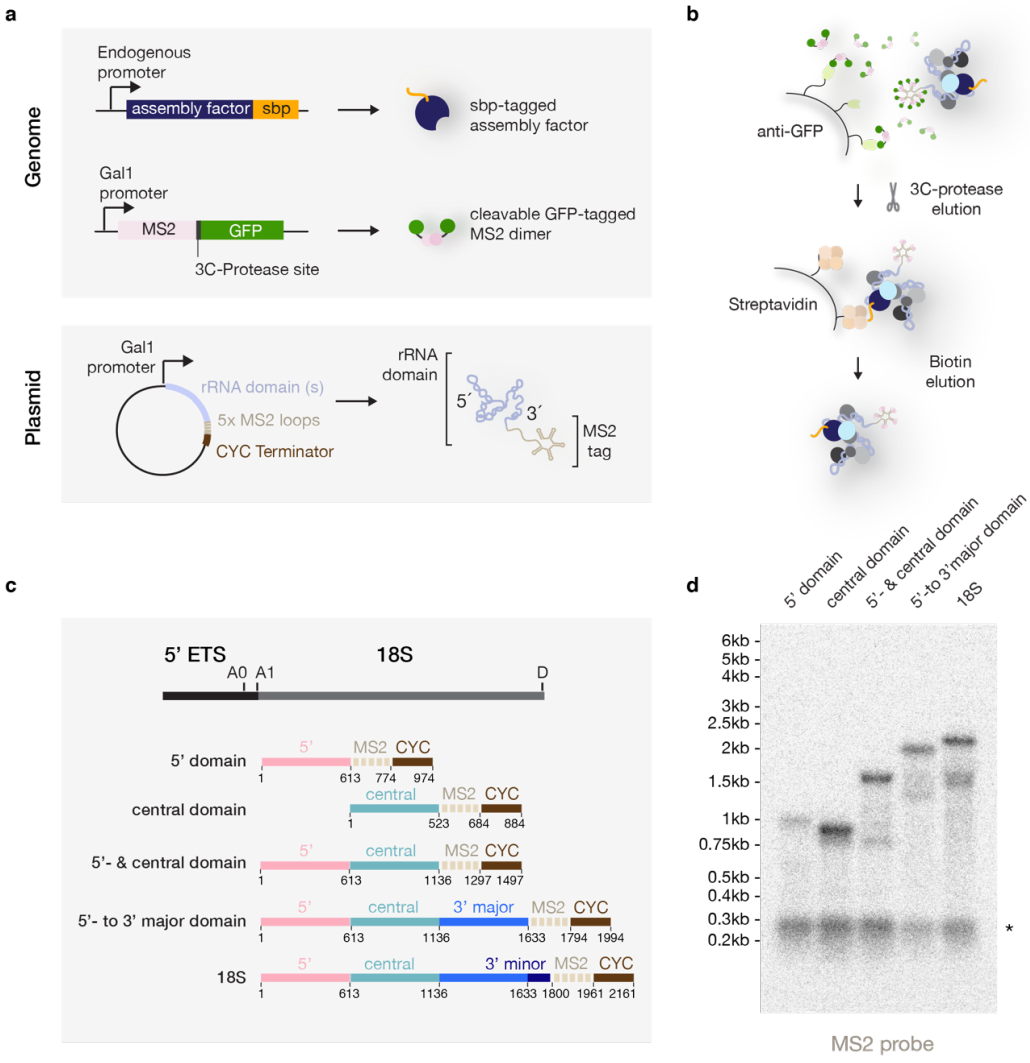
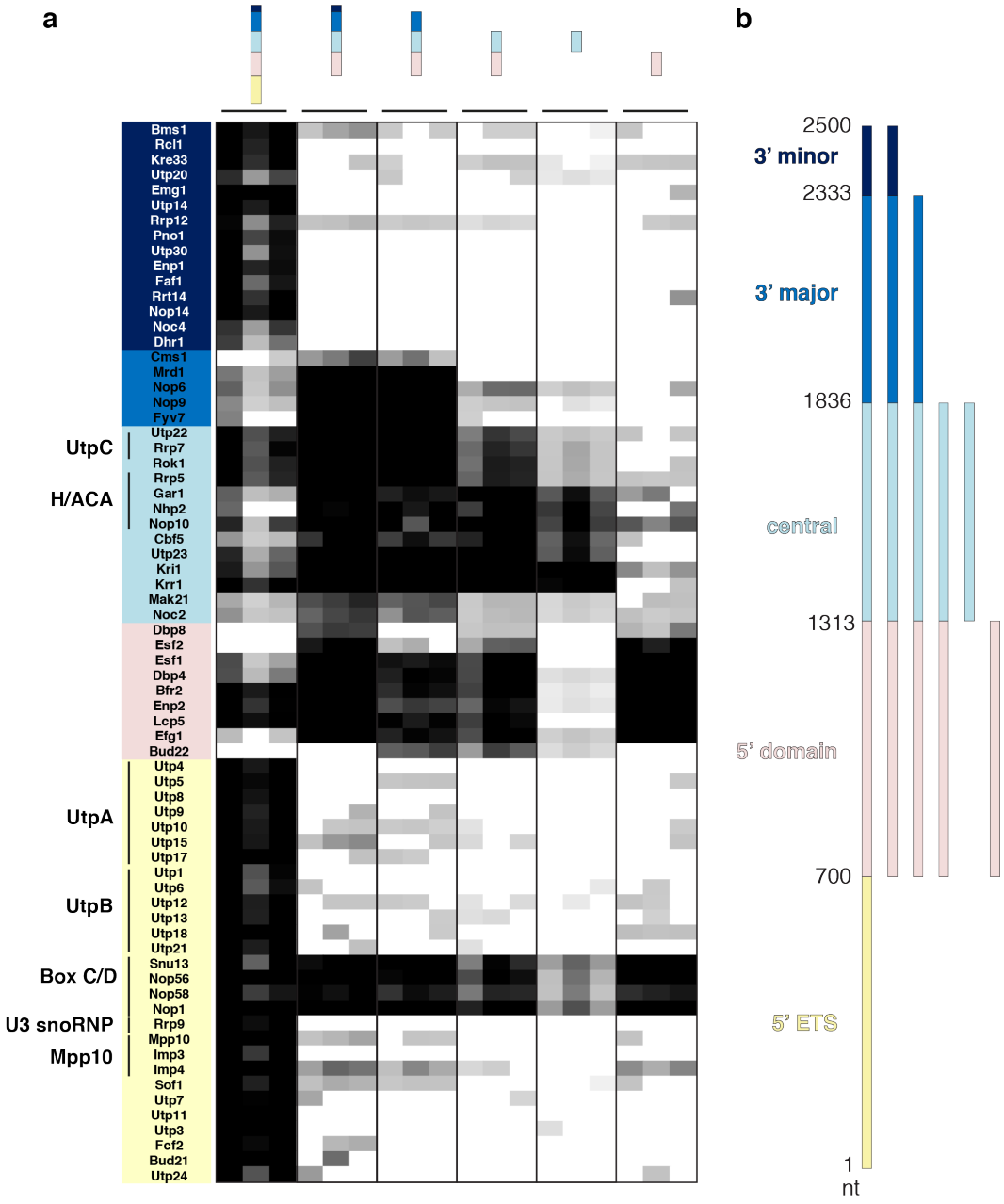

a
Genome
Endogenous promoter
assembly factor
sbp
sbp-tagged assembly factor
Gal1 promoter
MS2
GFP
3C-Protease site
cleavable GFP-tagged MS2 dimer
Plasmid
Gal1 promoter
rRNA domain (s)
5x MS2 loops
CYC Terminator
rRNA domain
5´
3´
MS2 tag
b
anti-GFP
3C-protease elution
Streptavidin
Biotin elution
c
5' ETS
18S
A0
A1
D
5' domain
central domain
5'- & central domain
5'- to 3' major domain
18S
MS2
CYC
3' major
3' minor
d
5' domain
central domain
5'- & central domain
5'-to 3'major domain
18S
6kb
5kb
4kb
3kb
2.5kb
2kb
1.5kb
1kb
0.75kb
0.5kb
0.4kb
0.3kb
0.2kb
*
MS2 probe

Figure 4.2 | 18S rRNA domains recruit ribosome assembly factors independent of the 5' ETS.

[a,b] Ribosome biogenesis factors identified by *in solution* mass spectrometry analysis of purified RNPs [a]. Truncated pre-rRNA constructs used as baits for the analyzed pulldowns are shown as schematic drawings with their nucleotide numbers and domain boundaries (5' ETS, 5' domain, central domain, 3' major domain, 3' minor domain) indicated in [b]. Identified ribosome assembly factors are listed on the left in [a] and ordered and color-coded according to their previously determined recruitment order in the presence of the 5' ETS (Chaker-Margot et al. 2015; Zhang et al. 2016). A heat map of protein abundances in all pulldowns was generated where black means high abundance, shades of grey indicate lower abundance and white boxes mark absent components. RNP purifications were carried out in triplicates.


a
b
Bms1
Rcl1
Kre33
Utp20
Emg1
Utp14
Rrp12
Pno1
Utp30
Enp1
Faf1
Rrt14
Nop14
Noc4
Dhr1
Cms1
Mrd1
Nop6
Nop9
Fyv7
UtpC
Utp22
Rrp7
Rok1
H/ACA
Rrp5
Gar1
Nhp2
Nop10
Cbf5
Utp23
Kri1
Krr1
Mak21
Noc2
Dbp8
Esf2
Esf1
Dbp4
Bfr2
Enp2
Lcp5
Efg1
Bud22
UtpA
Utp4
Utp5
Utp8
Utp9
Utp10
Utp15
Utp17
UtpB
Utp1
Utp6
Utp12
Utp13
Utp18
Utp21
Box C/D
Snu13
Nop56
Nop58
U3 snoRNP
Nop1
Rrp9
Mpp10
Mpp10
Imp3
Imp4
Sof1
Utp7
Utp11
Utp3
Fcf2
Bud21
Utp24
2500
3' minor
2333
3' major
1836
central
1313
5' domain
700
5' ETS
1
nt

All purifications were carried out in technical triplicates and the protein compositions of the purified RNPs were analyzed by *in solution* mass spectrometry. Protein levels in each experiment were normalized using the MS2-protein value of the respective purification to account for varying rRNA fragment expression levels (**Figure 4.1d**). The control pre-rRNA construct containing the 5' ETS and 18S rRNA recruited the expected SSU processome components, indicating that the used purification conditions yielded similar results as previously reported (Chaker-Margot et al. 2015; Zhang et al. 2016).

Surprisingly, the isolated 18S rRNA domains were not only bound by ribosome assembly factors stably associating with the respective domain in the SSU processome but also by transient factors previously reported to bind in the context of transcriptional stages including the 5' ETS (**Figure 4.2, see also Figure 1.6 for comparison**).

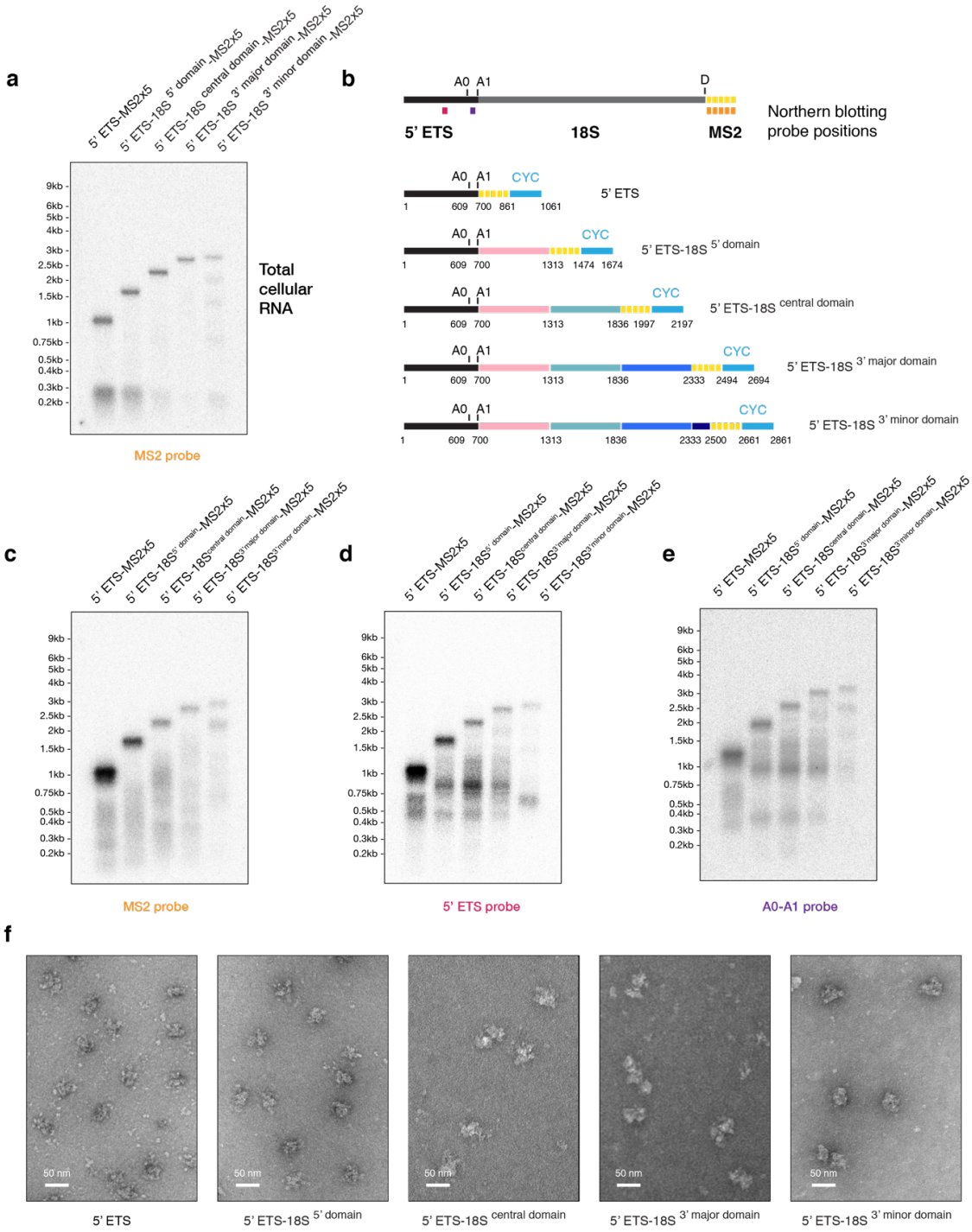
In addition to domain-specific ribosome assembly proteins, the isolated 18S rRNA domains were associated with distinct snoRNAs. The common box C/D snoRNP core proteins (Nop1, Nop56, Nop58, Nop1) were identified in all pre-rRNA fragment purifications (**Figure 4.2**). Yet the U3 snoRNA specific Rrp9 was only co-purified with the 5' ETS-containing control construct, indicating the presence of additional box C/D snoRNPs in the isolated individual 18S rRNA domain samples. The 5' domain could possibly be bound by the transiently associated U14 (Zhang et

al. 2016), a box C/D snoRNA that has been shown to base-pair with sequences in the 5' domain (Liang & Fournier 1995). The central domain RNP was associated with the common box C/D snoRNA proteins as well as the common H/ACA snoRNP subunits (Nhp2, Gar1, Nop10, Cbf5) (**Figure 4.2a**). snR30, a 600-nucleotide long RNA, base-pairs with expansion segment 6 (ES6) in the central domain and associated with constructs spanning the 5' ETS and central domain (Zhang et al. 2016).

While the expression of isolated 18S rRNA sequences up to the 3' major domain lead to the independent association of specific ribosome assembly factors, the expression of all four 18S rRNA domains without the 5' ETS failed to co-purify proteins associating at the 3' minor stage and factors shown to dissociate from the pre-ribosomal particle upon SSU processome formation continued to be bound to the 18S rRNA (**Figure 4.2a**). These findings suggest that the sequence information of the 18S rRNA is sufficient to recruit initial 18S rRNA-domain specific maturation factors in a non-hierarchical, modular and 5' ETS-independent manner, but that the presence of the 5' ETS particle and the 3' minor domain are required for the formation of the SSU processome and the dissociation of transient factors.

Figure 4.3 | Biochemical characterization of small subunit assembly stages preceding SSU processome formation.

[a] Northern blotting analysis of total cellular RNA extracted from yeast cells overexpressing truncated, MS2 tagged pre-rRNA constructs mimicking transcription of the pre-18S rRNA gene using an MS2 aptamer complementary probe. **[b]** Schematic representation of the pre-rRNA constructs expressed in [a] and used for purification of assembly stages preceding the SSU processome. On top the positions of the northern blotting probes used in [a, c-e] are indicated and color-coded on a full length pre-rRNA locus. rRNA processing sites in the 5' ETS (A0, A1) and on the 18S (D) are shown above each drawing if applicable. Nucleotide numbers are labeled below the border of each pre-rRNA domain (5' ETS, 5' domain, central domain, 3' major domain, 3' minor domain), cleavage site and design element (MS2 tags, CYC terminator). **[c-e]** Comparative northern blotting analysis of RNA extracted from purified pre-ribosomal particles. The particles were isolated from the cells expressing the MS2-tagged RNA constructs in [a] using the purification strategy shown in (Figure 4.1). Probes used in each blot are indicated below and color-coded according to [b]. **[f]** Purified pre-ribosomal particles containing the respective pre-rRNA constructs shown in [b] were stained with 2% uranyl acetate and imaged at a magnification of 39'000x. An approximate size bar is shown in the lower left corner of each cropped micrograph.

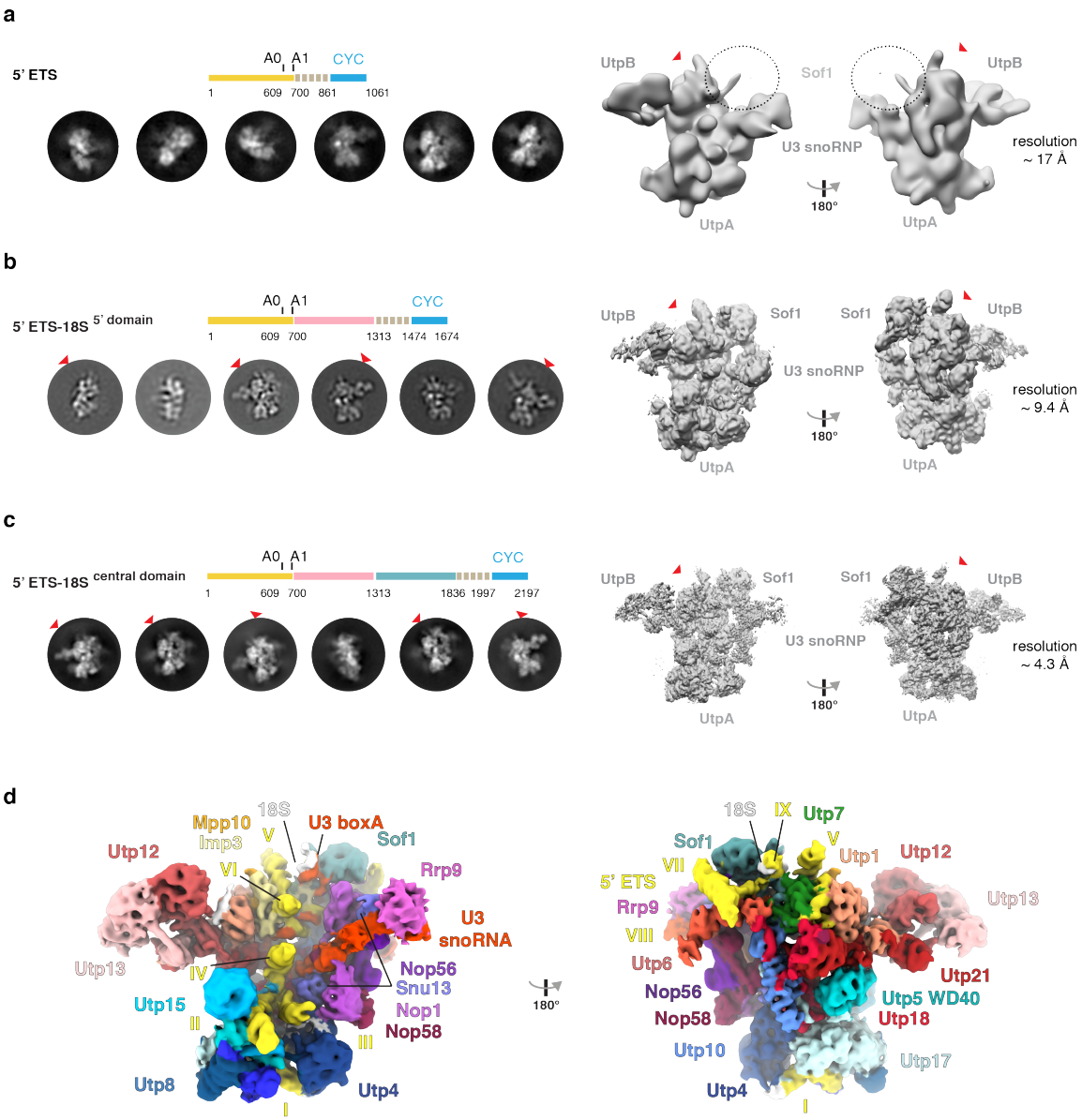

a
5' ETS-MS2x5
5' ETS-18S 5' domain-MS2x5
5' ETS-18S central domain-MS2x5
5' ETS-18S 3' major domain-MS2x5
5' ETS-18S 3' minor domain-MS2x5
9kb
6kb
5kb
4kb
3kb
2.5kb
2kb
1.5kb
1kb
0.75kb
0.5kb
0.4kb
0.3kb
0.2kb
Total cellular RNA
MS2 probe
b
A0 A1
D
5' ETS
18S
MS2
Northern blotting probe positions
CYC
5' ETS
1 609 700 861 1061
5' ETS-18S 5' domain
1 609 700 1313 1474 1674
5' ETS-18S central domain
1 609 700 1313 1836 1997 2197
5' ETS-18S 3' major domain
1 609 700 1313 1836 2333 2494 2694
5' ETS-18S 3' minor domain
1 609 700 1313 1836 2333 2500 2661 2861
c
5' ETS-MS2x5
5' ETS-18S 5' domain-MS2x5
5' ETS-18S central domain-MS2x5
5' ETS-18S 3' major domain-MS2x5
5' ETS-18S 3' minor domain-MS2x5
MS2 probe
d
5' ETS probe
e
A0-A1 probe
f
50 nm
5' ETS
5' ETS-18S 5' domain
5' ETS-18S central domain
5' ETS-18S 3' major domain
5' ETS-18S 3' minor domain

4.2 Initial biochemical and structural characterization of transcriptional stages preceding the SSU processome

The critical role of the 5' ETS particle for SSU processome formation prompted us to structurally characterize it in the context of several of the nucleolar precursors. Pre-ribosomal particles mimicking the transcriptional stages preceding and including SSU processome formation were purified using the strategy outlined for the purification of the individual 18S rRNA domains (**Figure 4.1** and **Figure 4.3**) (Chaker-Margot et al. 2015). Protein compositions of these particles were analyzed by mass spectrometry (not shown). Comparison with the previously published components list of these particles (Chaker-Margot et al. 2015), confirmed the robustness of the purifications and the presence of all expected ribosome assembly factors. To ensure the integrity of the overexpressed pre-rRNA fragments, the RNA content of the purified samples, as well as total RNA from cells expressing the tagged constructs, were analyzed by northern blotting (**Figure 4.3 a-e**). The majority of the overexpressed pre-rRNA constructs are present as full-length, uncut products *in vivo* (**Figure 4.3a**). Only in cells expressing the pre-rRNA encompassing the 5' ETS and all four 18S rRNA domains, a distinct degradation or processing pattern can be observed (**Figure 4.3a**). A small (200-300 nucleotide) degradation product, likely containing the MS2 loops and the CYC-terminator sequence, is also discernible. Northern blots of RNA extracted from the RNP samples after tandem affinity purification, indicated that degradation during isolation was minor and the majority of the pre-rRNA is intact in the purified RNPs (**Figure 4.3 a,c**).

Figure 4.4 | cryo-EM analysis of the 5' ETS particle during different transcriptional stages.

[a-c] 2D class averages and 3D reconstructions obtained from cryo-EM datasets of the 5' ETS alone [a], 5' ETS with 5' domain [b], and the 5' ETS with the 5'- and central domains [c]. A schematic drawing above the 2D class averages represents the pre-rRNA content of the particles. A red arrowhead indicates the position on the 5' ETS particle in the 2D class averages and on the 3D reconstructions where additional density for the 18S rRNA domains would be expected. Estimated resolutions of all three cryo-EM reconstructions are indicated as well as the position of architectural landmark proteins. **[d]** The cryo-EM density of the 5' ETS-$^{\text{central domain}}$ particle (4.4 Å) [c], is segmented and color-coded according to its protein and RNA components. UtpA subunits are shown in shades of blue, UtpB subunits in shades of red, U3 snoRNP proteins in shades of purple and the U3 snoRNA in bright red. The 5' ETS helices are numbered with roman numerals in yellow.


a
5' ETS
A0
A1
CYC
1
609
700
861
1061
UtpB
Sof1
UtpB
U3 snoRNP
resolution
~ 17 Å
180°
UtpA
UtpA
b
5' ETS-18S
5' domain
A0
A1
CYC
1
609
700
1313
1474
1674
UtpB
Sof1
Sof1
UtpB
U3 snoRNP
resolution
~ 9.4 Å
180°
UtpA
UtpA
c
5' ETS-18S
central domain
A0
A1
CYC
1
609
700
1313
1836
1997
2197
UtpB
Sof1
Sof1
UtpB
U3 snoRNP
resolution
~ 4.3 Å
180°
UtpA
UtpA
d
Utp12
Mpp10
Imp3
18S
V
U3 boxA
Sof1
VI
Rrp9
U3
snoRNA
Utp13
IV
Nop56
Snu13
Utp15
II
Nop1
Nop58
III
Utp8
Utp4
I
180°
18S
IX
Utp7
V
Sof1
VII
Utp1
Utp12
5' ETS
Utp13
Rrp9
VIII
Utp6
Utp21
Nop56
Utp5 WD40
Nop58
Utp18
Utp10
Utp17
Utp4
I

The isolated particles were subjected to a limited negative stain EM analysis (**Figure 4.3f**). While visualization of the particle containing the 5' ETS and the complete 18S rRNA (5' ETS-18S $^{3'minor}$) revealed a compacted structure with SSU processome features, the particles of the early transcription stages (5' ETS, 5' ETS-18S$^{5'domain}$, 5' ETS-18S^{central}, 5' ETS-18S$^{3'major}$) adopted flexible and open architectures.

The 5' ETS, 5' ETS-18S$^{5'\ domain}$ and 5' ETS-18S ^{central} particles were further analyzed by cryo-EM (**Figure 4.4**). 2D class averages for all three stages showed similar views with characteristics of the 5' ETS moiety of the particle (**Figure 4.4 a,b,c**). Despite the rRNA being intact (**Figure 4.3c**) and the clear presence of 5'- and central domain specific assembly proteins in the sample (**Appendix 7.3**), structured density for the 5' and central domain was not visible in the cryo-EM 2D averages (**Figure 4.4 a,b,c**). This suggests, that they are either unfolded or highly flexible. 3D classification and refinement led to reconstructions at ~17 Å (5' ETS), ~9.4 Å (5' ETS-18S $^{5'\ domain}$) and 4.4 Å (5' ETS-18S ^{central}) of the 5' ETS particle respectively (**Figure 4.4** and **Figure 7.1**). The lower resolutions of the reconstructions from the 5' ETS and 5' ETS-18S $^{5'\ domain}$ particles may be a result of lower numbers of micrographs (see **6.26**), resulting lower particle numbers, the quality of the datasets and probable conformational flexibility.

As expected from the 2D class averages, and consistent with the notion that the 18S rRNA domains are initially biochemically independent modules, ordered

density could not be observed for the 18S rRNA domains in the 5' ETS-18S$^{5'domain}$ and 5' ETS-18S^{central} structures. All reconstructions of the 5' ETS particle during the three maturation stages are highly similar and the 4.4 Å 5' ETS-18S^{central} density was used to interpret the particle and to place, fit and adjust previously built models of individual components (**Figure 4.4d**) (Barandun et al. 2017). This cryo-EM reconstruction also forms the basis for the analysis of the ensuing section.

4.3 The functional architecture of the 5' ETS particle during early transcription events facilitates 18S rRNA domain independence

Surprisingly, the 5' ETS particle and its protein components adopt conformations significantly different from their SSU processome states. In particular UtpB assumes an unexpected retracted conformation during these early stages of ribosome biogenesis (**Figure 4.5**). In the 5' ETS particle the two tandem β-propellers of Utp12 and Utp13 are stacked upon each other, close to the tetrameric α-helical bundle formed by their CTDs and the Utp1 and Utp21 CTDs (**Figure 4.5a**). In the SSU processome the CTD tetramer and the β-propellers of Utp12 and Utp13 adopt a fanned-out structure and form interaction sites with late factors (Pno1, Rcl1) and the 3' minor and major domain RNAs themselves (H27, H44, H45) (**Figure 4.5b**).

Helix 45 of the 3' minor domain is positioned on top of Pno1 and chaperoned by UtpB and Mpp10 in the SSU processome. In the 5' ETS particle the bipartite binding site of Pno1 on UtpB is not formed, thus inhibiting premature stabilization of

Pno1 (**Figure 4.5 c,d**). An almost 90-degree rotation of the tetrameric CTDs with respect to the β-propellers of Utp21 is needed for the Pno1-binding loop of Utp21 to be layered on top of the β-propeller of Utp1 and thus enable Pno1 and helix 45 stabilization.

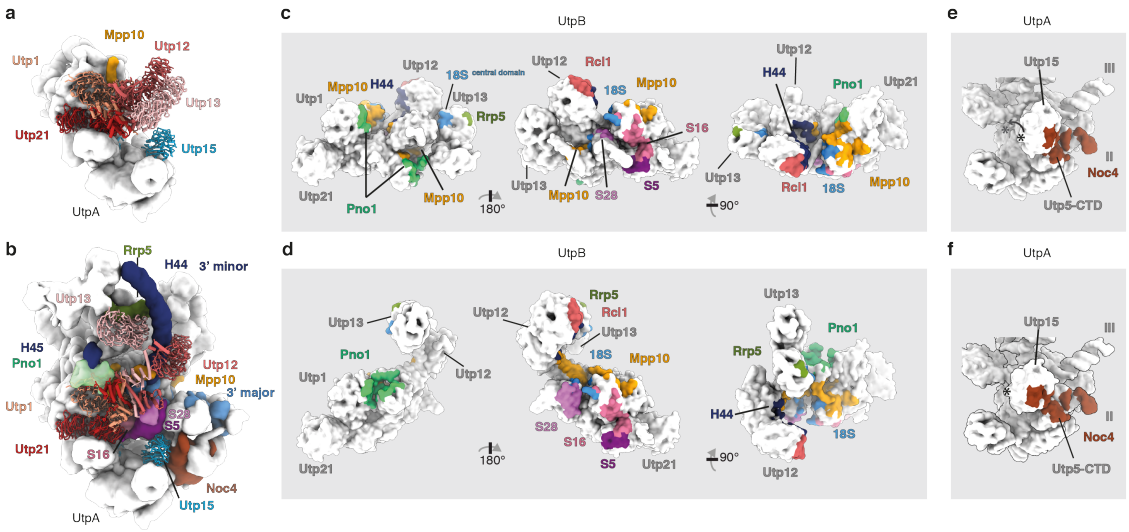


Figure 4.5 | A conformational switch in UtpB prevents premature exposure of late factor binding-sites.

[a] UtpB (as cartoon, shades of red) is in a retracted state in the 5' ETS particle (as surface, white). Individual subunits of UtpB are labeled and color-coded. The helix of Mpp10 resolved in the 5' ETS particle is shown as orange surface. Utp15, the only UtpA subunit, positioned differently in the 5' ETS particle is shown as blue cartoon. **[b]** In the SSU processome (as surface, white) UtpB (as cartoon, shades of red) and Utp15 (as cartoon, blue) interact with ribosome assembly factors (Pno1, Noc4, Rrp5, Mpp10, color-coded), 18S rRNA elements (3' minor: dark-blue, 3' major and central domain: light blue, helices labeled with H) and ribosomal proteins (S5, S16, S25, in shades of purple). **[c, d]** Three views of UtpB (as surface, white) in its retracted state as observed in the 5' ETS particle [c] and in the SSU processome conformation [d]. Sequence elements of UtpB binding to factors in the SSU processome are colored and labeled. **[e, f]** The β-propeller of Utp15 (white) partially occludes the binding site of Noc4 (dark brown) in the 5' ETS particle [e]. Two Asterix and an arrow indicate the position of Utp15 in the SSU processome as seen in [f]. Roman numerals indicate 5' ETS helices (white) and the CTD of Utp5 (white) is labeled.

This 90-degree rotation of the CTD-tetramer also elongates the binding interface of Mpp10 on UtpB. UtpB is an important platform for the 5' ETS-bound Mpp10 complex (Mpp10, Imp3, Imp4). In the 5' ETS particle, as in the SSU processome, Imp3 and an α-helix of Mpp10 are placed between Utp7 and Utp1 to stabilize the 3' hinge of the U3 snoRNA and 5' ETS (**Figure 4.4d**).The flexible and 593 residue-long Mpp10 harbors multiple conserved binding sites and spans from the 3' hinge in the back of the SSU processome to the front where it associates with Bms1 (**Figure 3.19**). Binding sites on UtpB used by Mpp10 along its path in the SSU processome are set apart and partially occluded in the 5' ETS particle due to the rotated state of the CTDs and placement of the tandem β-propellers (**Figure 4.5 a-d**). The coordinated interaction of Mpp10 and UtpB facilitates the stabilization of the remodeled bases of helix 44 and 45 and a long single stranded RNA linker connecting these distantly located helices in the SSU processome (**Figure 3.13c**).

In the SSU processome the 3' minor domain helix 44 is cradled between the two β-propellers of Utp12 (**Figure 3.10f**). There Utp12 also contacts the proximally located remodeled helix 27 and thereby coordinates the beginning of both, the 3' major- and 3' minor domain, on the highly conserved surface of its tandem β-propeller (Barandun et al. 2018). In addition to these RNA binding sites Utp12 also interacts with Rcl1, the co-factor of the GTPase Bms1. The Rcl1-Bms1 complex is only recruited upon completion of the 18S rRNA gene (Chaker-Margot et al. 2015; Zhang et al. 2016). The RNA and Rcl1 binding sites on Utp12 are accessible in the

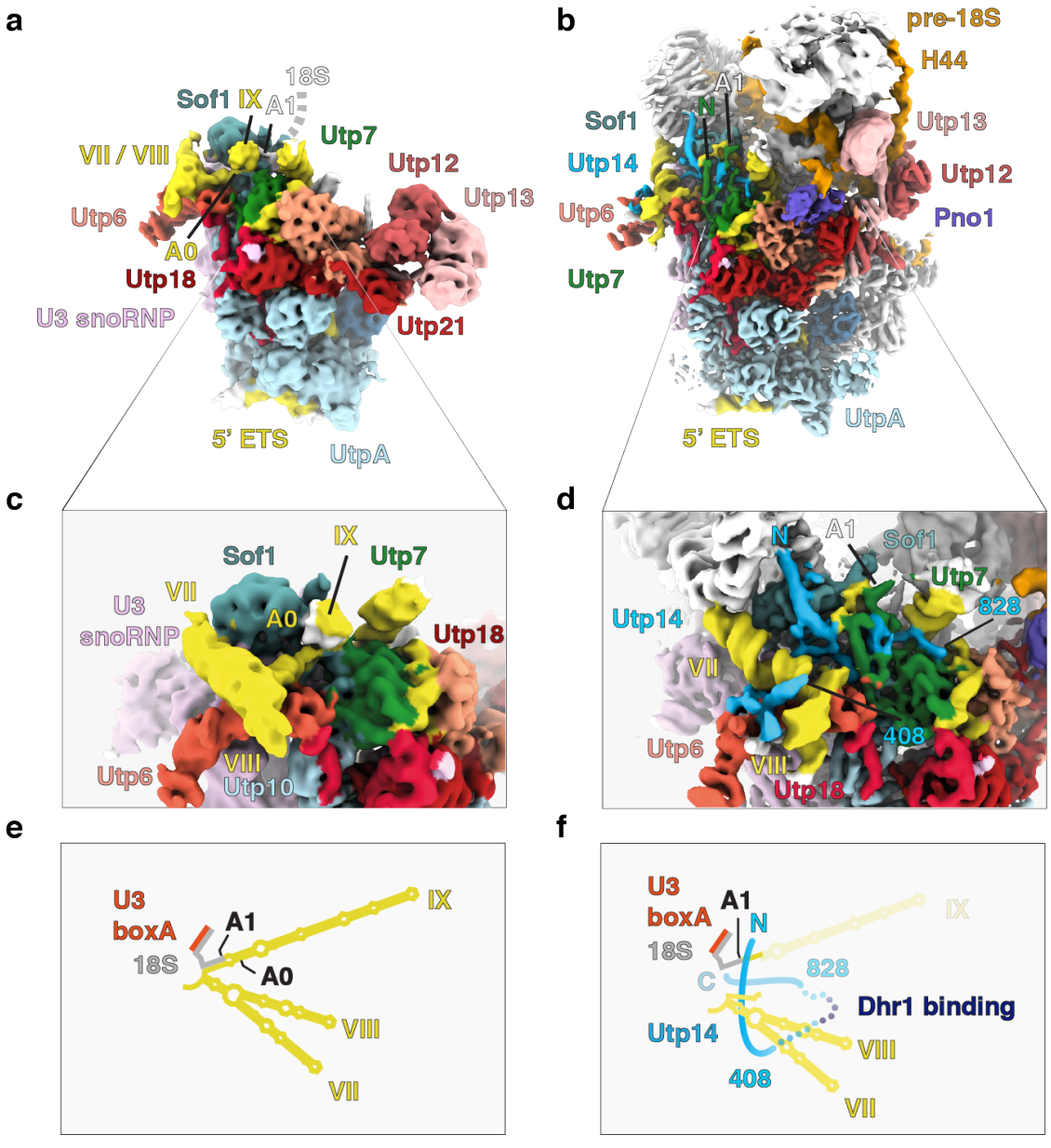
5’ ETS particle, thus making it possible for Utp12 to initially recognize helix 44. A long flexible linker between the CTD and the β-propeller of Utp12 would allow sampling of the rRNA transcription status (**Figure 4.5 c,d**). Binding of Utp12 to all three of its SSU processome interaction partners could potentially help to displace the 3’ major domain-recruited transient assembly factors Mrd1 and Nop9, which have been shown to bind either to helix 27 or shortly downstream (Segerstolpe et al. 2013; Wang & Ye 2016). In contrast to Utp12, the small central-domain-binding site of Utp13 is facing towards the solvent in the 5’ ETS particle and has to undergo a dramatic repositioning to adopt the SSU processome conformation (**Figure 4.5 c,d**). Aside from contacting the central domain, Utp13 binds ITS1 sequences *in vivo* (**Figure 2.7a**).

Taken together, the retracted conformation of UtpB during early stages of transcription prevents the pre-mature stabilization of 3’ minor factors. However, the binding site for the first helix of the 3’ minor domain, helix 44, is accessible on Utp12. This interaction could trigger displacement of the WD40 domains of Utp12 followed by a rotation of the tetrameric CTD domain, which leads to the repositioning of Utp13 close to the central domain and ITS1 as well as the formation of the contact sites for Pno1, Mpp10 and the 3’ major ribosomal proteins rpS16, rpS28 and rpS5 (**Figure 4.5 c,d**).

In contrast to UtpB, the conformation and positioning of UtpA in the early 5' ETS particles is similar to the SSU processome (**Figure 4.5 a,b**). Only Utp15, which binds helix II of the 5' ETS is shifted in the absence of the 3' major domain and the late binding proteins Noc4 and Nop14 (**Figure 4.5 e,f**). Even though Noc4 and Nop14 bind directly to the 3' major domain they are only recruited upon 3' minor domain transcription. The β-propeller of Utp15 covers part of the Noc4 binding site on the CTD of Utp5 and helix II of the 5' ETS (**Figure 4.5e**). Upon SSU processome formation Nop14 and Noc4 likely displace Utp15 (**Figure 4.5f**).

Similar to most of the protein components, the 5' ETS pre-rRNA adopts an SSU processome-like structure in the 5' ETS particle. While the majority of its helices are formed but slightly more flexible than in the SSU processome, two differences in the 5' ETS particle were observed. Helix VI, which serves as a platform for Bms1, Utp11 and Fcf2 (**Figure 3.20**), is pointing upwards into the solvent in the 5' ETS particle (**Figure 4.4d**) whereas in the SSU processome it points downwards and contacts helix IV. Upon compaction during SSU processome formation this helix is presumably locked down by the positioning of the 5' domain and Bms1.

Figure 4.6 | The intact A0 site of the 5' ETS occupies the Utp14 binding site.
[a] Segmented cryo-EM density map of the 5' ETS particle. UtpA subunits are shown in light blue, U3 snoRNP in light-purple, UtpB components in shades of red and the 5' ETS in yellow with helices labeled with roman numerals. The A1 and A0 sites are indicated. **[b]** Segmented and low-pass filtered density of the SSU processome with selected subunits and complexes colored as in [a]. SSU processome specific factors such as Utp14 (blue) and Pno1 (purple) as well as RNA elements of the 18S rRNA (orange) are colored and labeled. **[c]** Zoomed in section of [a] showing the intact A0 site on the 5' ETS particle. **[d]** Zoomed in section of [b] highlighting the binding site of Utp14 in the SSU processome. **[e]** Secondary structure diagram of the RNA components seen in [c]. **[f]** Secondary structure diagram of RNA components seen in [d] with the unusual split structure adopted by Utp14 schematically depicted. The Dhr1 binding sequence of Utp14 is located on a flexible part of the protein and not resolved in the SSU processome structure.


a
18S
Sof1 IX
A1
Utp7
VII / VIII
Utp12
Utp13
Utp6
A0
Utp18
U3 snoRNP
Utp21
5' ETS
UtpA
b
pre-18S
H44
A1
N
Sof1
Utp13
Utp14
Utp12
Utp6
Pno1
Utp7
UtpA
5' ETS
c
Sof1
IX
Utp7
VII
U3
snoRNP
A0
Utp18
Utp6
VIII
Utp10
d
N
A1
Sof1
Utp7
828
Utp14
VII
408
Utp6
VIII
Utp18
e
U3
boxA
18S
A1
A0
IX
VIII
VII
f
U3
boxA
18S
A1
N
C
828
Dhr1 binding
Utp14
VIII
408
VII
IX

The second structural difference in the 5' ETS pre-rRNA during early transcription stages is a result of its processing state. In the density maps of the 5' ETS particle an RNA linker connecting the end of helix VIII with the base of helix IX, which was not resolved in the SSU processome structure, was observed (**Figure 4.6**). The 5' ETS in the SSU processome is cleaved at site A0 (**Figure 3.2**) (**Figure 4.6 b,d,f**), which results in a flexibly attached and unresolved helix IX. Hence, this linker likely leads to the uncut A0 site at the base of helix IX. To confirm that the 5' ETS is indeed uncut during pre-SSU processome transcription stages, comparative northern blotting analysis was performed.

Analyzing the processing state of the pre-rRNAs expressed for the purification of the assembly stages preceding SSU processome formation *in vivo* has shown that the constructs spanning up to the 3' major domain of the 18S rRNA were full length (**Figure 4.3 a,b**). This suggests, that the A0 site indeed remains intact during the early stages of transcription *in vivo* and that the observed additional density in the 5' ETS particle structure contains the A0 site. To confirm that the A0-site remains intact during purification we analyzed the pre-rRNA of the isolated RNPs with probes complementary to the MS2-loops, 5' ETS and the sequence segment between the A0 and A1 sites (**Figure 4.3 c,d,e**). As stated before, despite some degradation artefacts, the pre-rRNAs remain largely intact during purification (**Figure 4.3 a,c**).

Blotting with the 5' ETS probe showed that the majority of the 5' ETS sequences from the transcription states preceding SSU processome formation were running at the full-length transcript size suggesting they are unprocessed. However, a band around ~780 nucleotides was observed in the 5', central and 3' major constructs which probably represents a clipping product within the 18S sequence containing the 700-nucleotide long 5' ETS (**Figure 4.3 d**).

Even though the four pre-rRNA constructs encompassing the 5' ETS and 18S rRNA domains up to the 3' major domain were unprocessed *in vivo* and not site specifically cut but rather unspecifically degraded during purification, the pre-rRNA containing the complete 18S rRNA gene (3' minor) showed a distinct pattern of smaller MS2-tagged RNAs *in vivo* (**Figure 4.3 a,b**). This could indicate either unspecific degradation targeting only the complete pre-rRNA transcript or specific processing of the precursor RNA after the complete transcription of the 18S rRNA.

Consistent with the observations that the protein composition of this RNP was almost identical to the protein composition of the SSU processome and that the particle adopted a compacted SSU processome-like structure under negative stain EM (**Figure 4.3f**), A0 cleavage is occurring in this construct (**Figure 4.3 a,c,d,e**). In MS2- probe northern blots of the purified particles, a band which is ~600 nucleotide shorter than the full-length construct can be observed (**Figure 4.3 c**).

This band is already present *in vivo* (**Figure 4.3 a**) but gets enriched during purification as the 3' MS2 loops as well as the 5' located Utp10 are used as baits. When blotting with the 5' ETS probe this sample shows a characteristic band at ~600 nucleotides (**Figure 4.3 d**), which is not recognized by the A0-A1 probe (**Figure 4.3 e**). Hence confirming that the 600-nucleotide band is the A0 cut 5' ETS.

The RNA linker connecting helix VIII and IX runs in between Sof1, Utp6 and Utp7 in the 5' ETS particle (**Figure 4.6 a,c**). In the SSU processome this region is occupied by the late binding factor Utp14 (**Figure 4.6 b,d** and **Figure 3.11b)**. The uncut A0 site during early transcription stages sterically hinders premature Utp14 binding. Either the A0 site has to be processed in order for Utp14 to bind or the pre-rRNA has to adopt a different conformation to accommodate the N- and C-terminus of Utp14. Integration of Utp14 into the SSU processome is likely accompanied by the recruitment of its binding partner, the RNA helicase Dhr1. Dhr1 unwinds the U3 snoRNA duplexes (Sardana et al. 2015).

In summary, the cryo-EM analysis of three assembly stages preceding the SSU processome and the RNA analysis of these constructs has shown that A0-cleavage in the 5' ETS only occurs after transcription of the 18S rRNA gene is completed and that the independent maturation of the 18S rRNA domains is structurally maintained until conformational changes in 5' ETS protein components and the A0-site support the recruitment of 3' minor factors such as Utp14-Dhr1, Noc4-Nop14, Bms1-Rcl1

and Pno1. Thus, SSU processome formation might represent the first inter-domain checkpoint during the early stages of small subunit assembly.

4.4 The presence of all pre-rRNA domains is read out during SSU processome formation

While each individual 18S rRNA domain can recruit specific ribosome assembly factors independently, the presence of the 5' ETS particle and the 3' minor domain are needed for the binding of 15 proteins associated with the SSU processome. To assess whether only the presence of the first and last pre-rRNA domain is sensed during SSU processome formation or if the assembly of all domains is needed, an RNA fragment with a deleted 18S rRNA domain was constructed.

The 5' ETS bound U3 snoRNA is a central organizer within the SSU processome and base-pairs with segments of the pre-18S rRNA. The RNA duplexes formed by U3 snoRNA and the pre-18S rRNA (Box A, Box A'), do not only prevent premature central pseudoknot formation (**Figure 3.13 a,b**), but also define the positions of the 18S rRNA domains within segregated compartments of the particle.

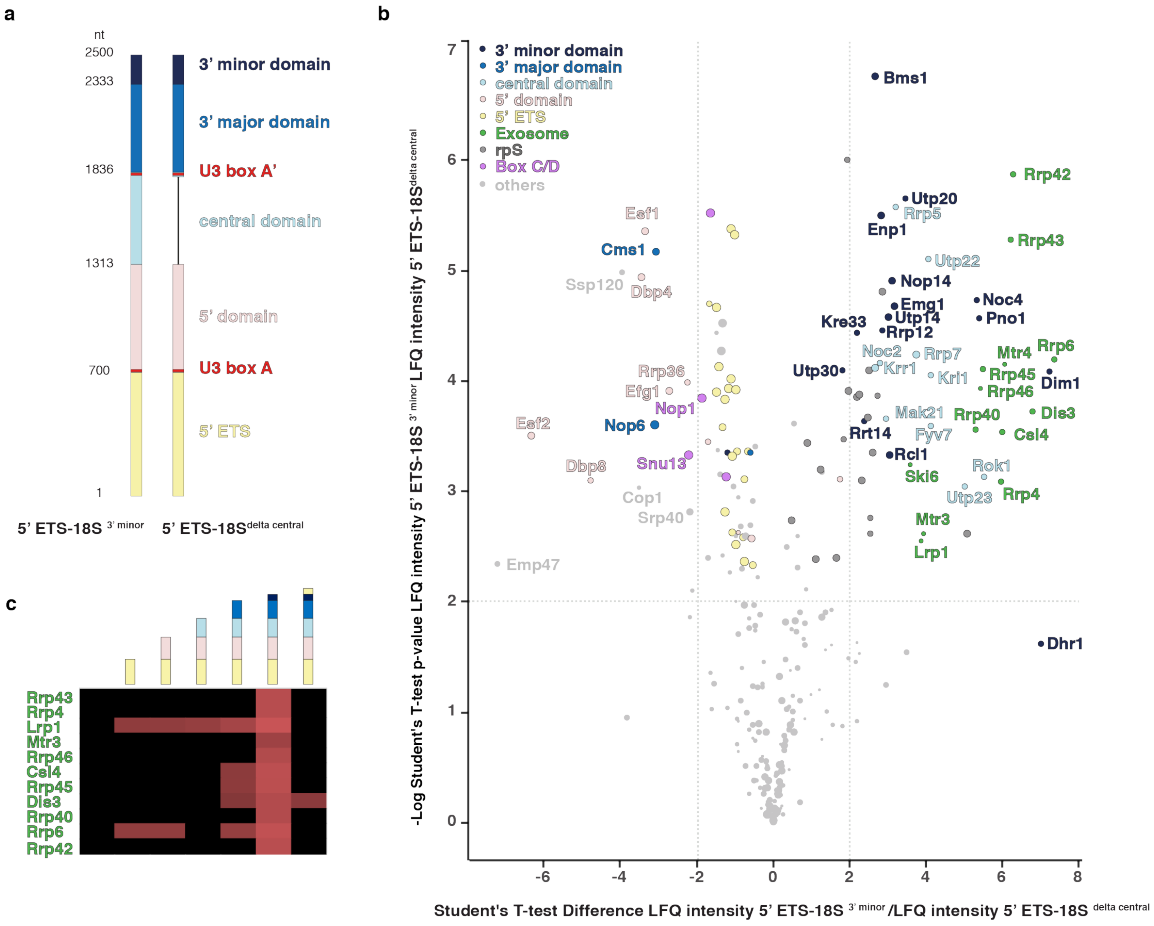
While the 5' domain and the 3' major domain are encapsulated by assembly factors, the central domain is more flexibly associated with the SSU processome and contacted by relatively few proteins including the UtpC complex, Rrp5 and Krr1. To test whether SSU processome formation can occur when both U3 snoRNA

interactions can be formed and yet a 18S rRNA domain is missing, the least integrated domain, the central domain, was deleted in a pre-rRNA construct spanning the otherwise complete transcript (**Figure 4.7a**). RNPs containing the complete pre-rRNA up to the 3' minor domain and the delta-central domain construct were purified as in Figure 4.1. Protein compositions were analyzed by comparative *in solution* mass spectrometry. The comparison of proteins that copurify with the two pre-rRNA segments illustrated the high fidelity with which SSU processome assembly is controlled.

While the pre-rRNA containing all 18S rRNA domains clearly formed SSU processomes as highlighted by the presence of late factors including Bms1, Utp20 and Kre33, the truncated pre-rRNA lacking the central domain associated with transient factors that bind the 5' domain (Dbp8, Esf2, Dbp4, Esf1 and Efg1) and 3' major domain (Nop6, Cms1) and failed to recruit any late factors (**Figure 4.7b**). Surprisingly, the complete nuclear exosome was enriched in purifications using the pre-rRNA with all 18S rRNA domains but not in the construct missing the central domain (**Figure 4.7b**). Hence the binding pattern of the exosome was similar to late recruited factors such as Bms1, Kre33, Utp14 and Pno1.

Figure 4.7 | SSU processome formation represents a checkpoint for the presence of all pre-rRNA domains.

[a] Schematic depiction of the pre-rRNA constructs used for RNP purifications analyzed in [b]. Domain boundaries are indicated with nucleotide numbers and individual domains as well as base-pairing regions of the U3 snoRNA (Box A, Box A') are indicated. **[b]** Volcano plot of label-free quantification (LFQ) values for proteins identified in purified RNPs containing the constructs shown in [a]. The x-axis indicates the log2-fold change of protein abundance of the 5' ETS-18S $^{3' \text{ minor}}$/ 5' ETS-18S $^{\text{delta central}}$ and the y-axis shows the negative log10 (p-value). Protein identities are labeled next to their data points and color coded as shown in the legend in the top left corner. **[c]** Heat map of nuclear exosome components in pull-downs of pre-rRNA transcripts mimicking the transcription of the pre-18S rRNA gene previously published in (Chaker-Margot et al. 2015). High abundance is indicated by shades of red and absence of signal is shown as black fields. Truncated pre-rRNA species used as baits in these pulldowns are shown schematically above each heat map column. Color-coding of the pre-rRNA domains is the same as in [a].


a
nt
2500
2333
1836
1313
700
1
3' minor domain
3' major domain
U3 box A'
central domain
5' domain
U3 box A
5' ETS
5' ETS-18S 3' minor
5' ETS-18S delta central
b
3' minor domain
3' major domain
central domain
5' domain
5' ETS
Exosome
rpS
Box C/D
others
Bms1
Rrp42
Utp20
Rrp5
Enp1
Rrp43
Utp22
Nop14
Noc4
Emg1
Pno1
Utp14
Kre33
Rrp12
Noc2
Rrp7
Mtr4
Rrp6
Krr1
Utp30
Krl1
Rrp45
Dim1
Rrp46
Mak21
Rrp40
Dis3
Rrt14
Fyv7
Csl4
Rcl1
Rok1
Ski6
Utp23
Rrp4
Mtr3
Lrp1
Dhr1
Esf1
Cms1
Ssp120
Dbp4
Rrp36
Efg1
Nop1
Nop6
Esf2
Dbp8
Snu13
Cop1
Srp40
Emp47
-Log Student's T-test p-value LFQ intensity 5' ETS-18S 3' minor LFQ intensity 5' ETS-18S delta central
Student's T-test Difference LFQ intensity 5' ETS-18S 3' minor /LFQ intensity 5' ETS-18S delta central
c
Rrp43
Rrp4
Lrp1
Mtr3
Rrp46
Csl4
Rrp45
Dis3
Rrp40
Rrp6
Rrp42

To investigate if association of the exosome with truncated pre-rRNA transcripts mimicking different transcriptional stages also exhibited the same pattern as late recruited factors, we re-analyzed previously published data (Chaker-Margot et al. 2015). Recruitment of the complete exosome only occurred in transcripts encompassing the 5' ETS and all 18S rRNA domains (**Figure 4.7c**). The addition of ITS1 to the transcript did not result in exosome binding (**Figure 4.7c**).

These data suggest, that either the exosome is recruited specifically during SSU processome formation, which represents a checkpoint for the presence of all 18S rRNA domains or that the accumulation of this specific precursor in the cell leads to the activation of the RNA surveillance machinery. The exosome could be involved in pre-rRNA processing events occurring during SSU processome formation or it might be recruited to degrade particles that do not mature further after this initial checkpoint because ITS1 is missing.

4.5 Conclusions

The *in vivo* analysis of artificial pre-rRNA species combined with the structural characterization of the 5' ETS particle has elucidated the process of SSU processome formation and its quality control mechanisms. Rather than functioning as a structural mold for the maturing 18S rRNA domains, the 5' ETS and 18S rRNA domains remain functionally independent during the initial steps of ribosome biogenesis. Completion of 18S rRNA gene transcription represents a quality control checkpoint during which the presence of all pre-rRNA domains is sensed and transient rRNA domain maturation factors dissociate. Cleavage or remodeling of the 5' ETS at position A0, a conformational switch in UtpB and the recruitment of late binding factors accompany this process. Protein components recruited upon completion of the 18S rRNA gene comprise a large number of enzymes and their co-factors. It remains unclear if these factors act during SSU processome formation or if they are positioned to act downstream in the maturation process. The dissociation of domain-specific enzymes such as Dpb4 and Rok1 and the association of the centrally located GTPase Bms1 and helicase Dhr1 could represent a switch from intra-domain specific maturation to inter-domain maturation facilitated by a swap of enzymatic components.

Chapter 5 | Discussion

Our understanding of the early steps in eukaryotic ribosome biogenesis has progressed from a list of involved proteins (Woolford & Baserga 2013) toward a three-dimensional assembly model for the SSU processome (**Figure 5.1 and Figure 5.2**). More than 50 components facilitating this process have been structurally characterized in the context of different assembly stages (Chapter III and IV, (Kornprobst et al. 2016; Chaker-Margot et al. 2017; Sun et al. 2017; Barandun et al. 2017; Cheng et al. 2017)), and biochemical studies have elucidated their RNA binding sites (Chapter II, (Granneman et al. 2011; Hunziker et al. 2016; Lin et al. 2013; Black et al. 2018; Kornprobst et al. 2016; Lebaron et al. 2013; Segerstolpe et al. 2013; Sardana et al. 2015)) recruitment order (Chaker-Margot et al. 2015; Zhang et al. 2016) and binding dependencies (Chapter IV) *in vivo*.

5.1 An assembly model for the 5' ETS particle - Initially flexible complexes form the base of the earliest precursor particles in eukaryotic ribosome biogenesis

UtpA initiates the assembly process of the SSU processome by binding to the first two hundred nucleotides of the 5' ETS (**Figure 5.1 a-c**) (Zhang et al. 2016; Hunziker et al. 2016). Individual subunits bind distinct sections within these first

nucleotides of the pre-rRNA (Chapter II, (Hunziker et al. 2016; Kornprobst et al. 2016). Assuming that the 5' ETS sequence is bound co-transcriptionally and independently by UtpA subunits, Utp9, Utp8 and Utp17 would be the first components to bind the nascent pre-rRNA as their RNA-protein cross-linking peaks cover its first 50 nucleotides. Utp8, Utp4 and Utp17 constitute the second group of overlapping RNA-protein cross-linking peaks and stabilize the single stranded linker between the first and second helix of the 5' ETS (**Figure 5.1b**). Early 5' ETS binding of Utp8 and Utp9 brings Utp15 and Utp5 close to the pre-rRNA through the flexibly-linked tetrameric complex formed by their CTDs (**Figure 5.1b**). Helix II of the 5' ETS is then clamped on top of the CTD-tetramer through the β-propeller of Utp15 (**Figure 5.1c**). The composite RNA binding site of UtpA defined by the ensemble of CRAC data sets, matches the binding site observed in the 5' ETS particle and SSU processome structures well (**Appendix 7.6**), despite being obtained from actively growing yeast and therefore likely representing a mix of assembly states. This suggests, that once UtpA associates with the pre-rRNA it remains bound to the same sequences throughout subsequent assembly stages. Structures of the 5' ETS particle in three assembly intermediates and the SSU processome support this overall static positioning of UtpA.

Prior to pre-rRNA and protein partner binding, UtpA is a highly flexible complex (**Figure 5.1a**) (Chapter II, (Hunziker et al. 2016). The inherent flexibility of UtpA is enabled by long inter-domain linkers which connect the compact globular folds within

its subunits (Chapter III, (Barandun et al. 2017; Cheng et al. 2017)). These linkers allow for the recognition of RNA and protein binding partners in distant locations by one subunit. While the formation of few binding interfaces might still preserve the high degree of flexibility observed for UtpA in isolation during the initial stages of pre-rRNA binding (**Figure 5.1 a-c**), the sequential formation of more interaction sites and addition of protein binding partners leads to a gradual stabilization of the complex as seen in the structure of the 5' ETS particle (Chapter IV) and the SSU processome (Chapter III, (Kornprobst et al. 2016; Chaker-Margot et al. 2017; Sun et al. 2017; Barandun et al. 2017; Cheng et al. 2017)).

UtpB is likely one of the first interactors of UtpA on the 5' ETS as it shares an early RNA binding site and several protein-protein contacts with UtpA (Chapter II-IV, (Barandun et al. 2017; Cheng et al. 2017)). Utp18 is anchored with its β-propeller moiety in between UtpA subunits, while its long and flexible N-terminus links UtpA and UtpB components (**Figure 5.1d**). Similar to UtpA, UtpB is dynamic in the absence of interacting molecules (**Figure 5.1d**) (Hunziker et al. 2016; Sun et al. 2017). However, while the high degree of conformational heterogeneity of UtpA prohibited the generation of 2D class averages in negative stain EM datasets, UtpB seems to adopt more distinct states in isolation that can be observed in 2D class averages (Hunziker et al. 2016; Sun et al. 2017). While some of these classes show the complex in a contracted state with Utp18 close to its core, others show a more elongated conformation of the complex. Yet, none of the states observed for UtpB in

isolation seem to be identical to the two different conformations it adopts in the 5' ETS particle or the SSU processome (J. Barandun, personal communication). Upon integration of UtpB in the 5' ETS particle, Utp18, Utp21 and Utp6 engage in interactions with UtpA subunits and the tandem β-propellers of Utp12 and Utp13 are stacked on each other at the periphery of the particle (**Figure 5.1e**).

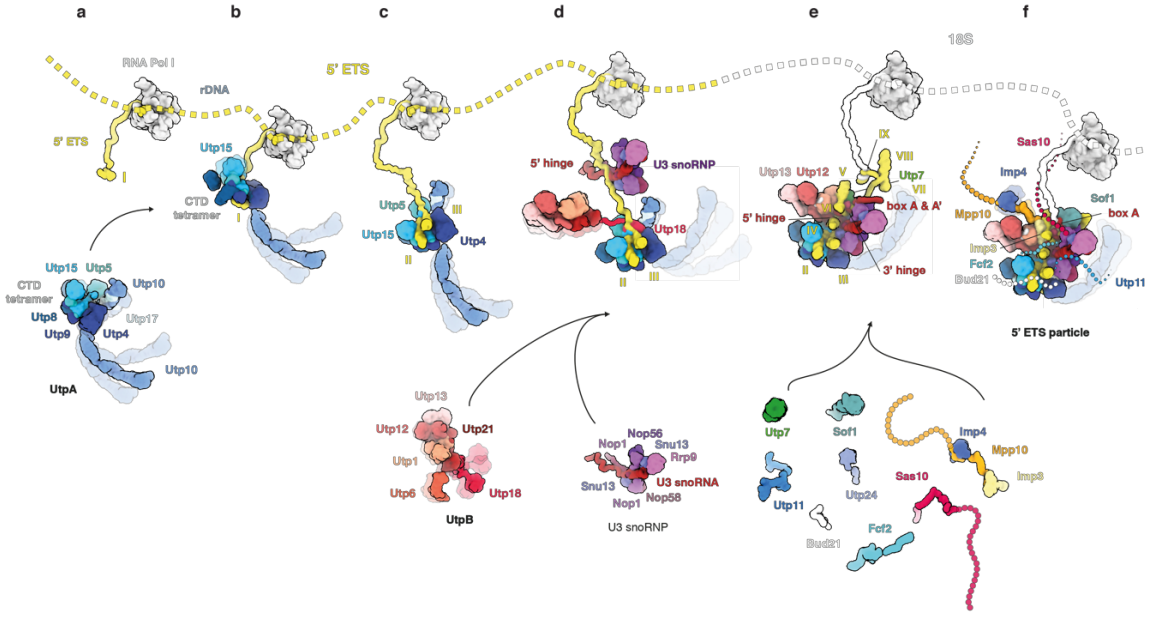


Figure 5.1 | Assembly model for the 5' ETS particle.

[a] A highly dynamic UtpA (shades of blue, subunits labeled) recognizes helix I of the 5' ETS (yellow) co-transcriptionally. **[b]** Helix I-binding of Utp8 and Utp9 brings Utp5 and Utp15 close to the 5' ETS through their complexed CTDs (grey label). **[c]** The CTD-tetramer and Utp15 are stabilized on helix II. **[d]** The β-propeller of Utp18 from the UtpB complex (shades of red, subunits labeled) is wedged between UtpA subunits, while the rest of the complex is still flexible. U3 snoRNA (in red with associated proteins in shades of purple) forms the 5' hinge duplex with the 5' ETS. **[e]** Utp7 (green) stabilizes the U3:5' ETS 3' hinge. **[f]** Multiple other proteins associate with the nascent RNA to form the 5' ETS particle. The extended flexible proteins (Sas10, Utp11, Fcf2, Bud21 and Mpp10) are shown as dotted lines in the 5' ETS particle.

The U3 snoRNP is the third large sub-complex of the 5' ETS particle. Its 3' and 5' hinges rigidify the scaffold of the 5' ETS (**Figure 5.1 d,e**) and the common Box C/D core proteins intertwine the snoRNP further with the 5' ETS particle, as the C-terminus of Nop58 binds UtpB and the two copies of Nop1 serve as binding hubs for Sas10, Fcf2, Bud21, Utp11 and the nuclease Utp24 (**Figure 5.1f**). The Mpp10 complex, consisting of Imp4, Imp3 and Mpp10, is bound to the 5' ETS close to the 3' hinge (**Figure 5.1f**). In the 5' ETS particle structure Mpp10 and Imp4 are flexible and not resolved. At the back of the 5' ETS particle the β-propeller-containing subunits Utp7 and Sof1 are associated (**Figure 5.1 e,f**).

Once the ~2 MDa 5' ETS particle is assembled, the initially flexible subcomplexes UtpA and UtpB as well as U3 snoRNP, Utp7 and Sof1 are stabilized (**Figure 5.1f**). While these architectural folds form a stable core, the 5' ETS also recruits the majority of the very long and flexible peptide-like protein subunits which connect the rRNA domains with the 5' ETS in the SSU processome (**Figure 5.1f**). Mpp10, Sas10, Fcf2, Utp11 and Bud21 have multiple binding sites on the 5' ETS and 5'- and 3' domains. Some of these proteins span more than 100 Å to mediate inter pre-rRNA domain contacts in the SSU processome (Barandun et al. 2017; Cheng et al. 2017).

5.2 An assembly model for the SSU processome

The transcription of the pre-18S rRNA domains and their maturation succeeds the assembly of the 5' ETS particle (**Figure 5.2**). Contrary to prior models which included a hierarchical system of assembly factor binding (Pérez-Fernández et al. 2007; Pérez-Fernández et al. 2011) or suggested molding of pre-rRNA domains (Kornprobst et al. 2016) on the 5' ETS moiety, the cryo-EM reconstruction of the 5' ETS particle of different assembly intermediates and *in vivo* binding assays of isolated 18S rRNA domains have shown that the early steps in eukaryotic ribosome assembly are instead governed by initial functional independence of the 5' ETS and all rRNA domains (**Figure 5.2 b-d**) (**Chapter IV**).

UtpB is a particularly important module of the 5' ETS particle as it can act as a molecular switch that prevents the premature association of other factors until all rRNA domains have come together to form the SSU processome (**Figure 5.2e**) (**Chapter IV**). The extended proteins (Mpp10, Utp11, Sas10, Fcf2, Bud21) recruited at the 5' ETS stage, and the single stranded 5' end of the U3 snoRNA harboring the 18S rRNA base-pairing regions box A and box A', may play important roles in connecting the pre-rRNA segments to the 5' ETS particle in a structurally unrestrained manner, which allows for their independent maturation until they are compacted onto the 5' ETS particle during SSU processome formation (**Figure 5.2f**).

Recruitment of SSU processome factors following the complete transcription of the 18S rRNA gene represents a checkpoint for the presence of all pre-rRNA domains (**Figure 5.2 e,f**).

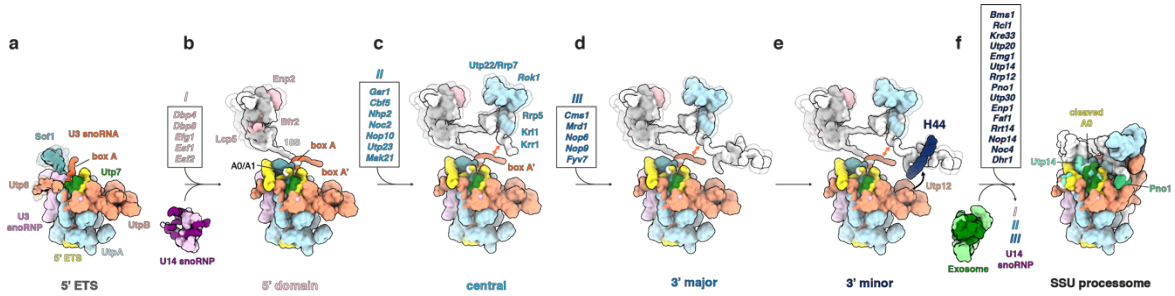


Figure 5.2 | Assembly model for the SSU processome.

[a] The assembled 5'-ETS particle with its components and subcomplexes is shown and labeled in a color-coded manner. **[b-d]** The 5'-, central- and 3' major domain of the 18S rRNA (white) are flexible and recruit transient ribosome biogenesis factors (boxed, color coded names listed) as well as proteins present in the SSU processome (as surface in color of domain) independent of the 5' ETS particle. The box C/D snoRNP U14 (shades of purple) may bind to the maturing 5' domain. The U3 snoRNA:18S pre-rRNA duplex box A (red) is likely formed. **[e]** Helix 44 (dark-blue) of the 3' minor domain is transcribed and bound by Utp12 (red) which triggers a conformational switch in UtpB that exposes the Pno1 and other late factor binding sites on UtpB. **[f]** During SSU processome formation the 18S rRNA subdomains are folded onto the 5' ETS particle. Transient factors recruited during earlier stages are released (I, II, III, U14) and SSU processome specific factors are recruited (dark-blue, boxed). The previously intact A0 site is remodeled or cleaved to allow for Utp14 binding. This stage represents a checkpoint for the presence of all pre-rRNA domains and can trigger exosome (shades of green) binding.

5.3 Modular flexibility as an emerging principle in the assembly of the SSU processome

The concept of distributing multiple flexibly-linked, conserved binding sites on the structure of a single protein subunit is repeated in many SSU processome assembly factors (Barandun et al. 2018). This characteristic allows for a guided co-transcriptional assembly process of the modular pre-rRNA. During this process individual binding events can occur independently at first (**Figure 5.1** and **Figure 5.2 a-d**), but the cumulative formation of all binding interfaces leads to a sequential reduction in conformational freedom and eventually the formation the SSU processome (**Figure 5.2f**).

The inherent properties of proteins and RNA molecules facilitating such dynamic assembly processes make them challenging to study with structural biology methods. Additionally, manipulations of endogenous ribosome biogenesis factors *in vivo* are often lethal, and the large number of identical rDNA repeats prohibits targeted genetic alterations in them. The early nucleolar stages in ribosome biogenesis are also highly transient and therefore hard to endogenously isolate from growing yeast cells. Hence, the artificial systems mimicking these early nucleolar steps in yeast ribosome biogenesis that have contributed greatly to our current understanding of these events (Chaker-Margot et al. 2015; Zhang et al. 2016; Chen et al. 2017) (Chapter IV), could further be used to address outstanding mechanistic questions.

It remains unclear whether the 5' ETS-bound extended proteins bind to their SSU processome interaction sites during the assembly process of the 18S rRNA domains or if they remain unbound until the transcription of the 18S rRNA is completed. Similarly, the timing of Box A' formation by U3 snoRNA during transcription is unknown. It could be that transient domain specific assembly factors block the base-pairing sites until SSU processome formation. To delineate the role of the highly flexible protein and RNA segments involved in SSU processome assembly, transcription stage-specific protein-protein, RNA-protein and RNA-RNA interaction maps could be determined using DSS- and UV-cross-linking analysis respectively.

The functional independence of individual 18S rRNA domains, as demonstrated by their *in vivo* expression and purification, has also laid out a potential route to structurally study these maturing domains in complex with transient ribosome assembly factors that are absent in the SSU processome.

The precise steps occurring after pre-18S rRNA transcription and during SSU processome formation as well as their order or dependencies remain poorly understood. While the recognition of the 3' minor domain by UtpB could trigger a conformational change leading to the recruitment of SSU processome factors, the potential contribution of enzymes such as Bms1 or RNA helicases and nucleases has not been determined yet. It will further be important to fully understand the

significance of A0-site cleavage and why it occurs only in 3' minor encompassing precursors despite being present at earlier transcription stages. The recruitment of the nuclear exosome to these specific pre-ribosomal particles and its potential function there as a pre-rRNA processing or aberrant RNA degradation machinery should be assessed. Protein-protein cross-linking studies of this precursor could elucidate which of the three exosome recruitment motifs in the SSU processome is used by this molecular RNA degradation machinery.

5.4 Maturation of the small ribosomal subunit beyond the SSU processome stage

The cryo-EM reconstructions of the SSU processome have elucidated the structure of more than 50 ribosome assembly factors for the first time (Chapter III, (Kornprobst et al. 2016; Chaker-Margot et al. 2017; Sun et al. 2017; Barandun et al. 2017; Cheng et al. 2017). In this nucleolar precursor the individual 18S rRNA domains are folded onto the base of the 5' ETS particle and kept in spatially segregated compartments by the concerted action of ribosome biogenesis factors and U3 snoRNA based remodeling of central pseudoknot elements. While the 5'-, central- and 3' minor domain are located towards the top of the particle, the 3' major domain is positioned close to UtpA at the base of the particle. The spatial separation of the 3' major domain from the three other domains is reminiscent of the mature small subunit in which the 3' major domain forms the head and the other domains together the body.

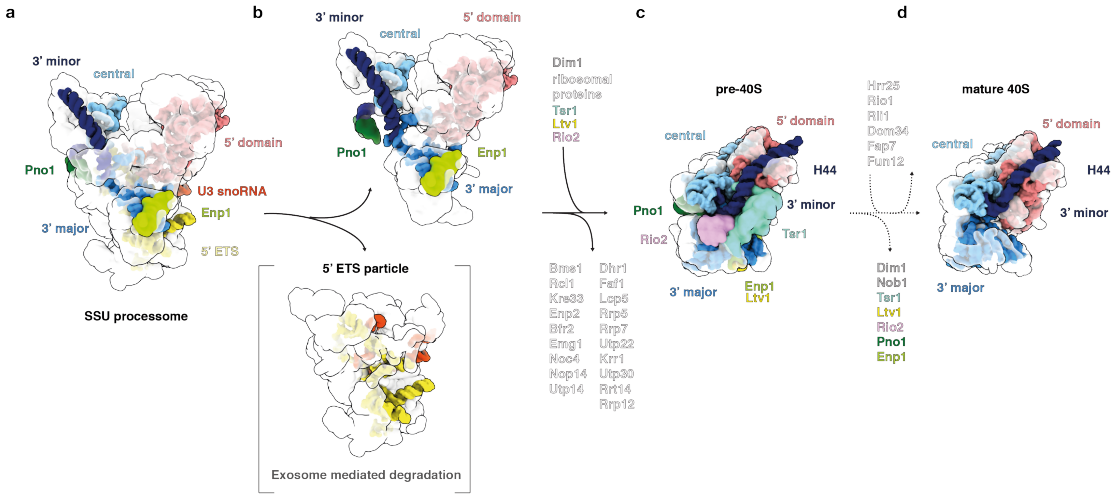


Figure 5.3 | Model for small subunit maturation steps following the SSU processome stage

[a] In the SSU processome (PDB 5WLC) the four 18S rRNA domains (5'-, central-, 3' major- and 3' minor, color-coded) are segregated into different compartments by ribosome assembly factors and U3 snoRNA (red) base-pairing. Pno1 and Enp1 (in shades of green) are two ribosome biogenesis factors staying associated with the pre-18S rRNA throughout the subsequent maturation steps. **[b]** For the SSU processome to transition to the pre-40S [c], the 5' ETS particle (5' ETS: yellow, U3 snoRNA:red, ribosome biogenesis proteins: white) has to be removed and degraded by the exosome. Pre-18S rRNA bound SSU processome proteins need to dissociate (protein names listed in white) while new assembly factors (Dim1, Rio2, Ltv1, Tsr1) and ribosomal proteins bind. **[c]** In the pre-40S particle (PDB 6FAI, (Scaiola et al. 2018)) the four pre-18S rRNA domains (labeled and color-coded as in [a-d]) adopt a near mature confirmation while the functionally important sites are blocked by ribosome assembly factors (Enp1, Pno1, Ltv1, Rio2, Tsr1). Ribosomal proteins are shown in transparent white. **[d]** Multiple quality control and maturation steps mediated by different proteins (Hrr25, Rio1, Rfl1, Dom34, Fap7, Fun12) lead to the formation of the compacted 40S structure (PDB 4V88, (Ben-Shem et al. 2011))

For the SSU processome to transition into the pre-40S structure (Heuer et al. 2017; Scaiola et al. 2018) multiple steps have to occur (**Figure 5.3a-d**). The 5' ETS particle moiety has to be removed and degraded, A1-cleavage has to be catalyzed, the U3:18S duplexes have to be unwound for the central pseudoknot to form and several ribosome assembly factors need to dissociate while new ones such as Tsr1, Rio2 and Ltv1, will bind.

This transition likely requires the concerted effort of multiple enzymes. Dhr1, the Utp14-bound DEAH-box RNA helicase which is recruited upon SSU processome formation has been shown to be involved in the unwinding of the U3 RNA duplexes (Sardana et al. 2015). Unwinding of these duplexes as well as Utp24 catalyzed A1-cleavage (Bleichert et al. 2006; Tomecki et al. 2015; Wells et al. 2016) would facilitate the separation of the 5' ETS particle from the SSU processome (**Figure 5.3b**).

Removal and degradation of the 5' ETS moiety of the SSU processome are thought to be facilitated by the exosome (**Figure 5.3b**) (Allmang et al. 2000). The 5' ETS particle harbors two proteins with distinct exosome recruitment motifs. Utp18 contains an AIM motif (Thoms et al. 2015) bound by the exosome-associated helicase Mtr4 and Sas10 harbors a sequence highly homologous to the exosome co-factor Rrp47 (Mitchell 2010; Costello et al. 2011).

The mechanism and timing of exosome binding to the SSU processome is not understood. In the large subunit maturation pathway, the AIM motif in Nop53 is bound by Mtr4, which unwinds ITS2 and feeds it into the catalytic core of the exosome to achieve precise pre-rRNA processing of the 5.8S rRNA (Thoms et al. 2015; Schuller et al. 2018).

To bring the 18S rRNA domains into near-mature positions as observed in the pre-40S structures, conformational changes are required (**Figure 5.3b-c**). These may occur due to the formation of the central pseudoknot after U3:18S duplex unwinding and the dissociation of ribosome assembly factors currently blocking binding sites of 18S rRNA domains. The centrally located GTPase Bms1 could further help induce these changes in a nucleotide-dependent manner. Bms1 is positioned in between 18S rRNA domains in the SSU processome and is connected to several proteins that pierce through the core of the particle. While GTP-hydrolysis by Bms1 has not been directly characterized, it binds GTP and mutations in its active site result in growth defects in yeast (Gelperin et al. 2001; Wegierski et al. 2001). A GTPase activating protein (GAP), typically required for the stimulation of GTPase activity, has not been identified for Bms1 so far. While it is tempting to speculate that the GTPase activity of Bms1 could facilitate conformational changes during post-SSU processome maturation, its potential enzymatic function could also be involved in domain compaction upon SSU processome formation.

Since the cryo-EM structure of SSU processome described in this study has been obtained from particles isolated from starved yeast cells, it could be that this particle represents a non-productive assembly intermediate that will not mature further after stress release but rather be degraded. A pulse and chase study of accumulated 23S pre-rRNA in yeast has suggested that the 23S precursor represents a non-productive assembly intermediate (Kos-Braun et al. 2017). However, all cryo-EM reconstructions obtained to date (Kornprobst et al. 2016; Chaker-Margot et al. 2017; Sun et al. 2017; Barandun et al. 2017; Cheng et al. 2017) exhibit a high degree of similarity in composition, structure and pre-rRNA state (A1 uncut, A0 cut), despite being obtained under different growth conditions. Thus, the fate of these particles remains unclear and further experiments are needed to determine if they are in fact incompetent to mature or if they are just stable intermediates of a functional and otherwise highly dynamic pathway. To date it is unknown which mechanisms and signaling pathways downstream of TOR lead to the increase in 23S and why precisely this intermediate accumulates.

5.5 Emerging similarities and differences in small- and large subunit assembly

Decades of genetic and biochemical research have revealed the identity of the plethora of proteins and snoRNAs facilitating ribosome biogenesis in yeast. While the introduction of affinity purifications and mass spectrometry has enabled the compositional definition of pre-ribosomal particles in the early 2000's, the recent advent of single particle cryo-EM has allowed for the high-resolution structural characterization of these dynamic assemblies. Together with biochemical and cell biology studies the solved structures (Barrio-Garcia et al. 2016; Heuer et al. 2017; Bradatsch et al. 2012; Kater et al. 2017; Sanghai et al. 2018; Zhou et al. 2018; Wu et al. 2016; Scaiola et al. 2018; Kornprobst et al. 2016; Chaker-Margot et al. 2017; Sun et al. 2017; Barandun et al. 2017; Cheng et al. 2017) have helped to understand the role of a multitude of ribosome biogenesis proteins in yeast in a detailed way. However, further research is needed to fully understand all the steps involved in the maturation of the small and large ribosomal subunit. The earliest nucleolar stages as well as events occurring in the transition from nucleolar to early nuclear maturation states are still not fully characterized for both subunits. Nevertheless, the recent advances in the ribosome biogenesis field allow us to begin to deduce the basic principles in small and large subunit assembly.

The earliest co-transcriptional assembly steps which lead to the formation of the small subunit processome are marked by the independent maturation of all pre-

rRNA domains (**Chapter IV**, **Figure 5.2**). snoRNAs, helicases and transient ribosome assembly factors facilitate the maturation of each rRNA domain. Upon complete transcription of the small subunit rRNA gene, compaction of the pre-ribosomal particle is achieved, and a large number of assembly factors bind (**Figure 5.2**). In the formed SSU processome the pre-18S rRNA is encapsulated by ribosome assembly factors. The 5' ETS particle plays an important role in the spatial arrangement of the pre-18S domains and their encapsulation in the SSU processome.

Contrary to the small subunit, the pre-25S rRNA is not encapsulated by assembly factors in the nucleolus (Sanghai et al. 2018; Kater et al. 2017). Rather, the pre-25S domain I and II and the 5.8S rRNA, which form the solvent exposed side of the mature large subunit, fold first and form a stable shell bound by ribosome assembly factors. The other domains, also associated with a distinct set of assembly factors, are then folded onto this ribosomal RNA shell. Unlike the 5' ETS, the spacer regions of the pre-rRNA of the large subunit, ITS1, ITS2 and the 3' ETS, and the assembly factors recruited by them do not act as an enclosing mold for the pre-25S rRNA domains, but rather provide a structural scaffold for domains I and II.

Aside from these differences there are also similarities for the nucleolar stages of the small and large subunit. In both pathways the complete transcription of the pre-rRNA for the respective subunit is sensed and a large number of assembly factors bind to the first and last transcribed element of the pre-rRNAs (**Figure 1.6, Figure**

1.8). It is only after completed transcription that the irreversible steps of pre-rRNA cleavage occur: The A0 and A2 or A3 sites are cleaved after transcription of all 18S rRNA domains and processing of ITS1 and the 3' ETS are coupled in the large subunit pathway (Lebaron et al. 2013; Gamalinda et al. 2014). Furthermore, both subunits require binding by snoRNAs and numerous helicases during the co-transcriptional assembly steps.

The transition from the nucleolus to the nucleoplasm is still poorly understood for both subunits. However, it includes the dissociation of a large number of assembly factors in the large and small subunit (**Figure 1.8**, **Figure 5.3**). While little is known about the maturation events in the nucleoplasm for the small subunit precursors, the pre-60S particles undergo processing of ITS2 and rotation of the 5S RNP. Ensuing nuclear assembly intermediates of both subunits adopt near-mature conformations with respect to the arrangement of ribosomal RNA domains. In the cytoplasm quality control steps of the newly produced subunits are performed. While the small subunit is tested by the formation of an 80S-like ribosome through binding to a mature large subunit, the functional centers of the pre-60S are quality controlled by individual ribosome assembly factors.

5.6 Perspectives

Ribosome biogenesis is a central and vital process that has a multitude of interfaces with other cellular pathways. Understanding which signaling networks

regulate ribosome biogenesis in response to changing internal and external factors, will be essential. The interplay of cell cycle regulation and ribosome biogenesis and its links to the development and progression of human diseases are of broad interest.

In addition to expanding the study of ribosome biogenesis from the model organism yeast to the more complicated mammalian systems, a clearer understanding of this process in archaea and bacteria as well as more diverse eukaryotes could yield interesting insights into the evolution of this essential pathway.

Chapter 6 | Methods

6.1 Molecular cloning

All cloning was carried out with standard restriction enzymes and T4 DNA Ligase from New England Biolabs, Phusion DNA polymerase and DH5α and Stbl3 *E. coli* cells. Constructs used in this study are listed in Table 6-1.

6.2 C-terminal tagging of endogenous DNA loci in yeast

DNA fragments containing the sequence of C-terminal affinity tags coupled to an antibiotic resistance gene were amplified from tagging plasmids (B029, B045, B086, B370, B372) with primers harboring 50 base-pair regions homologous to the C-terminus of the target locus. 10ml yeast cultures were grown to an optical density of 0.5 at 30 ºC, washed twice with ddH_2O and once with 1X TE (100 mM Tris-HCl, pH 7.5, 10 mM EDTA) supplemented with 100 mM Lithium Acetate (pH 7.5), before resuspension of the pelleted cells in 50 μl of the latter buffer. Transforming DNA fragments, 50 ug of salmon sperm DNA and 300 μl of 40% PEG 4000 in 1X TE supplemented with 100 mM Lithium acetate were added to the cell suspension. After incubation for 30 minutes at 30 ºC, cells were heat shocked at 42 ºC for 15 minutes and recovered for 2 hours in non-selective media before plating on plates containing selection markers. All yeast strains used in this study are based on *S. cerevisiae* strain BY4741 and listed in Table 6-2.

Table 6-1 | List of plasmids used in this study

Name	Description (base plasmid_promoter - (gene(s) of interest) - integration locus)	Bacterial resistance	Yeast selection marker	Source
B029	pFA6a_(sbp-H14-3c-GFP)	Ampicillin	G418	Chapter 2
B045	pFA6a_(3myc-tev-mCherry-3flag)	Ampicillin	HygromycinB	Chapter 2
B079	pESC_GAL1-(NLS-HA-MS2-3C-GFP)-PEP4	Ampicillin	G418	Chapter 4
B086	pFA6a_(3myc-tev-GFP-3flag)	Ampicillin	HygromycinB	Chapter 3
B117	pESC_GAL1-(5' ETS-5xMS2)	Ampicillin	URA	Chaker-Margot et al. 2015, Chapter 4
B133	pESC_(GAL10-Utp8,GAL1-Utp9)-PEP4	Ampicillin	G418	Chapter 2
B178	pESC_(GAL10-Utp4, GAL1-Utp5, GAL10-Utp15-3myc-tev-mCherry-3flag)-PRB1	Ampicillin	HygromycinB	Chapter 2
B217	pESC_(GAL10-Utp10-sbp-H14-3c-GFP, GAL1-Utp17)-YORWdelta22	Ampicillin	HIS	Chapter 2
B221	pESC_GAL1-(5' ETS-5' domain of 18S rRNA-5xMS2)	Ampicillin	URA	Chaker-Margot et al. 2015, Chapter 4
B222	pESC_GAL1-(5' ETS-central domain of 18S rRNA-5xMS2)	Ampicillin	URA	Chaker-Margot et al. 2015, Chapter 4
B223	pESC_GAL1-(5' ETS-3' major domain of 18S rRNA-5xMS2)	Ampicillin	URA	Chaker-Margot et al. 2015, Chapter 4
B224	pESC_GAL1-(5' ETS-3' minor domain of 18S rRNA-5xMS2)	Ampicillin	URA	Chaker-Margot et al. 2015, Chapter 4
B370	pFA6a_(linker-sbp)	Ampicillin	G418	Chapter 3
B372	pFA6a_(linker-sbp)	Ampicillin	Nourseothricin	Chapter 4
B495	pESC_GAL1-(5' domain, central domain and 3' major domain of 18S rRNA-5xMS2)	Ampicillin	URA	Chapter 4
B502	pESC_GAL1-(central domain of 18S rRNA-5xMS2)	Ampicillin	URA	Chapter 4
B503	pESC_GAL1-(5' domain and central domain of 18S rRNA-5xMS2)	Ampicillin	URA	Chapter 4
B504	pESC_GAL1-(18S rRNA-5xMS2)	Ampicillin	URA	Chapter 4
B506	pESC_GAL1-(5' domain of 18S rRNA-5xMS2)	Ampicillin	URA	Chapter 4
B514	pESC_GAL1-(5' ETS- 18S rRNA delta central domain-5xMS2)	Ampicillin	URA	Chapter 4

6.3 Yeast strain construction for galactose driven overexpression of UtpA subunits

To generate a yeast strain overexpressing the seven subunits of UtpA (Y143), pairs of subunits (Utp8-Utp9, Utp5- Utp15-3myc-TEV-mCherry-3FLAG, Utp17-Utp10-sbp-H14-3C-GFP, Utp4) were cloned into derivates of the pESC_URA vector (Agilent Technologies) under gal1 and gal10 promoters respectively. Utp4 was first cloned alone under a gal10 promoter and then combined with Utp5 and -3myc-TEV-mCherry-3FLAG. UtpA subunits were either amplified from genomic DNA of BY4741 or yeast strains carrying a C-terminal affinity tag on selected UtpA members (Y32). The plasmids with galactose inducible copies of UtpA subunits and homology regions for integration in the genomic loci of *pep4*, *prb1* and *YORWδ22* were linearized and transformed as described for endogenous tagging (**see 6.2**) into a yeast strain containing a C-terminal tag (-sbp-H14-3C-GFP) on its endogenous Utp10 copy (Y116). Sequential integration of the three plasmids (B133, B178, B217) generated several yeast strains which can be used to express sub-complexes of UtpA (Y122) or the holo-complex (Y143).

Table 6-2 | List of yeast strains used in this study

Name	Genotype	Yeast selection marker	Source
Y32	MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 UTP10--sbp-H14-3C-GFP::kanMX6, UTP15-3myc-TEV-mCherry-3FLAG::HphMX4	G418, HygromycinB	Chapter 2
Y122	MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 UTP10--sbp-H14-3C-GFP::URA PEP4::kanMX6-B133(GAL1-Utp9, GAL10-Utp8)- PRB1::HphMX4-B178(GAL10-Utp15-3myc-TEV-mCherry-3FLAG, GAL1-Utp5, GAL10-Utp4)	URA, G418, HygromycinB	Chapter 2
Y133	MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 UTP10-HTP::HphMX4	Hygromycin B	Chapter 2
Y134	MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 UTP15-HTP::HphMX4	Hygromycin B	Chapter 2
Y137	MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 UTP5-HTP::HphMX4	Hygromycin B	Chapter 2
Y138	MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 UTP9-HTP::HphMX4	Hygromycin B	Chapter 2
Y139	MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 UTP4-HTP::HphMX4	Hygromycin B	Chapter 2
Y140	MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 UTP8-HTP::HphMX4	Hygromycin B	Chapter 2
Y143	MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 UTP10-sbp-H14-3C-GFP::URAPEP4::B133(GAL1-Utp9,GAL10-Utp8)-PRB1::HphMX4-HphMX4-B178(GAL10-Utp15-3myc-TEV-mCherry-3FLAG, GAL1-Utp5, GAL10-Utp4) Yorwdelta22::HIS::KB217(GAL10-Utp10---sbp-H14-3C-GFP, GAL1-Utp17)	HIS, URA, G418, HygromycinB	Chapter 2
Y144	MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 UTP17-HTP::HphMX4	Hygromycin B	Chapter 2
Y158	MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 UTP1-HTP::HphMX4	Hygromycin B	Chapter 2
Y159	MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 UTP13-HTP::HphMX4	Hygromycin B	Chapter 2
Y160	MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 UTP18-HTP::HphMX4	Hygromycin B	Chapter 2
Y186	MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 PEP4::kanMX6-NLS-HA-MS2-3c-GFP UTP10-linker-sbp::HphMX4	G418, HygromycinB	Chapter 4
Y232	MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 UTP1-3myc-TEV-GFP-3FLAG::HphMX4 Kre33-linker-sbp::kanMX6	Hygromycin B,G418	Chapter 3
Y367	MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 PEP4::kanMX6-NLS-HA-MS2-3c-GFP ESF1-linker-sbp::ClonNat	G418, Nourseothricin	Chapter 4
Y372	MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 PEP4::kanMX6-NLS-HA-MS2-3c-GFP KRI1-linker-sbp::ClonNat	G418, Nourseothricin	Chapter 4
Y374	MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 PEP4::kanMX6-NLS-HA-MS2-3c-GFP MRD1-linker-sbp::ClonNat	G418, Nourseothricin	Chapter 4

6.4 Purification of UtpA from endogenous sources

Yeast strains used for endogenous complex purifications, Y32 (Utp10-sbp-H14-3C-GFP, Utp15-3myc-TEV-mCherry-3FLAG), were grown to saturation in yeast peptone dextrose (YPD) media and harvested by centrifugation at *4000 x g*. Cell pellets were washed twice with ice-cold water and once with a volume of water supplemented with protease inhibitors (PMSF, Pepstatin A, E64) equal to the weight of each pellet. The final cell paste was flash-frozen as "noodles" by pushing it through a syringe into liquid nitrogen. Cell disruption was performed by cryo-milling using a Retsch Planetary Ball Mill PM 100, and the cryo-ground powder was stored at -80 °C until further use.

30 grams of cryo-milled powder from strain Y32 were resuspended in 30 ml of binding buffer (50mM HEPES-NaOH pH 7.6 (4 ºC), 200mM NaCl, 1mM EDTA, 1mM DTT, 10% glycerol) supplemented with protease inhibitors (PMSF, Pepstatin A, E64), DNase I and RNAseA. The solution was incubated on ice for 30 minutes and cleared by centrifugation at 40,000 g, 4 ºC for 30 minutes. The cleared lysate was incubated with 500 µl of anti-GFP nanobody-coupled sepharose for 4 hours at 4 ºC on a nutator and washed six times with binding buffer. Bound protein complexes were eluted by incubation with TEV and 3C protease for 150 minutes at 4 ºC. Eluted protein complexes were further purified by size-exclusion chromatography (Superose 6 10/300 GL, GE Healthcare) in binding buffer lacking glycerol and DTT.

6.5 Purification of overexpressed UtpA

A pre-culture of Y143 was grown in -URA, -HIS Synthetic Dropout (SD) media supplemented with HygromycinB (final: 0.4mg/ml), G418 (final 0.3mg/ml), Ampicillin (50ug/ml) and 2% Glucose (w/v) at 30 ºC overnight. The pre-culture was scaled up in YPD media containing Ampicillin (50ug/ml) and grown for 6 hours at 30 ºC. Cells were induced after dilution of the pre-culture into YP media supplemented with 2% Galactose (w/v) and Ampicillin (50ug/ml). Induced cells were grown overnight at 30 ºC or until they reached an optical density (OD) of 5 and harvested by centrifugation at *4000 x g*. The harvesting procedure, cryo-milling and storage were performed as described in (**6.4**).

10 grams of cryo-milled Y143 powder was resuspended in 30 ml of binding buffer (50mM HEPES-NaOH pH 7.6 (4 ºC), 100mM NaCl, 1mM EDTA, 1mM DTT, 10% glycerol) supplemented with protease inhibitors (PMSF, Pepstatin A, E64). The lysate was incubated with RNAseA and DNase I for 30 minutes at 4 ºC, prior to centrifugation for 20 minutes at 40,000 g, 4 ºC. The supernatant of the cleared lysate was mixed with 500 µl of anti-GFP nanobody coupled sepharose and incubated overnight at 4 ºC on a nutator.

Following overnight binding the beads were washed four times with 50ml wash and cleavage buffer (50mM HEPES-NaOH pH 7.6 (4 ºC), 200mM NaCl, 1mM EDTA, 1mM DTT, 10% glycerol). To elute the over-expressed complex, washed beads were

transferred into a 2 ml tube and incubated with 15 μl of 3C protease (1mg/ml) for 120 minutes at 4 ºC. Eluted protein complexes were further purified by size-exclusion chromatography (Superose 6 10/300 GL, GE Healthcare) in binding buffer lacking glycerol, EDTA and DTT.

6.6 Purification of UtpA sub-complexes

6.6.1 UtpAΔUtp4, Utp10-Utp17 dimer

11 grams of cryo-milled Y32 (Utp10-sbp-H14-3C-GFP, Utp15-3myc-TEV-mCherry-3FLAG) powder were resuspended in 44 ml binding buffer (50mM HEPES-NaOH pH 7.6 (4 ºC), 200mM NaCl, 1mM EDTA, 1mM DTT, 10% glycerol) supplemented with protease inhibitors (PMSF, Pepstatin A, E64) and RNAseA and DNase I. The lysate was cleared for 30 minutes at 40,000 g, 4 ºC. 560 μl of anti-GFP nanobody coupled sepharose was added to the supernatant and incubated for 3.5 hours at 4 ºC. Beads were washed twice with binding buffer and distributed in 14 tubes (40 μl of beads each). The aliquots were washed and incubated with binding buffer lacking glycerol and containing concentrations of NaCl ranging from 200mM to 800mM in 100mM steps (50mM HEPES-NaOH pH 7.6 (4 ºC), 200–800mM NaCl, 1mM EDTA, 1mM DTT) for 7 hours at 4 ºC. Protein complexes were eluted by 3C protease cleavage overnight at 4 ºC. Elutions were spun at 15,000 g for 6 minutes and supernatants were analyzed on a 4–12% SDS-PAGE gel by Coomassie-blue staining.

6.6.2 UtpAΔUtp4ΔUtp10ΔUtp17

To overexpress Utp4, Utp5, Utp8, Utp9 and Utp15-3myc-TEV-mCherry-3FLAG, a pre-culture of Y122 was grown overnight in YP medium containing 2% (w/v) glucose at 30 °C. The pre-culture was used to inoculate a larger volume of YP medium supplemented with 2% (w/v) raffinose. Cells were grown at 30 °C in raffinose-containing medium until OD600 = 0.9. Expression of proteins under galactose promoters was induced by the addition of 2% (w/v) galactose overnight at 30 °C, and cells were harvested by centrifugation at 4000 x g. Cell pellets were washed twice with ice-cold water and once with a volume of water supplemented with protease inhibitors (PMSF, Pepstatin A, E64) equal to the weight of each pellet. The final cell paste was flash-frozen as “noodles” by pushing it through a syringe into liquid nitrogen. Cell disruption was performed by cryo-milling using a Retsch Planetary Ball Mill PM 100, and the cryo-ground powder was stored at -80 °C until further use.

20 grams of cryo-milled Y122 powder were resuspended with 80 ml of binding buffer (50mM HEPES-NaOH pH 7.6 (4 ºC), 100mM NaCl, 1mM EDTA, 1mM DTT, 10% glycerol) supplemented with protease inhibitors (PMSF, Pepstatin A, E64), DNase I and RNAseA. The lysate was cleared by centrifugation at 40,000 g, 4 ºC for 30 minutes. 800 µl of anti-mCherry nanobody-coupled sepharose were added to the supernatant and incubated for 4 hours at 4 ºC. Beads were washed six times with

binding buffer. The protein complex was eluted by TEV cleavage at 4 ºC for 150 minutes. Endogenous UtpA was removed through Utp10-sbp-H14-3C-GFP by incubating the elution with anti-GFP nanobody-coupled sepharose for 30 minutes. The flow through was stored overnight at 4 ºC and centrifuged for 10 minutes at 15,000 rpm at 4 ºC before injection on a Superose 6 Increase 10/300 GL column (GE Healthcare) equilibrated with 50mM HEPES-NaOH pH 7.6 (4 ºC), 200mM NaCl, 1mM EDTA.

6.7 Negative-stain electron microscopy analysis of isolated UtpA

Purified UtpA was applied to glow discharged home-made carbon-coated copper grids and negatively stained with 0.75% uranyl formate as previously described (Ohi et al. 2004). Images were recorded on a Philips CM10 operated at an acceleration voltage of 100 kV equipped with a XR16-ActiveVu (AMT) camera at a nominal magnification of 52,000 and a calibrated pixel size of 2.8 Å at the specimen level.

6.8 On-column glutaraldehyde cross-linking of UtpA for negative stain analysis

UtpA was purified from Y143 with anti-GFP nanobody coupled beads as described in (**6.5**). To perform on-column cross-linking (Shukla et al. 2014), 500 µl of 0.25% glutaraldehyde solution was pre-injected onto a Superose 6 Increase 10/300

GL column (GE Healthcare) equilibrated with 50mM HEPES-NaOH pH 7.6 (4 ºC), 200mM NaCl, 1mM EDTA and run for 20 minutes at a flowrate of 0.2 ml/minute. The run was paused and after stringent washing of the injection loop, UtpA was loaded and the run continued at the same flow rate. Cross-linked UtpA fractions were used for preparation of negative stain electron microscopy grids.

6.9 DSS cross-linking analysis of UtpA

Peak fractions of size-exclusion chromatography purified UtpA (in 50mM HEPES-NaOH pH 7.6, 200mM NaCl, 1mM EDTA) were pooled (total volume 2 ml) and split into 200 µl cross-linking reactions. To each 200-µl aliquot 0.8 µl of DiSuccinimidylSuberate (DSS; 50 mM, 1:1 molar ratio mixture of DSS-H12 and DSS-D12, Creative Molecules Inc.) was added to yield a final DSS concentration of 0.2 mM. Samples were incubated at 25 ºC for 30 minutes with 400 r.p.m. constant shaking. The cross-linking reaction was quenched with 50mM ammonium bicarbonate. Cross-linked samples were precipitated by adding methanol to a final concentration of 90% and overnight incubation at -80 ºC. Precipitated cross-linked UtpA was collected in one tube by repeated centrifugation of the precipitated solution at 21,000 g, 4 ºC for 15 minutes. The resulting pellet was washed once with 1ml cold 90% methanol, air-dried and resuspended in 50 ml of 1X NuPAGE LDS buffer (Thermo Fisher Scientific).

DSS cross-linked UtpA complexes in LDS buffer were reduced with 25mM DTT, alkylated with 100mM 2-chloroacetamide, separated by SDS-PAGE using several lanes of a 4–12% Bis-Tris gel, and stained with Coomassie-blue. The gel region corresponding to the cross-linked complexes was sliced and digested in-gel overnight with trypsin to generate cross-linked peptides. After digestion, the peptide mixture was acidified and extracted from the gel as previously described (Shi et al. 2014; Shi et al. 2015). Analyses by LC-ESI-MS were performed either directly on the extracted peptides or following fractionation by size-exclusion chromatography (Leitner et al. 2012) or high pH reverse-phase chromatography. Peptides were loaded onto an EASY-Spray column (Thermo Fisher Scientific, either ES800: 15 cm 75 μm ID, PepMap C18, 3 μm or ES801: 15 cm 50 μm ID, PepMap RSLC C18, 2 μm) via an EASY-nLC 1000. MS and MS/MS analyses were carried out on a Q Exactive Plus mass spectrometer (Thermo Fisher Scientific). MS/MS analyses of the top 8 precursors in each full scan used the following parameters: resolution: 17,500 (at 200 Th); AGC target: 2×10^5; maximum injection time: 800 ms; isolation width: 1.4m/z; normalized collision energy: 29%; 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 pLink (Yang et al. 2012). All spectra reported here were manually verified (Shi et al. 2014; Shi et al. 2015). All Figures showing DSS cross-links found for UtpA were prepared using xiNET (Combe et al. 2015).

6.10 UV cross-linking analysis of cDNA (CRAC)

Yeast strains carrying C-terminal HTP tags (His_6-TEV-2xProteinA) on endogenous UtpA and UtpB subunits (Y133, Y134, Y137, Y138, Y139, Y140, Y144, Y158, Y159, Y160, Y162) that were actively growing in SD –TRP medium at 30 °C to an OD600 of 0.5, were irradiated at 254 nm UV for 100 seconds as described (Granneman et al. 2011). Cross-linked yeast strains were pelleted at 4000 x g at 4°C, for 5 minutes, washed with 1x PBS and flash frozen in liquid nitrogen. Pellets were stored at -80 ºC until use. Purification procedure of RNA-protein complexes and RT-PCR amplification of associated RNA fragments is described (Granneman et al. 2009; Granneman et al. 2011). cDNA libraries were sequenced on an Illumina HiSeq2500 at Edinburgh Genomics, University of Edinburgh. Illumina sequencing data were aligned to the yeast genome using Novoalign (http://www.novocraft.com). Bioinformatics analyses were performed as described using PyCRAC (Wlotzka et al. 2011; Webb et al. 2014).

6.11 Purification of the SSU processome

The small subunit processome was purified from Y232, a *Saccharomyces cerevisiae* BY4741 strain harboring a TEV-protease cleavable C-terminal GFP-tag on Utp1 (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 dropout media containing 2% raffinose (w/v) at 30 ºC prior to the addition

of 2% galactose (w/v) and subsequently grown to saturation. Cells were harvested by centrifugation at 4000 x *g* for 5 minutes at 4 ºC. The pellet was washed with ice cold ddH_2O, 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, 40,000 x *g* for 10 min and incubated with anti-GFP nanobody beads for 2 hours at 4 °C. Beads were washed three times in buffer A before bound protein complexes were eluted through TEV-protease cleavage (1 hour, 4 °C). The eluted supernatant was subjected to a second affinity purification step by incubation with streptavidin beads in buffer B (50 mM Tris-HCl, pH 7.7 (20 °C), 150 mM NaCl, 1 mM EDTA) for 1 hour 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.

6.12 Transcription and purification of 5' ETS-18S[1-194nt] control RNA

A DNA template for T7 RNA Polymerase dependent *in vitro* transcription encompassing nucleotides 1-894 of rDNA was generated by PCR from a plasmid containing the 5' ETS and the 5' domain of the 18S pre-rRNA (pSK_B221). The PCR product was isolated from a native 1% agarose (VWR) gel using the NucleoSpin Gel and PCR Clean Up Kit (Macherey-Nagel) and used as a template for an overnight T7 RNA Polymerase transcription reaction. The DNA template was removed by digestion with RNase-free DNase I for 1 hour at 37 °C. Subsequently, the transcription reaction was run on a preparative 4% urea-polyacrylamide gel and the band corresponding to the 5' ETS-18S[1-194nt] control RNA was cut out and soaked in 0.3 M NaOAc, pH 5.2 overnight at 4 °C. Extracted RNA was ethanol precipitated and stored at -80 °C until use.

6.13 SSU processome RNA isolation and analysis by northern blotting

RNA was extracted from the final D-Biotin elution of an SSU processome purification with 1 ml TRIzol (Life Technologies) according to the manufacturer's instructions. 0.5 μg of isolated SSU processome RNA and 0.4 μg of purified 5' ETS-18S[1-194nt] control RNA was separated on a denaturing 1.2% Formaldehyde-Agarose gel (SeaKem LE, Lonza). After staining the gel in a 2X SYBR Green II (Lonza) buffer solution (20 mM MOPS-NaOH pH 7.5, 5 mM sodium acetate, 1 mM EDTA) for 60 minutes, RNA species were visualized with a Gel Doc EZ Imager (Bio-Rad) and

subsequently transferred onto a cationized nylon membrane (Zeta-Probe GT, Bio-Rad) using downward capillary transfer. RNA was cross-linked to the membrane for northern blot analysis by UV irradiation at 254 nm with a total exposure of 120 millijoules/cm^2 in a UV Stratalinker 2400 (Stratagene). Cross-linked membranes were incubated with hybridization buffer (750 mM NaCl, 75 mM trisodium citrate, 1 % (w/v) SDS, 10% (w/v) dextran sulfate, 25% (v/v) formamide) at 65 °C for 30 minutes prior to addition of γ-32P-end-labeled DNA oligo nucleotide probes (**Table 6-3**). Probes were hybridized at 65 °C for 1 hour and then at 37 °C overnight. Membranes were washed once with wash buffer 1 (300 mM NaCl, 30 mM trisodium citrate, 1% (w/v) SDS) and once with wash buffer 2 (30 mM NaCl, 3 mM trisodium citrate, 1% (w/v) SDS) for 20 minutes each at 45 °C. Radioactive signal was detected by exposure of the washed membranes to a storage phosphor screen which was scanned with a Typhoon 9400 variable-mode imager (GE Healthcare).

Table 6-3 | List of northern blotting probes used in this study

Name	Sequence (5'-3')	Source
MS2 probe	ACATGGGTGATCCTCATG	Chapter 4
5' ETS probe	GGAATGGTACGTTTGATATCGCTGATTTGAGA	Chapter 3, Chapter 4
A0-A1 probe	CCCACCTATTCCCTCTTGC	Chapter 3, Chapter 4
A2-A3 probe	TGTTACCTCTGGGCCCCGATTG	Chapter 3
D-A2 probe	TCTTGCCCAGTAAAAGCTCTCATGC	Chapter 3
A3-5.8S probe	CCAGTTACGAAAATTCTTG	Chapter 3
U3 snoRNA probe	TAGATTCAATTTCGGTTTCTC	Chapter 3
18S (5'-domain)	GCTTAATCTTTGAGACAAGCATATGA	Chapter 3
18S (3'-minor)	GGAAACCTTGTTACGACTTTTAG	Chapter 3

6.14 Negative stain electron microscopy analysis of the SSU processome

4 µl of pure SSU processome at an absorbance of 0.6 mAU at 260 nm (Nanodrop 2000, Thermo Scientific) was applied on glow-discharged carbon coated grids (EMS, CF200-Cu). Grids were subsequently washed three times with water, twice with 2% (w/v) uranyl acetate and dried. Micrographs were collected on a Tecnai G2 spirit (FEI Company) operating at a voltage of 120 kV mounted with an UltraScan 895 CCD camera (Gatan Inc.). From 200 micrographs, 36,000 particles were picked using “swarm” and subjected to 2D classification in EMAN2 (Tang et al. 2007).

6.15 SSU processome cryo-EM grid preparation

Grids were prepared from separate SSU processome purifications to collect four separate datasets (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 mAU to 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 to 4 µl sample was applied onto glow-discharged grids (30 seconds 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), while 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.

6.16 SSU processome 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 previously (Chaker-Margot et al. 2017) and have been included in the larger dataset and re-processed together with ds2, ds3 and ds4 (**Figure 3.4** and **Appendix 7.2**). SerialEM (Mastronarde 2005) was used to automatically acquire micrographs with a defocus range of 0.6 to 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 seconds resulting in a total dose of 50 electrons per $Å^2$. Data collection parameters can be found in **Appendix 7.2**.

All 32 movie frames were gain corrected, aligned and dose weighted using MotionCor2 (Zheng et al. 2017). CTFFIND 4.1.5 was used for estimating the contrast transfer function (CTF) (Rohou & Grigorieff 2015). 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 (Kimanius et al. 2016) 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 5 classes using EMD-8473 (Chaker-Margot et al. 2017), low-pass filtered to 60 Å, as an input model (**Figure 3.4**). This 3D classification produced two good classes, both

combined containing a total of 284,213 particles. Auto-refinement 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 (**Figure 3.5**). 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 (**Figure 3.4** and **Figure 3.5**). 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 improved the density for UtpA significantly (**Figure 3.4** and **Figure 3.5**).

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 (**Figure 3.4** and **Figure 3.5**).

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 lead to an improvement of the resolution of this domain to 7.2 Å, allowing for the docking of crystal structures (**Figure 3.17** and **Appendix 7.1**).

Local resolution was estimated using Resmap (Kucukelbir et al. 2014). All computation was performed on a single Thinkmate SuperWorkstation 7048GR-TR equipped with four NVIDIA QUADRO P6000 video cards, 2 x Twenty-two Core Intel Xeon 2.40 GHz Processors, and 512 GB RAM.

6.17 DSS cross-linking of the SSU processome

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 minutes 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% followed by overnight incubation at -80 °C. Precipitated cross-linked SSU processomes were collected in one tube by repeated centrifugation at 21,000 x *g*, 4 °C for 30 minutes. The resulting pellet was washed three times with 1 ml cold 90% methanol, air-dried and resuspended in 50 µl of 1X 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 (Shi et al. 2014; Shi et al. 2015). 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 8 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 pLink (Yang et al. 2012). All spectra reported here were manually verified as previously (Shi et al. 2015) and all cross-links are listed in the Supplementary data table 1 of (Barandun et al. 2017). Cross-links were visualized using xiNET (Combe et al. 2015).

6.18 Building of the atomic model of the SSU processome and refinement

The poly-alanine model of the SSU processome provided by PDB 5TZS (Chaker-Margot et al. 2017) 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 initial template for the model building of some proteins (Kelley & Sternberg 2009). All model building was done in *Coot* (Emsley & Cowtan 2004). A complete list of templates, crystal structures and maps used to build the model can be found in **Appendix 7.1**.

The model was refined against overall map 1 (3.8 Å) in PHENIX with phenix.real_space_refine and secondary structure restraints for proteins and RNAs (Adams et al. 2010). Model statistics can be found in **Appendix 7.2**.

6.19 Map and model visualization for SSU processome figures

Structure analysis and figure preparation was performed using PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC and Chimera (Pettersen et al. 2004).

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).

6.20 Cloning of MS2-tagged 18S-rRNA domains and the MS2-3C-GFP construct

Defined fragments of the rDNA locus of *S. cerevisiae* strain BY4741 were cloned into a derivative of the pESC_URA vector (Agilent Technologies). The rDNA fragments were tagged with five MS2-aptamer stem loops at their 3´ ends and cloned downstream of a *gal1* promoter and upstream of a CYC terminator (**Table 6-1**). A modified MS2-coat protein (LeCuyer et al. 1995) fused to an N-terminal nuclear localization signal (NLS) as well as a hemagglutinin (HA) tag and a C-terminal 3C-protease-cleavable GFP (NLS-HA-MS2-3C-GFP) was cloned into a modified pESC plasmid suitable for genome integration in yeast (B079: *gal1* promotor, Geneticin resistance) (**Table 6-2).**

6.21 Generation of yeast strains for the purification of isolated MS2-tagged 18S rRNA domains

B079 was linearized and integrated into the *pep4* locus of *S. cerevisiae* strain BY4741 by homologous recombination to yield Y180. This strain was used to create all subsequent strains harboring C-terminal streptavidin-binding-peptide (SBP) tags on endogenous ribosome biogenesis factors (Utp10, Esf1, Kri1 and Mrd1) (**Table 6-2**). Homologous recombination-based integration events of C-terminal SBP tags were selected for with Hygromycin or Nourseothricin resistance cassettes. Yeast

transformation and genetic manipulations were performed according to established procedures.

6.22 Growth and expression of MS2-tagged pre-rRNA fragments

Yeast strains harboring genome-integrated galactose-inducible NLS-HA-MS2-3C-GFP in the *pep4* locus and a streptavidin binding peptide (SBP)-tagged ribosome assembly factor were transformed with plasmids carrying an URA marker and rDNA fragment constructs: strain Y186 (Utp10-SBP) was used for all experiments with plasmids containing the 5' ETS sequence (B117, B221 to B224 and B514), Y367 (Esf1-SBP) for the 5' domain containing plasmid (B506) and the 18S rRNA plasmid (B504), Y372 (Kri1-SBP) for the central domain plasmids (B502, B503) and Y374 (Mrd1-SBP) for the 3' major domain containing plasmid (B495) **(Table 6-1**).

Transformed cells were grown on -URA synthetic dropout media plates supplemented with 2% glucose (w/v) and the yeast strain corresponding antibiotics for 2 days at 30 °C. Selected colonies were picked and grown in pre-cultures (100 ml) of -URA synthetic dropout media supplemented with 2% raffinose (w/v) and the selection antibiotics to an optical density (OD) of 2 at 600 nm. Large-scale cultures were inoculated with the pre-cultures and grown at 30 °C for 16 hours to an OD of 4.5-7 in the presence of 2 % galactose. Yeast cells were pelleted, washed twice with ddH_2O and once with ddH_2O supplemented with protease inhibitors (PMSF,

Pepstatin A, E-64). The cells were flash frozen in liquid N_2 and subsequently cryo-ground using a planetary ball mill (Retsch PM100).

6.23 Purification of MS2-tagged pre-rRNA fragment containing RNPs

10-40 grams of cryo-ground yeast strain powder was resuspended in binding buffer (50 mM Tris-HCl, pH 7.5 (RT), 150 mM NaCl, 5 mM $MgCl_2$, 5% glycerol, 0.1% NP40) supplemented with protease inhibitors (PMSF, Pepstatin A, E-64). Insoluble fractions were pelleted at 40'000 *x g* and the supernatant was incubated with anti-GFP nanobody coupled sepharose (Chromotek) for 3 to 4 hours at 4 °C. Pre-rRNA fragments and their associated proteins were eluted by 3C-protease cleavage at 4 °C for 1 hour. Subsequently, eluted RNPs were applied to Streptavidin-coupled sepharose resin for 1 hour at 4 °C, washed four times and released from beads by incubation in 50 mM Tris-HCl, pH 7.5 (RT), 150 mM NaCl, 5 mM $MgCl_2$, 5 mM D-biotin for 30 minutes at 4 °C. Eluted samples where either directly used for RNA extraction, mass-spectrometry analysis, negative stain electron microscopy sample preparation or supplemented with 5 mM Putrescine, 1 mM Spermidine and 0.03% Triton X-100 for subsequent cryo-electron microscopy studies.

6.24 Negative-stain electron microscopy analysis of purified pre-SSU processome assembly intermediates

3.5 µl of pre-ribosomal particles purified from Y186 transformed with B117 (5' ETS), B221 (5' ETS-18S$^{5'\text{-domain}}$), B222 (5' ETS-18S$^{\text{central-domain}}$), B223 (5' ETS-18S$^{3'\text{major-domain}}$) or B224 (5' ETS-18S$^{3'\text{minor}}$ domain) were applied on glow-discharged carbon coated grids (EMS, CF200-Cu). Subsequently, grids were washed three times with ddH_2O, twice with 2% (w/v) 0.2 µm-filtered uranyl acetate and air and vacuum dried. Micrographs were acquired on a Philips CM10 operated at an acceleration voltage of 100 kV equipped with a XR16-ActiveVu (AMT) camera at a nominal magnification of 39,000 and a calibrated pixel size of 3.4 Å at the specimen level.

6.25 Cryo-EM grid preparation of pre-SSU processome assembly intermediates

Samples in elution buffer (50 mM Tris-HCl, pH 7.5 (RT), 150 mM NaCl, 5mM $MgICl_2$, 5 mM D-biotin) with absorbances of 0.35 mAU (5' ETS, B117), 0.65 mAU (5' ETS-18S$^{5'\text{-domain}}$, B221), 0.8 mAU (5' ETS-18S$^{\text{central-domain}}$, B222) at 260 nm (Nanodrop 2000, Thermo Scientific) were supplemented with 5 mM Putrescine, 1 mM Spermidine and 0.03% Triton X-100. 3.5 µl of sample was applied on glow-discharged lacey-carbon grids containing a thin carbon film (TED PELLA, Inc, Prod. No. 01824). Following a 15 second sample incubation period at close to 100 %

humidity, grids were blotted for 1.5 - 2.5 seconds with a blotting force of 0 and flash frozen in liquid ethane using a Vitrobot Mark IV (Thermo Fisher Scientific).

6.26 Cryo-EM data collection and processing for the 5' ETS, 5' ETS-18S$^{5'\text{-domain}}$ and 5' ETS-18S$^{\text{central domain}}$ particles

Cryo-EM data collection was performed either on a Talos Arctica or Titan Krios (Thermo Fisher Scientific) operated at 200 kV or 300 kV respectively, both mounted with a K2 Summit detector (Gatan, Inc). SerialEM (Mastronarde 2005) was used for automated data collection. Datasets of the 5' ETS (1199 micrographs, 36'000 x magnification, 1.2 Å pixel size, 8 electron per pixel and second) and the 5' domain particle (697 micrographs, 22'000 x magnification, 1.9 Å pixel size, 15 electron per pixel and second) were collected on the Talos Arctica and processed using Motioncor2 (Zheng et al. 2017) and RELION 2.1 (Kimanius et al. 2016). As a starting model, a Cryosparc (Punjani et al. 2017) generated initial model obtained from the 5' domain data set was used. While the 5' ETS particle could not be refined to a high resolution (~17 Å) due to heterogeneity in the sample, we were able to obtain a ~ 10 Å reconstruction of the 5' domain particle with similar overall structure but better resolved density for the A1 binding site of Sof1 and Utp7.

The central-domain particle dataset was acquired on the Titan Krios: 2750 movies with 32 frames over an exposure time of 8 seconds at a dose rate of 10 electrons per pixel and second (total dose of 31.25 $e^-/Å^2$) over a defocus range of 1

- 3.5 µm at 1.6 Å pixel size. Motioncor2 (Zheng et al. 2017) was used for gain normalization, beam-induced motion correction and dose-weighting. The contrast transfer function was estimated with CTFFIND-4.1.5 (Rohou & Grigorieff 2015). Removal of micrographs with bad CTF fits resulted in a total of 2592 micrographs used for reference-free particle picking with gautomatch (http://www.mrc-lmb.cam.ac.uk/kzhang/) yielding 275'080 particles. Particles were extracted with a box size of 360 pixel (576 Å) and subjected to 3D classification in RELION 2.1 with 3, 4, 5 and 7 classes and the 5' domain structure low-pass filtered to 60 Å as initial model (**Appendix 7.4**). From the different 3D classification runs, top classes were selected, combined and checked for duplicates resulting in 180'274 unique particles. These were used for focused refinement and postprocessing (b-factor -108.704) resulting in a final masked map (EMDB XXX) with an overall resolution of 4.4 Å (**Appendix 7.4**).

6.27 Model building of the 5' ETS particle

The structure of the 5' ETS particle moiety of the small subunit processome (PDB 5WLC, (Barandun et al. 2017)) was used as initial coordinates for model building in the 4.4 Å map of the central-domain particle. The entire starting coordinates of the 5' ETS particle part from the SSU processome were docked as one entity into the density using UCSF Chimera (Pettersen et al. 2004). All subunits were then individually rigid body fitted and trimmed in *COOT* (Emsley & Cowtan 2004). Major differences were observed in the six-subunit complex UtpB, which required rigid

body docking of individual subunit domains (C-terminal tetramerization domains and tandem β-propellers). Additional helical density next to Utp12 could not be unambiguously assigned and therefore a poly-Alanine helix was placed. The structure was refined using phenix.real_space_refine (Adams et al. 2010) with secondary structure restraints obtained from the model and (PDB 5WLC, (Barandun et al. 2017)). Removing of side-chains resulted in a poly-alanine model with residue information. Data collection and processing information as well as model statistics can be found in (**Appendix 7.5**).

Molecular graphics and analyses were performed with UCSF ChimeraX (Pettersen et al. 2004), developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH R01-GM129325 and P41-GM103311.

6.28 RNA extraction and northern blots for SSU processome assembly intermediates and isolated 18S rRNA domains

Total cellular RNA was extracted from 0.2 gram of frozen yeast cells after lysis by glass-beads beating in 1 ml Trizol® (Ambion). To isolate RNA from purified pre-ribosomal particles, 0.5 ml Trizol® (Ambion) was added to the final D-biotin elutions and the extraction was performed according to the manufacturer's instructions. For the analysis of pre-rRNA processing states by northern blotting, 3 µg total cellular RNA or ~100 ng of RNA extracted from purified RNPs, were loaded in each lane of

a 1.2 % agarose formaldehyde-formamide gel and separated at 75V for 2.5 hours. After running, the separated RNA was transferred onto a cationized nylon membrane (Zeta-Probe GT, Bio-Rad) using downward capillary transfer and cross-linked to the membrane for northern blotting analysis by UV irradiation at 254 nm with a total exposure of 120 millijoules/cm2 in a UV Stratalinker 2400 (Stratagene).

Prior to the addition of γ-32P-end-labeled DNA oligo nucleotide probes (table S1), cross-linked membranes were incubated with hybridization buffer (750 mM NaCl, 75 mM trisodium citrate, 1 % (w/v) SDS, 10% (w/v) dextran sulfate, 25% (v/v) formamide) for 30 minutes at 65 °C Labeled hybridization probes were incubated with the membrane first at 65 °C for 1 hour and then at 37 - 45 °C overnight. Blotted membranes were washed once with wash buffer 1 (300 mM NaCl, 30 mM trisodium citrate, 1% (w/v) SDS) and once with wash buffer 2 (30 mM NaCl, 3 mM trisodium citrate, 1% (w/v) SDS) for 30 minutes each at 45 °C, before radioactive signal was read out by exposure of the washed membranes to a storage phosphor screen, which was subsequently scanned with a Typhoon 9400 variable-mode imager (GE Healthcare).

6.29 Mass spectrometry sample processing and data analysis of MS2-tagged RNPs

Purified RNP samples were dried and dissolved in 8 M urea/0.1 M ammonium bicarbonate/10mM DTT. After reduction, cysteines where alkylated in 30 mM iodoacetamide. Proteins where digested with LysC (LysC, Endoproteinase LysC, Wako Chemicals) in less than 4 M urea followed by trypsination (Trypsin Gold, Promega) in less than 2 M Urea. Digestions were halted by adding TFA and digests was desalted (Rappsilber et al. 2007) and analyzed by reversed phase nano-LC-MS/MS using a Fusion Lumos (Thermo Scientific) operated in high/high mode.

Data were quantified and searched against the *S. cerevisiae* Uniprot protein database (October 2018) concatenated with the MS2-coat-protein sequence and common contaminations. For the search and quantitation, MaxQuant v. 1.6.0.13 (Cox & Mann 2008) was used. Oxidation of methionine and protein N-terminal acetylation were allowed as variable modifications and all cysteines where treated as being carbamidomethylated. Peptide matches were filtered using a false discovery ratios (FDR) for peptide spectrum matches and proteins of 2% and 1% respectively.

Data analysis: Log2 transformed intensity Based Absolute Quantitation (iBAQ) values (Schwanhäusser et al. 2011) were used for the analysis. To access loading across the 6 conditions, each in technical triplicate, three metabolic enzymes (Enolase 2, Galactokinase and Glyceraldehyde-3-phosphate dehydrogenase) which

we considered to be ‘innocent bystanders’ where examined. The signal for the three proteins were comparable in-between the 18 samples. Hereafter we used the MS2-protein signal to adjust the iBAQ values for each of the 78 ribosome biogenesis proteins of interest. Data were processed using Perseus v 1.6.0.7 (Tyanova et al. 2016).

Chapter 7 | Appendix

7.1 Organization of the PDB model of the SSU processome (PDB 5WLC)

Subcomplex	Chain ID	SegID	Molecule name	Total residues	Domains	Modeling	Initial PDB template	Maps used for building
RNA	L0	L0	ETS rRNA	700		atomic (488 residues)	Manual building	Core map
	L1	L1	pre-18S rRNA	1800		atomic (1025 residues)	4V88	All
	L2	L2	U3 snoRNA	333		atomic (169 residues)	5GIN	Core map and overall map 2
ribosomal proteins	L3	L3	rpS18_US13	146		atomic (2-7, 11-87, 99-120), poly-Alanine (121-128)	4V88	Core map
	L4	L4	rpS4_ES4	261		atomic (14-241)	4V88	Overall map 2
	L5	L5	rpS5_US7	225		atomic (13-225)	4V88	Core map
	L6	L6	rpS6_ES6	236		atomic (1-78, 91-125)	4V88	Overall map 2
	L7	L7	rpS7_ES7	190		atomic (4-97, 117-186)	4V88	Overall map 2
	L8	L8	rpS8_ES8	200		atomic (2-123, 150-197)	4V88	Overall map 2
	L9	L9	rpS9_US4	197		atomic (10-184)	4V88	Overall map 2
	LC	LC	rpS16_US9	143		atomic (3-127)	4V88	Core map
	LD	LD	rpS11_US17	156		atomic (6-19, 31-143)	4V88	Overall map 2
	LE	LE	rpS22_US8	130		atomic (4-130)	4V88	Overall map 2
	LF	LF	rpS24_ES24	135		atomic (4-93)	4V88	Overall map 2
	LG	LG	rpS28_ES28	67		atomic (5-67)	4V88	Core map
	SR	SR	rpS23_US12	145		atomic (41-144)	4V88	Core map
	NF	NF	rpS13_US15	151		side-chain trimmed crystal structure (28-151)	4V88	Central domain map
	NG	NG	rpS14_US11	137		side-chain trimmed crystal structure (10-120)	4V88	Central domain map
UtpA	LH	LH	Utp17	896	tandem-WD40	atomic (9-101, 104-159, 166-730, 734-788)	4NSX	Core map
					CTD	poly-Alanine (818-831), atomic (846-896)	Manual building	
	LI	LI	Utp8	713	WD40	poly-Alanine (27-358)	5KDO	Core map and 3' domain map
					CTD	poly-Alanine (457-548), atomic (549-617, 631-681)	Manual building	
					Strand shared with Utp4	atomic (711-723)	Manual building	
	LJ	LJ	Utp15	513	WD40	atomic (2-334)	Phyre model based on 3DM0	Core map and 3' domain map
					Linker	atomic (353-385)	Manual building	
					CTD	atomic (386-512)	Manual building	
	LK	LK	Utp9	575	CTD	poly-Alanine (391-426), atomic (428-515)	Manual building	Core map and 3' domain map
	LL	LL	Utp5	643	WD40	atomic (5-36,46-67, 72-262, 269-332)	Phyre model based on 3DM0	Core map and 3' domain map
					Linker	atomic (333-372)	Manual building	
					CTD	atomic (434-557)	Manual building	
	LN	LN	Utp4	776	tandem-WD40	atomic (29-405, 413-500, 508-600, 608-666, 731-776), poly-Alanine (676-689)	2YMU	3' domain map
	LM	LM	Utp10 N-term	1769	Helical repeat	atomic (2-425), poly-Alanine (426-432)	Manual building	Core map
UtpB	LO	LO	Utp1	923	tandem-WD40	atomic (2-223, 245-706)	Phyre model based on 4NSX	Core map
					CTD	atomic (708-856)	Phyre model based on 5ICA	
	LP	LP	Utp6	440	Helical repeat	atomic (2-207, 264-290, 296-326), poly-Alanine (327-431)	Manual building	Core map
	LQ	LQ	Utp12	943	tandem-WD40	atomic (5-28, 38-264, 270-287, 321-504, 520-687), poly-Alanine (288-317)	1NR0	Core map
					CTD	atomic (745-775, 787-943), poly-Alanine (776-786)	5ICA	
	LR	LR	Utp13	817	tandem-WD40	poly-Alanine (7-650)	1NR0	Core map
					CTD	atomic (649-707, 717-817)	5ICA	
	LS	LS	Utp18	594	Utp6 interaction domain	atomic (13-28)	Manual building	Core map
					Utp10 interaction domain	atomic (29-44)	Manual building	
					Utp21 interaction domain	atomic (122-184)	Manual building	
					Linker	atomic (202-226)	Manual building	
					WD40	atomic (227-326, 334-594)	5IC7	
	LT	LT	Utp21	939	tandem-WD40	atomic (19-673)	4NSX	Core map
					Linker	atomic (702-766)	Manual building	
					CTD	atomic (810-939)	5ICA	
U3 snoRNP	SA	SA	Nop56	504	NTD	atomic (3-163)	Phyre model based on 3ID5 and 2OZB	Core map
					CTD	atomic (164-227, 264-336, 370-414), poly-Alanine (231-259)		
	SB	SB	Nop58	511	NTD	poly-Alanine (1-125), atomic (133-154)	Phyre model based on 3ID5 and 2OZB	Core map
					CTD	atomic (159-403)		
					C-terminal extension	atomic (404-437)	Manual building	
	SC	SC	Nop1	327		atomic (80-128, 134-326)	Phyre model based on 2IPX	Core map
	SD	SD	Nop1	327		atomic (85-125,138-324)	Phyre model based on 2IPX	Core map
	SE	SE	Snu13	126		atomic (5-125)	2ALE	Core map
	SF	SF	Snu13	126		atomic (5-125)	2ALE	Core map
	SG	SG	Rrp9	573	WD40	atomic (107-166, 184-375, 394-570)	4J0X	Overall map 2

Subcomplex	Chain ID	SegID	Molecule name	Total residues	Domains	Modeling	Initial PDB template	Maps used for building
other ribosome assembly factors	LU	LU	Sof1	489	NTD	atomic (1-55)	Manual building	Overall map 2
					WD40	atomic (56-159, 168-377)	3DM0	
					CTD	atomic (378-465)	Manual building	
	LV	LV	Enp2	707	WD40	atomic (2-22, 43-363), sidechain-trimmed (23-42)	3DM0	Overall map 2
	LW	LW	Utp7	554	NTD	atomic (36-102)	Manual building	Overall map 2
					WD40	atomic (103-408)	Phyre model based on 3DM0	
					CTD	atomic (409-473)	Manual building	
	LX	LX	Kre33	1056	NTD	atomic (3-154, 180-208), poly-Alanine (155-172, 209-221)	Phyre model based on 2ZPA	Overall map 2
					Helicase	poly-Alanine (225-475)		
					N-acetyltransferase	poly-Alanine (476-820)		
					CTD	poly-Alanine (821-932)		
	LY	LY	Kre33	1056	NTD	poly-Alanine (3-66, 92-172,180-221)	Phyre model based on 2ZPA	Overall map 2
					Helicase	poly-Alanine (225-475)		
					N-acetyltransferse	poly-Alanine (476-820)		
					CTD	poly-Alanine (821-932)		
	LZ	LZ	Imp3	183		atomic (2-183)	5A2Q chain J	Core map
	SH	SH	Rcl1	367		atomic (7-366)	4CLQ	Overall map 2
	SI	SI	Bms1	1183	GTPase	atomic (42-312, 749-1035)	Phyre model based on 5IW7	Core map and Overall map 2
					Kre33 interaction peptide	atomic (344-376), poly-Alanine (389-417)	Phyre model based on 5IW7	
					Rcl1 interaction peptide	atomic (548-569, 606-636)	4CLQ	
					CTD	atomic (1036-1164)	Phyre model based on 5IW7	
	SJ	SJ	Emg1	252		atomic (29-55, 63-251)	3OIJ	3' domain map
	SK	SK	Emg1	252		atomic (16-56, 63-251)	3OIJ	3' domain map
	SL	SL	Utp24	189		atomic (7-31, 41-189)	Phyre model based on 4MJ7	Core map
	SM	SM	Imp4	290		atomic (1-63,72-290)	5A53	Core map
	SN	SN	Utp30	274	ribosomal protein L1-like	atomic (11-257)	4LQ4	Core map
	SO	SO	Pno1	274		sidechain-trimmed KH domain (96-236), atomic (237-274)	4QMF (Krr1)	Core map
	SP	SP	Utp20	2493		poly-Alanine helices (907-2256)	Manual building	Overall map 2
	SQ	SQ	Fcf2	217		atomic (50-84, 101-154, 171-216)	Manual building	Core map
	SS	SS	Utp14	899		atomic (276-350, 828-897), poly-Alanine (351-363, 369-408)	Manual building	Core map and Overall map 2
	ST	ST	Nop14	810	NTD	atomic (5-16, 43-66, 78-116), poly-Alanine (201-252, 264-316)	Manual building	3' domain map
					helical repeat	poly-Alanine (382-418), atomic (419-578, 585-696)		
					CTD	atomic (697-806)		
	SU	SU	Noc4	552	helical repeat	poly-Alanine (1-147, 162-169, 514-534), atomic (148-161, 170-513)	Manual building	3' domain map
	SV	SV	Rrt14	206		atomic (178-203), poly-Alanine (112-148)	Manual building	Core map
	SY	SY	Utp11	250		atomic (3-172, 180-250)	Manual building	Core map
	SZ	SZ	Enp1	483	helical repeat	sidechain-trimmed (205-465)	Crystal structure 5WYK	3' domain map
	NA	NA	Mpp10	593		atomic (295-386, 425-539)	Manual building	Core map
	NB	NB	Sas10 (aka Utp3)	610		atomic (431-441,493-610), poly-Alanine (442-454)	Manual building	Core map and 3' domain map
	NC	NC	Lcp5	357		atomic (222-357)	Manual building	Overall map 2
	ND	ND	Bud21	214		atomic (155-214), poly-Alanine (101-108, 110-122)	Manual building	3' domain map
	NE	NE	Faf1	346		atomic (200-330), poly-Alanine (91-98,101-119)	Manual building	Overall map 2
	NH	NH	Utp22	1237		sidechain-trimmed crystal structure (97-1237)	Crystal structure 4M5D	Central domain map
	NI	NI	Rrp7	297		sidechain-trimmed crystal structure (3-189)	Crystal structure 4M5D	Central domain map
	NJ	NJ	Rrp5	1729	TRP repeat	sidechain-trimmed crystal structure (1457-1721)	Crystal structure 5WWM	Central domain map
	NK	NK	Krr1	316	KH domain	sidechain-trimmed crystal structure (38-212)	Crystal structure 4QMF	Central domain map
	SX	SX	unassigned peptides	126		poly-Alanine	Manual building	Core map

7.2 Data collection table for the cryo-EM reconstruction of the SSU processome

	Overall map 1 (5WLC, 8859)	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 (um)	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 B-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 (°)	0.87				
Validation					
MolProbity score	1.86				
Clashscore	7.68				
Poor rotamers (%)	0.72				
Good sugar puckers (%)	97.74				
Ramachandran					
Favored (%)	94.60				
Allowed (%)	5.34				
Outliers (%)	0.06				

7.3 Mass spectrometry analysis of the 5' ETS, 5' ETS-18S$^{5' \text{ domain}}$ and 5' ETS-18S$^{\text{central domain}}$ particles

5' ETS

Protein Name	Σ# Unique Peptides	Area
PABP	63	3.856E10
NOP1	37	2.260E10
SMT3	9	1.928E10
RRP9	60	1.695E10
BUD21	21	1.547E10
UTP13	54	1.531E10
UTP7	60	1.269E10
UTP18	33	1.199E10
UTP4	91	1.154E10
UTP15	43	8.894E9
IMP4	36	8.743E9
MPP10	58	8.463E9
UTP5	33	8.072E9
UTP9	55	7.971E9
SOF1	51	6.995E9
SNU13	9	6.837E9
UTP12	58	6.796E9
UTP8	40	6.209E9
Strepavidin	14	5.546E9
UTP17	51	5.510E9
UTP10	71	5.489E9
UTP21	75	5.462E9
SAS10	44	5.284E9
UTP1	63	5.227E9
UTP6	41	5.091E9
NOP56	28	3.646E9
NOP58	29	3.226E9
UTP11	27	3.222E9
IMP3	28	2.484E9
UTP24	10	7.645E8
PYC2	17	5.376E8
PYC1	12	5.376E8
HSP71	13	5.032E8
FCF2	22	3.572E8
ADH1	16	3.321E8
ACAC	70	3.097E8
G3P3	13	3.064E8
G3P1	4	2.883E8
EF1A	22	2.732E8
HSP76	3	2.719E8
HSP75	2	2.719E8
HSP72	7	2.547E8
YRA1	20	2.506E8
RS5	8	2.127E8
HSP74	1	2.012E8
RL7A	10	1.718E8
RS6A	7	1.697E8
RRT14	11	1.693E8
EMP47	14	1.589E8
H2B1	6	1.527E8
RL19A	5	1.521E8
RL4A	2	1.409E8
RL4B	2	1.409E8
SS120	8	1.380E8
RL401	1	1.380E8
RL44A	3	1.234E8
RS19A	8	1.212E8
RS16A	8	1.118E8
RL11A	4	1.049E8
RL8A	2	1.005E8
RL8B	2	1.005E8

5' ETS - 18S 5' domain

Protein Name	Σ# Unique Peptides	Area
SMT3	9	1.655E9
UTP18	18	1.501E9
PABP	45	1.072E9
IMP4	17	7.449E8
UTP5	24	6.557E8
SOF1	37	6.243E8
SAS10	36	5.923E8
UTP13	31	5.316E8
UTP4	43	4.907E8
MPP10	36	4.781E8
UTP12	51	4.736E8
UTP17	42	3.916E8
UTP15	27	3.815E8
RRP9	31	3.758E8
BFR2	30	3.647E8
UTP8	29	3.443E8
UTP21	46	3.392E8
DBP4	36	3.379E8
UTP10	48	3.037E8
ESF1	40	2.597E8
Strepavidin	10	2.446E8
NOP56	27	2.341E8
UTP7	43	2.210E8
UTP11	21	2.132E8
NOP58	26	2.059E8
FCF2	10	1.790E8
SNU13	3	1.734E8
UTP1	30	1.700E8
NOP1	23	1.667E8
UTP9	30	1.599E8
ENP2	29	1.388E8
RS4A	17	1.368E8
BUD21	16	1.204E8
IMP3	15	1.203E8
DBP8	14	1.134E8
UTP6	31	1.101E8
RS11A	7	1.058E8
ESF2	23	9.384E7
UTP24	5	7.060E7
LCP5	19	6.830E7
RS8A	9	6.166E7
RRP36	12	6.095E7
EFG1P	10	5.069E7
YRA1	5	3.331E7
HSP71	8	3.181E7
HSP75	18	2.632E7
BUD22	13	2.452E7
HSP72	3	2.406E7
ACAC	28	1.944E7
RL35A	7	1.745E7
HAS1	8	1.596E7
RL7A	3	1.520E7
RL7B	1	1.520E7
RRT14	5	1.227E7
RL8A	10	1.211E7
G3P1	3	1.189E7
RS16A	10	1.181E7
G3P3	6	1.159E7
PYC2	5	1.115E7
PYC1	2	1.006E7

5' ETS - 18S central domain

Protein Name	Σ# Unique Peptides	Area
PABP	63	1.546E10
NOP1	37	9.892E9
Strepavidin	14	7.863E9
RRP9	60	7.091E9
UTP13	54	6.743E9
SOF1	51	6.370E9
UTP7	60	5.293E9
BUD21	21	5.108E9
UTP4	91	4.617E9
SNU13	9	4.224E9
ESF1	45	4.191E9
KRI1	33	4.028E9
MPP10	58	3.952E9
BFR2	36	3.745E9
UTP5	33	3.692E9
IMP4	36	3.210E9
UTP17	51	3.122E9
UTP12	58	3.037E9
UTP11	27	3.016E9
KRR1	34	2.983E9
SAS10	44	2.979E9
UTP18	33	2.925E9
UTP9	55	2.918E9
UTP15	43	2.816E9
FCF2	22	2.800E9
NHP2	9	2.784E9
DBP4	44	2.508E9
UTP8	40	2.403E9
RS14A	1	2.299E9
RS14B	1	2.299E9
UTP6	41	2.298E9
UTP21	75	2.259E9
NOP56	28	2.180E9
UTP10	71	2.075E9
UTP1	63	1.818E9
IMP3	28	1.736E9
LCP5	27	1.686E9
UTP23	20	1.662E9
NOP58	29	1.590E9
UTP24	10	1.550E9
SMT3	9	1.541E9
CBF5	37	1.383E9
ENP2	57	1.352E9
ESF2	29	1.239E9
RS4A	20	1.201E9
RRP36	24	1.183E9
EFG1P	21	1.148E9
NOP10	5	1.110E9
YRA1	20	1.106E9
RRP5	111	1.102E9
GAR1	5	9.360E8
RS8A	21	8.556E8
UTP22	40	8.528E8
ROK1	27	8.388E8
RPP7	23	7.306E8
RS11A	14	6.883E8
PYC2	17	5.056E8
PYC1	12	4.981E8
RL18A	3	4.912E8
HSP71	13	4.550E8
H2B1	6	3.754E8
ACAC	70	3.610E8
HSP76	3	3.470E8
HSP75	2	3.470E8
RL7A	10	3.452E8
BUD22	21	3.311E8
HSP72	7	3.201E8
RL4A	2	2.741E8
RL4B	2	2.741E8
RS13	10	2.719E8
EF1A	22	2.463E8
RL25	6	2.201E8
G3P3	13	2.106E8
RS6A	7	2.055E8
SRP40	6	2.019E8
RL26B	1	1.902E8
RL26A	1	1.902E8
G3P1	4	1.879E8
RLA2	4	1.852E8
FYV7	9	1.792E8
MAK21	26	1.688E8
RL401	1	1.684E8
IML1	2	1.658E8
RL3	22	1.616E8
HAS1	12	1.611E8
ADH1	16	1.610E8
DBP8	10	1.584E8
RL34B	4	1.564E8
NOC2	17	1.539E8
RL13B	9	1.482E8
H2A1	3	1.479E8
RL12A	7	1.432E8
RS5	8	1.370E8
GLO2	18	1.368E8
HSP74	1	1.237E8
SSBP1	13	1.143E8
HSP77	17	1.077E8

7.4 Cryo-EM processing strategy for the 5’ ETS particle

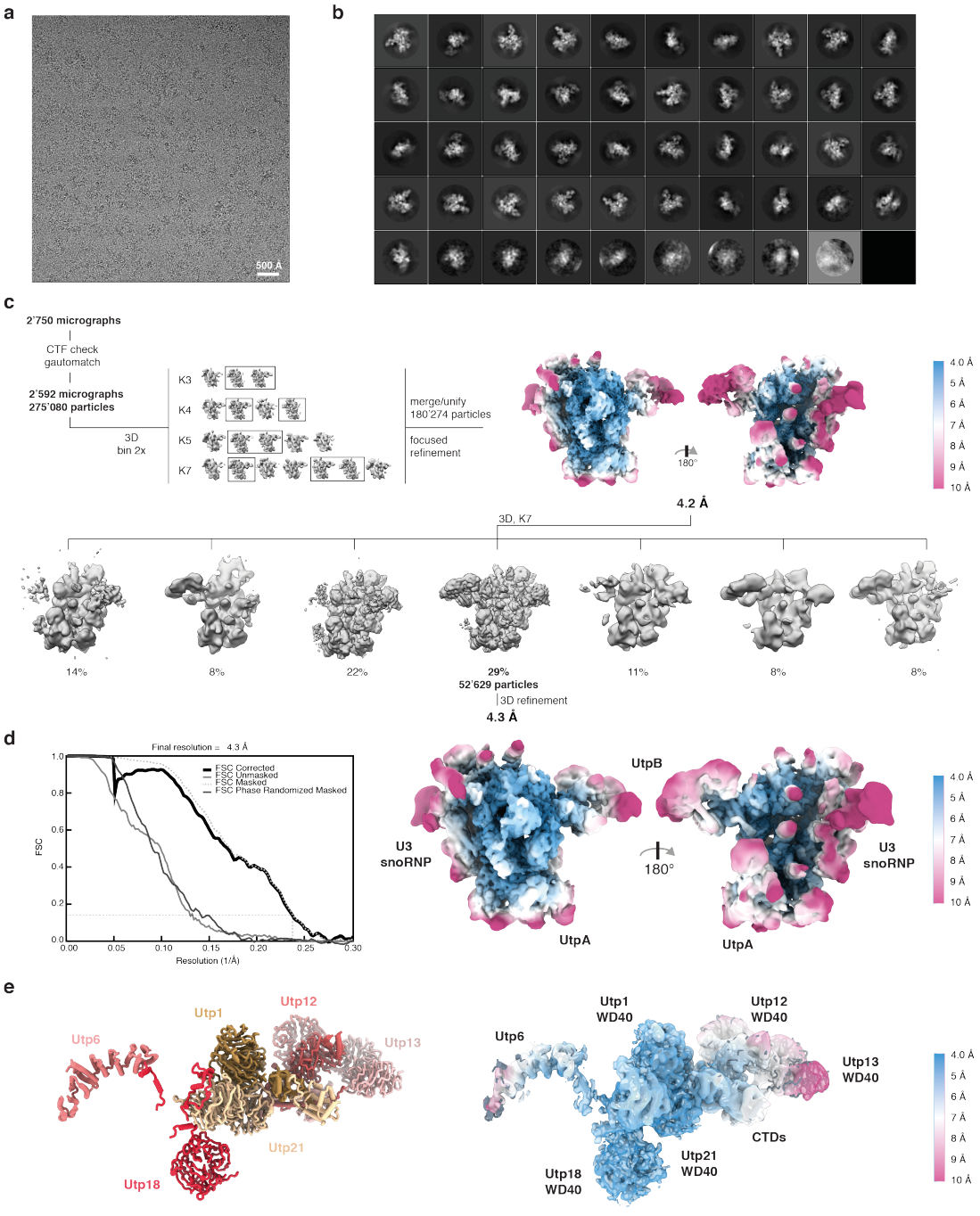


Figure 7.1 | Processing strategy and resolution estimation for the 5' ETS particle cryo-EM reconstruction.

[**a**] Representative micrograph and [**b**] 2D class averages of the 5' ETS particle. [**c**] Manual inspection of 2,750 collected micrographs resulted in 2,592 micrographs of good quality used to extract 275,080 particles picked with gautomatch. These particles were subjected to 3D classification with varying numbers of classes. Particles from boxed out classes (180,275 particles combined) were used for 3D refinement. A subsequent 3D classification with seven classes resulted in a single class (52'629 particles) with better-resolved features in the periphery of the particle, which yielded a reconstruction at 4.3 Å. [**d**] Overall and local resolution estimation of the 5' ETS particle. An FSC value of 0.143 is indicated. RELION-3 (Zivanov et al. 2018) was used to estimate local resolution. A volume filtered to the local resolution is shown. [**e**] UtpB model (left) with subunits colored and labeled, and the corresponding density fit of the model with local resolution color-coded (right).

7.5 Preliminary data collection table for the cryo-EM structure of the 5' ETS particle

	Structure of the 5' ETS particle
Data collection and processing	
Magnification	
Voltage (kV)	300
Pixel size (Å)	1.6
Electron exposure (e- / Å^2)	31.25
Defocus range (um)	1-3.5
Symmetry imposed	C1
Initial particle images	275'080
Final particle images	52'629
Resolution (Å)	4.3
FSC threshold	0.143
Map sharpening B-Factor (Å^2)	-64.77
Refinement	
Initial model used	5WLC
Model composition	
Non hydrogen atoms	64,186
Protein residues	10309
RNA bases	569
Ligands	0
R.m.s. deviations	
Bond length (Å)	0.01
Angles (°)	1.25
Validation	
MolProbity score	1.71
Clashscore	5.27
Poor rotamers (%)	0.0
Good sugar puckers (%)	98.41
Ramachandran	
Favored (%)	93.39
Allowed (%)	6.56
Outliers (%)	0.05

7.6 Re-evaluation of RNA-protein cross-linking data confirms binding sites of individual UtpA and UtpB subunits

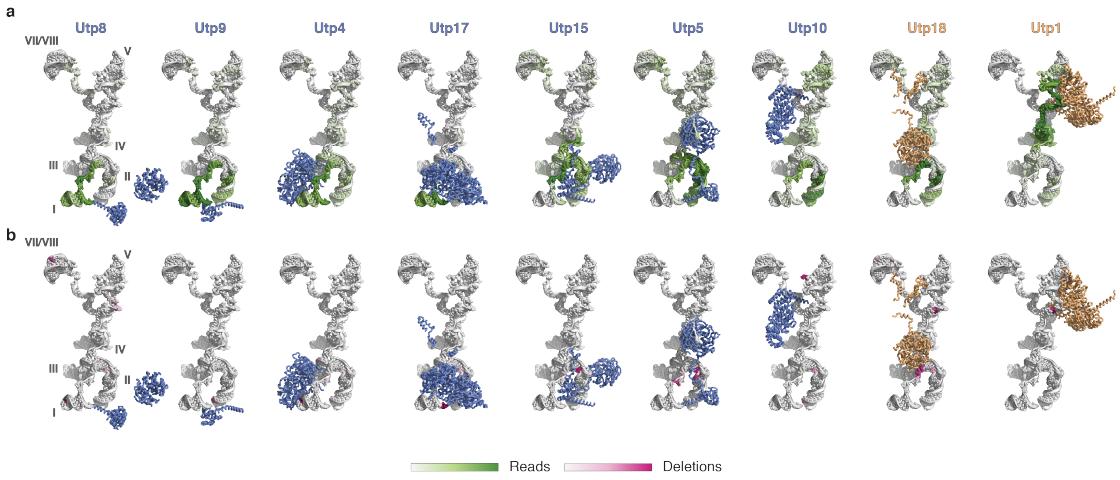


Figure 7.2 | RNA-protein cross-linking sites of UtpA and UtpB displayed on the SSU processome structure.

De-duplicated RNA sequencing reads **[a]** (green) and number of deletions **[b]** (pink) obtained previously (Chapter II, (Hunziker et al. 2016)) are plotted onto the 5' ETS RNA (white). A color gradient is used to indicated low (white) or high numbers of reads (green) or deletions (pink). 5' ETS helices are labeled with roman numerals in the first panels of [a] and [b]. The subunits used for RNA-protein cross-linking are shown as ribbon representation (UtpA subunits in blue, UtpB subunits in orange) in their SSU processome position.

Chapter 8 | References

Adams, P.D. et al., 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography*, 66(Pt 2), pp.213–21.

Albert, B. et al., 2016. A Molecular Titration System Coordinates Ribosomal Protein Gene Transcription with Ribosomal RNA Synthesis. *Molecular Cell*, 64(4), pp.720–733.

Albert, V. & Hall, M.N., 2015. mTOR signaling in cellular and organismal energetics. *Current Opinion in Cell Biology*, 33(1), pp.55–66.

Allmang, C. et al., 2000. Degradation of ribosomal RNA precursors by the exosome. *Nucleic acids research*, 28(8), pp.1684–91.

Atzorn, V., Fragapane, P. & Kiss, T., 2004. U17/snR30 Is a Ubiquitous snoRNA with Two Conserved Sequence Motifs Essential for 18S rRNA Production. *Molecular and Cellular Biology*, 24(4), pp.1769–1778.

Ban, N. et al., 2000. The complete atomic structure of the large ribosomal subunit at 2.4 Å Resolution. *Science*, 289(5481), pp.905–920.

Barandun, J. et al., 2017. The complete structure of the small-subunit processome. *Nature Structural & Molecular Biology*, 24(11), pp.1–36.

Barandun, J., Hunziker, M. & Klinge, S., 2018. Assembly and structure of the SSU processome — a nucleolar precursor of the small ribosomal subunit. *Current Opinion in Structural Biology*, 49(Macromolecular assemblies), pp.85–93.

Barrio-Garcia, C. et al., 2016. Architecture of the Rix1-Rea1 checkpoint machinery during pre-60S-ribosome remodeling. *Nature Structural and Molecular Biology*, 23(1), pp.37–44.

Baßler, J. et al., 2010. The AAA-ATPase Rea1 Drives Removal of Biogenesis Factors during Multiple Stages of 60S Ribosome Assembly. *Molecular Cell*, 38(5), pp.712–721.

Baudin-Baillieu, A. et al., 1997. Functional analysis of Rrp7p, an essential yeast protein involved in pre-rRNA processing and ribosome assembly. *Molecular and cellular biology*, 17(9), pp.5023–32.

Baudin-Baillieu, A. et al., 2009. Nucleotide modifications in three functionally important regions of the Saccharomyces cerevisiae ribosome affect translation accuracy. *Nucleic Acids Research*, 37(22), pp.7665–7677.

Baxter-Roshek, J.L., Petrov, A.N. & Dinman, J.D., 2007. Optimization of ribosome structure and function by rRNA base modification. *PLoS ONE*, 2(1).

Beltrame, M. & Tollervey, D., 1995. Base pairing between U3 and the pre-ribosomal RNA is required for 18S rRNA synthesis. *The EMBO Journal*, 14(17), pp.4350–4356.

Ben-Shem, A. et al., 2011. The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science (New York, N.Y.)*, 334(6062), pp.1524–9.

Birkedal, U. et al., 2015. Profiling of ribose methylations in RNA by high-throughput sequencing. *Angewandte Chemie - International Edition*, 54(2), pp.451–455.

Black, J.J. et al., 2018. Utp14 interaction with the small subunit processome. *RNA (New York, N.Y.)*, 24(9), pp.1214–1228.

Bleichert, F. et al., 2006. The PINc domain protein Utp24, a putative nuclease, is required for the early cleavage steps in 18S rRNA maturation. *Proceedings of the National Academy of Sciences*, 103(25), pp.9464–9469.

Bradatsch, B. et al., 2012. Structure of the pre-60S ribosomal subunit with nuclear export factor Arx1 bound at the exit tunnel. *Nature Structural and Molecular Biology*, 19(12), pp.1234–1241.

Braglia, P. et al., 2010. Role of the RNA/DNA kinase Grc3 in transcription termination by RNA polymerase I. *EMBO Reports*, 11(10), pp.758–764.

Braglia, P., Kawauchi, J. & Proudfoot, N.J., 2011. Co-transcriptional RNA cleavage provides a failsafe termination mechanism for yeast RNA polymerase I. *Nucleic Acids Research*, 39(4), pp.1439–1448.

Chaker-Margot, M. et al., 2017. Architecture of the yeast small subunit processome. *Science (New York, N.Y.)*, 355(6321), pp.1–9.

Chaker-Margot, M. et al., 2016. Architecture of the yeast small subunit processome. *Science*.

Chaker-Margot, M., 2018. Assembly of the small ribosomal subunit in yeast: Mechanism and regulation. *RNA (New York, N.Y.)*, 24(7), pp.881–891.

Chaker-Margot, M. et al., 2015. Stage-specific assembly events of the 6-MDa small-subunit processome initiate eukaryotic ribosome biogenesis. *Nat Struct Mol Biol*, 22(11), pp.920–923.

Chen, W. et al., 2017. Stepwise assembly of the earliest precursors of large ribosomal subunits in yeast. *Nucleic Acids Research*, 45(11), pp.6837–6847.

Chen, Z. et al., 2018. Structural Insights into Mdn1, an Essential AAA Protein Required for Ribosome Biogenesis. *Cell*, 175, pp.822–834.

Cheng, J. et al., 2017. 3.2-Å-resolution structure of the 90S preribosome before A1 pre-rRNA cleavage. *Nature Structural and Molecular Biology*, 24(11), pp.954–964.

Choe, S.Y., Schultz, M.C. & Reeder, R.H., 1992. *In vitro* definition of the yeast RNA polymerase I promoter. *Nucleic Acids Research*, 20(2), pp.279–285.

Chu, S. et al., 1994. The RNA of RNase MRP is required for normal processing of ribosomal RNA. *Proc Natl Acad Sci U S A*, 91(2), p.659–63.

Clemons, W.M.J. et al., 1999. Structure of a bacterial 30S ribosomal subunit at 5.5 A resolution. *Nature*, 400(August), pp.833–840.

Combe, C., Fischer, L. & Rappsilber, J., 2015. xiNET: cross-link network maps with residue resolution. *MCP*, (410), pp.1–32.

Costello, J.L. et al., 2011. The C-terminal region of the exosome-associated protein Rrp47 is specifically required for box C/D small nucleolar RNA 3′-maturation. *Journal of Biological Chemistry*, 286(6), pp.4535–4543.

Cox, J. & Mann, M., 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, 26(12), pp.1367–1372.

Crick, F.H.C., 1958. On protein synthesis. In *Symp Soc Exp Biol*. p. 8.

Decatur, W.A. & Fournier, M.J., 2002. rRNA modifications and ribosome function. *Trends in Biochemical Sciences*, 27(7), pp.344–351.

Delprato, A. et al., 2014. Crucial role of the Rcl1p-Bms1p interaction for yeast pre-ribosomal RNA processing. *Nucleic Acids Research*, 42(15), pp.10161–10172.

Dever, T.E., Dinman, J.D. & Green, R., 2018. Translation elongation and recoding in eukaryotes. *Cold Spring Harbor Perspectives in Biology*, 10(8), pp.1–20.

Dever, T.E., Kinzy, T.G. & Pavitt, G.D., 2016. Mechanism and regulation of protein synthesis in Saccharomyces cerevisiae. *Genetics*, 203(1), pp.65–107.

Dragon, F. et al., 2002. A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature*, 417(6892), pp.967–70.

Dutca, L.M., Gallagher, J.E.G. & Baserga, S.J., 2011. The initial U3 snoRNA:pre-rRNA base pairing interaction required for pre-18S rRNA folding revealed by in vivo chemical probing. *Nucleic Acids Research*, 39(12), pp.5164–5180.

Emsley, P. & Cowtan, K., 2004. Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography*, 60(12 I), pp.2126–2132.

Eppens, N.A. et al., 1999. The roles of Rrp5p in the synthesis of yeast 18S and 5.8S rRNA can be functionally and physically separated. *RNA (New York, N.Y.)*, 5(6), pp.779–793.

Fatica, A., Tollervey, D. & Dlakić, M., 2004. PIN domain of Nob1p is required for D-site cleavage in 20S pre-rRNA. *RNA*, 10(11), pp.1698–1701.

Fayet-Lebaron, E. et al., 2009. 18S rRNA processing requires base pairings of snR30 H/ACA snoRNA to eukaryote-specific 18S sequences. *EMBO Journal*, 28(9), pp.1260–1270.

Ferreira-Cerca, S. et al., 2014. Dominant Rio1 kinase/ATPase catalytic mutant induces trapping of late pre-40S biogenesis factors in 80S-like ribosomes. *Nucleic Acids Research*, 42(13), pp.8635–8647.

French, S.L. et al., 2003. In Exponentially Growing Saccharomyces cerevisiae Cells, rRNA Synthesis Is Determined by the Summed RNA Polymerase I Loading Rate Rather than by the Number of Active Genes. *Molecular and cellular biology*, 23(5), pp.1558–1568.

Fromm, L. et al., 2017. Reconstitution of the complete pathway of ITS2 processing at the pre-ribosome. *Nature Communications*, 8(1), pp.1–11.

Gamalinda, M. et al., 2014. A hierarchical model for assembly of eukaryotic 60S ribosomal subunit domains. *Genes & Development*, 28(2), pp.198–210.

Ganot, P., Bortolin, M.L. & Kiss, T., 1997. Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell*, 89(5), pp.799–809.

Gasse, L., Flemming, D. & Hurt, E., 2015. Coordinated Ribosomal ITS2 RNA Processing by the Las1 Complex Integrating Endonuclease, Polynucleotide Kinase, and Exonuclease Activities. *Molecular Cell*, 60(5), pp.808–815.

Geerlings, T.H. et al., 2003. Rio2p: an evolutionarily conserved, low abundant protein kinase essential for processing of 20S pre-rRNA in Saccharomyces cerevisiae. *Journal of Biological Chemistry*, 278(25), pp.22537–22545.

Gelperin, D. et al., 2001. Bms1p, a novel GTP-binding protein, and the related Tsr1p are required for distinct steps of 40S ribosome biogenesis in yeast. *RNA (New York, N.Y.)*, 7(9), pp.1268–1283.

Ghalei, H. et al., 2015. Hrr25/CK1δ-directed release of Ltv1 from pre-40S ribosomes is necessary for ribosome assembly and cell growth. *Journal of Cell Biology*, 208(6), pp.745–759.

Ghalei, H. et al., 2017. The ATPase Fap7 Tests the Ability to Carry Out Translocation-like Conformational Changes and Releases Dim1 during 40S Ribosome Maturation. *Molecular Cell*, 67(6), pp.990–1000.

Graille, M. & Séraphin, B., 2012. Surveillance pathways rescuing eukaryotic ribosomes lost in translation. *Nature Reviews Molecular Cell Biology*, 13(11), pp.727–735.

Grandi, P. et al., 2002. 90S pre-ribosomes include the 35S pre-rRNA, the U3 snoRNP, and 40S subunit processing factors but predominantly lack 60S synthesis factors. *Molecular Cell*, 10(1), pp.105–115.

Granneman, S. et al., 2010. Cracking pre-40S ribosomal subunit structure by systematic analyses of RNA-protein cross-linking. *EMBO Journal*, 29(12), pp.2026–2036.

Granneman, S. et al., 2009. Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proceedings of the National Academy of Sciences*, 106(24), pp.9613–9618.

Granneman, S. et al., 2006. The nucleolar protein Esf2 interacts directly with the DExD/H box RNA helicase, Dbp8, to stimulate ATP hydrolysis. *Nucleic Acids Research*, 34(10), pp.3189–3199.

Granneman, S., Nandineni, M.R. & Baserga, S.J., 2005. The putative NTPase Fap7 mediates cytoplasmic 20S pre-rRNA processing through a direct interaction with Rps14. *Mol. Cell Biol.*, 25(23), pp.10352–10364.

Granneman, S., Petfalski, E. & Tollervey, D., 2011. A cluster of ribosome synthesis factors regulate pre-rRNA folding and 5.8S rRNA maturation by the Rat1 exonuclease. *EMBO Journal*, 30(19), pp.4006–4019.

Greber, B.J. et al., 2016. Insertion of the Biogenesis Factor Rei1 Probes the Ribosomal Tunnel during 60S Maturation. *Cell*, 164(1–2), pp.91–102.

El Hage, A. et al., 2008. Efficient termination of transcription by RNA polymerase I requires the 5′ exonuclease Rat1 in yeast. *Genes and Development*, 22(8), pp.1069–1081.

Harms, J. et al., 2001. High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell*, 107(5), pp.679–688.

Hashem, Y. & Frank, J., 2018. The Jigsaw Puzzle of mRNA Translation Initiation in Eukaryotes: A Decade of Structures Unraveling the Mechanics of the Process. *Annual Review of Biophysics*, 47(1), pp.125–151.

Henry, Y. et al., 1994. The 5' end of yeast 5.8S rRNA is generated by exonucleases from an upstream cleavage site. *The EMBO Journal*, 13(10), pp.2452–2463.

Heuer, A. et al., 2017. Cryo-EM structure of a late pre-40S ribosomal subunit from Saccharomyces cerevisiae. *eLife*, 6, pp.1–14.

Hoareau-Aveilla, C. et al., 2012. Utp23p is required for dissociation of snR30 small nucleolar RNP from preribosomal particles. *Nucleic Acids Research*, 40(8), pp.3641–3652.

van Hoof, A., Lennertz, P. & Parker, R., 2000. Yeast exosome mutants accumulate 3'-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs. *Mol Cell Biol*, 20(2), pp.441–452.

Hughes, J.M. & Ares, M., 1991. Depletion of U3 small nucleolar RNA inhibits cleavage in the 5' external transcribed spacer of yeast pre-ribosomal RNA and impairs formation of 18S ribosomal RNA. *The EMBO Journal*, 10(13), pp.4231–4239.

Hunziker, M. et al., 2016. UtpA and UtpB chaperone nascent pre-ribosomal RNA and U3 snoRNA to initiate eukaryotic ribosome assembly. *Nature communications*, (May), pp.1–10.

Ito, S. et al., 2014. A single acetylation of 18S rRNA is essential for biogenesis of the small ribosomal subunit in Saccharomyces cerevisiae. *Journal of Biological Chemistry*, 289(38), pp.26201–26212.

Jansa, P. & Grummt, I., 1999. Mechanism of transcription termination: PTRF interacts with the largest subunit of RNA polymerase I and dissociates paused transcription complexes from yeast and mouse. *Molecular and General Genetics*, 262(3), pp.508–514.

Jansen, R., Tollervey, D. & Hurt, E.C., 1993. A U3 snoRNP protein with homology to splicing factor PRP4 and G beta domains is required for ribosomal RNA processing. *The EMBO Journal*, 12(6), pp.2549–2558.

Kallstrom, G., Hedges, J. & Johnson, A., 2003. The putative GTPases Nog1p and Lsg1p are required for 60S ribosomal subunit biogenesis and are localized to the nucleus and cytoplasm, respectively. *Molecular and cellular biology*, 23(12), pp.4344–55.

Kappel, L. et al., 2012. Rlp24 activates the AAA-ATPase Drg1 to initiate cytoplasmic pre-60S maturation. *Journal of Cell Biology*, 199(5), pp.771–782.

Karbstein, K., 2013. Quality control mechanisms during ribosome maturation. *Trends in Cell Biology*, 23(5), pp.242–250.

Karbstein, K., Jonas, S. & Doudna, J.A., 2005. An essential GTPase promotes assembly of preribosomal RNA processing complexes. *Molecular Cell*, 20(4), pp.633–643.

Kater, L. et al., 2017. Visualizing the Assembly Pathway of Nucleolar Pre- Article Visualizing the Assembly Pathway of Nucleolar Pre-60S Ribosomes. *Cell*, 171(7), p.1599–1610.e13.

Kawauchi, J. et al., 2008. Budding yeast RNA polymerases I and II employ parallel mechanisms of transcriptional termination. *Genes and Development*, 22(8), pp.1082–1092.

Kelley, L. a & Sternberg, M.J.E., 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nature protocols*, 4(3), pp.363–71.

Kempers-Veenstra, A.E. et al., 1986. 3'-End formation of transcripts from the yeast rRNA operon. *The EMBO journal*, 5(10), pp.2703–2710.

Keys, D.A. et al., 1996. Multiprotein transcription factor UAF interacts with the upstream element of the yeast RNA polymerase I promoter and forms a stable preinitiation complex. *Genes and Development*, 10(7), pp.887–903.

Khoshnevis, S. et al., 2016. 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 Biology*, 14(6), pp.1–25.

Kimanius, D. et al., 2016. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *eLife*, 5(November), pp.1–21.

King, T.H. et al., 2003. Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. *Molecular Cell*, 11(2), pp.425–435.

Kiss-Laszlo, Z. et al., 1996. Site-Specific Ribose Methylation of Preribosomal RNA: A Novel Function for Small Nucleolar RNAs. *Cell*, 85(June), pp.1077–1088.

Klinge, S. et al., 2011. Crystal structure of the eukaryotic 60S ribosomal subunit in complex with initiation factor 6. *Science (New York, N.Y.)*, 334(6058), pp.941–8.

Konikkat, S. & Woolford, J.L., 2017. Principles of 60S ribosomal subunit assembly emerging from recent studies in yeast. *Biochemical Journal*, 474(2), pp.195–214.

Kornprobst, M. et al., 2016. Architecture of the 90S Pre-ribosome: A Structural View on the Birth of the Eukaryotic Ribosome. *Cell*, 166(2), pp.380–393.

Kos-Braun, I.C., Jung, I. & Koš, M., 2017. Tor1 and CK2 kinases control a switch between alternative ribosome biogenesis pathways in a growth-dependent manner. *PLoS Biology*, 15(3).

Koš, M. & Tollervey, D., 2005. The putative RNA helicase Dbp4p is required for release of the U14 snoRNA from preribosomes in Saccharomyces cerevisiae. *Molecular Cell*, 20(1), pp.53–64.

Koš, M. & Tollervey, D., 2010. Yeast Pre-rRNA Processing and Modification Occur Cotranscriptionally. *Molecular Cell*, 37(6), pp.809–820.

Kressler, D. et al., 2008. The AAA ATPase Rix7 powers progression of ribosome biogenesis by stripping Nsa1 from pre-60S particles. *Journal of Cell Biology*, 181(6), pp.935–944.

Krogan, N.J. et al., 2004. High-definition macromolecular composition of yeast RNA-processing complexes. *Molecular cell*, 13(2), pp.225–39.

Kucukelbir, A., Sigworth, F.J. & Tagare, H.D., 2014. Quantifying the local resolution of cryo-EM density maps. *Nature Methods*, 11(1), pp.63–65.

Kufel, J., Dichtl, B. & Tollervey, D., 1999. Yeast Rnt1p is required for cleavage of the pre-ribosomal RNA in the 3' ETS but not the 5' ETS. *RNA (New York, N.Y.)*, 5(7), pp.909–917.

Kulkens, T. et al., 1991. The yeast RNA polymerase I promoter: Ribosomal DNA sequences involved in transcription initiation and complex formation *in vitro*. *Nucleic Acids Research*, 19(19), pp.5363–5370.

Lafontaine, D. et al., 1994. The DIM1 gene responsible for the conserved m26Am26A dimethylation in the 3′-terminal loop of 18 S rRNA is essential in yeast. *Journal of Molecular Biology*, 241(3), pp.492–497.

Lafontaine, D.L.J. et al., 1998. The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. *Genes and Development*, 12(4), pp.527–537.

Lebaron, S. et al., 2012. Proofreading of pre-40S ribosome maturation by a translation initiation factor and 60S subunits. *Nature Structural and Molecular Biology*, 19(8), pp.744–753.

Lebaron, S. et al., 2013. Rrp5 binding at multiple sites coordinates Pre-rRNA processing and assembly. *Molecular Cell*, 52(5), pp.707–719.

LeCuyer, K.A., Behlen, L.S. & Uhlenbeck, O.C., 1995. Mutants of the Bacteriophage MS2 Coat Protein That Alter Its Cooperative Binding to RNA. *Biochemistry*, 34(33), pp.10600–10606.

Lee, S.J. & Baserga, S.J., 1999. Imp3p and Imp4p, two specific components of the U3 small nucleolar ribonucleoprotein that are essential for pre-18S rRNA processing. *Molecular and cellular biology*, 19(8), pp.5441–52.

Leidig, C. et al., 2014. 60S ribosome biogenesis requires rotation of the 5S ribonucleoprotein particle. *Nature communications*, 5, p.3491.

Leitner, A. et al., 2012. Expanding the Chemical Cross-Linking Toolbox by the Use of Multiple Proteases and Enrichment by Size Exclusion Chromatography. *Molecular & Cellular Proteomics*, 11(3), p.M111.014126.

Lemay, V. et al., 2011. Identification of novel proteins associated with yeast snR30 small nucleolar RNA. *Nucleic Acids Research*, 39(22), pp.9659–9670.

Leulliot, N. et al., 2008. Structure of the Yeast tRNA m7G Methylation Complex. *Structure*, 16(1), pp.52–61.

Liang, W.Q. & Fournier, M.J., 1995. U14 base-pairs with 18S rRNA: A novel snoRNA interaction required for rRNA processing. *Genes and Development*, 9(19), pp.2433–2443.

Liang, X., Liu, Q. & Fournier, M.J., 2007. rRNA Modifications in an Intersubunit Bridge of the Ribosome Strongly Affect Both Ribosome Biogenesis and Activity. *Molecular Cell*, 28(6), pp.965–977.

Liang, X.H., Liu, Q. & Fournier, M.J., 2009. Loss of rRNA modifications in the decoding center of the ribosome impairs translation and strongly delays pre-rRNA processing. *RNA (New York, N.Y.)*, 15(9), pp.1716–1728.

Lin, J. et al., 2013. An RNA-Binding Complex Involved in Ribosome Biogenesis Contains a Protein with Homology to tRNA CCA-Adding Enzyme. *PLoS Biology*, 11(10).

Lindahl, L., Archer, R.H. & Zengel, J.M., 1991. A new rRNA processing mutant of *Saccharomyces cerevisiae*. *Nucleic acids research*, 20(2), pp.295–301.

Ling, C. & Ermolenko, D.N., 2016. Structural insights into ribosome translocation. *Wiley Interdisciplinary Reviews: RNA*, 7(5), pp.620–636.

Lygerou, Z. et al., 1996. hPop1: an autoantigenic protein subunit shared by the human RNase P and RNase MRP ribonucleoproteins. *Embo J*, 15(21), pp.5936–5948.

Lygerou, Z. et al., 1994. The POP1 gene encodes a protein component common to the RNase MRP and RNase P ribonucleoproteins. *Genes and Development*, 8(12), pp.1423–1433.

Ma, C. et al., 2017. Structural snapshot of cytoplasmic pre-60S ribosomal particles bound by Nmd3, Lsg1, Tif6 and Reh1. *Nature Structural and Molecular Biology*, 24(3), pp.214–220.

Makarova, K.S. et al., 2005. Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucleic Acids Research*, 33(14), pp.4626–4638.

Malyutin, A.G. et al., 2017. Nmd3 is a structural mimic of eIF5A, and activates the cpGTPase Lsg1 during 60S ribosome biogenesis. *The EMBO Journal*, 36(7), pp.854–868.

Marmier-Gourrier, N. et al., 2011. 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 Research*, 39(22), pp.9731–9745.

Martin, R. et al., 2014. A pre-ribosomal RNA interaction network involving snoRNAs and the Rok1 helicase. *RNA (New York, N.Y.)*, 20(8), pp.1173–1182.

Mastronarde, D.N., 2005. Automated electron microscope tomography using robust prediction of specimen movements. *Journal of Structural Biology*, 152(1), pp.36–51.

Matsuo, Y. et al., 2014. Coupled GTPase and remodelling ATPase activities form a checkpoint for ribosome export. *Nature*, 505(7481), pp.112–116.

Mayer, C. & Grummt, I., 2006. Ribosome biogenesis and cell growth: mTOR coordinates transcription by all three classes of nuclear RNA polymerases. *Oncogene*, 25(48), pp.6384–6391.

McCaughan, U.M. et al., 2016. Pre-40S ribosome biogenesis factor Tsr1 is an inactive structural mimic of translational GTPases. *Nature Communications*, 7(June), pp.1–8.

Miller, O.L. & Beatty, B.R., 1969. Visualization of nucleolar genes. *Science*, 164(3882), pp.955–957.

Mitchell, P., 2010. Rrp47 and the function of the Sas10/C1D domain. *Biochemical Society transactions*, 38(4), pp.1088–92.

Mitchell, P. et al., 1997. The exosome: A conserved eukaryotic RNA processing complex containing multiple 3'→5' exoribonucleases. *Cell*, 91(4), pp.457–466.

Mougey, E.B. et al., 1993. The terminal balls characteristic of eukaryotic rRNA transcription units in chromatin spreads are rRNA processing complexes. *Genes and Development*, 7(8), pp.1609–1619.

Ni, J., Tien, A.L. & Fournier, M.J., 1997. Small Nucleolar RNAs Direct Site-Specific Synthesis of Pseudouridine in Ribosomal RNA. *Cell*, 89(May), pp.565–573.

Nissen, P. et al., 2000. The Structural Basis of Ribosome Activity in Peptide Bond Synthesis. *Science*, 289(August), p.493.

Noller, H.F., Hoffarth, V. & Zimniak, L., 1992. Unusual resistance of peptidyl transferase to protein extraction procedures. *Science*, 256(5062), pp.1416–1419.

Nomura, M. et al., 2004. Transcription of rDNA in the yeast *Saccharomyces cerevisiae*. In *Molecular Biology Intelligence …*. Landes Bioscience, pp. 1–25.

Oakes, M. et al., 1998. Mutational analysis of the structure and localization of the nucleolus in the yeast *Saccharomyces cerevisiae*. *Journal of Cell Biology*, 143(1), pp.23–34.

Oakes, M. et al., 1993. Structural alterations of the nucleolus in mutants of *Saccharomyces cerevisiae* defective in RNA polymerase I. *Molecular and Cellular Biology*, 13(4), pp.2441–2455.

Oeffinger, M. et al., 2009. Rrp17p Is a Eukaryotic Exonuclease Required for 5′ End Processing of Pre-60S Ribosomal RNA. *Molecular Cell*, 36(5), pp.768–781.

Ohi, M. et al., 2004. Negative Staining and Image Classification – Powerful Tools in Modern Electron Microscopy. *Biological Procedures Online*, 6(1), pp.23–34.

Osheim, Y.N. et al., 2004. Pre-18S Ribosomal RNA Is Structurally Compacted into the SSU Processome Prior to Being Cleaved from Nascent Transcripts in *Saccharomyces cerevisiae*. *Molecular C*, 16, pp.943–954.

Palade, G.E., 1955. A small particulate component of the cytoplasm. *J Biophys Biochem Cytol.*, 1(1), pp.59–68.

Pérez-Fernández, J. et al., 2007. The 90S preribosome is a multimodular structure that is assembled through a hierarchical mechanism. *Molecular and cellular biology*, 27(15), pp.5414–29.

Pérez-Fernández, J., Martín-Marcos, P. & Dosil, M., 2011. Elucidation of the assembly events required for the recruitment of Utp20, Imp4 and Bms1 onto nascent pre-ribosomes. *Nucleic acids research*, 39(18), pp.8105–21.

Pertschy, B. et al., 2007. Cytoplasmic Recycling of 60S Preribosomal Factors Depends on the AAA Protein Drg1. *Molecular and Cellular Biology*, 27(19), pp.6581–6592.

Pettersen, E.F. et al., 2004. UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), pp.1605–1612.

Peyroche, G., 2000. The recruitment of RNA polymerase I on rDNA is mediated by the interaction of the A43 subunit with Rrn3. *The EMBO Journal*, 19(20), pp.5473–5482.

Philippi, A. et al., 2010. TOR-dependent reduction in the expression level of Rrn3p lowers the activity of the yeast RNA Pol I machinery, but does not account for the strong inhibition of rRNA production. *Nucleic Acids Research*, 38(16), pp.5315–5326.

Polikanov, Y.S. et al., 2015. Structural insights into the role of rRNA modifications in protein synthesis and ribosome assembly. *Nature Structural and Molecular Biology*, 15(4), pp.1–8.

Prieto, J.L. & McStay, B., 2007. Recruitment of factors linking transcription and processing of pre-rRNA to NOR chromatin is UBF-dependent and occurs independent of transcription in human cells. *Genes and Development*, 21(16), pp.2041–2054.

Punjani, A. et al., 2017. CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*, 14(3), pp.290–296.

Rabl, J. et al., 2011. Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. *Science (New York, N.Y.)*, 331(6018), pp.730–6.

Raman, N., Weir, E. & Müller, S., 2016. The AAA ATPase MDN1 Acts as a SUMO-Targeted Regulator in Mammalian Pre-ribosome Remodeling. *Molecular Cell*, 64(3), pp.607–615.

Rappsilber, J., Mann, M. & Ishihama, Y., 2007. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature Protocols*, 2(8), pp.1896–1906.

Rohou, A. & Grigorieff, N., 2015. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology*, 192(2), pp.216–221.

Rudra, D. et al., 2007. Potential Interface between Ribosomal Protein Production and Pre-rRNA Processing. *Molecular and Cellular Biology*, 27(13), pp.4815–4824.

Sahasranaman, A. et al., 2011. Assembly of *Saccharomyces cerevisiae* 60S ribosomal subunits: Role of factors required for 27S pre-rRNA processing. *EMBO Journal*, 30(19), pp.4020–4032.

Sandmeier, J.J. et al., 2002. RPD3 is required for the inactivation of yeast ribosomal DNA genes in stationary phase. *EMBO Journal*, 21(18), pp.4959–4968.

Sanghai, Z.A. et al., 2018. Modular assembly of the nucleolar pre-60S ribosomal subunit. *Nature*, 556(7699), pp.126–129.

Sardana, R. et al., 2015. The DEAH-box Helicase Dhr1 Dissociates U3 from the Pre-rRNA to Promote Formation of the Central Pseudoknot. *PLoS Biology*, 13(2), pp.1–25.

Saveanu, C. et al., 2001. Nog2p, a putative GTPase associated with pre-60S subunits and required for late 60S maturation steps. *EMBO Journal*, 20(22), pp.6475–6484.

Scaiola, A. et al., 2018. Structure of a eukaryotic cytoplasmic pre-40S ribosomal subunit. *The EMBO Journal*, p.e98499.

Schäfer, T. et al., 2006. Hrr25-dependent phosphorylation state regulates organization of the pre-40S subunit. *Nature*, 441(7093), pp.651–655.

Schäfer, T. et al., 2003. The path from nucleolar 90S to cytoplasmic 40S pre-ribosomes. *EMBO Journal*, 22(6), pp.1370–1380.

Schluenzen, F. et al., 2000. Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. *Cell*, 102(5), pp.615–623.

Schmitt, M.E. & Clayton, D.A., 1993. Nuclear RNase MRP is required for correct processing of pre-5.8S rRNA in *Saccharomyces cerevisiae*. *Molecular and cellular biology*, 13(12), pp.7935–41.

Schneider, C. et al., 2012. Transcriptome-wide Analysis of Exosome Targets. *Molecular Cell*, 48(3), pp.422–433.

Schuller, J.M. et al., 2018. Structure of the nuclear exosome captured on a maturing preribosome. *Science*, 360(6385), pp.219–222.

Schwanhäusser, B. et al., 2011. Global quantification of mammalian gene expression control. *Nature*, 473(7347), pp.337–342.

Segerstolpe, Å. et al., 2013. Multiple RNA interactions position Mrd1 at the site of the small subunit pseudoknot within the 90S pre-ribosome. *Nucleic Acids Research*, 41(2), pp.1178–1190.

Sharma, K. & Tollervey, D., 1999. Base pairing between U3 small nucleolar RNA and the 5' end of 18S rRNA is required for pre-rRNA processing. *Molecular and cellular biology*, 19(9), pp.6012–9.

Sharma, S. et al., 2017. Specialized box C/D snoRNPs act as antisense guides to target RNA base acetylation. *PLoS Genetics*, 13(5), pp.1–23.

Sharma, S. et al., 2015. Yeast Kre33 and human NAT10 are conserved 18S rRNA cytosine acetyltransferases that modify tRNAs assisted by the adaptor Tan1/THUMPD1. *Nucleic Acids Research*, 43(4), pp.2242–2258.

Shi, Y. et al., 2015. A strategy for dissecting the architectures of native macromolecular assemblies. *Nature Methods*, 12(12), pp.1135–1138.

Shi, Y. et al., 2014. Structural characterization by cross-linking reveals the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex. *Molecular & Cellular Proteomics*.

Shuai, K. & Warner, J.R., 1991. A temperature sensitive mutant of *Saccharomyces cerevisiae* defective in pre-rRNA processing. *Nucleic Acids Research*, 19(18), pp.5059–5064.

Shukla, A.K. et al., 2014. Visualization of arrestin recruitment by a G-protein-coupled receptor. *Nature*, 512(7513), pp.218–222.

Sloan, K.E. et al., 2017. Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biology*, 14(9), pp.1138–1152.

Soltanieh, S. et al., 2015. DEAD-Box RNA Helicase Dbp4 Is Required for Small-Subunit Processome Formation and Function. *Molecular and Cellular Biology*, 35(5), pp.816–830.

Soltanieh, S., Lapensée, M. & Dragon, F., 2014. Nucleolar proteins Bfr2 and Enp2 interact with DEAD-box RNA helicase Dbp4 in two different complexes. *Nucleic Acids Research*, 42(5), pp.3194–3206.

Strunk, B.S. et al., 2012. A translation-like cycle is a quality control checkpoint for maturing 40S ribosome subunits. *Cell*, 150(1), pp.111–121.

Strunk, B.S. et al., 2011. Ribosome assembly factors prevent premature translation initiation by 40S assembly intermediates. *Science*, 333(6048), pp.1449–1453.

Sun, Q. et al., 2017. Molecular architecture of the 90S small subunit pre-ribosome. *eLife*, 6, pp.1–28.

Talkish, J. et al., 2016. Disruption of ribosome assembly in yeast blocks cotranscriptional pre-rRNA processing and affects the global hierarchy of ribosome biogenesis. *RNA (New York, N.Y.)*, 22(6), pp.852–866.

Tang, G. et al., 2007. EMAN2: An extensible image processing suite for electron microscopy. *Journal of Structural Biology*, 157(1), pp.38–46.

Taoka, M. et al., 2016. The complete chemical structure of *Saccharomyces cerevisiae* rRNA: Partial pseudouridylation of U2345 in 25S rRNA by snoRNA snR9. *Nucleic Acids Research*, 44(18), pp.8951–8961.

Thoms, M. et al., 2015. The Exosome Is Recruited to RNA Substrates through Specific Adaptor Proteins. *Cell*, 162(5), pp.1029–1038.

Thomson, E. & Tollervey, D., 2010. The Final Step in 5.8S rRNA Processing Is Cytoplasmic in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*, 30(4), pp.976–984.

Tishchenko, S. et al., 2012. High-resolution crystal structure of the isolated ribosomal L1 stalk. *Acta Crystallographica Section D: Biological Crystallography*, 68(8), pp.1051–1057.

Tomecki, R. et al., 2015. hUTP24 is essential for processing of the human rRNA precursor at site A1, but not at site A0. *RNA Biology*, 12(9), pp.1010–1029.

Torchet, C., Jacq, C. & Hermann-Le Denmat, S., 1998. Two mutant forms of the S1/TPR-containing protein Rrp5p affect the 18S rRNA synthesis in *Saccharomyces cerevisiae*. *RNA (New York, N.Y.)*, 4(12), pp.1636–1652.

Torreira, E. et al., 2017. The dynamic assembly of distinct RNA polymerase i complexes modulates rDNA transcription. *eLife*, 6, pp.1–23.

Turowski, T.W. et al., 2014. Rio1 mediates ATP-dependent final maturation of 40S ribosomal subunits. *Nucleic Acids Research*, 42(19), pp.12189–12199.

Tyanova, S. et al., 2016. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods*, 13(9), pp.731–740.

Ulbrich, C. et al., 2009. Mechanochemical Removal of Ribosome Biogenesis Factors from Nascent 60S Ribosomal Subunits. *Cell*, 138(5), pp.911–922.

Venema, J. et al., 2000. Yeast Rrp9p is an evolutionarily conserved U3 snoRNP protein essential for early pre-rRNA processing cleavages and requires box C for its association. *RNA (New York, N.Y.)*, 6(11), pp.1660–1671.

Wai, H. et al., 2001. Yeast RNA Polymerase I Enhancer Is Dispensable for Transcription of the Chromosomal rRNA Gene and Cell Growth , and Its Apparent Transcription Enhancement from Ectopic Promoters Requires Fob1 Protein. *Molecular and cellular biology*, 21(16), pp.5541–5553.

Wang, B. & Ye, K., 2016. Nop9 binds the central pseudoknot region of 18S rRNA. *Nucleic Acids Research*, 45(6), pp.3559–3567.

Warner, J.R., 1999. The economics of ribosome biosynthesis in yeast. *Trends in biochemical sciences*, 24(11), pp.437–40.

Webb, S. et al., 2014. PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. *Genome Biology*, 15(1), pp.1–15.

Wegierski, T. et al., 2001. Bms1p, a G-domain-containing protein, associates with Rcl1p and is required for 18S rRNA biogenesis in yeast. *RNA (New York, N.Y.)*, 7(9), pp.1254–1267.

Wehner, K.A., Gallagher, J.E. & Baserga, S.J., 2002. Components of an interdependent unit within the SSU processome regulate and mediate its activity. *Mol Cell Biol*, 22(20), pp.7258–7267.

Weis, F. et al., 2015. Mechanism of eIF6 release from the nascent 60S ribosomal subunit. *Nature Structural and Molecular Biology*, 22(11), pp.914–919.

Wells, G.R. et al., 2016. The PIN domain endonuclease Utp24 cleaves pre-ribosomal RNA at two coupled sites in yeast and humans. *Nucleic Acids Research*, 44(11), pp.5399–5409.

Wlotzka, W. et al., 2011. The nuclear RNA polymerase II surveillance system targets polymerase III transcripts. *EMBO Journal*, 30(9), pp.1790–1803.

Woolford, J.L. & Baserga, S.J., 2013. Ribosome Biogenesis in the Yeast *Saccharomyces cerevisiae*. *Genetics*, 195(3), pp.643–81.

Wu, S. et al., 2016. Diverse roles of assembly factors revealed by structures of late nuclear pre-60S ribosomes. *Nature*, 534(7605), pp.133–137.

Yang, B. et al., 2012. Identification of cross-linked peptides from complex samples. *Nature methods*, 9(9), pp.904–6.

Yusupov, M.M. et al., 2001. Crystal structure of the ribosome at 5.5 Å resolution. *Science*, 292(5518), pp.883–896.

Yusupova, G. & Yusupov, M., 2014. High-Resolution Structure of the Eukaryotic 80S Ribosome. *Annual Review of Biochemistry*, 83(1), pp.467–486.

Zhang, L. et al., 2016. Stepwise and dynamic assembly of the earliest precursors of small ribosomal subunits in yeast. *Genes & development*, 30(6), pp.718–32.

Zhang, Y. et al., 2010. The RNA polymerase-associated factor 1 complex (Paf1C) directly increases the elongation rate of RNA polymerase I and is required for efficient regulation of rRNA synthesis. *Journal of Biological Chemistry*, 285(19), pp.14152–14159.

Zheng, S.Q. et al., 2017. MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods*, 14(4), pp.331–332.

Zhou, D. et al., 2018. Cryo-EM structure of an early precursor of large ribosomal subunit reveals a half-assembled intermediate. *Protein and Cell*, pp.1–11.

Zhou, X. et al., 2015. Ribosomal proteins: Functions beyond the ribosome. *Journal of Molecular Cell Biology*, 7(2), pp.92–104.

Zhu, J. et al., 2016. Utp14 Recruits and Activates the RNA Helicase Dhr1 To Undock U3 snoRNA from the Preribosome. *Molecular and Cellular Biology*, 36(6), pp.965–978.

Zivanov, J. et al., 2018. RELION-3: new tools for automated high-resolution cryo-EM structure determination. *bioRxiv*, p.421123.