Rockefeller University
Digital Commons @ RU

Student Theses and Dissertations

2018

Making Ribosomes: Biochemical and Structural Studies of Early Ribosome Biogenesis in Yeast

Malik Chaker-Margot

Follow this and additional works at: https://digitalcommons.rockefeller.edu/student_theses_and_dissertations

 Part of the Life Sciences Commons

Recommended Citation
Chaker-Margot, Malik, "Making Ribosomes: Biochemical and Structural Studies of Early Ribosome Biogenesis in Yeast" (2018). *Student Theses and Dissertations*. 472.
https://digitalcommons.rockefeller.edu/student_theses_and_dissertations/472

This Thesis is brought to you for free and open access by Digital Commons @ RU. It has been accepted for inclusion in Student Theses and Dissertations by an authorized administrator of Digital Commons @ RU. For more information, please contact nilovao@rockefeller.edu.

MAKING RIBOSOMES:

BIOCHEMICAL AND STRUCTURAL STUDIES OF EARLY RIBOSOME BIOGENESIS 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

Malik Chaker-Margot

June 2018

© Copyright by Malik Chaker-Margot 2018

MAKING RIBOSOMES: BIOCHEMICAL AND STRUCTURAL STUDIES OF EARLY RIBOSOME BIOGENESIS IN YEAST

Malik Chaker-Margot, Ph.D.

The Rockefeller University 2018

The ribosome is a complex macromolecule responsible for the synthesis of all proteins in the cell. In yeast, it is made of four ribosomal RNAs and 79 proteins, asymmetrically divided in a small and large subunit. In a growing yeast cell, more than 2000 ribosomes are assembled every minute. The ribosome is assembled through a highly complex process involving more than 200 *trans*-acting factors. Ribosome assembly begins in the nucleolus where RNA polymerase I transcribes a long polycistronic RNA, the 35S pre-ribosomal RNA which contains the sequences for three of the four ribosomal RNAs, as well as spacer sequences which are transcribed and removed during assembly. Ribosome assembly factors bind co-transcriptionally to the nascent chains of pre-ribosomal RNA and coordinate its correct folding, modification and cleavage. While most ribosome assembly factors have been identified, the function of numerous factors is still unknown. The timing of their involvement in ribosome assembly has not been characterized, limiting our understanding of their function.

To further our knowledge of this essential cellular process, we set out to characterize the order of assembly of ribosomal assembly factors on the first half of the nascent pre-ribosomal RNA, which forms the earliest precursors of the small subunit (Chapter II). Moreover, using state-of-the-art cryo-electron microscopy we determined the structure

at near-atomic resolution of the earliest yet intermediate of small subunit assembly, the small subuit processome (Chapter III). The combined insights from the structure of this early intermediate and co-transcriptional assembly of pre-ribosomal complexes has redefined our understanding of early ribosome assembly. The cell invests more than 70 factors into the early events of small subunit assembly, with a combined molecular weight of more than 5 megadalton, four times the size of the mature small subunit. This impressive number of factors form a structural blueprint for the spatial segregation and individual maturation of the domains of the 18S. Ribosome assembly factors perform a multitude of functions within this platform, including specific modification of the ribosomal RNA and the concerted coordination of important RNA elements. We have also determined a new mechanism by which cells regulate ribosome biogenesis in response to nutritional depletion.

We finally set out characterize the timing of recruitment of assembly factors to the second half of the 35S pre-ribosomal RNA, which leads to the formation of the earliest large subunit precursors (Chapter IV). This has allowed us to complete a new model for the co-transcriptional assembly of pre-ribosomal complexes. Our work has not only provided new insights into the role of more than 100 factors of early ribosome biogenesis but will serve as a platform for the further characterization of individual factors, as well as the regulation of ribosome assembly.

Acknowledgements

A doctoral thesis is, in many ways, a collective project. I have many individuals to thank for their contributions during these last few years, perhaps too many for an exhaustive list here. I will try, nonetheless, to acknowledge as many as possible. First and foremost, I am thankful to Sebastian Klinge, my advisor, who has been an extraordinarily devoted mentor. At every turn, Sebastian has provided me with the resources and guidance needed to make my thesis project move forward.

I also had the chance of sharing my work environment with truly exceptional and likeable people. Mirjam Hunziker has been my partner-in-crime since my first days in the lab and I could not have asked for a better comrade for a PhD. I was lucky to work closely with Jonas Barandun, whose relentlessness and creativity was a fantastic model for a young scientist. I also had the chance to work with Linamarie Miller, the best rotation student one could hope for, and my bench mate, Zahra Assur Sanghai, whose affability and experience made her an amazing colleague. Lastly, I am thankful to past members of the lab, Camille Cheng and Antoine, and our newest addition, Sameer Singh, whose cheerfulness have made the lab a very positive environment in which to work.

I would like to thank my thesis committee, Dr. Seth Darst, Dr. Brian Chait and Dr. Scott Blanchard, who have been very supportive during my time at Rockefeller and provided me with much needed insight concerning my scientific endeavors. I also thank Sarah Woodson for serving as my external referee.

I am indebted to the many employees of Rockefeller and the Tri-institutional campus that made my work possible, including collaborators and scientists from the research centers. Particularly, I am grateful to Mark Ebrahim and Johanna Sotiris for their exceptional support at the cryo-EM center, Brian D. Dill, Henrik Molina and Milica Tesic Mark from the Proteomics facility, Kelly Molloy from the Chait lab and Devrim Acehan from the electron microscopy resource center.

I am grateful to the many scientists with whom I've had insightful conversations, particularly Deena Oren, Tom Walz, Yixiao Zhang, Hiroshi Suzuki, Kelly Hyojin Kim, Greg Alushin, Richard Hite, James Chen and Stefanie Gerstberger.

Lastly, I could not have done it without tremendous personal support from friends and family. My parents, Hamid Chaker and Joëlle Margot, have been extraordinarily supportive and have provided me with not only love but also invaluable career advice. I would be remiss if I didn't mention the many wonderful people I've had the pleasure of befriending and who have made New York City a new home I will dearly miss, to name a few: my classmates Zac Hann and John Zinder, Emma Garst, Rachel Leicher, Jakob Litke, Emily Rundlet, and Elizabeth Wasmuth, along with the friends I've already mentioned. I wish them all the best in their future projects, scientific and otherwise. Finally, I am eternally grateful to my partner Claire Lengaigne, who has been exceedingly patient and caring during these last five years, and has offered me some much-needed peace of mind.

Table of Contents

List of Figures

List of Tables

Chapter I – Introduction: the ribosome and its assembly

1.1 Ribosomes: Discovery, function and structure

The central dogma of molecular biology states that DNA encodes RNA, which encodes protein. Throughout the domains of life, ribosomes are the molecules that translate the information contained in RNA to synthesize proteins. They are large macromolecular complexes made of both RNA and proteins, which were first discovered in the 1950s as dense cellular particles located in the cytoplasm (Palade 1955). Ribosomes were later identified to be ribozymes, meaning that all the catalytic activity, i.e. messenger RNA decoding and peptide-bond formation, is performed by the RNA components of the ribosomes (Noller et al. 1992; Ban et al. 2000; Schluenzen et al. 2000; Wimberly et al. 2000). Ribosomal proteins have a structural role in the particle.

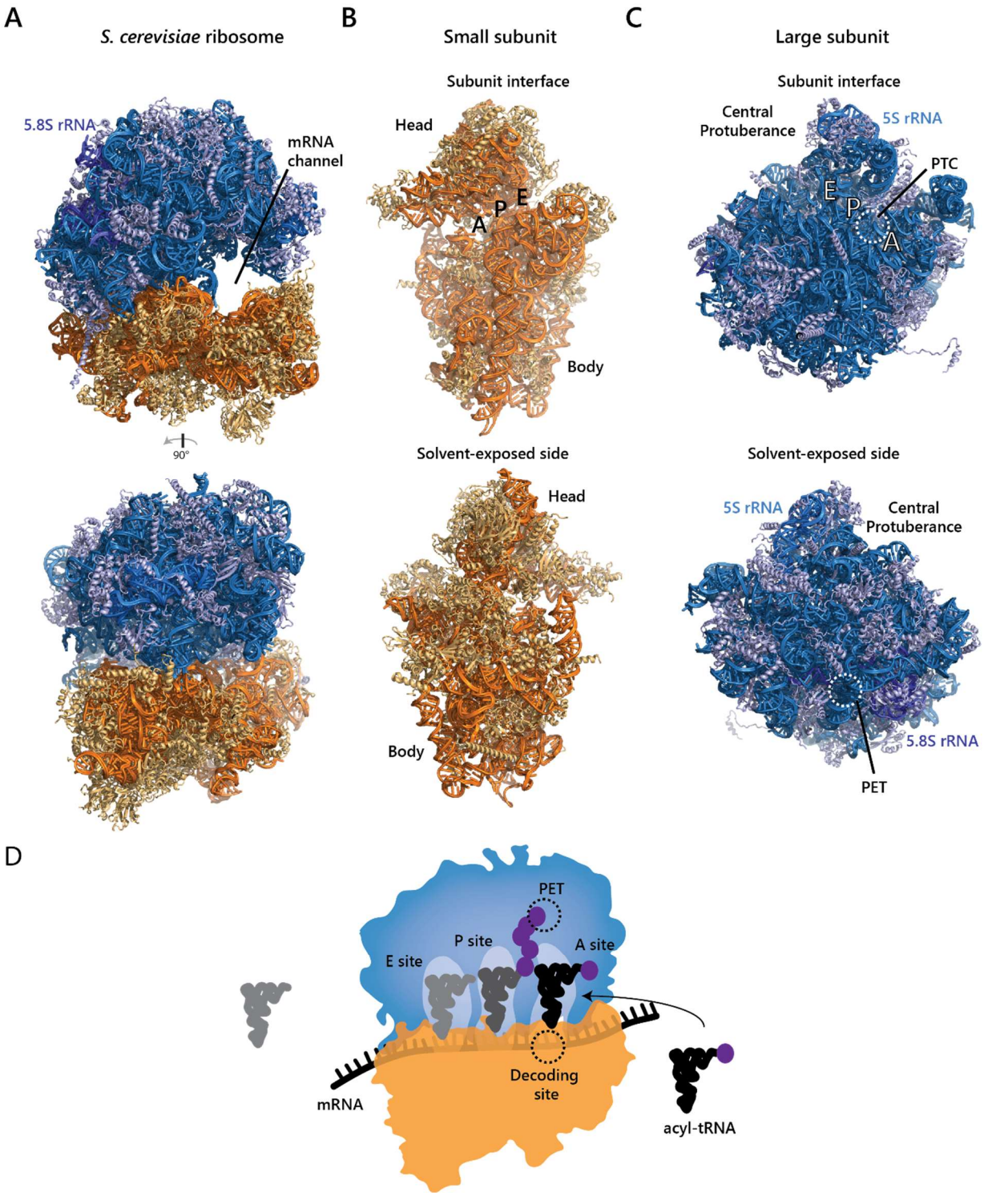
Ribosomes are organized into 2 asymmetric subunits: the large subunit (LSU), and the small subunit (SSU) (Fig. 1.1A). In the yeast *Saccharomyces cerevisiae*, the large subunit, or 60S, is made of three RNAs, the 25S, the 5.8S and the 5S, as well as 46 proteins. The small subunit, or 40S, is composed of one RNA, the 18S, and 33 proteins.

High-resolution structures of the bacterial ribosome were determined in the early 2000's (Clemons et al. 1999; Ban et al. 2000; Schluenzen et al. 2000; Yusupov et al. 2001; Harms et al. 2001). Later work then revealed the architecture of eukaryotic ribosomes (Rabl et al. 2011; Klinge et al. 2011; Ben-Shem et al. 2011) (Fig. 1.1A). The 18S RNA was shown to be composed of 4 domains, the 5' domain, the central and the 3' major and minor domain. These domains are organized to form a "body", made of the 5',

central and 3’ minor domains, and a “head”, made of the 3’ major domain (Fig. 1.1B). Domains are kept together in part by a tertiary structure element at the core of the small subunit: the central pseudoknot. The central pseudoknot involves base-pairing by sequences from the 5’ end of the 18S and in the middle of the RNA. The messenger RNA (mRNA) binds the groove between the body and the head during translation.

The 25S RNA large subunit (Fig. 1.1C) was similarly shown to be composed of 6 structural domains, I through VI. The polypeptide exit tunnel, by which nascent protein chains exit the ribosome during translation, sits at the interface of these domains. Meanwhile on the small-subunit facing side, a cavity allows for the presence of the transfer RNAs (tRNAs). This cavity, which corresponds with the groove in the small subunit, can be subdivided in 3 functional sites: the A, P and E sites (Fig. 1.1D). The A site is where the aminoacyl tRNA enters the ribosome and where tRNA selection occurs. The peptidyl transfer and formation of the peptidyl-tRNA then leads to translocation, when the peptidyl-tRNA is transferred to the P site. Lastly, the uncharged tRNA is translocated to the E site, which represents the exit site of the tRNA. Initiation and elongation factors bind near the A site and regulate aminoacyl tRNA entry, selection and translocation, as reviewed in (Dever et al. 2016).

Figure 1.1. Architecture and functional sites of the ribosome. (**A**) Two views of the *S. cerevisiae* ribosome (PDB 4V88), the first showing the mRNA channel, and the second, rotated by 90°. The small and large subunits are colored in orange and blue respectively. (**B**) Structure of the small ribosomal subunit showing the subunit interface and solvent-exposed views, with the characteristic “head” and “body” sections. A, P and E site are indicated(**C**) Structure of the large ribosomal subunit showing the subunit interface and solvent-exposed views. The polypeptide exit tunnel (PET), the peptidyl transferase site (PTC), the central protuberance and the A, P and E sites are indicated. (**D**) Schematic view of translation, including acyl-tRNA binding and decoding, polypeptide formation and exit through the polypeptide exit tunnel (PET) and tRNA dissociation.


A
S. cerevisiae ribosome
5.8S rRNA
mRNA channel
90°
B
Small subunit
Subunit interface
Head
E
P
A
Body
Solvent-exposed side
Head
Body
C
Large subunit
Subunit interface
Central Protuberance
5S rRNA
PTC
E
P
A
Solvent-exposed side
5S rRNA
Central Protuberance
5.8S rRNA
PET
D
PET
P site
A site
E site
Decoding site
mRNA
acyl-tRNA

1.2 The eukaryotic ribosome: differences with bacteria

As many other conserved cellular structures, ribosomes have seen a dramatic increase in complexity from bacteria to eukaryotes (Fig. 1.2). The 16S rRNA in bacteria is homologous to the 18S rRNA, in eukaryotes, while the 23S rRNA is homologous to the 5.8S and the 25S, in yeast (28S in humans). The 5S has stayed approximately the same size. This increase in size for the two large rRNAs is due to the emergence of large eukaryotic expansion segments (Gerbi 1996). These large expansion segments are located on the surface of the RNA and do not affect the main functional sites of the ribosome (Ben-Shem et al. 2011). The segments are seemingly highly variable sites as they have continued to expend throughout evolution: while in yeast they span 50-200 nucleotides, they have expanded to several hundred nucleotides in metazoan, and even more so in mammalians (Anger et al. 2013; Khatter et al. 2015).

Simultaneously, new ribosomal proteins have appeared in eukaryotes and conserved proteins have gotten more complex through evolution. While there are 20 and 33 proteins in the small and large bacterial ribosomal subunit respectively, these numbers increase to 33 and 46 for the eukaryotic ribosome. Conserved proteins also feature new domains and extensions sometimes spanning several tens of angstroms and connecting distant modules in the eukaryotic ribosomes (Rabl et al. 2011; Klinge et al. 2011; Ben-Shem et al. 2011). These new elements combined account for an increase the molecular weight of the ribosome of approximately 1 megadalton, that is from 2.1 to 3.2 MDa (Fig. 1.2B).

The emergence of these new elements can be surmised to allow, first, a great degree of stability. The new protein elements interconnect distant regions of the ribosome thereby

providing a higher degree of interconnectedness. Some of these eukaryotic specific elements also bridge small and large subunits, for instance RPL19 or ES41. Furthermore, it has been proposed that the new eukaryotic elements assist in the regulation of translation, such as facilitating the recruitment of translation factors to the ribosome. Lastly, expansion segments are also suggested to have a role in ribosome biogenesis (Ramesh and Woolford 2016). The increase of complexity of the ribosome has been accompanied by the emergence of an elaborate assembly process. Whereas only 20 factors are used for bacterial ribosome biogenesis, most of which are dispensable (Shajani et al. 2011), in excess of 200 factors are required to make a eukaryotic ribosome, 130 of which are essential (Woolford and Baserga 2013). Furthermore, cells have a critical need for ribosome production: in growing yeast cells, approximately 2000 ribosomes must be synthesized every minute (Warner 1999).

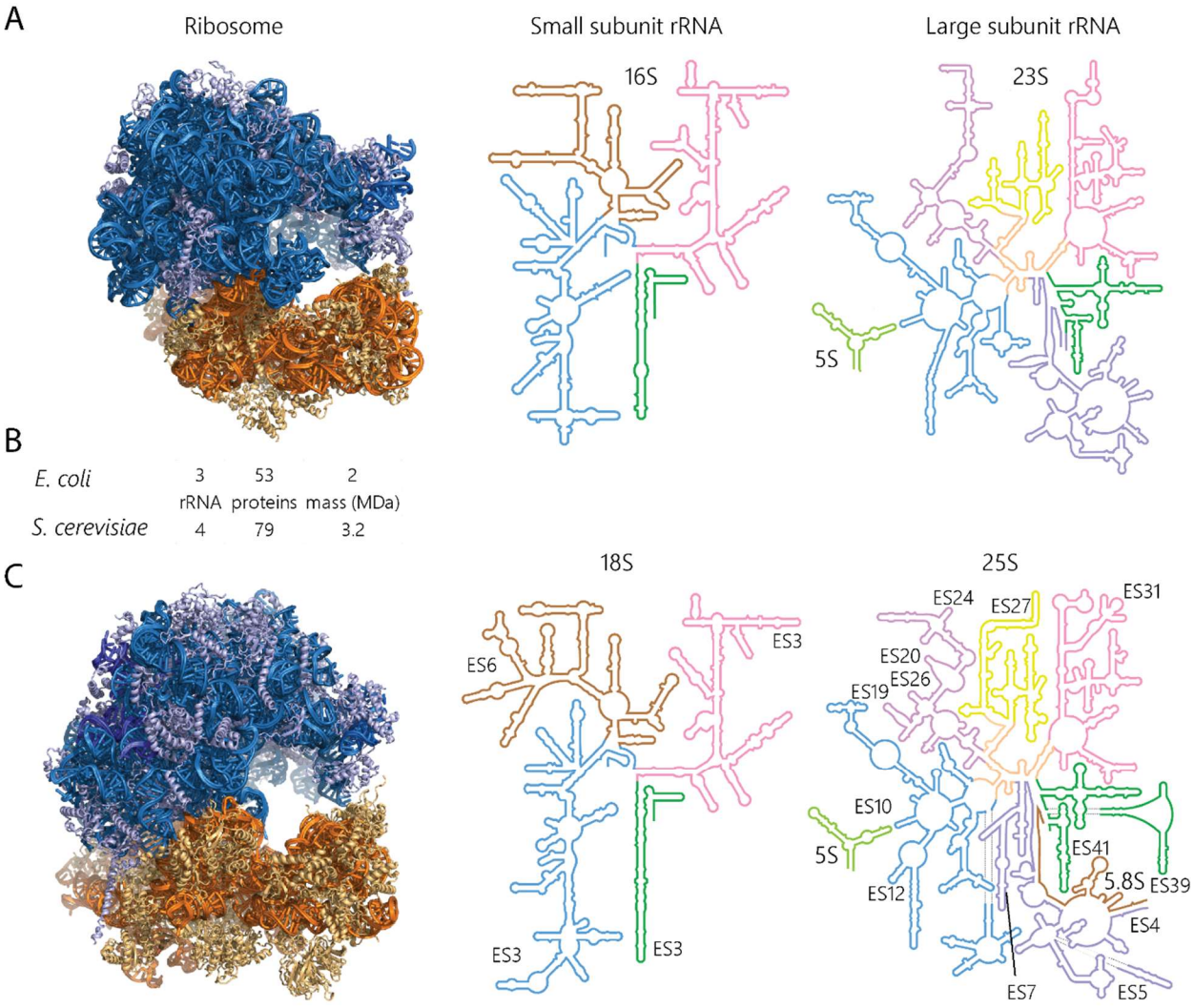


Figure 1.2. Differences between bacterial and eukaryotic ribosomes. Secondary structure diagram adapted from (Petrov et al. 2014). (**A**) Side view of the *E. coli* ribosome (PDB 5AFI) with the large and small subunit colored in blue and orange respectively. Secondary structure diagrams of the small and large subunit ribosomal RNA are also shown with 16S and 23S subdomains shown in different colors. (**B**) Comparative table of RNA and protein components, and molecular weight of the ribosome from *E. coli* and *S. cerevisiae*. (**C**) Same as in (A) but for the *S. cerevisiae* ribosome structure (PDB 4V88) and ribosomal RNA secondary structure.

1.3 Ribosomal RNA transcription

Ribosome assembly requires the concerted activity of all three eukaryotic RNA polymerases (Woolford and Baserga 2013). RNA polymerase (RNAP I) is dedicated to ribosomal RNA synthesis. RNA polymerase II (RNAP II) must transcribe the messenger RNA encoding all ribosomal protein genes, and ribosome assembly factors. RNA polymerase III transcribes the 5S and several small nucleolar RNA (snoRNA) involved in ribosomal RNA modification (Watkins and Bohnsack 2012). Ribosome assembly begins with the transcription of rDNA in the nucleolus. Eukaryotic cells have 100 to 200 copies of rDNA loci, one hundred of which can be actively transcribed simultaneously. In yeast, rDNA loci encode two genes: the pre-5S and the 35S.

RNAP I is responsible for the transcription of the 35S pre-rRNA. This polycistronic RNA contains the 18S, 5.8S and 25S rRNAs as well as long spacers that flank the rRNA sequences: two external transcribed spacers (ETS), the 5' and 3' ETS, and two internal transcribed spacers (ITS), ITS1 and ITS2. While these spacers are not part of the translational machinery, they have an essential role in recruiting ribosome assembly factors during ribosome biogenesis.

RNAP I requires few factors to initiate transcription at the rDNA promoter. RNA polymerase I is composed of 14 subunits, in contrast to the 12 subunits of RNAP II. These additional subunits, Rpa34 and Rpa49, are homologous to the TFIIF dimer that binds RNAP II during pre-initiation complex formation, suggesting their integration in RNAP I is to facilitate efficient transcription initiation (Engel et al. 2013). Recent structures of RNA polymerase I combined with biochemical data have detailed its initiation mechanism, as reviewed in (Khatter et al. 2017). Transcription initiation first

requires binding of RNA polymerase I to Rrn3, which renders the enzyme transcription competent (Engel et al. 2016; Pilsl et al. 2016) (Fig. 1.3A). The complex subsequently binds to the 3-subunit Core Factor (CF) complex leading to promoter binding and melting (Engel et al. 2017; Han et al. 2017; Sadian et al. 2017).

RNA polymerases I and II initiation complexes adopt strikingly different structures when bound to open promoter DNA, with a 60° difference in promoter bend. This was suggested to contribute to RNAP I promoter specificity, more so than sequence recognition (Engel et al. 2017). The upstream activating factors (UAF) and the TATA-binding protein (TBP), which are also involved in RNAP II initiation, can further stimulate transcription, but they are not required for rDNA transcription (Schneider 2012). During elongation, RNAP I adopts a closed conformation similar to other RNA polymerases (Neyer et al. 2016). Studies by electron tomography of actively transcribed rDNA showed that the distance between RNAP I molecules were likely too high for inter-molecular interactions (Neyer et al. 2016), contrary to what was previously proposed.

Similar to RNAP I, RNAP III has a higher number of subunits than RNAP II, 17 subunits in total (Vannini and Cramer 2012). These additional subunits have homology to general transcription factors and are thought to facilitate the rapid initiation of the polymerase on its target genes. For example, the PolIII-specific heterotrimer C82-C34-C31 stabilizes downstream elements in a manner reminiscent of TFIIE and TFIIF in RNAPl II initiation (Hoffmann et al. 2015). RNAP III initation only relies a single transcription factor: TFIIIB, a 3-subunit complex which has functions similar to that of the CF complex or TFIIB (Fig. 1.3B). It binds the promoter DNA to promote melting and transcription initiation (Vorländer et al. 2018).

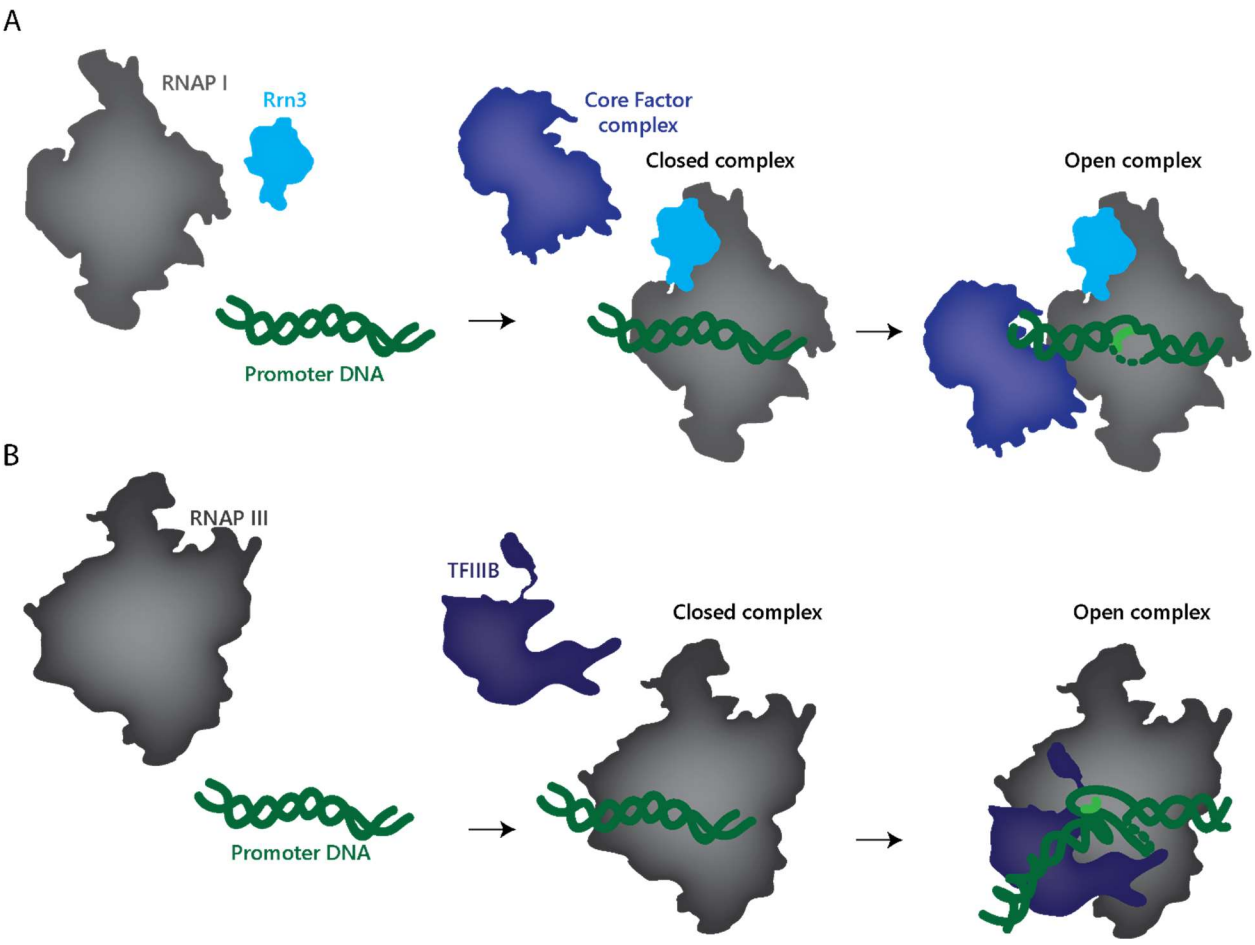


Figure 1.3 Ribosomal RNA transcription by RNA polymerase I and III. (**A**) Schematic view of the minimal initiation process for RNA polymerase I, involving first Rrn3 binding, closed promoter DNA binding, and subsequent core factor complex association and promoter melting (**B**) Schematic view of the minimal initiation for RNA polymerase III, with a similar mechanism.

1.4 Ribosomal RNA processing

Ribosome assembly requires the coordinated processing of the 35S into mature ribosomal RNA. The 35S can undergo two main maturation pathways (Fig. 1.4). In the main pathway, the 35S gets co-transcriptionally cleaved in the 5' ETS at site A0 first followed by cleavage at A1 and A2. Cleavage at the A2 leads to the separation of small and large subunit maturation, which then proceed independently. Alternatively, the 35S can be cleaved first at site A3 by RNase MRP (Lygerou et al. 1994) in a post-transcriptional fashion, which immediately separates large and small subunit processing. This leads to the emergence of the 27S and 23S species for large and small subunit precursors, respectively.

Cleavage at A0 has been shown to occur first while A2 and A1 seem to happen in a coordinated fashion (Hughes and Ares 1991). Cleavage at the A1 site is catalyzed in yeast, and in humans, by the PinC nuclease Utp24, also known as Fcf1 (Bleichert et al. 2006; Tomecki et al. 2015; Wells et al. 2016). Utp24 is also thought to catalyze A2 cleavage in yeast. The enzyme responsible for A0 site cleavage is unknown. The cleavage at the A1 site defines the 5' end of the RNA and leads to the formation of the 20S in small subunit assembly. Removal of the 5' ETS by cleavage at A0 is followed by its degradation which is catalyzed by the RNA exosome and its cofactor, the RNA helicase Mtr4 (Allmang et al. 2000). The RNA exosome is a 3' to 5' exonuclease which possesses two catalytic subunits, the processive nuclease Rrp44, and the distributive nuclease Rrp6 (Mitchell et al. 1997; Wasmuth and Lima 2012; Zinder and Lima 2017).The 20S species gets further cleaved in the cytoplasm by the endonuclease

Nob1 at the D-site, thereby defining the 3’ end of the 18S (Fatica et al. 2004). This represents one of the last steps of small subunit assembly.

In large subunit assembly, 3’ processing of the 27S happens in two steps: the endonuclease Rnt1 catalyzes the cleavage of a stem-loop in the 3’ ETS at site B0 (Kufel et al. 1999). The 3’ end of the 25S is trimmed to its mature end, possibly by the exonuclease Rex1 (Kempers-Veenstra et al. 1986). The 27SA is processed at its 5’ end by 5’ to 3’ exonucleases, such as Rrp17, Xrn1 or Rat1, to the 5’ mature end of the 5.8S (Henry et al. 1994; Oeffinger et al. 2004). The 5.8S exists in two similarly present isoforms, $5.8S_S$ and $5.8S_L$, which only differ in their 5’ end. This is possibly a consequence of a lack of specificity from the 5’ to 3’ nucleases during the 5’ processing of the 27SA.

Later in large subunit maturation, ITS2 is cleaved by the Las1 complex at the C2 site (Gasse et al. 2015). The endonuclease Las1 cleaves in the middle of ITS2 and leaves a 2’3’ cyclic phosphate end of the 5’ side, and a 5’ phosphate of the 3’ side. The 5’ end of the resulting 25S precursor gets processed to its mature end by the 5’ to 3’ exonuclease Rat1. Concomitantly, the 3’ end of the 7S, which harbors the cyclic phosphate is degraded by the RNA exosome (Mitchell et al. 1997), leaving a small unprocessed overhang (6S pre-rRNA). Removal of this last element occurs in the cytoplasm and is catalyzed by Ngl2 (Thomson and Tollervey 2010).

The 5S RNA is transcribed as a precursor RNA by RNA polymerase III, the pre-5S, which possesses a 3’ tail that needs to be trimmed. This process is catalyzed by Rex1, the same enzyme involved in 25S and 5.8S 3’ end processing (van Hoof 2000). It is currently unknown how this process is regulated.

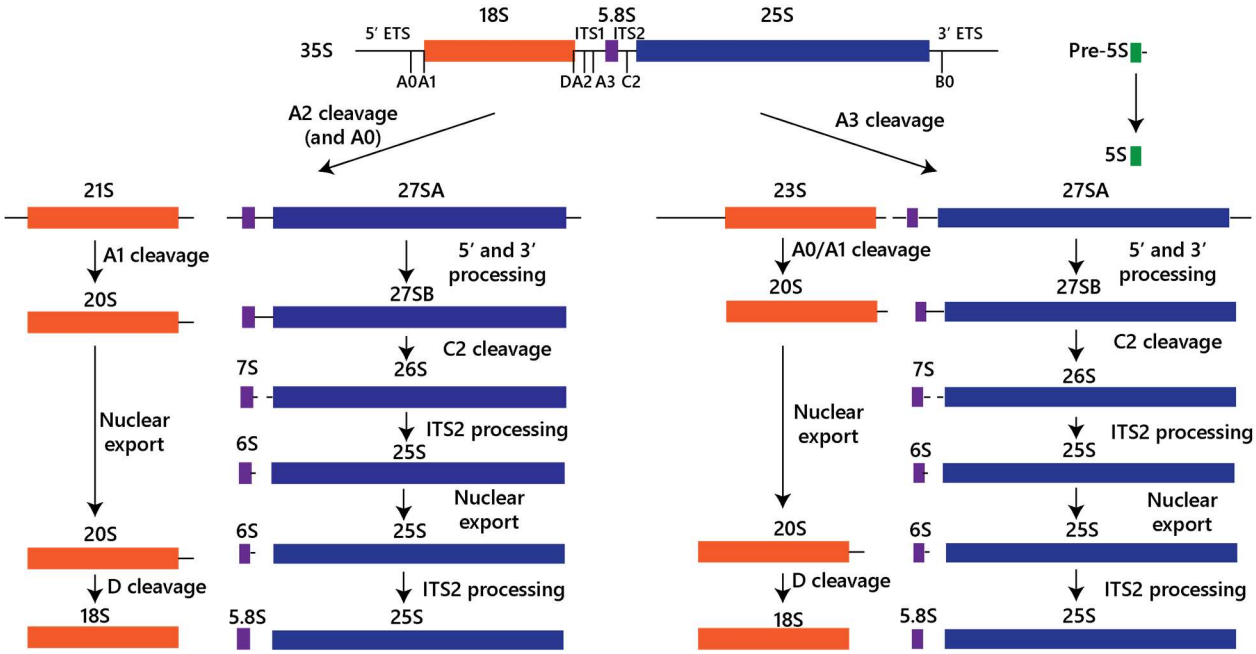


Figure 1.4. Pre-rRNA processing in ribosome assembly. Schematic view of the two main RNA processing pathways for the 35S involving either A2 or A3 cleavage first. 5S processing (top right) happens separately in the nucleolus.

1.5 Ribosome assembly: Modification of the pre-rRNA

In all domains of life, ribosomal RNA has to undergo essential chemical modification. RNA chemical modification takes two main forms: 2'-O methylation of the ribose moiety of the RNA backbone, and isomerization of uridines into pseudouridines. While for many RNAs, chemical modification is catalyzed by standalone proteins, ribosomal RNA modifications are often catalyzed by enzymes guided by small nucleolar RNA, as reviewed in (Watkins and Bohnsack 2012). These guide RNAs typically interact with distinct proteins, which both stabilize the guide RNA and perform the chemical modification. The guide RNA base pairs with the target RNA and thereby directs the modification of the target nucleotide.

As the chemical modification of the pre-rRNA mainly consists in two types, there are, accordingly, two broad families of snoRNA that guide the modification of the pre-rRNA: box C/D snoRNAs and box H/ACA snoRNAs. They are named after the secondary structure elements, which define their overall topology. Box C/D snoRNAs guide the 2'-O methylation of the pre-rRNA. They interact with four proteins, Nop56 and Nop58, two copies of fibrillarin/Nop1, the catalytic subunit, and two copies of Snu13. Structures of the archaeal box C/D small nucleolar ribonuclear particle (snoRNP) suggest that box C/D snoRNP can function as dimers to perform their modifying function (Fig. 1.5A). However, whether box C/D snoRNP act as monomer or dimer to perform methylation in eukaryotes is still debated (Bleichen et al. 2009; Lin et al. 2011).

The best-known box C/D snoRNA is the U3 snoRNA, but unlike most others, it is not thought to guide the methylation of pre-rRNA. Rather, it has a role in the earliest stages of small subunit biogenesis by base pairing the 5' ETS and 18S (Beltrame and Tollervey

1995; Sharma and Tollervey 1999; Dutca et al. 2011; Marmier-Gourrier et al. 2011). Another example of such snoRNA is U14, which is involved in the methylation of C414 in the 5' domain (Liang and Fournier 1995; Morrissey and Tollervey 1997).

Similarly, box H/ACA snoRNAs catalyze the pseudouridylation of the pre-RNA and interact with four proteins: Cbf5, Nhp2, Gar1 and Nop10. Cbf5 catalyzes the pseudoruridylation while the other proteins stabilize the snoRNA (Fig. 1.5B). snR30 is an example of an essential box H/ACA snoRNA, although, like U3 snoRNA, has not been shown to direct modification of pre-rRNA. Instead, snR30 base pairs with eukaryotic expansion segment 6 (ES6) and is required for subsequent cleavage at the A1 and A2 site (Fayet-Lebaron et al. 2009).

However, as mentioned above, there are a few rRNA modifications that are catalyzed by proteins directly, seemingly without requiring snoRNA guide. Three examples of such proteins are the essential methylase Emg1, which catalyzes modification of nucleotide U1191 in the 3' domain (Leulliot et al. 2008), Kre33, an essential acetyltransferase that modifies two nucleotides: 1280 and 1773, also in the 3' domain (Sharma et al. 2015), and the essential protein Dim1, that methylates the 3' loop of the 18S rRNA (Lafontaine et al. 1994).

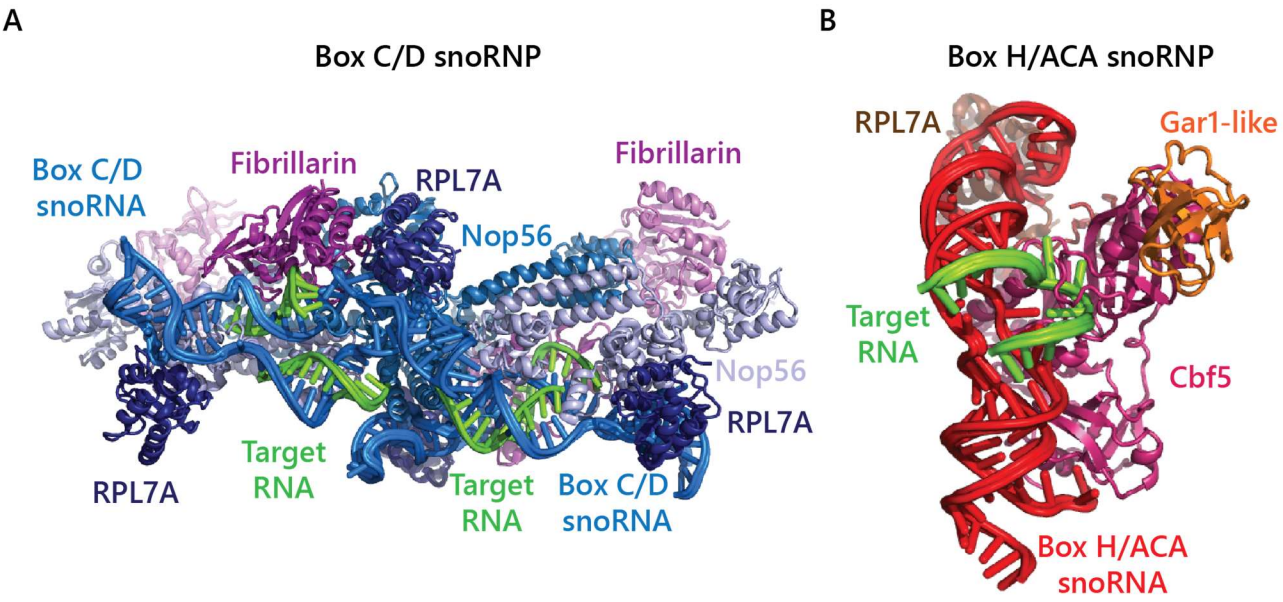


Figure 1.5. Small nucleolar RNA guide chemical modification of the ribosomal RNA. (**A**) Structure of the archaeal box C/D snoRNP from *Pyrococcus furiosa* (PDB 4BY9) forming a homodimer. The snoRNAs are bound to target RNA (**B**) Structural of the archaeal box H/ACA snoRNP from *Pyrococcus furiosa* (PDB 3HAY) bound to its target RNA.

1.6 Co-transcriptional ribosome assembly: the small subunit processome

Ribosome assembly starts in the nucleolus where RNAP I transcribes the 35S. There, ribosome assembly factors assemble co-transcriptionally on the nascent RNA chains and coordinate the folding and processing of the pre-rRNA. This leads to the formation of large multi-component complexes, corresponding to early precursors of the small and large subunits.

Pre-ribosomal complexes were initially visualized as a terminal structure that forms on nascent pre-rRNA of actively transcribed rDNA (Miller and Beatty 1969). These terminal structures were shown to be ribosome assembly complexes (Mougey et al. 1993). Later work then showed that these structure were early small subunit assembly complexes containing the U3 snoRNA (Grandi et al. 2002; Dragon et al. 2002), an essential snoRNA involved in ribosome assembly (Hughes and Ares 1991). Given the presence of many factors essential for small subunit maturation and the absence of large subunit assembly factors in these U3-containing complexes, they were named small subunit (SSU) processomes.

Northern blot analysis of these particles suggested they contained a 35S pre-rRNA, which was erroneously thought to represent “90S pre-ribosomes”, containing both small and large subunit pre-rRNA, as had been described previously by metabolic labeling (Trapman et al. 1975; Grandi et al. 2002). These complexes included several unknown proteins essential in the assembly of the small subunit, which were called U three proteins (UTP) (Dragon et al. 2002). Included in those assembly factors were large multi-protein complexes, which could also be isolated individually: UtpA, UtpB, and UtpC (Krogan et al. 2004).

UtpA and UtpB are large complexes of approximately 600 and 500 kDa of molecular weight, respectively. UtpA contains 7 subunits: Utp4, Utp5, Utp8, Utp9, Utp10, Utp15 and Utp17. Two of these, Utp17 and Utp4, are composed of two β-propeller domains. Utp5, Utp8, Utp9 and Utp15 are all composed of one β-propeller domain with a C-terminal helical repeat. Utp10 is a long helical repeat protein with a molecular weight of 200 kDa. Similarly, UtpB is made of 6 subunits, Utp1, Utp6, Utp12, Utp13, Utp18 and Utp21. Utp1, Utp12, Utp13 and Utp21 are double β-propeller proteins with a C-terminal helical repeat. Utp6 is a helical repeat protein while Utp18 contains a β-propeller domain flanked by N- and C- terminal extensions.

The U3 snoRNA is a box C/D snoRNA interacting with box C/D proteins Nop56, Nop58, Nop1 and Snu13 (Jansen et al. 1993; Lafontaine and Tollervey 2000) and a U3 specific chaperone, Rrp9 (Venema et al. 2000). Four segments of the U3 snoRNA were shown to base pair with the pre-ribosomal RNA: the 3' and 5' hinges base-pair with the 5' external transcribed spacer (5' ETS) whereas box A and A' may form a duplex with sequences of the 5' end of the 18S (Beltrame and Tollervey 1995; Sharma and Tollervey 1999; Marmier-Gourrier et al. 2011; Dutca et al. 2011).

UtpA, UtpB and the U3 snoRNP were proposed to be the first factors recruited to the pre-rRNA during co-transcriptional assembly of the SSU processome (Fig. 1.6). Depletion of UtpA component Utp17 was further shown to prevent co-precipitation of UtpB, U3 snoRNA with small subunit precursors, while the reverse was not the case. This strongly suggested that UtpA was the complex initiating the process followed by UtpB and U3 snoRNP (Perez-Fernandez et al. 2007; Pérez-Fernández et al. 2011).

Other early factors in small subunit assembly include the complex UtpC, a two-subunit complex, Utp22 and Rrp7, which has also been shown to associate with Casein Kinase 2. Utp20, Nop14, Noc4, Enp1 and Rrp12 are helical repeat proteins that are also involved in the early stages of small subunit assembly. U1191 in the 3' domain (Leulliot et al. 2008), Kre33, an essential acetyltransferase that modifies two nucleotides: 1280 and 1773, also in the 3' domain

Several enzymes are required for early small subunit assembly. As mentioned previously, the proteins Kre33 and Emg1 catalyze the acetylation of nucleotides 1280 and 1773, and the methylation of U1191, respectively. Bms1 is an essential *bonafide* GTPase whose GTP-binding, and its interaction with its cofactorRcl1, are required for A2 cleavage (Wegierski et al. 2001; Gelperin et al. 2001; Karbstein et al. 2005; Delprato et al. 2014).

The helicases Dbp4, Dbp8, Has1, Rok1, Dhr1 and Dhr2 have been shown to be essential for small subunit assembly and A1/A2 cleavage. While Dbp4 was associated with removal of U14 snoRNA from the 5' domain, along with its proposed interactors, Enp2, Lcp5 and Bfr2 (Koš and Tollervey 2005; Soltanieh et al. 2015, 2014), another DEAD–box protein, Rok1 was associated with the removal of snR30 (Martin et al. 2014; Khoshnevis et al. 2016). Similarly, the DEAH helicase Dhr1 was shown to unwind the U3-18S duplex, with the help of its cofactor Utp14 (Sardana et al. 2015; Zhu et al. 2016). Unlike other helicases here, Has1 is also involved in large subunit assembly but its roles both in small and large subunit biogenesis are still unclear (Emery et al. 2004). Finally, the PinC nuclease Utp24 is involved in A1 and A2 cleavage (Tomecki et al. 2015; Wells et al. 2016).

Maturation of the SSU processome into pre-40S particles requires removal of the 5' ETS. After the 5' ETS is cleaved from the pre-18S, the RNA exosome and its cofactor the DEAH helicase Mtr4 are responsible for the degradation of the spacer RNA, and presumable the recycling of ribosome assembly factors (Allmang et al. 2000). This complex is aided in its function by the RNA helicase Mtr4. Biochemical evidence has shown that UtpB subunit Utp18 is involved in the recruitment of the Mtr4-exosome by interacting directly with the Mtr4 arch-domain through an arch-interacting motif (AIM) (Thoms et al. 2015).

Prior to the work described herein, the assembly and function of SSU processome factors was poorly understood. Genetic and proteomics study had detailed the list of proteins and snoRNA involved in those early steps of ribosome assembly but the order of their assembly on the pre-rRNA and their specific function remained mostly unknown. Later intermediates of small subunit assembly resemble the mature small subunit, but how these particles have matured was still unclear.

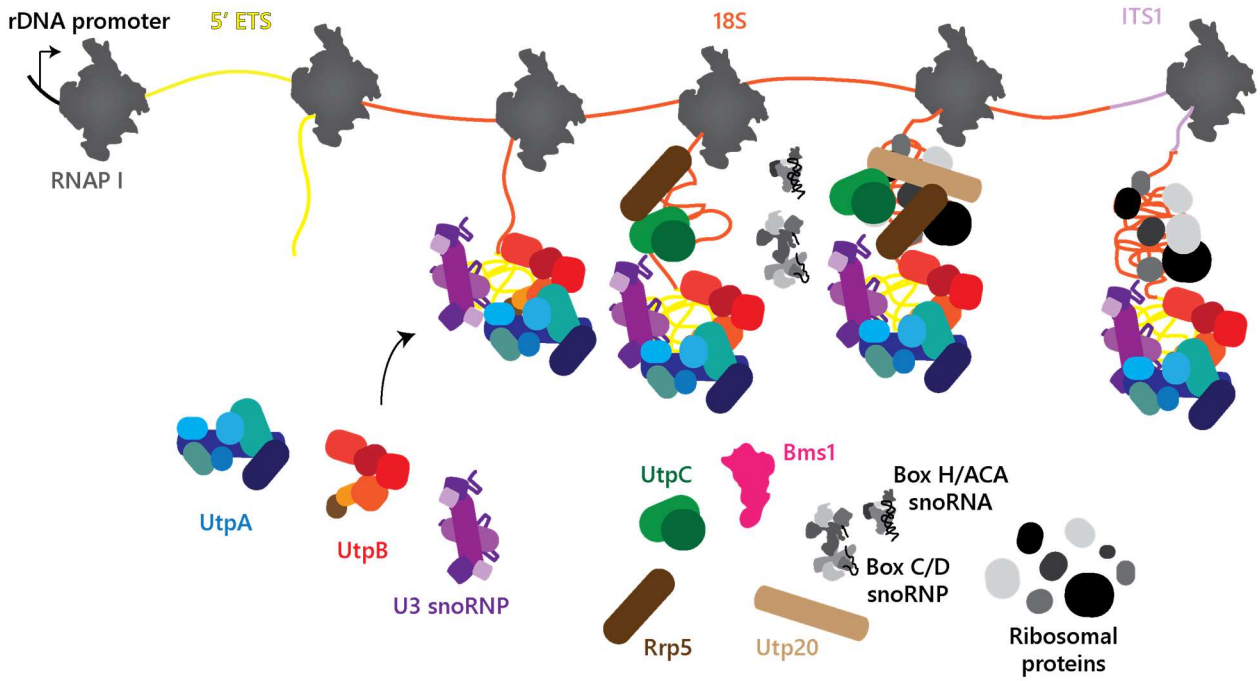


Figure 1.6. Co-transcriptional assembly of the small subunit processome. Schematic view of early ribosome assembly factors binding nascent pre-rRNA as it is being transcribed by RNA polymerase I, in the nucleolus. These include the early complexes UtpA, UtpB and U3 snoRNP.

1.7 Nuclear export and cytoplasmic maturation of the small subunit precursor

Prior to nuclear export, the core of pre-40S particles is almost identical to the mature small subunit. However, these particles remain associated with several ribosome assembly factors that protect important functional sites of the maturing small subunit. These factors include the early factors components Enp1, Pno1 and Rrp12, as well as later proteins such as Rio2, Ltv1, Dim1 or Tsr1, which bind the pre-40S particle in the nucleus (Schäfer et al. 2003) (Fig. 1.7). Notably, Tsr1 is an inactive GTPase (McCaughan et al. 2016) that has high homology to Bms1 but lacks its catalytic activity. High-resolution structures of pre-40S particles have revealed the interactions between some of these factors and the maturing small subunit (Heuer et al. 2017; Scaiola et al. 2018). Rio2, Tsr1 and Dim1 are located on the subunit interface of the pre-40S, near the decoding site. Enp1 is placed on the solvent facing side, interacting with the beak structure. A peptide of Ltv1 is bound on the solvent exposed side of Enp1. Finally, Pno1 is positioned at the end of helix 45, on the shoulder of the maturing small subunit. Despite being bound to the particle, Nob1 is flexibly attached and not resolved in the structures of the pre-40S.

Export of pre-40S particles to the cytoplasm has been shown to rely on multiple overlapping pathways (Fig. 1.7). Several assembly factors such as Rio2, Ltv1, and Pno1 have predicted nuclear export signals (NES) that can recruit the nuclear export machinery. The main exportin in this pathway is the karyopherin Crm1/Xpo1. Crm1 typically binds NES-containing proteins and RanGTP, thereby forming a ternary complex (Fung et al. 2015). Crm1, as any karyopherin, then interacts with the FG nucleoporins, which leads to cargo-export (Hayama et al. 2017). Recent studies have

shown that the small subunit assembly factor Slx9 binds RanGTP and NES-containing protein Rio2, which cooperatively recruits Crm1. The Mex67-Mtr2 dimer, which has been shown to be important for pre-60S export also has a role in pre-40S export (Faza et al. 2012). It was also proposed that Rrp12 can interact with FG nucleoporins (Oeffinger et al. 2004). With multiple proteins having the ability to interact with Xpo1 and RanGTP, Mex67-Mtr2 or with the FG-Nucleoporins (Nups) directly, the pre-40S can use multiple pathways to achieve nuclear export. Such redundancy in the presence of nuclear export signals and interaction with the nuclear pore complex may help efficient export of pre-40S subunits, given that NES-exportin interactions are notoriously weak (Kutay and Guttinger 2005). The presence of multiple weakly binding NES may also insure the presence of all NES containing factor for the efficient export of pre-40S, as quality control step.

Once in the cytoplasm, pre-40S particles undergo two main last maturation steps: association with mature 60S subunits, and cleavage at the D-site. Pre-40S particles associate with mature 60S subunits to form 80S-like complexes (Lebaron et al. 2012; Strunk et al. 2012). The formation of these 80S-like complexes is promoted by the initiation factor and GTPase Fun12 (eIF5B), which binds the large ribosomal subunit as well as the pre-40S particle. During translation, Fun12 enhances subunit joining and Met-tRNA binding at the beginning of translation. However, here, Fun12 is involved in the quality control of the pre-40S subunit.

Prior to subunit joining, several factors leave the pre-40S particles. Hrr25 phosphorylates Enp1 and Ltv1, leading to Ltv1 dissociation (Schäfer et al. 2006). Mutations in the phosphorylated residues of Ltv1 prevent its dissociation and subunit

joining (Ghalei et al. 2015). Rio2 likely also dissociates from pre-40S prior to 80S-like complex formation, in an ATPase dependent manner (Ferreira-Cerca et al. 2012). The mechanism by which Ltv1 dissociation is coupled to 80S-like complex formation is not understood.

In 80S-like complexes, dissociation is dependent on the activity of the adenylate cyclase Fap7 (Loc'h et al. 2014; Ghalei et al. 2017), the ATPase/kinase Rio1 and GTP hydrolysis by Fun12. Fap7 has been suggested to induce pseudo-translocation of this 80S-like complex (Ghalei et al. 2017). Its departure from the complex is promoted by ATP hydrolysis, which may also induce Dim1 dissociation. 80S-like complex separation is promoted by Rli1 and Dom34, which are involved in mature 80S dissociation during translation termination (Strunk et al. 2012). Complex dissociation is followed by cleavage at the D-site. Pno1 is strategically located near the cleavage site in the pre-40S structure while Nob1 is flexibly attached to the particle (Heuer et al. 2017; Scaiola et al. 2018). This suggests that Pno1 may protect the site until the particle is mature, at which point it either presents the substrate to Nob1, or dissociates from the particle, leading to cleavage, and incorporation of rpS26, whose binding site is blocked by Pno1 (Ferreira-Cerca et al. 2007). Nob1, which binds ITS1 (Granneman et al. 2010) is thought to leave with the cleaved substrate.

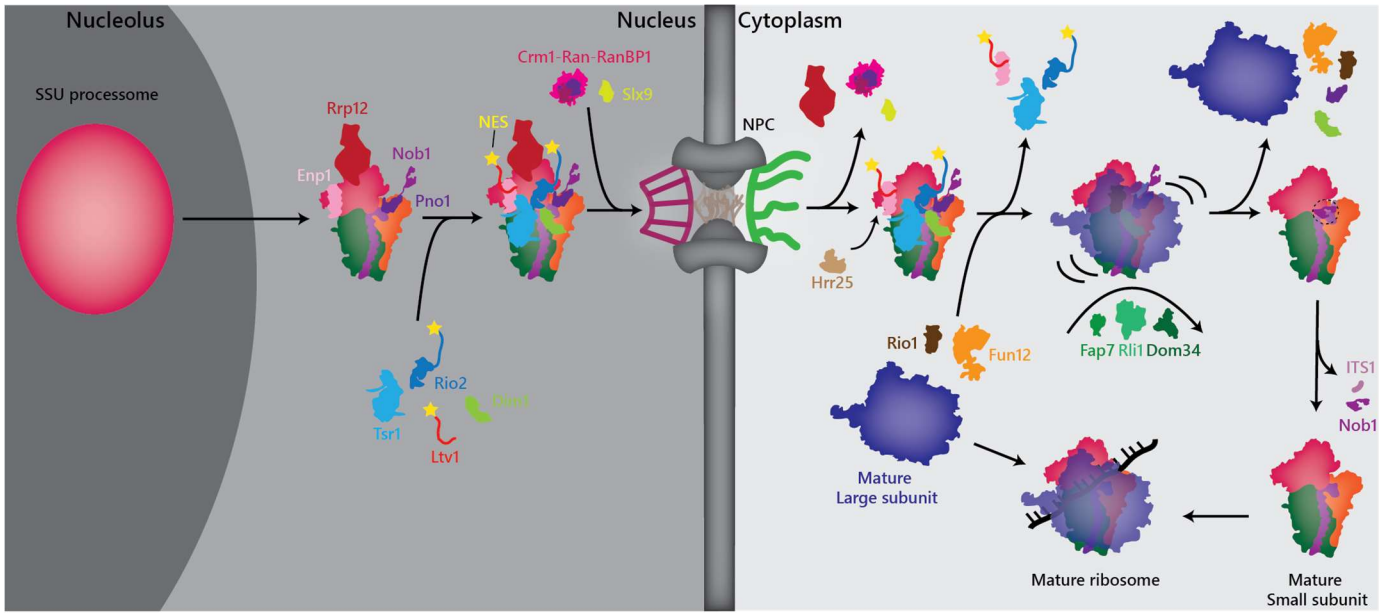


Figure 1.7. Later stages of small ribosomal subunit assembly. Overview of small subunit biogenesis from late nucleolar stage to the final cytoplasmic step. Pre-40S particles form in the nucleus and are exported by several redundant pathways. After several factors leave in the cytoplasm, 80S-like complexes form. Their dissociation leads to D-site cleavage and mature small subunit formation.

1.8 Nucleolar and nuclear maturation of the large subunit precursors

Co-transcriptional assembly of factors onto the second half of the 35S pre-rRNA lead to the formation of the earliest large subunit precursors. After the pre-rRNA is cleaved at sites A2 or A3, large subunit processing occurs in parallel to small subunit biogenesis. In the nucleolus, the earliest pre-60S particles contain the 27SA RNA species, bound by several assembly factors, including factors associated with ITS2, Cic1, Erb1, Nop7 and Nop15 (Granneman et al. 2011; Wu et al. 2016; Kater et al. 2017; Sanghai et al. 2018) as well as 27SA specific factors Pwp1 and Nop12 (Granneman et al. 2011; Talkish et al. 2014). There, it undergoes 3' and 5' processing to lead to the formation of the nucleolar pre-60S, a 27SB containing particle which is typically characterized by the presence of the assembly factors Nsa1 (Kressler et al. 2008). In these early stages, the 5S rRNA is also incorporated to the pre-60S, although structures of nuclear pre-60S particles have shown that it is in a rotated state relative to its mature structure (Wu et al. 2016).

The transition from nucleolus to nucleus is still poorly understood but is known to require the activity of the large AAA+ protein Mdn1 (also known as Rea1/Midasin). Mdn1, a dynein-related AAA+ protein, can bind ubiquitin-like domains through its Midas domain and use its ATPase activity to remove factors harboring these domains, such as the nucleolar assembly factor Ytm1, or the nuclear assembly factor Rsa4 (Ulbrich et al. 2009; Baßler et al. 2010; Raman et al. 2016). These ubiquitin-like folds can either be added as a result of post-translational modification, i.e SUMOylation, or be inherent parts of the polypeptide sequence of certain factors, such as Ytm1.

In the nucleus, pre-60S particles are defined by the presence of the assembly Nog2, a putative GTPase (Saveanu et al. 2001). The structure of the Nog2 particle has revealed

that the pre-rRNA adopts a conformation similar to that of the mature ribosome (Wu et al. 2016), since domains of the 25S are folded in a near-mature state. However, this particle is still heavily chaperoned by many assembly factors, mostly located on what will become the subunit interface, which contains most of the important functional sites of the maturing large subunit. Furthermore, the polypeptide exit tunnel, which has already formed in this nuclear intermediate, is nonetheless plugged by an extension of the protein Nog1 and covered by the protein Arx1, which may ensure its stabilization. At this stage, the immature 5S remains bound by Rrs1 and Rpf2, and needs to undergo a 180-degree rotation to adopt its mature state (Leidig et al. 2014). ITS2 has not been cleaved and processed at this stage (Wu et al. 2016).

Nuclear maturation of the pre-60S involves predominantly the rotation of the 5S rRNA and the removal or nearby factors, as well as ITS2 processing. The remodeling of the central protuberance, of which the 5S rRNA is the main component, is triggered by the action of several assembly factors: the Rix1 complex, the factor Sda1 as well the AAA+ ATPase Mdn1 (Ulbrich et al. 2009; Barrio-Garcia et al. 2016). Mdn1, here is involved in the removal of the factor Rsa4, which is located near the 5S. The structure of pre-60S particles containing the Rix1 complex and Mdn1 revealed that the 5S rRNA had undergone rotation to a near-mature state but that Rsa4 was still present (Barrio-Garcia et al. 2016). This structure further showed that the binding sites of Sda1 and Rpf2 were overlapping. It is unclear if the removal of the 5S chaperones Rpf2 and Rrs1 happens prior to Sda1/Rix1 complex binding or as a result of it. Nonetheless, these data point to a model of central protuberance maturation where the 5S rotates in its near mature

state as Rpf2 and Rrs1 dissociate and Sda1, Mdn1 and the Rix1 complex bind the nuclear pre-60S. This is followed by the ATP-driven removal of Rsa4 by Mdn1.

Interestingly, structures of nuclear pre-60S particles have shown that the remodeling of the central protuberance can either precede (Wu et al. 2016) or succeed (Barrio-Garcia et al. 2016) the removal of ITS2. However, mutation of Rix1 which prevent the recruitment the Mdn1 to these particles prevented ITS2 processing (Barrio-Garcia et al. 2016). These data suggest that Mdn1 binding is required for ITS2 processing.

As described previously, removal of ITS2 involves initial cleavage at the C2 site by the Las1 complex, which happens in the nucleus (Gasse et al. 2015). This results in the emergence of the so-called 7S, the precursor of the 5.8S, which still requires 3' end processing. Nop53, an assembly factor which binds near ITS2, was shown to recruit the helicase Mtr4, through the AIM motif (Thoms et al. 2015), like the small subunit assembly factor Utp18. This leads to exosome recruitment and 3' to 5' exonucleolytic degradation of the 3' end of the 7S (Schuller et al. 2018). Mutations in Nop53 results in the accumulation of ITS2-containing large subunit-like particles in the cytoplasm. These faulty large subunits are targeted by the nonsense-mediated decay proteins (Sarkar et al. 2017).

1.9 Nuclear export and cytoplasmic maturation of the large subunit precursor

In contrast to small subunit assembly, nuclear export of large subunit intermediates is directed by a specific protein: Nmd3 (Hedges et al. 2005; Ma et al. 2017; Malyutin et al. 2017). Nmd3 is an NES-containing adapter protein to the exportin Crm1 and is essential for nuclear export of the pre-60S. The Mex67-Mtr2 dimer facilitates efficient export of the large subunit precursor but like for small subunit export, the mechanism is unclear (Yao et al. 2007).

Nmd3 and Nog1/Nog2 have overlapping binding sites (Wu et al. 2016; Ma et al. 2017; Malyutin et al. 2017) on the pre-60S. Removal of Rsa4 by Mdn1 and release of Nog2, in a GTP hydrolysis dependent manner, was shown to be important for the recruitment of Nmd3 (Matsuo et al. 2014). Surprisingly, Nog1 was shown to be a component of Nmd3 containing particles suggesting that either Nog1 or Nmd3 may have dynamic conformations during ribosome biogenesis, allowing for their simultaneous presence (Kallstrom et al. 2003; Matsuo et al. 2014).

In the cytoplasm, the last stages of maturation occur. The main step is catalyzed by the AAA+ protein Drg1, which triggers Rlp24 release and the subsequent release of nearby factors, such as Nog1 (Pertschy et al. 2007; Kappel et al. 2012). This leads to the recruitment of Rpl24 and Rei1, which, like Nog1, inserts an extension in the PET, presumably to protect this important functional site, or otherwise prevent premature protein synthesis (Greber et al. 2016). This step is followed by Arx1 release in a Hsp70 and Jjj1 dependent manner, a mechanism reminiscent of nascent protein chaperoning, possibly suggesting that it may be a quality control step (Meyer et al. 2007).

The last step is the removal of the exporter Nmd3 and the factor Tif6, which blocks binding to the small subunit. The release of Nmd3 depends on the activity of the GTPase Lsg1, which is activated by binding to H69 of the 25S, possibly signaling proper assembly of the 25S (Hedges et al. 2005; Ma et al. 2017; Malyutin et al. 2017). Tif6 is released through a complex mechanism involving the factor Sdo1 and the GTPase Efl1 (Lo et al. 2010; Finch et al. 2011). Efl1 has homology to the translation factor eEF2, which induces translocation of the tRNA to the P site during translation.

A study of a human - *D. Discoideum* heterologous system, has revealed the structures of Tif6 pre-60S bound by either the human Sdo1 homolog, SBDS, or both SBDS and EFL1 (Weis et al. 2015). post-Tif6 release complex still bound by those two factors. In these structures, Sdo1 is located in the P site, near the peptidyl transferase center, and covers the entrance of the polypeptide exit tunnel, thereby protecting the main functional sites of the large subunit. Efl1 binds near the canonical translational GTPase center, which induces Tif6 release. While the timing of GTP hydrolysis is currently unclear, it is thought that it may be coupled to dissociation of those factors. This mechanism in reminiscent of a translocation cycle, where Sdo1 acts as a tRNA mimic and Efl1 is the eEF2 mimic (Weis et al. 2015). This would represent a quality control step analogous to the formation of 80S-like complexes in small subunit maturation.

Release of Tif6 leads to the formation of the mature large ribosomal subunit, which can now bind the mature small subunit to form translation competent 80S complexes.

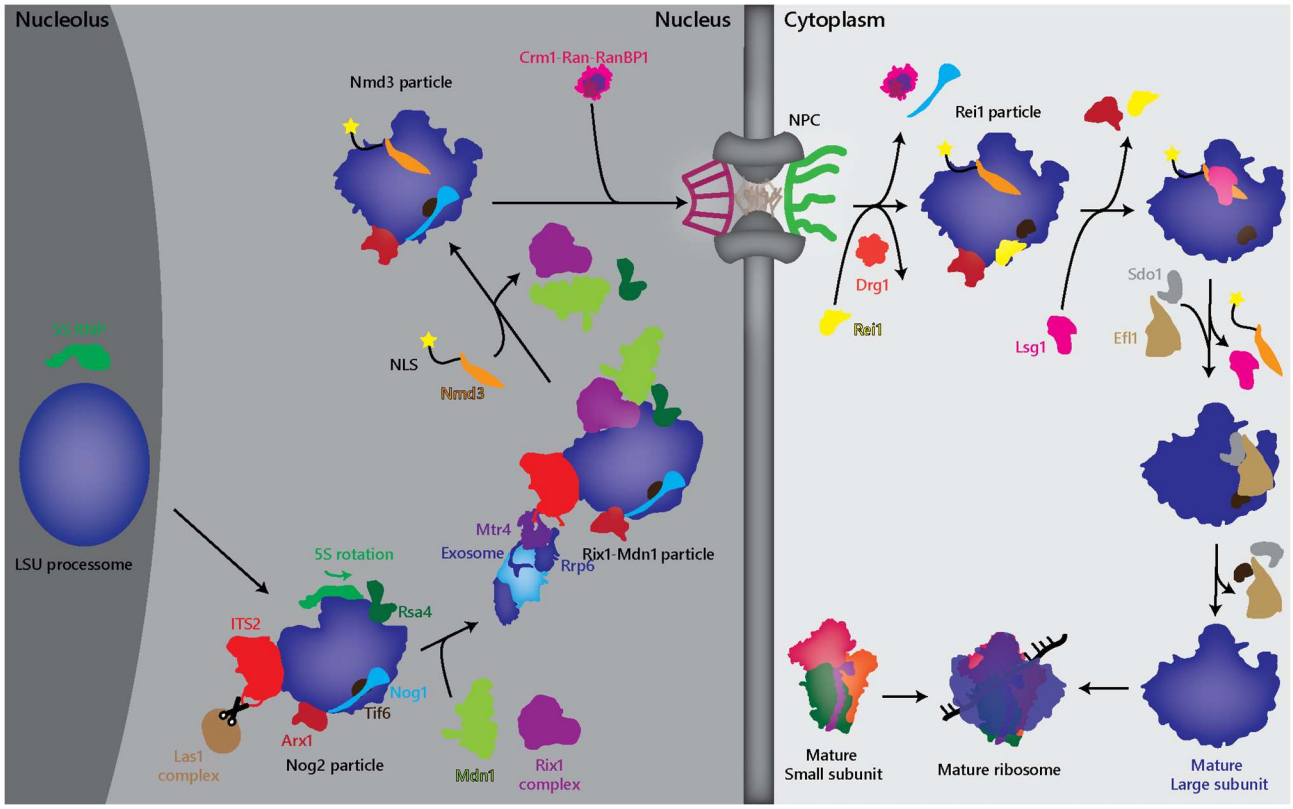


Figure 1.8. Large ribosomal subunit assembly. Overview of large subunit maturation in the nucleus and cytoplasm. After ITS2 processing by the Las1 complex and the exosome, remodeling of the 5S RNP, and removal of nuclear factors by Mdn1, pre-60S particles are exported to the cytoplasm, where Drg1, Lsg1 and Efl1 induce factor dissociation and lead to the formation of the mature large subunit.

1.10 Regulation of ribosome assembly

Given the importance of constant ribosome production in a growing cell, the biogenesis of ribosomes is tightly regulated to account for energetic needs and nutrient availability. Regulation of ribosome biogenesis occurs at several levels: the transcription of rRNA by RNAP I, the transcription of ribosomal protein genes by RNAP II, and further down during the assembly pathway (Philippi et al. 2010; Iacovella et al. 2015; Albert et al. 2016).

Nutrient availability regulates TOR activity: when nutrients are abundant, TOR is activated and phosphorylates several downstream effectors (Albert and Hall 2015). One indirect downstream target of TOR is RNA polymerase I, which is therefore regulated in response to the nutrient state (Philippi et al. 2010). In response to starvation, RNAP I can dimerize, preventing its association with the initiation factor Rrn3 and therefore transcription (Torreira et al. 2017). It was suggested that this is partly regulated by phosphorylation of Rrn3 (Mayer et al. 2004; Philippi et al. 2010; Torreira et al. 2017). This elegant mechanism permits a rapid response to changing environmental factors, while storing the enzyme. The assembly factor Rio1 is involved in the phosphorylation of a conserved tyrosine on Rpa43, a subunit of RNAP I. This phosphorylation promotes rDNA transcriptional repression prior to chromosome segregation, which has been suggested to prevent double strand breaks in the DNA during mitosis (Iacovella et al. 2015). Rio1's multiple roles in both RNA Pol I regulation and pre-40S particles might suggest a negative feedback mechanism, the details of which are still unclear.

Casein kinase 2 (CK2) is also implicated in the regulation of ribosomal biogenesis (Albert et al. 2016; Kos-Braun et al. 2017). UtpC, a component of the SSU processome,

was also shown to associate with both CK2 and transcriptional factor Ifh1 (Albert et al. 2016) to form the CURI complex. Ifh1 regulates ribosomal protein gene transcription but can be titrated off the promoters by binding to UtpC/CK2. This process is regulated by TOR activity through its downstream effector Sch9. This process highlights the multiple ways in which ribosome biogenesis can be regulated by the TOR pathway. Lastly, the kinase Kns1, which regulates RNAP III transcription, also links CK2 and TOR pathways. Kns1 is inactivated by TOR and can modulate the activity of CK2 (Sanchez-Casalongue et al. 2015).

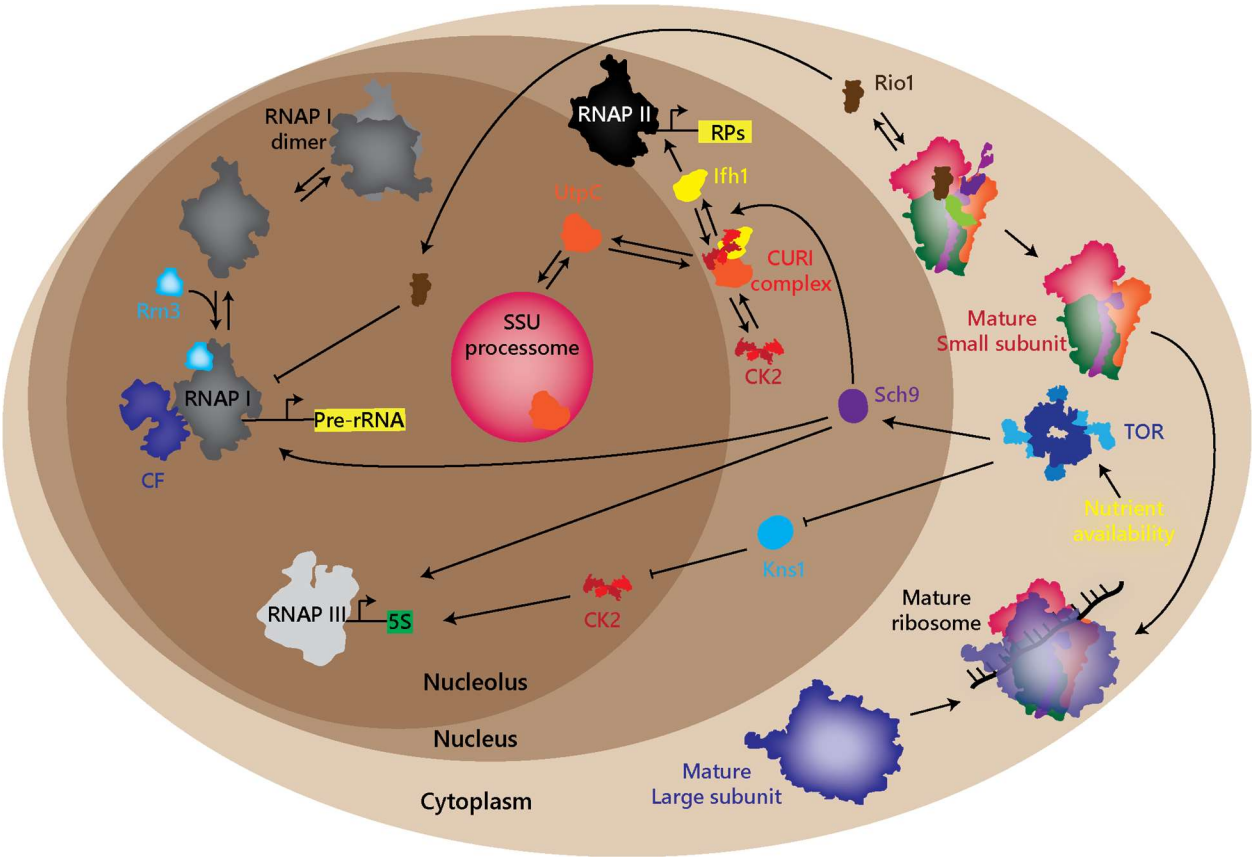


Figure 1.9. The TOR pathway regulates translation and ribosome assembly. TOR is the master regulator of ribosome biogenesis. TOR affects Kns1 and Sch9, which regulate RNAP III transcription, RNAP II transcription of ribosomal protein genes, and RNAP I transcription of pre-rRNA. RNAP I can be silenced by dimerization which prevents Rrn3 association. CK2 also regulates the SSU processome component UtpC.

1.11 Ribosome assembly and human diseases

Most genes involved in the ribosome assembly are essential and their deletion is lethal to cells (Woolford and Baserga 2013). However, rare genetic mutations in genes involved in ribosome biogenesis can lead to numerous developmental diseases. This group of diseases has been named ribosomopathies (McCann and Baserga 2013; Sondalle and Baserga 2014).

However, despite their common origin, these diseases can lead to various diseases. Diamond Blackfan Anemia is caused by mutations in various ribosomal proteins of both the small and large subunit and is characterized, among other symptoms, by low red blood cell counts, craniofacial malformations and cleft palate (Choesmel et al. 2007). A recent study has elucidated the pathogenesis of this disease (Khajuria et al. 2018)

Shwachman-Bodian-Diamond syndrome arises from mutations in a single protein, SBDS, homologous to Sdo1 in yeast, which is involved in the late stages of large subunit maturation (Menne et al. 2007). Symptoms include bone marrow dysfunction, pancreatic insufficiency and skeletal abnormalities. North American Indian Childhood Cirrhosis is caused by a single mutation in Utp4, a subunit of UtpA (Chagnon et al. 2002). A mutation in the GTPase Bms1 can lead to Aplasia cutis congenita, a skin/hair disease (Marneros 2013). Various mutations in rpSA (rpS0 in yeast) are associated with isolated congenital asplenia, the absence of spleen in individuals, with no other developmental defect (Bolze et al. 2013). Dyskeratosis congenita, which can be caused by mutations in the box H/ACA proteins, causes symptoms which mainly affect the skin but leads to bone marrow failure. Lastly, mutations in the RNase MRP RNA gene cause

the disease cartilage-hair hypoplasia, involving short stature, hair hypoplasia, and predisposition to lymphomas and other cancer.

Furthermore, ribosome biogenesis is heavily upregulated in cancer cells; presumably to account for higher needs in protein synthesis. This has led some to propose that ribosome assembly or RNA polymerase I could represent interesting therapeutic targets for cancer (Pelletier et al. 2017). Taken together, the number of diverse symptoms and diseases affected by faulty or excessive ribosome biogenesis highlights the need for a better mechanistic understanding of this process, which would provide insights into the pathogenesis of these conditions.

Chapter II - Small subunit processome assembly on pre-ribosomal RNA

Early steps in ribosome biogenesis involve the co-transcriptional assembly of ribosome assembly factors on the pre-rRNA. Our understanding of this process was limited by a lack of techniques which could provide a temporal view of these early assembly events. We therefore set out to design a new technique that would allow us to dissect the co-transcriptional assembly of biogenesis factors. The result of our findings were reported in this publication (Chaker-Margot et al. 2015). Mirjam Hunziker performed all the RNA anaylsis.

Our initial rationale was that the expression of truncated RNA could serve as a mimic for the process of pre-rRNA transcription (Fig. 2.1A). However, such an approach would require a way to isolate the expressed RNA specifically. The MS2 aptamer had been used to purify pre-ribosomal complexes (Gupta and Culver 2014), which suggested this type of biochemical tag could be used to isolate analyzable amounts of pre-ribosomal complexes from cells. Furthermore, RNA polymerase II based overexpression of rDNA was previously shown to yield viable ribosomes in yeast (Nogi et al. 1991). This suggested that the use an RNA polymerase II based promoter to overexpress our targeted RNA would not prevent the assembly of ribosome assembly factors. Initial efforts consisted in the co-overexpression of MS2-GFP with 5' ETS RNA tagged at its 5' end with 5 copies of the MS2 loops. GFP pulldown of the resulting lysate yielded an abundance of MS2-GFP and faint amounts of other proteins, which were thought to be contaminants (data not shown). This result was interpreted to mean that MS2-GFP was not able to specifically recognize the RNA to sufficient amounts.

To try to improve this method, we GFP-tagged an early assembly factor, UtpA subunit Utp10, and added the recombinant MS2 protein, tagged with an sbp tag, in large amounts into the lysate (Fig. 2.1A). We hypothesized that Utp10 might be better able to recognize the overexpressed RNA. We further surmised that the addition of MS2-sbp in the lysate in large amounts would allow its binding to the overexpressed RNA. Therefore, by pulling down the assembly factor and its associated factors, we would be able to pulldown the target RNA, along with other endogenous ribosome biogenesis complexes (Fig. 2.1B, Fig 2.2A). The targeted complex could then be recaptured using the sbp tag (Fig. 2.1C).

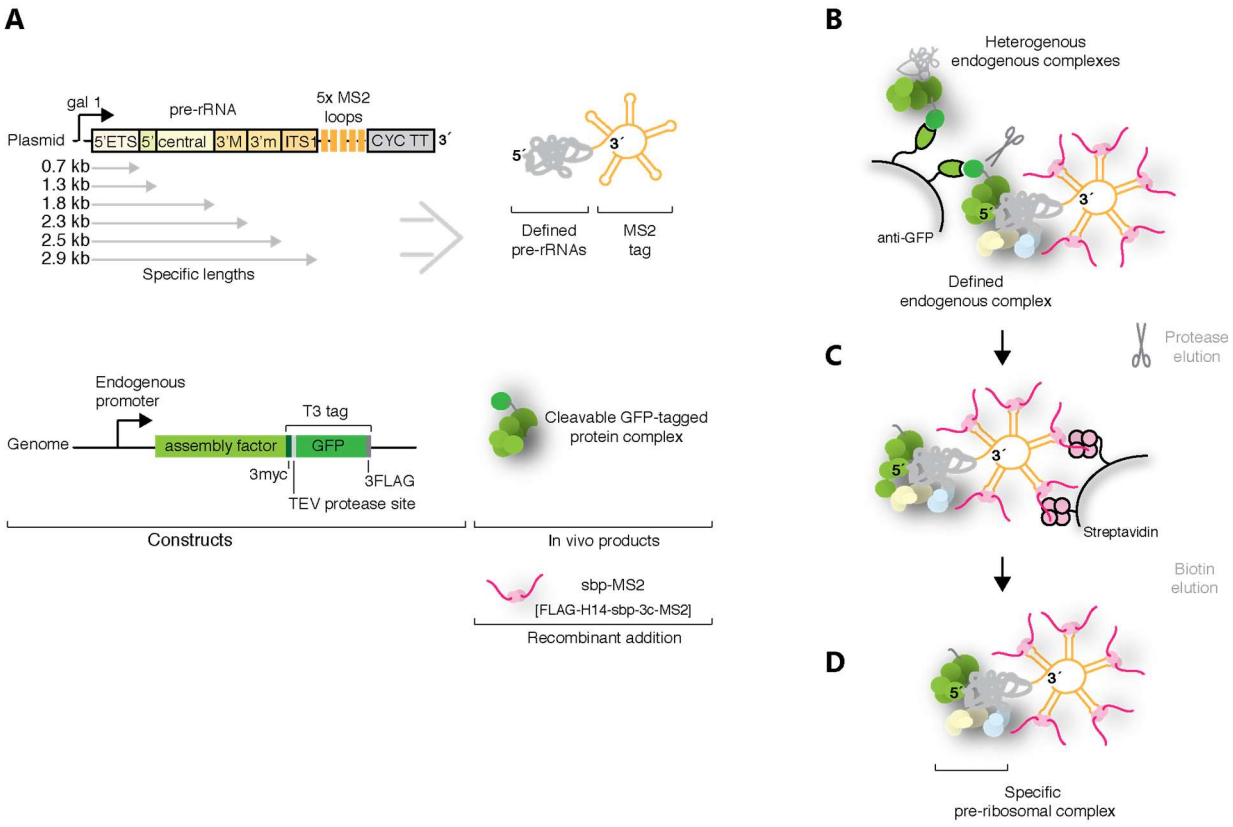


Figure 2.1 Purification of pre-ribosomal particles. (**A**) Plasmid overexpression of aptamer tagged pre-rRNAs of different length in genetically modified yeast strains containing a GFP-tagged ribosome assembly factor. Plasmids contain the gal1 promoter and CYC transcriptional terminator (CYC TT). Pre-rRNA domains, GFP-tagged ribosome assembly factor and recombinant sbp-MS2 protein are shown in shades of yellow, green and pink respectively. (**B-D**) Purification strategy using GFP (B) and sbp (C) as sequential baits followed by biotin elution (D).

Using the MS2 tagged 5' ETS as a target, this approach was successful is isolating a seemingly specific set of bands, which were not seen for the negative control (Fig. 2.2A). Protein capture with Utp10 only showed a more complex sample with several additional higher molecular weight proteins bands (Fig. 2.2A). Changing the tagged assembly factor to other early factors like Rrp9 or Utp1, or changing the position of the MS2 apatamer did not change the apparent composition of the isolated complex, which seemed to suggest the specificity of the technique (Fig. 2.2A). The overexpressed RNA could be observed at the correct size *in vivo*, showing that it was stable (Fig. 2.2B).

Mass spectrometry analysis of the particles obtained by overexpressing the tagged 5' ETS revealed the presence of a 2 megadalton particle, which in addition to UtpA, UtpB, the U3 snoRNP and MPP10 complexes also included several other pre-90S factors, such as Utp3, which by sequence homology with Rrp47 contains a an exosome interaction domain (Mitchell 2010; Schuch et al. 2014) (Fig. 2.2A, 2.3).

We were initially concerned with the possibility that MS2 tags may interfere with the normal assembly of ribosome biogenesis factor onto the pre-rRNA. To test this, we added the MS2 tag at the 5' end of the 35S pre-rRNA, on a plasmid carrying a complete rDNA locus (Fig. 2.2D). The plasmid was transformed into a yeast strain expressing all its pre-rRNA from a single plasmid. The strain expressing MS2-tagged pre-rRNA did not show any growth defect relative to a control strain expressing untagged pre-rRNA from the same plasmid (Fig. 2.2C). The MS2 tag at the 5´ end of pre-rRNA did not prevent transcription or pre-rRNA processing *in vivo* (Fig2.2E). These results suggested the MS2 tag could be used to capture truncated pre-rRNAs without interfering with the normal assembly of assembly factors on the RNA.

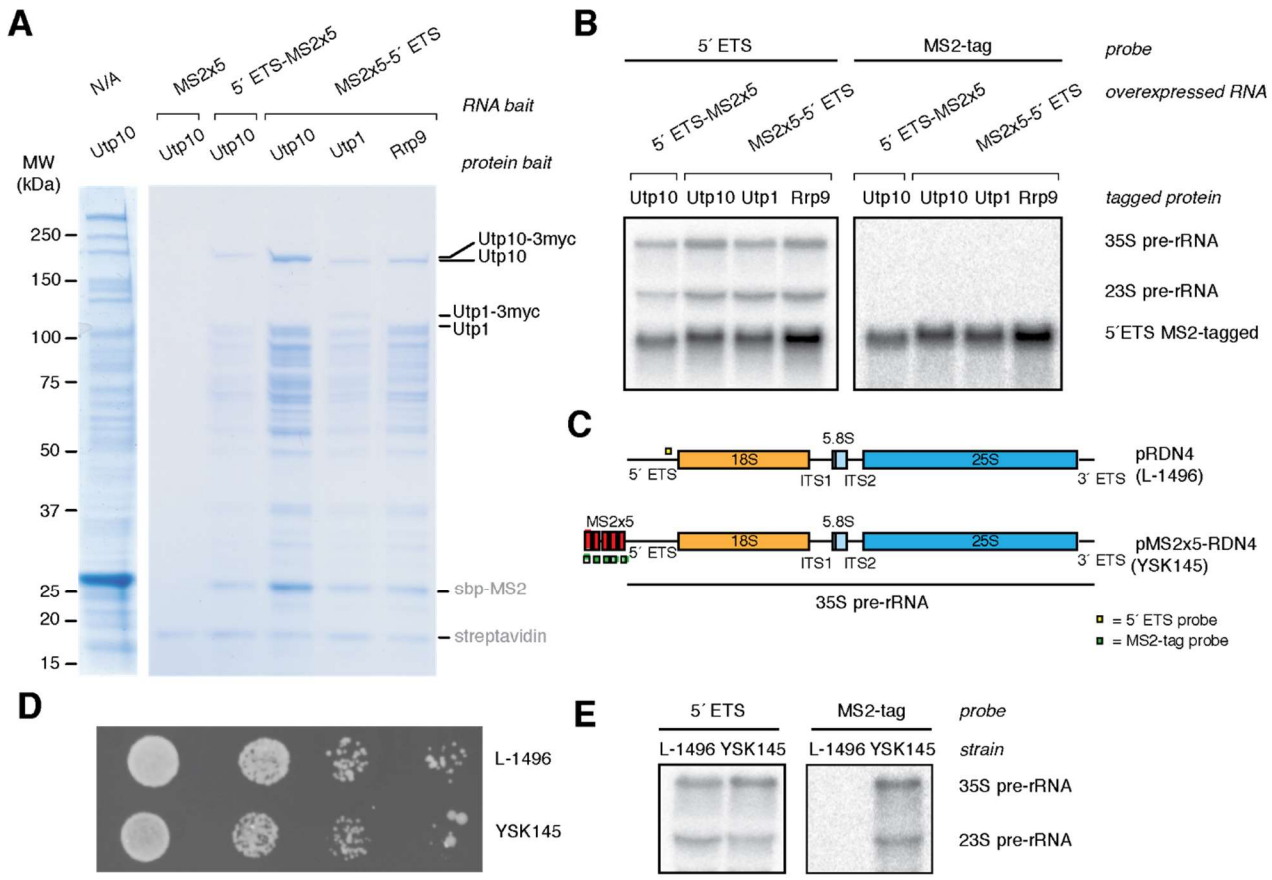


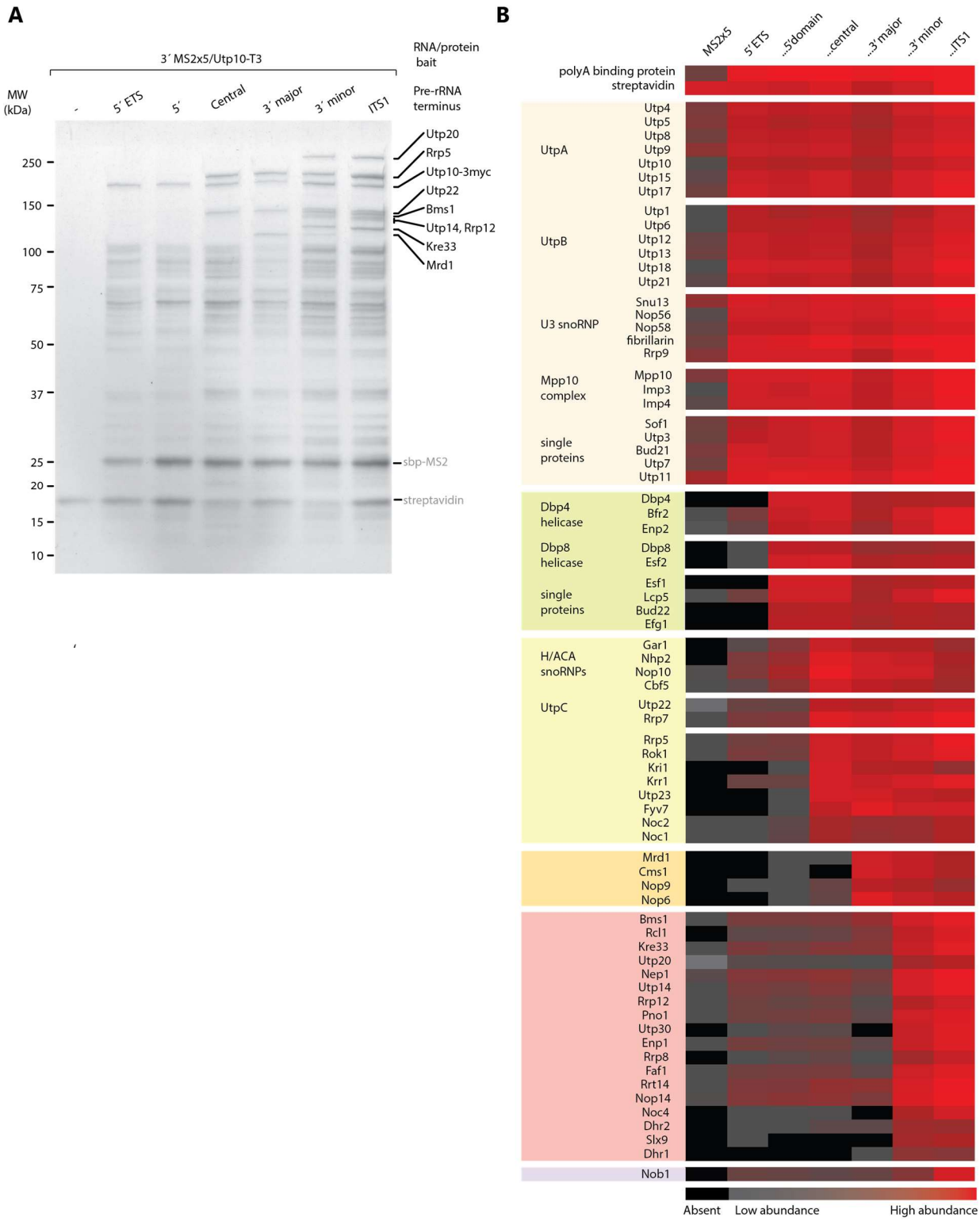
Figure 2.2 A complex is formed between 26 proteins and the 5´ ETS in the presence of the MS2 tag, which does not inhibit ribosome assembly in vivo. (**A**) SDS PAGE analysis of affinity purified RNA-protein complexes with either protein only or RNA and protein baits as indicated on the top. GFP-tagged UtpA (via Utp10), UtpB (via Utp1) and U3 snoRNP (via Rrp9) were used as baits to purify the complexes. Lanes 2 and 3 contain Utp10 with 3´ or 5´ tagged 5´ ETS sequences. Positions of protein baits, and proteins that are part of the purification scheme are indicated on the right. Corresponding mass spectrometry analysis is listed in Supplementary Table 1. (**B**) Northern blot analysis for 5´ MS2-tagged 5´ ETS transcripts shown in (A) using probes specific for the 5´ ETS (left) and the MS2 aptamer (right). Native 35S and 23S pre-rRNA intermediates are highlighted. (**C**) Schematic representations of the 35S pre-rRNA transcripts that are generated from pRDN4 (top) and pMS2x5-RDN4 (bottom). Boxes indicate positions of probes. (**D**) Growth phenotypes of yeast strains expressing a single copy of untagged (L-1496) or MS2-tagged (YSK145) rDNA. YPD plates were incubated for 1 day at 30 °C. (**E**) Northern blot analysis of pre-rRNAs from strains L-1496 and YSK145 using probes specific to the 5' ETS (left) and the MS2 aptamer (right). The RNA analysis shown here was the work of Mirjam Hunziker.

Stage-specific assembly of the small subunit processome

As UtpA associates with the 5´ ETS, using longer pre-rRNA transcripts with an MS2 tag at the 3´ end guarantees that intact pre-rRNPs are purified in the two-step protocol. The sequential addition of the four rRNA domains of the 18S rRNA (5´, central, 3´major and 3´minor domains) and ITS1 facilitates the identification of ribosome assembly factors as a function of rRNA folding (Fig. 2.1). This procedure does not elucidate direct binding sites of ribosome assembly factors within rRNA domains. Rather, the addition of folded rRNA domains and the resulting proximal and distal structural changes within the entire particle allow new ribosome assembly factors to bind at new sites.

High stringency was used for the analysis of ribosome assembly factors. We have excluded ribosomal proteins from our mass spectrometry analysis, as signals of highly abundant peptides from contaminating mature ribosomes prevent an unambiguous assignment of ribosomal proteins to a particular stage of ribosome assembly in our system (Tackett et al. 2005). This set of experiments was done in duplicate. The data for the second replicate can be found in Appendix 7.1.

Figure 2.3 Stage-specific associations of ribosome assembly factors. (**A**) SDS-PAGE analysis of defined stages of SSU processome assembly using Utp10 as protein bait and pre-rRNA baits with MS2 tags at the 3´ end. Identified proteins are indicated on the right. (**B**) Heat-map of protein abundance as determined by mass spectrometry and label free quantitation (iBAQ), as a function of transcript length for ribosome assembly factors for which clear stage-specific assignments could be made. High abundance proteins are shown in red and low abundance proteins in grey. Proteins absent from the sample are shown as a black box. Proteins are colored in beige (5´ ETS), green (5´ domain), light green (central domain), yellow (3´ major domain), red (3´ minor domain) and purple (ITS1).


A
3´ MS2x5/Utp10-T3
RNA/protein bait
MW (kDa)
5´ ETS
5´
Central
3´ major
3´ minor
ITS1
Pre-rRNA terminus
250
150
100
75
50
37
25
20
15
10
Utp20
Rrp5
Utp10-3myc
Utp22
Bms1
Utp14, Rrp12
Kre33
Mrd1
sbp-MS2
streptavidin
B
MS2x5
5´ETS
...5´domain
...central
...3´major
...3´minor
...ITS1
polyA binding protein
streptavidin
UtpA
Utp4
Utp5
Utp8
Utp9
Utp10
Utp15
Utp17
UtpB
Utp1
Utp6
Utp12
Utp13
Utp18
Utp21
U3 snoRNP
Snu13
Nop56
Nop58
fibrillarin
Rrp9
Mpp10 complex
Mpp10
Imp3
Imp4
single proteins
Sof1
Utp3
Bud21
Utp7
Utp11
Dbp4 helicase
Dbp4
Bfr2
Enp2
Dbp8 helicase
Dbp8
Esf2
single proteins
Esf1
Lcp5
Bud22
Efg1
H/ACA snoRNPs
Gar1
Nhp2
Nop10
Cbf5
UtpC
Utp22
Rrp7
Rrp5
Rok1
Kri1
Krr1
Utp23
Fyv7
Noc2
Noc1
Mrd1
Cms1
Nop9
Nop6
Bms1
Rcl1
Kre33
Utp20
Nep1
Utp14
Rrp12
Pno1
Utp30
Enp1
Rrp8
Faf1
Rrt14
Nop14
Noc4
Dhr2
Slx9
Dhr1
Nob1
Absent
Low abundance
High abundance

The consecutive addition of structured rRNA domains of the 18S rRNA results in the stepwise association of assembly factors forming a 6 megadalton SSU processome. Most factors associate with a sharp transition at a particular stage, whereas a few factors gradually increase with transcript length, thereby preventing direct assignments (Fig. 2.3 and 2.4). In addition to the 5´ ETS associated proteins, a pre-rRNA transcript extending to the 5´ domain contained two helicase complexes (Dbp4 and Dbp8) of which Dbp4 is known to bind to this domain (Soltanieh et al. 2014; Granneman et al. 2006). Lcp5, a protein containing an exosome interaction domain (Wiederkehr et al. 1998) is also highly enriched (Fig. 2.4B). In the presence of the central domain, H/ACA snoRNPs gradually become more abundant and known binders of expansion segment 6 (UtpC, Rok1) and several other essential proteins including Rrp5, Kri1, Krr1 and Utp23 associate (Lin et al. 2013; Lebaron et al. 2013) (Fig. 2.4C). The presence of the 3´ major domain, which later forms the head structure, results in the appearance of the essential proteins Mrd1 and Nop9 (Segerstolpe et al. 2013; Thomson et al. 2007) (Fig. 2.4D).

The structural completion of the 18S rRNA by the 3´ minor domain recruits numerous proteins including the large heat repeat protein Utp20, the cytosine acetyltransferase Kre33 and the helicase Dhr1, which later separates U3 from pre-rRNA (Sardana et al. 2015) (Fig. 2.3 and 2.4E). The final addition of ITS1 recruits the D site nuclease Nob1 (Turowski et al. 2014) (Fig. 2.4F).

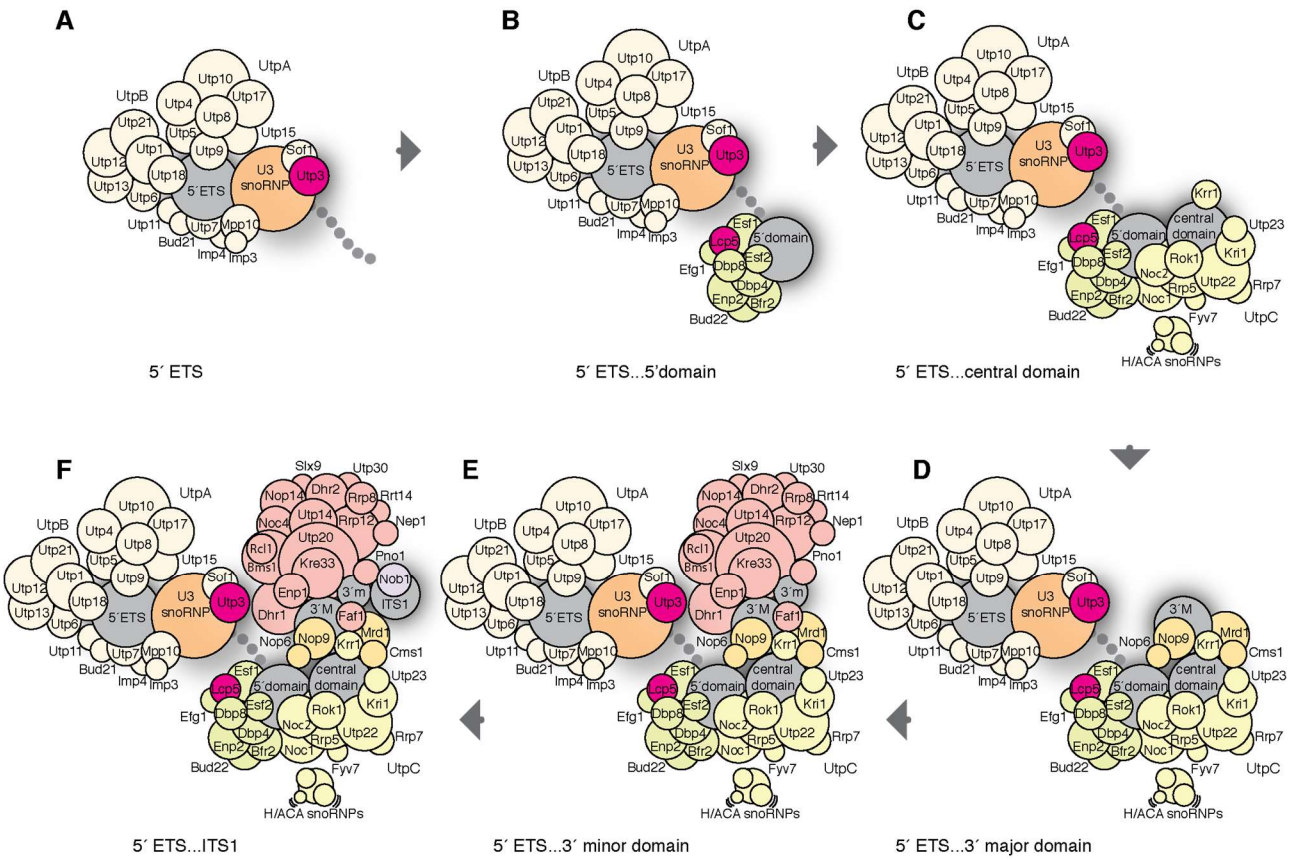


Figure 2.4 Model for stage-specific associations of ribosome assembly factors. (**A-F**) Schematic representation of mass spectrometry data showing the stepwise assembly of pre- ribosomal particles with color-coded assembly factors. Proteins are colored in beige (5´ ETS), green (5´ domain), light green (central domain), yellow (3´ major domain (3´M)), red (3´ minor domain (3´m)) and purple (ITS1). Predicted exosome interactors and pre-rRNAs are shown as circles in purple and grey respectively.

Conclusions

By using truncated pre-rRNA transcripts we have recapitulated and arrested early events of eukaryotic ribosome assembly. This technique has not only confirmed the association of previously mapped ribosome assembly factors, such as UtpC (Lin et al. 2013), Mrd1 (Segerstolpe et al. 2013) and Nob1 (Turowski et al. 2014), but also elucidated the stages at which numerous known ribosome assembly factors associate with pre-rRNA as a function of transcription. The specificity of this system is further underlined by the marked absence of late pre-40S factors such as Tsr1, Rio2 or Ltv1 (Ferreira-Cerca et al. 2012).

Our method nonetheless presents certain limitations. First and foremost, the overexpression and capture of RNA occurs over much longer time than the timespan of the most transient interactions in ribosome biogenesis. While mass spectrometry is sensitive enough to allow us to observe low abundance proteins, our method is unable to capture the more dynamic aspects of ribosome assembly. Furthermore, with new structural data of the SSU processome (Chaker-Margot et al. 2017; Barandun et al. 2017), it has now transpired that several factors recruited here, such as Mrd1 or Dbp4, among others, leave the particle, performing a more short-lived role in SSU processome formation. While this may have been intuited, it was not reflected in the mass spectrometry data. Similarly, structural data showed clearly the presence of factors such as Fcf1 and Fcf2 in the SSU processome, but their relatively low abundance in the mass spectrometry prevented their inclusion in the model of the SSU processome assembly.

Notwithstanding those caveats, our method has provided a new platform for the study of early ribosome biogenesis. At the earliest stage we have identified the 5´ ETS particle, a complex of 26 proteins at the 5´ ETS, which has a molecular weight of 2 megadalton if each protein is present as a single copy. The complete SSU processome will thus have a combined molecular weight of approximately 6 megadalton, assuming 1 copy of every protein present here.

We propose a revised model for early eukaryotic ribosome assembly events, which involves the consecutive recruitment of ribosome assembly factors on pre-rRNA domains. Ribosome assembly factors are positioned in vicinity to the rRNA domain that needs to be present for their association with the SSU processome, as our method does not provide precise binding sites (Fig. 2.4). The large number of proteins being incorporated with the completion of the 3' minor domain may be indicative of a conformation change in the particle relative to previous stage, which could create new bindings site for several new ribosome assembly factors.

Protein complexes can be formed prior to their association with rRNA (UtpA, UtpB, UtpC or U3 snoRNP) or by the sequential recruitment of proteins through different rRNA domains, as illustrated by the Krr1-Faf1 complex (Fig. 2.4B-F). The placement of exosome interactors within the 5´ ETS particle (Utp3) and the 5´ domain (Lcp5) further suggests an elegant mechanism for exosome-mediated degradation of the 5´ ETS with recycling of ribosome assembly factors or degradation of aberrant pre-40S particles.

Chapter III - Structural characterization of the small subunit processome

Our study of the assembly of SSU processome factors onto the pre-rRNA has revealed the timing of association of some 75 factors and provided some insights into their possible functions. However, many factors, which have no discernable enzymatic activity, still have uncharacterized functions. A cryo-EM structure of the *Chaetomium thermophilum* SSU processome has been obtained but the low resolution of 7.3Å, prevented the functional interpretation of the core of this particle (Kornprobst et al. 2016). We therefore set out to obtain the cryo-EM structure of *S. cerevisiae* SSU processome at a higher resolution, which would not only uncover the roles of many yet uncharacterized factors in this intermediate, but also provide a platform for further biochemical characterization within the model organism for the study of ribosome biogenesis. We were able to obtain successive structures of 5.1Å and 3.8Å of resolution of this particle which were reported in two publications (Chaker-Margot et al. 2017; Barandun et al. 2017). This work was a joint effort by Jonas Barandun, Mirjam Hunziker and I.

Purification and analysis of the small subunit processome

We initially set out to determine the best purification condition to isolate the SSU processome. Our previous work on the assembly of the SSU processome suggested that seemingly stoichiometric SSU processome particles could be obtained from endogenously tagged yeast, with a single capture step of the assembly factor Utp10 (Fig. 2.2A). However, preliminary efforts to reproduce this purification were not successful in obtaining a similarly stoichiometric sample (data not shown). The main difference between those purification was in the medium used to grow the cells. Whereas the Utp10-single capture had been done from cells grown in synthetic medium (Fig. 2.2A), the newer and unsuccessful purifications were from cells grown in “rich” Yeast-peptone-dextrose (YPD) medium.

This allowed us to determine that pre-ribosomal particles accumulate in response to nutrient starvation, which occurs when cells are grown to high optical density in minimal medium and with a non-glucose source of carbon, e.g. galactose. This suggests that the SSU processome may not only represent an early assembly intermediate, but can also function as storage particle or a non-productive assembly intermediate of the small ribosomal subunit in response to these stress conditions (Kos-Braun et al. 2017).

We therefore grew a yeast strain carrying two genomic protein tags under these conditions to obtain sufficient amounts of stoichiometric SSU processomes for subsequent cryo-EM studies. An early factor, UtpB subunit Utp1, and the later factor and acetyltransferase Kre33 were determined to be ideal protein baits (data not shown). Tandem affinity purified SSU processomes (Fig. 3.1) were analyzed by mass spectrometry to confirm the composition of the particles (Appendix 7.2)

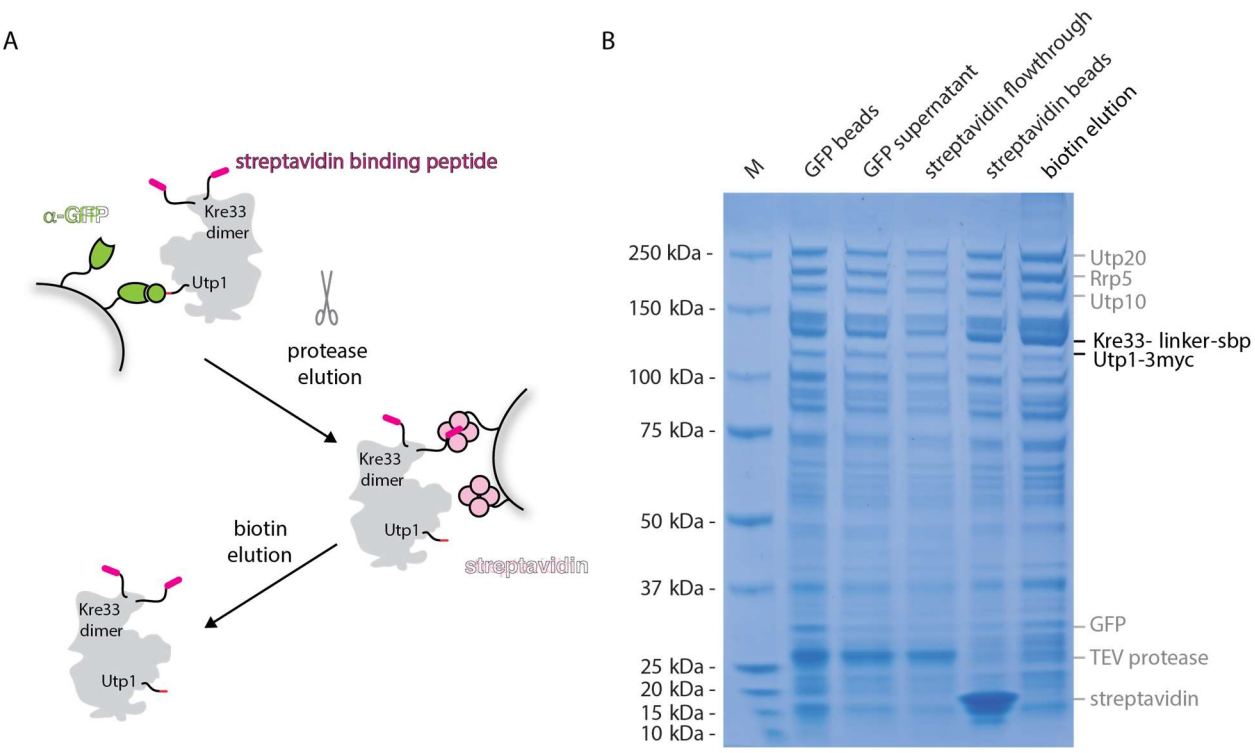


Figure 3.1. Purification of the small subunit processome. (**A**) Schematic overview of the tandem affinity purification procedure. SSU processomes are isolated from a yeast strain harboring a protease cleavable GFP tag (green) on Utp1 and a streptavidin binding protein tag (pink) on Kre33. Particles 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 (light-pink) beads. SSU processomes are eluted from streptavidin beads with biotin. (**B**) Coomassie blue stained 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 (M; marker).

Mass spectrometry analysis revealed that the composition of the purified SSU processome particles were very similar to the “ITS1” particle, as described in our study of SSU processome assembly (Chaker-Margot et al. 2015). The main difference was the definitive absence of several factors that were thought to be transient, like the RNA chaperones Nop9 and Mrd1, and the helicase Dbp4. A few more factors were seen in the analysis but were in very low abundance, comparable with large subunit assembly factors, suggesting they may not be present in the vast majority of particles. These included the H/ACA snoRNP proteins, and the helicase Dbp8 and its cofactor Esf2.

RNA was also extracted from pure particles to analyze the RNA species present in the complex. Analysis of the purified SSU processome RNA revealed the presence of U3 snoRNA, 18S rRNA precursors and a 600 nucleotide 5’ ETS. Comparative northern blotting analysis with an *in vitro* transcribed control RNA, showed that the 5’ ETS within the SSU processome has been cleaved at site A0, while cleavage at site A1 has not yet occurred in the major species of the 18S rRNA precursors (Fig. 3.2). Besides some degree of RNA degradation, heterogeneity of the major 18S rRNA precursor species is only observed in the 3’ ends beyond helix 44, which cannot be resolved in our cryo-EM reconstruction (Fig. 3.2). All RNA work was performed by Mirjam Hunziker.

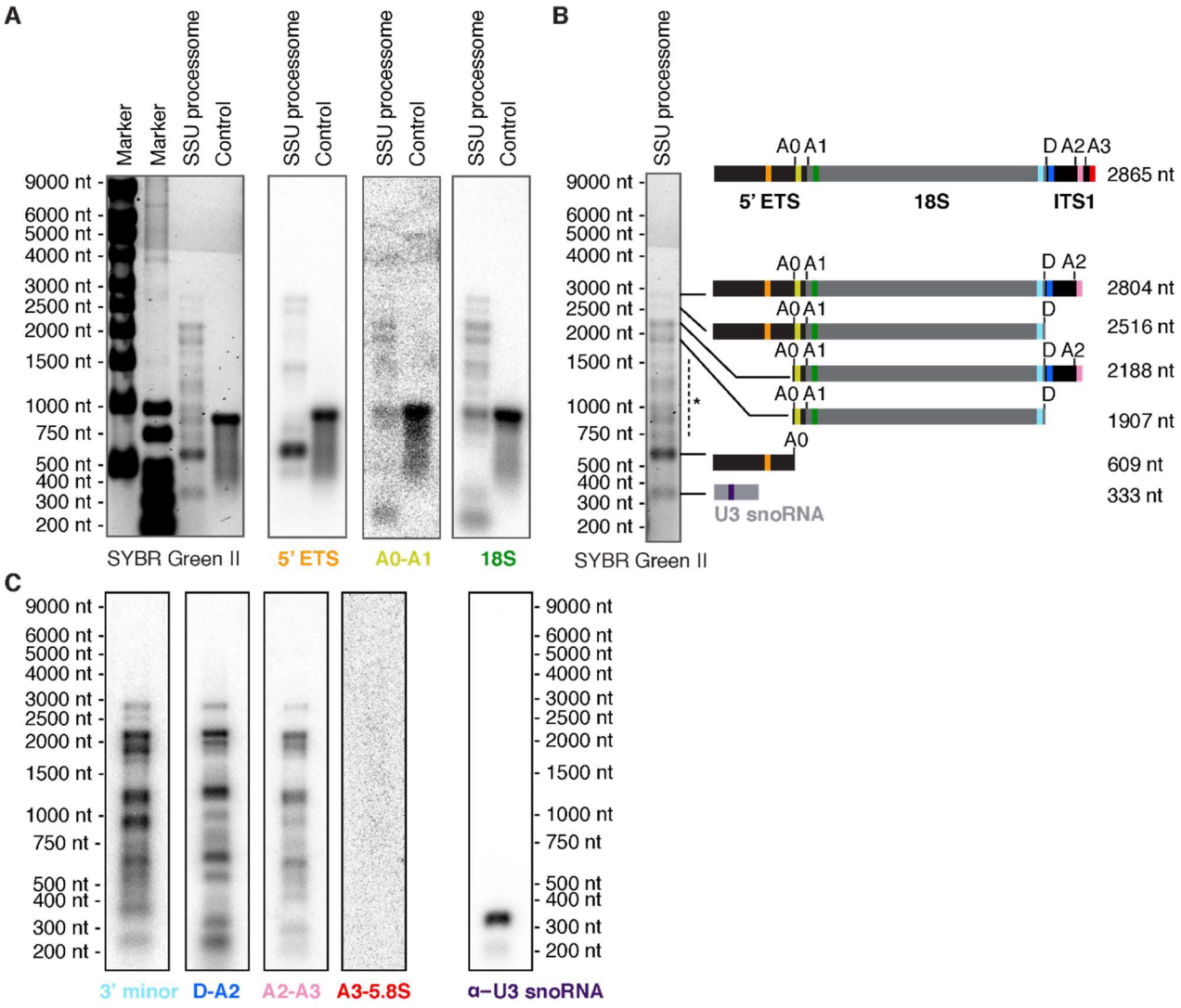


Figure 3.2 Analysis of RNA components present in the SSU processome sample. (**A**) Comparative analysis of SSU processome RNA and control pre-rRNA fragment (5' ETS to nucleotide (nt) 194 of 18S rRNA). RNAs are visualized by SYBR Green II and northern blotting with 5' ETS, 18S and A0-A1 probes. (**B**) Schematic representation of RNA species present in SSU processome sample. Northern blotting probes used in (A) and (C) are labeled and color-coded. An imaging artifact in the SYBR Green II stained gel (A and B) results in a color change above 4000 nt. The asterisk denotes unassigned degradation products of the pre-rRNA. (**C**) Northern blot analysis of SSU processome RNA components using 3'minor, D-A2, A2-A3, A3-5.8S and U3 snoRNA probes.

SSU processomes were analyzed by negative stain EM to study the homogeneity of the sample (Fig. 3.3A). The sample was mostly homogeneous and 2D classification of 36,000 particles yielded good 2D classes of a particle of the expected size (Fig 3.3B). 2D classes showed different view of the particle, which allowed us to obtain an initial model of the SSU processome (Fig 3.3C)

Subsequently, cryo-EM was used to determine the structure of the SSU processome (Fig. 3.3D). An initial dataset (Dataset 1) of approximately 1500 micrographs was collected, from which 70,000 particles were picked. 2D classification of the particle yielded classes similar to those obtained by negative stain EM, but some classes (Fig3.3E, middle line) likely represented new views of the particle. We also observed the presence of flexibility in the top region of the molecule in some of the classes.

After 3D classification and refinement of this first dataset, we were able to obtain a 5.1 Å reconstruction of the SSU processome (Fig. 3.4A). At that time, another cryo-EM reconstruction of an SSU processome, at a resolution of 7.3 Å, from the thermophilic eukaryote *Chaetomium Thermophlium*, had been reported (Kornprobst et al. 2016). While the structure had revealed the basic architecture of the complex, the low resolution had prevented the elucidation of the structure of the 5' ETS, and limited the functional interpretation of this assembly intermediate.

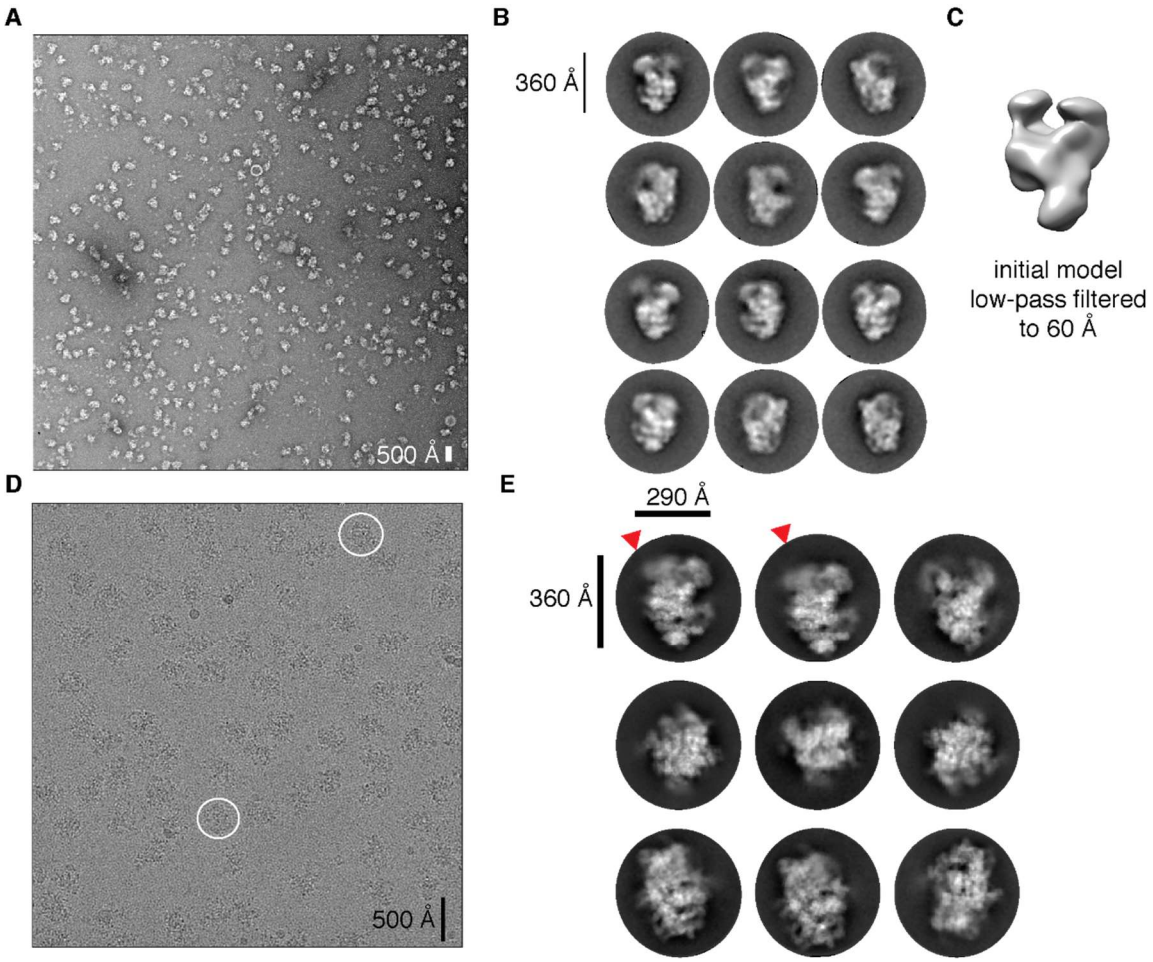


Figure 3.3 Negative stain and cryo-electron microscopy analysis of SSU processome particles. (**A**) Sample micrograph of SSU processome particles stained with 2% uranyl acetate collected at a magnification of 29,000x. (**B**) Selection of 2D class averages generated from 36,000 manually picked particles using EMAN2 (Tang et al. 2007). (**C**) Initial model generated using e2initialmodel.py (Tang et al. 2007) low-pass filtered to 60 Å. The reconstruction was generated using 11,700 selected particles. (**D**) Representative micrograph collected at a magnification of 22,500x. Two particles are circled. A scale bar is indicated at the lower right corner. (**E**) Selection of RELION-2 2D class averages. Dimensions of the small subunit processome are indicated on the left and flexible regions are highlighted with red markers.

The cryo-EM reconstruction of the SSU processome at 5.1 Å allowed us to build an architectural model of this early ribosome assembly intermediate (Chaker-Margot et al. 2017). The architecture of the particle was mostly similar to the *C. Thermophilum* structure. The subcomplex UtpA is located at the bottom of the structure covering the first half of the 5' ETS as had been observed by *in vivo* RNA-protein crosslinking (Hunziker et al. 2016). UtpB and the U3 snoRNP are located above UtpA on either side of the particle, and make extensive contact with the 5' ETS. Docking of the structural domains of the 18S allowed us to place the 5' domain and 3' domain at the top and the side of the structure respectively (Fig. 3.4). These domains are clearly separated with respect to the mature 18S rRNA, suggesting that spatial segregation allows for separate maturation within the SSU processome. Importantly, the 3' domain, which later gives rise to the head and helix 44 of the small ribosomal subunit, is largely unfolded within the SSU processome.

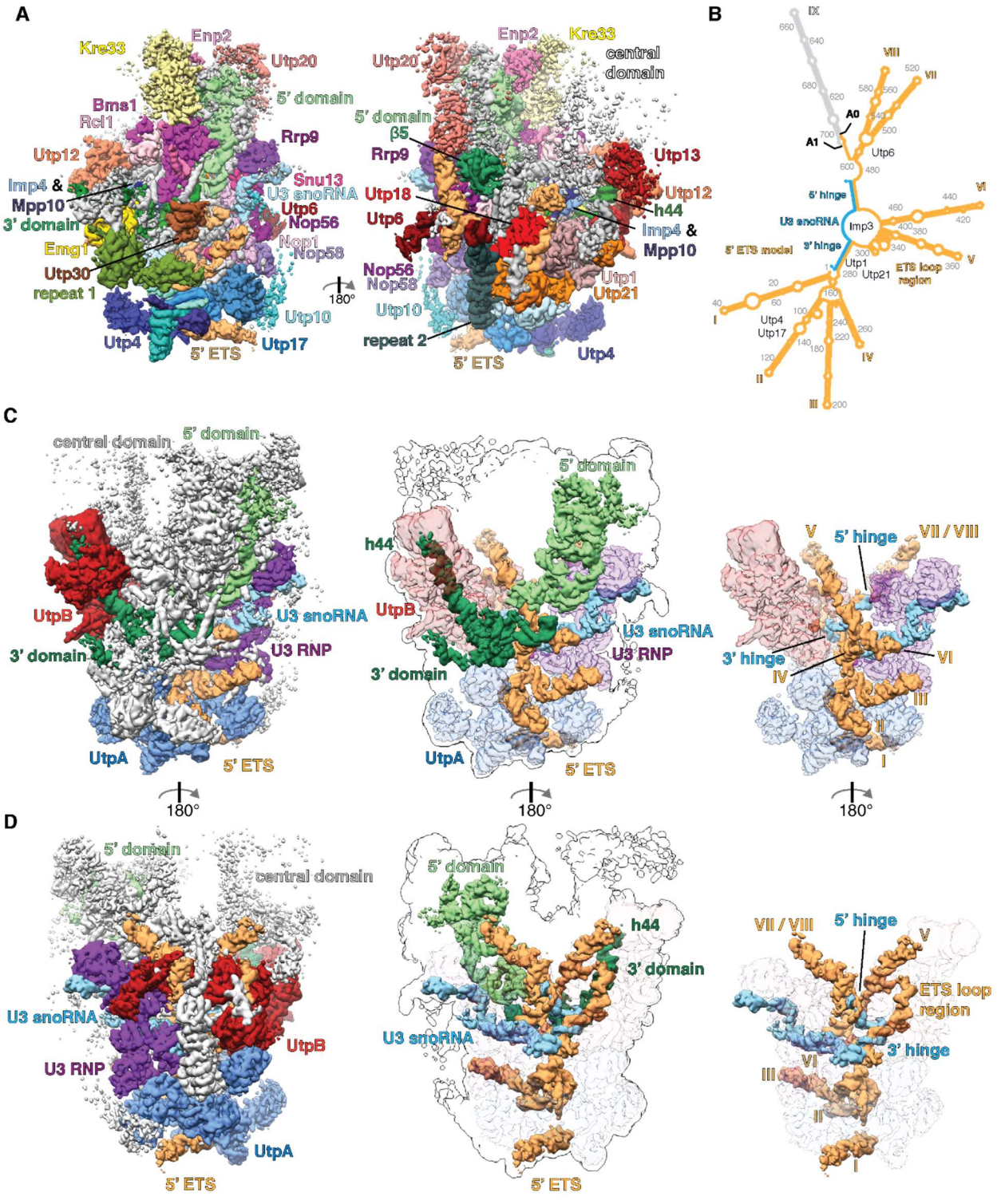
These extended conformations of RNA helices are presumably used to facilitate enzymatic access to distinct sites for subsequent covalent modifications within the RNA. The methyltransferase Emg1, is positioned to methylate base 1191 (Meyer et al. 2011) located in the still unfolded 3' major domain. By contrast, the homodimeric acetyltransferase Kre33, which is located in a solvent exposed region at the very top of the structure, is further away from its substrates at nucleotides 1280 and 1773 within the 3' domain (Sharma et al. 2015). While a network of ribosome assembly chaperones, such as UtpA and UtpB, ensures the correct folding of pre-rRNA, the separation of 18S rRNA domains during SSU processome formation allows for the later recruitment of

enzymes (Chaker-Margot et al. 2015), such as Kre33 and Emg1, which can subsequently covalently modify ribosomal RNA.

Between the 5’ and 3’ domains of the 18S rRNA, we also identified the GTPase Bms1 (Fig. 3.4). While the crystal structure of Rcl1 in complex with peptides of Bms1 (Delprato et al. 2014) had been placed previously (Kornprobst et al. 2016), the structural domains forming the bulk of the 135 kDa GTPase Bms1 had not been recognized. Bms1 is structurally related to translational GTPases as well as the ribosome assembly factor Tsr1 (McCaughan et al. 2016) and shares their classic four domain architecture (Fig. 3.16B) (Gelperin et al. 2001; Wegierski et al. 2001; Karbstein et al. 2005).

The *Saccharomyces cerevisiae* 5’ ETS base-pairs with the 3’ and 5’ hinges of U3 snoRNA, which also interact with regions of the 18S rRNA (Sharma and Tollervey 1999; Marmier-Gourrier et al. 2011; Dutca et al. 2011). However, previous secondary structure models of the 5’ ETS had been generated failing to take into account the additional restrains provided by the known base-pairing that occurs between the 5’ ETS and U3 snoRNA. In contrast, we fused interacting regions of the 5’ ETS, 18S, and U3 snoRNA into a single RNA for secondary structure prediction using LocARNA-P (Will et al. 2012). The analysis of fused 5’ ETS-18S-U3snoRNAs from yeast revealed the architecture of these RNAs (Fig. 3.4B). The secondary structure model (Fig. 3.4B) was in good agreement with the RNA density observed in the cryo-EM reconstruction (Fig. 3.4C-D). The 5’ ETS-U3 snoRNA form the structural blueprint for the spatial segregation of the ribosomal RNA domains. Importantly, cleavage at site A0 does not result in the dissociation of the 5’ ETS and its associated factors from the rest of the particle.

Figure 3.4 Architecture of the *S. Cerevisiae* SSU processome and the 5' ETS. (**A**) Front and back view of the segmented cryo-EM density map. (**B**) Secondary structure model of the 5' ETS (orange) and the 3' and 5' hinges of U3 snoRNA (light-blue) generated with LocARNA-P (22). Major helices are numbered I-IX. Helix IX, which is not observable in the EM density map, is shown in gray. A0 and A1 cleavage sites are indicated on helix IX. Proteins found proximal to the 5' ETS in the structure are labeled in black. (**C and D**) Front (C) and back (D) views of the SSU processome. Left: Segmented map of the complete particle with color-coded protein and RNA elements. Center: Outline of a 10 Å low-pass filtered map with only RNAs shown in full color. Right: Only the 5' ETS, UtpA, UtpB and U3 snoRNP are shown.


A
Kre33
Enp2
Utp20
5' domain
Bms1
Rcl1
Rrp9
Utp12
Snu13
U3 snoRNA
Imp4 &
Mpp10
Utp6
3' domain
Nop56
Nop1
Nop58
Emg1
Utp30
repeat 1
180°
Utp10
Utp4
5' ETS
Utp17
Enp2
Kre33
central
domain
Utp20
5' domain
β5
Rrp9
Utp13
Utp18
Utp12
Utp6
h44
Imp4 &
Mpp10
Nop56
Nop58
Utp1
Utp21
Utp10
repeat 2
5' ETS
Utp4
B
IX
VIII
VII
A0
A1
Utp6
5' hinge
VI
U3 snoRNA
Imp3
5' ETS model
3' hinge
Utp1
Utp21
ETS loop
region
V
Utp4
Utp17
I
IV
II
III
C
central domain
5' domain
UtpB
U3 snoRNA
3' domain
U3 RNP
UtpA
5' ETS
180°
5' domain
h44
UtpB
U3 snoRNA
U3 RNP
3' domain
UtpA
5' ETS
180°
V
5' hinge
VII / VIII
3' hinge
VI
IV
III
II
I
180°
D
5' domain
central domain
U3 snoRNA
UtpB
U3 RNP
UtpA
5' ETS
5' domain
h44
3' domain
U3 snoRNA
5' ETS
VII / VIII
5' hinge
V
ETS loop
region
3' hinge
VI
III
II
I

Lastly, we noticed that four subunits of UtpA, Utp5, Utp8, Utp9 and Utp15, as well as Utp1, Utp12, Utp13 and Utp21 of UtpB, all contained a C-terminal tetramerization domain. This feature, evident at the sequence as well as the structural level (Fig. 3.5), suggested a common evolutionary origin of the CTDs of UtpA and UtpB subunits. In contrast to UtpA, where single β-propeller containing subunits are used to form the tetramer, tandem β-propellers are used in the context of UtpB (Fig. 3.5).

While this 5.1 Å architectural model has allowed us to understand the role of the 5' ETS as the organizing platform of the SSU processome, the limited resolution has prevented the assignments of many protein elements in the structure. Several folds, such as β-propellers, helical repeats and KH-domains, could not be assigned to specific proteins, limiting the interpretation of this structure. For that reason, we set out to improve the resolution of the cryo-EM reconstruction of the SSU processome, both in the core of the particle and in the peripheral areas.

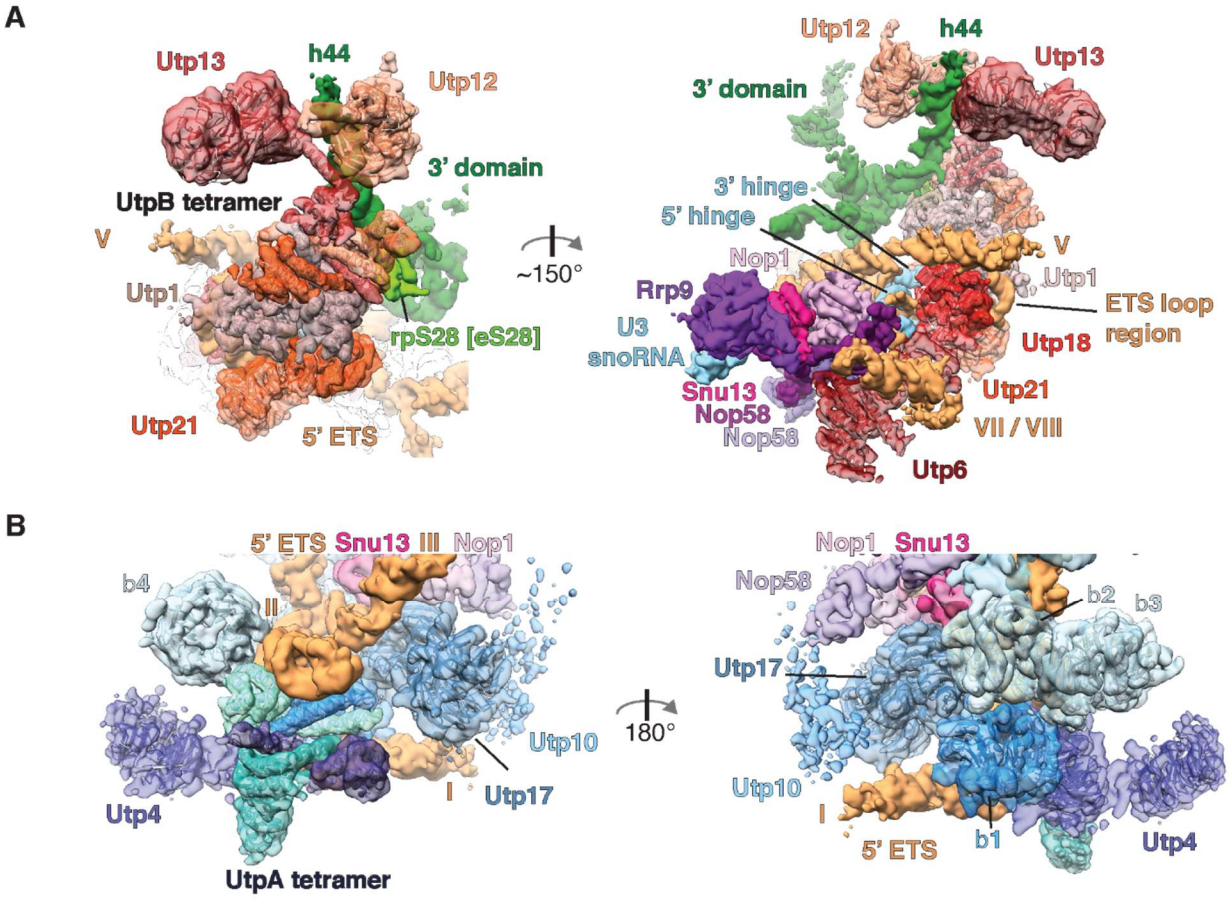


Figure 3.5 Subunits of UtpA and UtpB share an evolutionary origin. (**A**) Two views of the segmented cryo-EM density map of UtpB with its subunits colored in shades of red and color coded bound RNA elements (5' ETS: gold, 3' domain: green). U3 snoRNP proteins are depicted in shades of purple, rpS28 in light-green and the U3 snoRNA in light-blue. (**B**) As for (A) but with UtpA (in shades of blue). Assigned subunits (Utp4, Utp17, Utp10) are labeled and unassigned β-propellers are labeled with β2-4.

Additional data collection, as well as a more extensive 3D classification and refinement strategy, enabled us to improve the resolution of the entire SSU processome to 3.8 Ångstroms (Fig. 3.6 and 3.7, Table 3.1). Improvements in the resolution came from the additional number of particles but also from the use of MotionCor2 (Zheng et al. 2017) for movie alignment instead of Unblur (Brilot et al. 2012), which yielded an improvement of 0.4 A in resolution using the same set of particles (data not shown)

The core, which contains approximately 80% of the proteins present in the SSU processome, could be refined to an overall resolution of 3.6 Ångstroms. Large regions in the center of the particle are well resolved near 3 Ångstroms (Fig. 3.7) and clear density can be observed for side chains and bases (Fig. 3.8). Focused classification and refinement resulted in significantly improved maps of more peripheral areas of the structure, such as the head region, containing the 5' domain, or the 3' region, containing parts of the 3' domain (both at 4.1 Ångstroms overall resolution) (Fig. 3.6 and 3.7, Table 3.1). These improved maps also feature clear side chain density for peripheral and more exposed areas. In addition, RNA densities in different regions of the SSU processome show clear separation of nucleotides (Fig. 3.8).

As previously shown, the central domain is largely flexible under the growth and purification conditions we used to obtain SSU processomes. Extensive iterative 3D classification has enabled us to isolate one particular conformation of this domain, where density can be visualized at an overall resolution of 7.2 Ångstroms (Fig. 3.6 and 3.7).

Figure 3.6 Cryo-EM data processing strategy. 10,029 micrographs were collected in four 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 resulted in 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 improved maps for these regions.

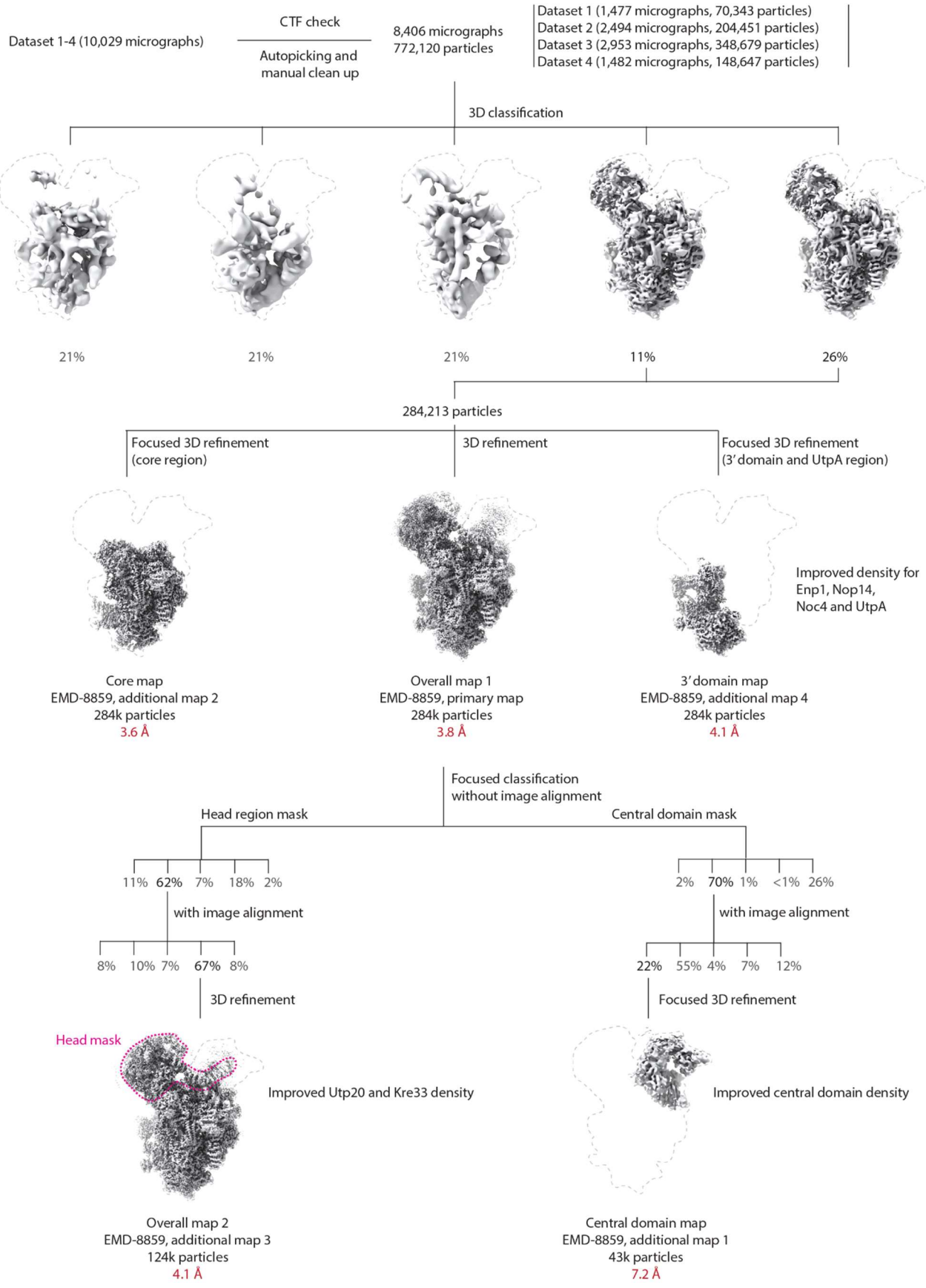

Dataset 1-4 (10,029 micrographs)
CTF check
Autopicking and manual clean up
8,406 micrographs
772,120 particles
Dataset 1 (1,477 micrographs, 70,343 particles)
Dataset 2 (2,494 micrographs, 204,451 particles)
Dataset 3 (2,953 micrographs, 348,679 particles)
Dataset 4 (1,482 micrographs, 148,647 particles)
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 3.7. 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). FSC curves calculated between the final model and map (light-blue), model2 and half map 1 (FSC test; red) or half map 2 (FSC work; yellow dashed line) are shown. 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).

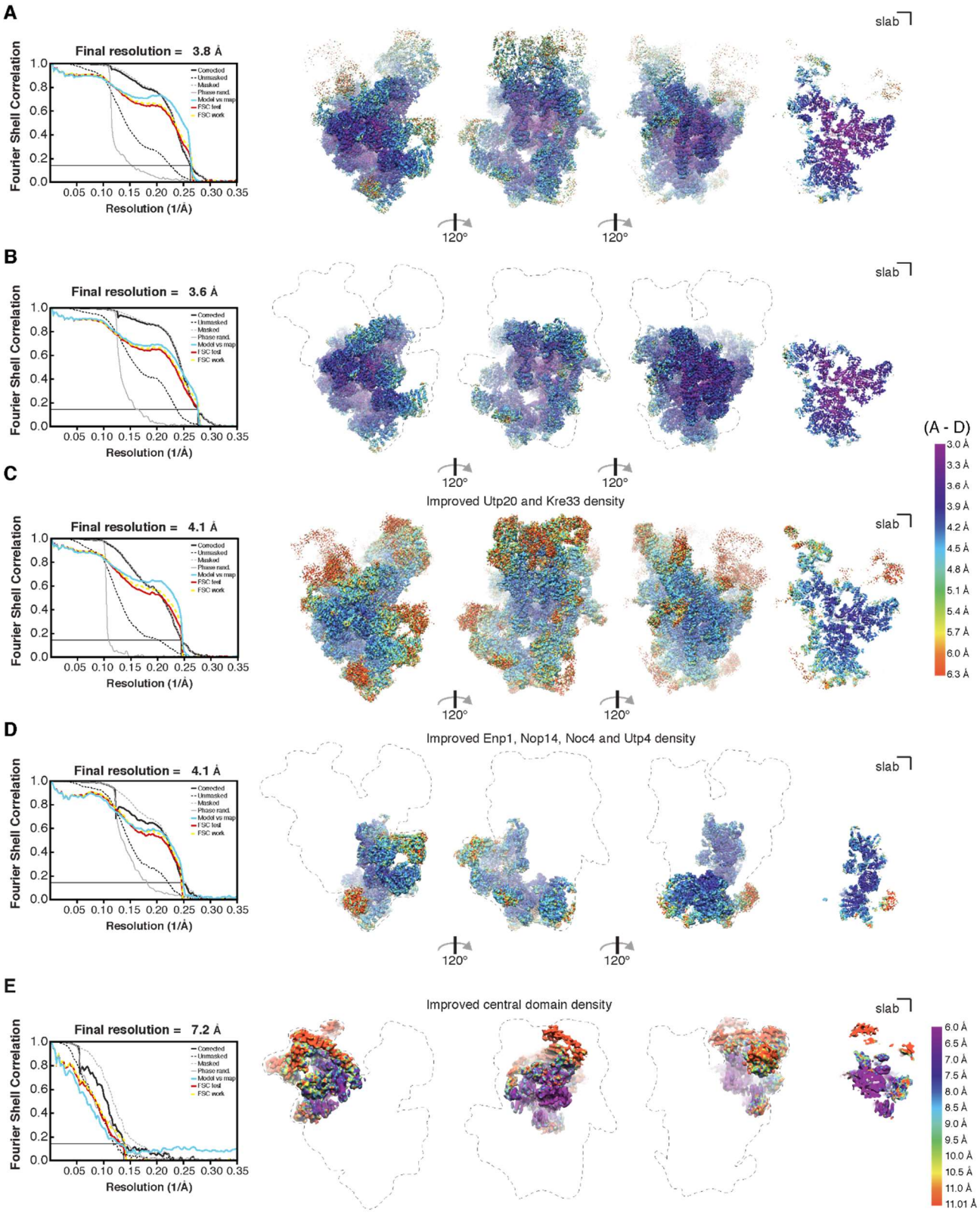

A
Final resolution = 3.8 Å
Fourier Shell Correlation
Resolution (1/Å)
Corrected
Unmasked
Masked
Phase rand.
Model vs map
FSC test
FSC work
slab
120°
120°
B
Final resolution = 3.6 Å
(A - D)
3.0 Å
3.3 Å
3.6 Å
3.9 Å
4.2 Å
4.5 Å
4.8 Å
5.1 Å
5.4 Å
5.7 Å
6.0 Å
6.3 Å
C
Final resolution = 4.1 Å
Improved Utp20 and Kre33 density
D
Final resolution = 4.1 Å
Improved Enp1, Nop14, Noc4 and Utp4 density
E
Final resolution = 7.2 Å
Improved central domain density
6.0 Å
6.5 Å
7.0 Å
7.5 Å
8.0 Å
8.5 Å
9.0 Å
9.5 Å
10.0 Å
10.5 Å
11.0 Å
11.01 Å

Figure 3.8 Representative electron density of selected protein and RNA elements.

A) Cryo-EM density and corresponding models of selected protein segments are shown. Labels below the density indicate the subunit and the corresponding residue-range visualized. While most elements show representative density from the core region of the SSU processome, Utp6 is located in the periphery and Utp15 is located towards the bottom of the particle. (**B-D**) Models of (**B**) the 5' ETS helices IV and VI, and the A1 cleavage site, (**C**) the 18S rRNA 5' and 3' domains, and helix 44, (**D**) the U3 snoRNA box A, 5' and 3' hinges and Box C'/D are depicted within their respective cryo-EM densities.

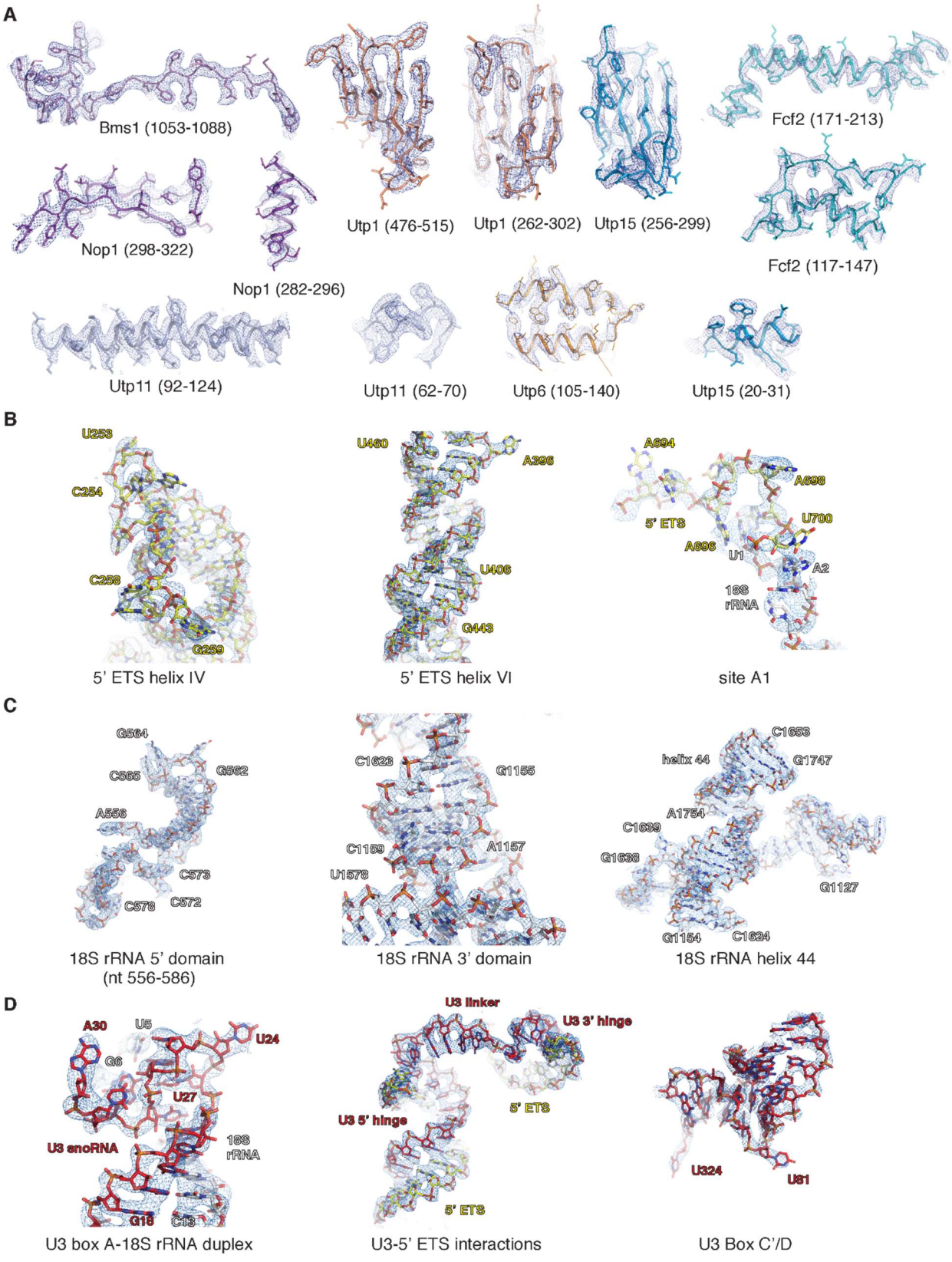

A
Bms1 (1053-1088)
Utp1 (476-515)
Utp1 (262-302)
Utp15 (256-299)
Fcf2 (171-213)
Nop1 (298-322)
Nop1 (282-296)
Fcf2 (117-147)
Utp11 (92-124)
Utp11 (62-70)
Utp6 (105-140)
Utp15 (20-31)
B
U253
C254
C258
G259
5' ETS helix IV
U460
A396
U406
G443
5' ETS helix VI
A694
A698
5' ETS
U700
A696
U1
A2
18S rRNA
site A1
C
G564
C565
G562
A556
C573
C578
C572
18S rRNA 5' domain (nt 556-586)
C1623
G1155
C1159
A1157
U1578
18S rRNA 3' domain
C1653
helix 44
G1747
A1754
C1639
G1638
G1127
G1154
C1624
18S rRNA helix 44
D
A30
U5
U24
G6
U27
U3 snoRNA
18S rRNA
G18
C13
U3 box A-18S rRNA duplex
U3 linker
U3 3' hinge
5' ETS
U3 5' hinge
5' ETS
U3-5' ETS interactions
U324
U81
U3 Box C'/D

Table 3.1: Cryo-EM data collection, refinement and validation statistic.

	Overall map 1 5WLC,8859	Core map	3' domain map	Overall map 2	Central domain map
Data collection and processing					
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	5JPQ				
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				

SSU processome were additionally analyzed by DSS cross-linking and mass spectrometry to provide additional distance restraints during model building and validation (Fig. 3.9). The combined use of the cryo-EM maps together with cross-linking and mass spectrometry data allowed us to unambiguously trace proteins not only in the core but also in the periphery of the particle to obtain a largely complete near-atomic model of the SSU processome (Fig. 3.10 and Appendix 7.3) that goes beyond previous models (Kornprobst et al. 2016; Chaker-Margot et al. 2017; Sun et al. 2017). In solvent-exposed regions, such as the central domain, where atomic resolution was not obtained, previously determined crystal structures have been fitted or poly-alanine models have been built *de novo*. We note that in these regions the sequence register of proteins is less reliable.

Analysis of crosslinks that could be mapped on the structural model show that more than 85% of crosslinks are between lysines whose alpha carbon are located less than 32 Å apart, which corresponds to the expected distance of DSS-linked lysines (Fig. 3.9C) (Shi et al. 2014). However, a subset of crosslinks was mapped to residues separated by more than 35 Å and sometimes up to more than 100 Å. Many of such crosslinks were emanating from the central domain proteins, which are located in a flexible region of the SSU processome, perhaps explaining the discrepancy between the model and the cross-linking data. A few cross-links remained unexplained, and may be the result of particle disintegration, or of a conformation which could not be observed by cryo-EM. It is also possible that these crosslinks are due to ambiguity in the model building but many of them are within proteins which form cross-links consistent with our model, suggesting they are more likely due to particle disintegration,

Figure 3.9 Overview and structural analysis of the DSS cross-linking mass spectrometry data. (**A**) 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 U-three-proteins (Utp) are labeled with their respective number. (**B**) 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 and are colored according to Figure 3.8. 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. (**C**) Histogram of all Cα cross-link distances in Å. 87.2% of all cross-link distances are within 32 Å. (**D**) 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.

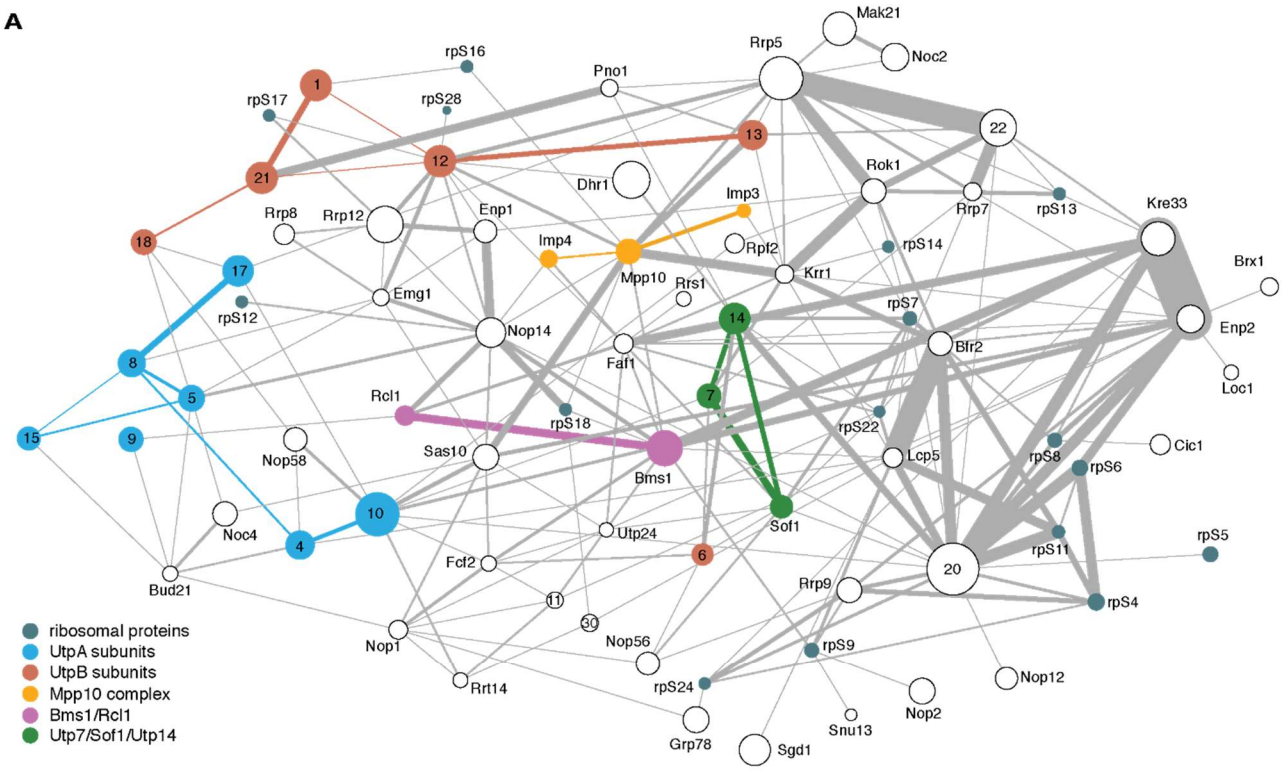
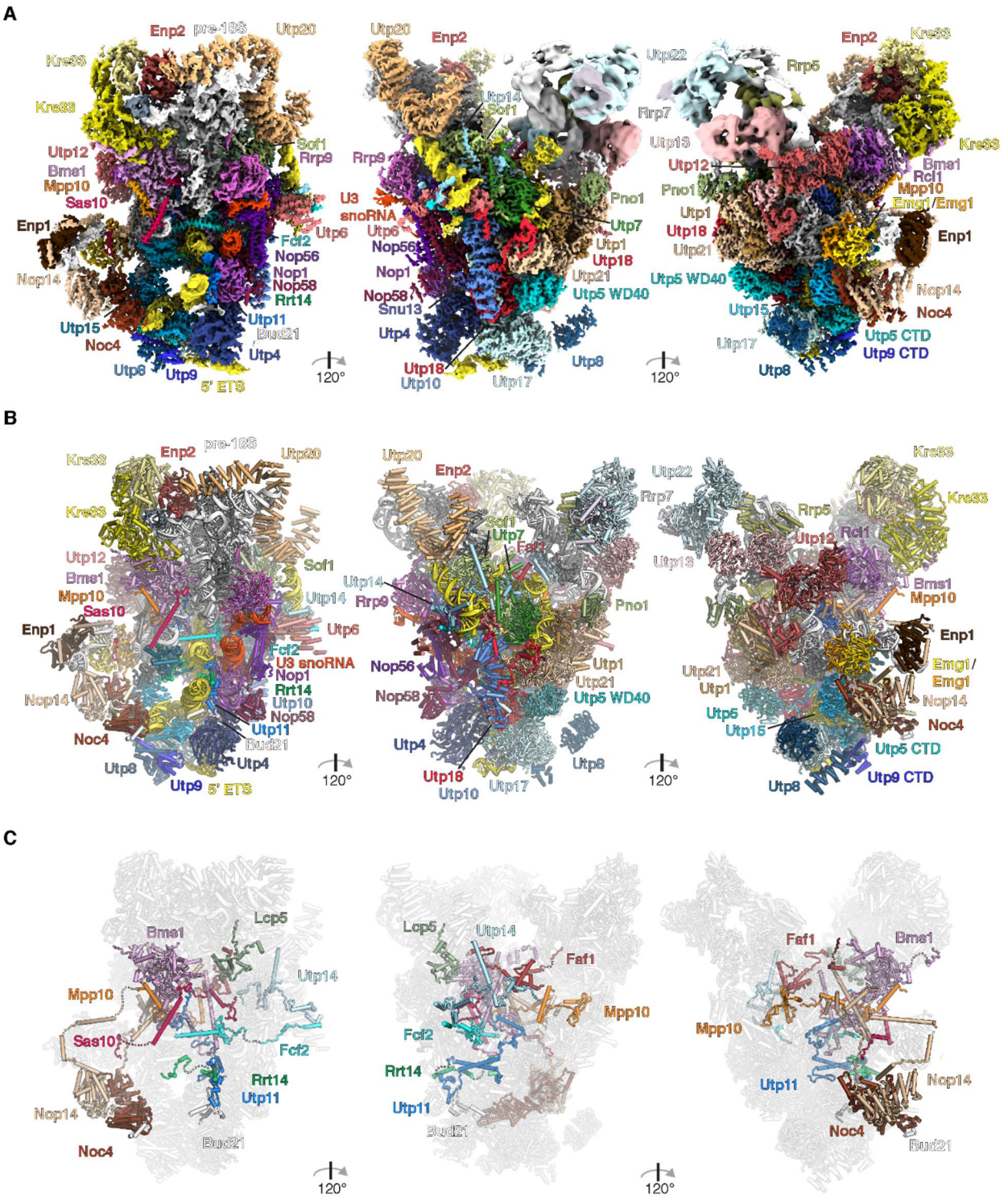

A
Mak21
Noc2
Rrp5
rpS16
Pno1
rpS17
rpS28
1
13
22
21
12
Dhr1
Rok1
Imp3
Rrp7
rpS13
Kre33
Rrp8
Rrp12
Enp1
18
Imp4
Rpf2
rpS14
Brx1
17
Mpp10
Krr1
Rrs1
Emg1
rpS12
rpS7
14
Enp2
Nop14
Bfr2
8
Faf1
Loc1
5
7
Rcl1
rpS18
rpS22
15
9
Nop58
Sas10
Lcp5
rpS8
Cic1
Bms1
rpS6
10
Sof1
Noc4
Utp24
rpS11
rpS5
4
6
Fcf2
20
Bud21
Rrp9
11
rpS4
30
Nop1
Nop56
rpS9
Nop12
Rrt14
rpS24
Nop2
Snu13
Grp78
Sgd1
ribosomal proteins
UtpA subunits
UtpB subunits
Mpp10 complex
Bms1/Rcl1
Utp7/Sof1/Utp14
B
120°
120°
ambiguous, shortest crosslink displayed
unambiguous
C
Number of crosslinks
0
10
20
30
40
50
60
70
80
8
16
24
32
40
48
56
64
72
80 more
Distance CaCa / Å
D
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

Figure. 3.10 Cryo-EM reconstruction and near-complete atomic model of the *S. cerevisiae* SSU processome. (**A**) Three views of a composite cryo-EM map consisting of the 3.6 Å core, the 4.1 Å head-focused, the 4.1 Å 3' domain and the 7.2 Å central domain maps. Densities for SSU processome components are color-coded with analogous labels. Subunits of complexes are shown in shades of blue (UtpA), red (UtpB), purple (U3 snoRNP), brown (Nop14-Noc4) and light-pink (Bms1-Rcl1). Ribosomal proteins are depicted in shades of grey. RNAs are colored in yellow (5' ETS), red (U3 snoRNA) and white (pre-18S rRNA). (**B**) Cartoon representation of the atomic model of the SSU processome with orientations and colors as in (A). (**C**) Cartoon representation of the atomic model of extended assembly factors, colored as in (A)


A
Enp2
pre-18S
Utp20
Kre33
Sof1
Rrp9
Utp12
Bms1
Mpp10
Sas10
Enp1
Fcf2
Utp6
Nop56
Nop1
Nop58
Rrt14
Nop14
Utp15
Noc4
Utp11
Bud21
Utp4
Utp8
Utp9
5' ETS
120°
Utp14
U3 snoRNA
Pno1
Utp7
Utp1
Utp18
Utp21
Utp5 WD40
Snu13
Utp10
Utp17
Rrp5
Utp22
Rrp7
Utp13
Rcl1
Emg1/Emg1
Utp5 CTD
Utp9 CTD
B
Lcp5
Faf1
C

The updated model of the SSU processome details the structure of the previously identified elements, the 5’ ETS, UtpA, UtpB and the U3 snoRNP (Fig. 3.11, Fig 3.13). With higher resolution data and cross-linking restrains, we have also identified and built models of 10 additional ribosome assembly factors within the SSU processome (Fig. 3.10C). These include the 5’ ETS particle proteins Utp11, Fcf2, Sas10 and Bud21 as well as later factors such as Faf1, Lcp5, Utp14 and Rrt14 (Chaker-Margot et al. 2015; Zhang et al. 2016). The Nop14-Noc4 sub-module (Kühn et al. 2009) was identified as a previously unassigned helical structure in the lower half of the particle. In addition, parts of Mpp10 that extend beyond the regions interacting with Imp3 and Imp4 (Lee and Baserga 1999; Zheng and Ye 2014) have now been identified.

Helical repeat elements are used towards the top of the structure near the 5’ and central domains to encapsulate and stabilize RNA helices in a particular conformation. In the context of the 5’ domain, Utp20 provides a structural support for RNA expansion segments ES3A and ES3B near the Kre33 heterodimer and Enp2 (3.12). Similarly, 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), is positioned in proximity to the UtpC complex (Lin et al. 2013), Krr1 (Zheng et al. 2014) and the central domain (Fig. 3.10). It provides a cradle to stabilize helix 24 in a different conformation with respect to the mature small subunit.

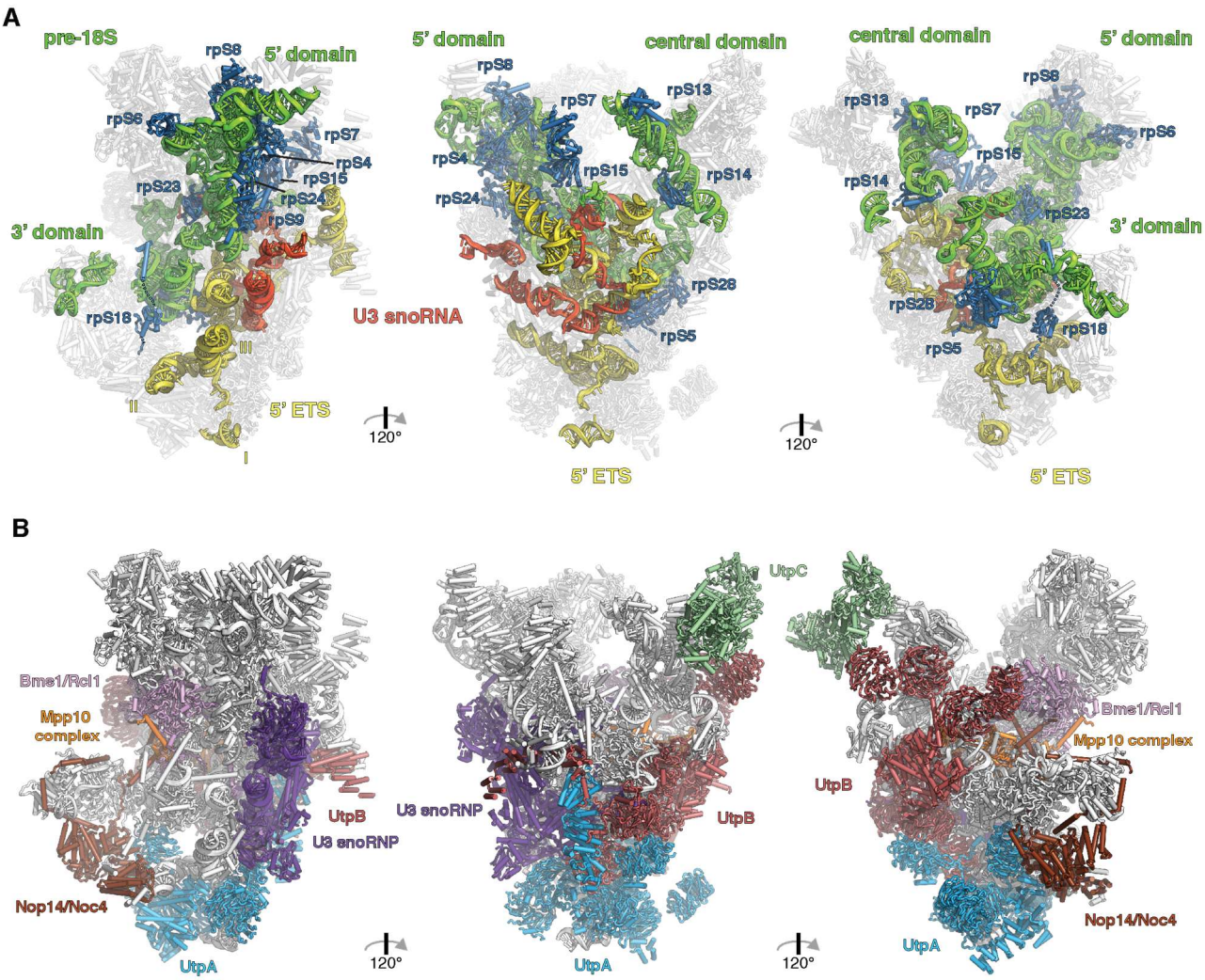


Figure 3.11. Structural elements of the SSU processome. (**A**) The SSU processome as transparent outline with superimposed RNAs (pre-18S in green, U3 snoRNA in red and 5' ETS in yellow) and ribosomal proteins in blue. (**B**) Same orientations as in (A) with complexes UtpA (blue), UtpB (red), U3 snoRNP (dark-purple), Bms1/Rcl1 (violet), the Mpp10 complex (orange) and UtpC (pale-green) highlighted.

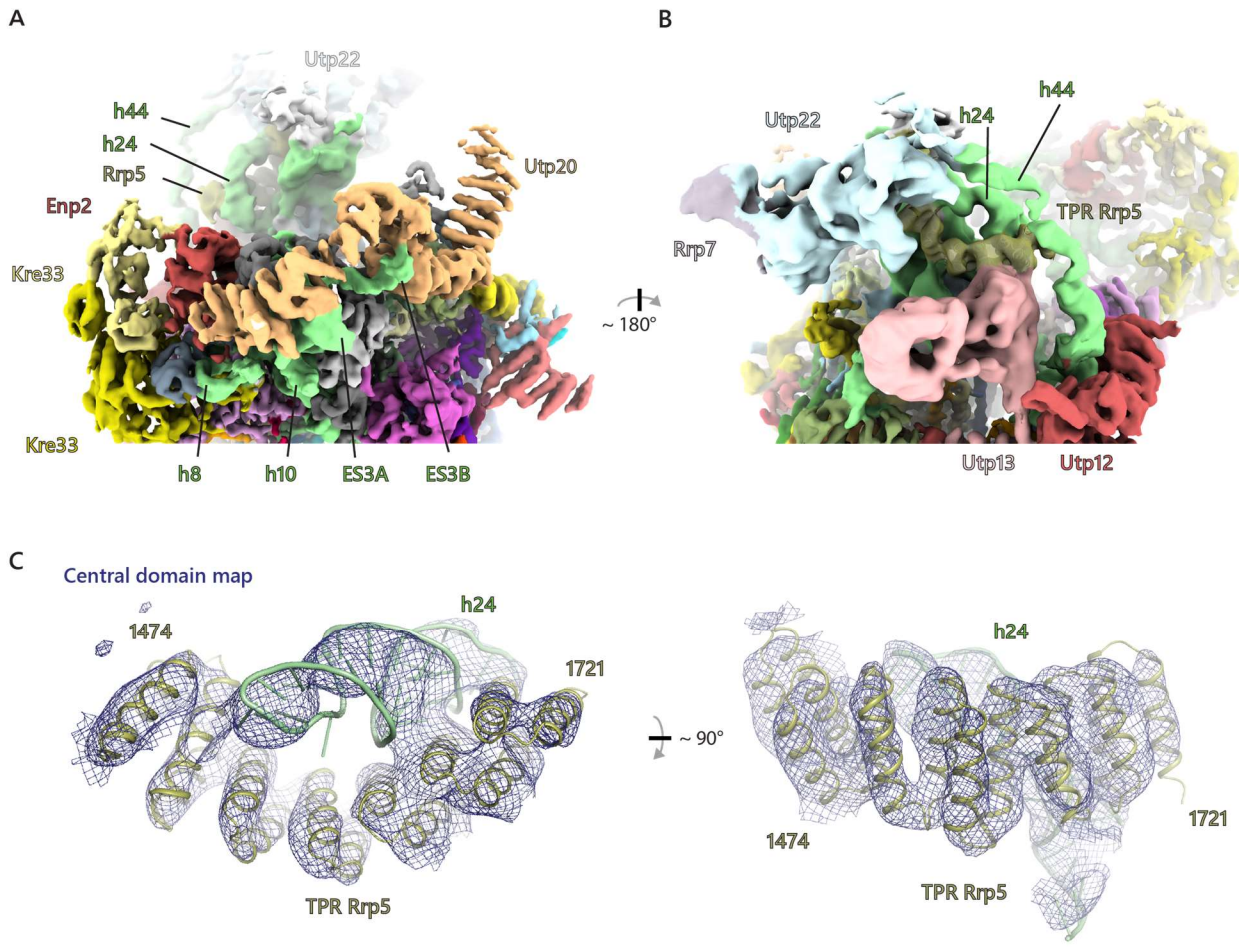
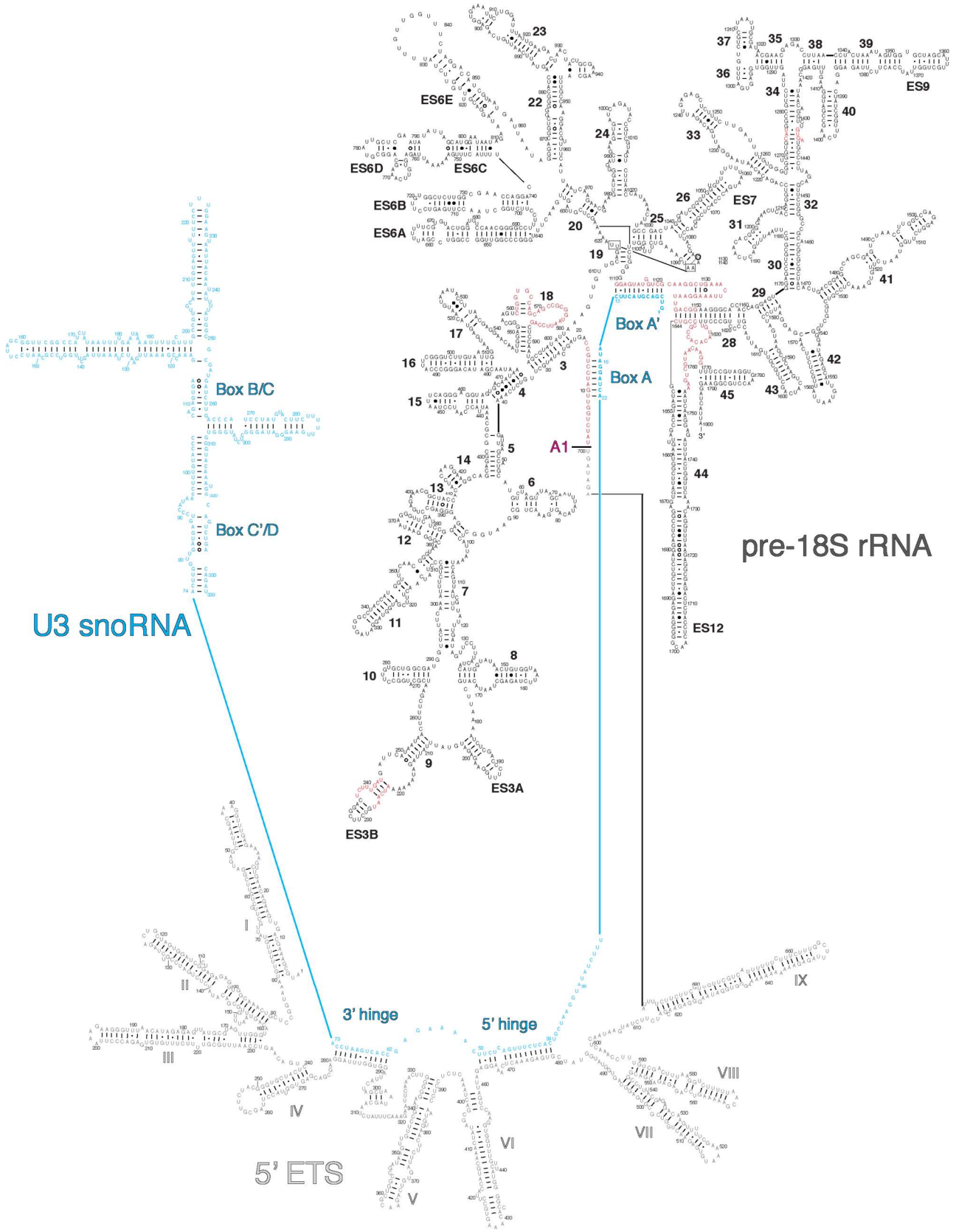


Figure 3.12. The helical repeat proteins Utp20 and Rrp5 chaperone rRNA in the 5'- and central domain. (**A** and **B**) Two views of a composite cryo-EM density map consisting of the 6 Å low-pass filtered overall map 2 and the 7.2 Å central domain map. The density is colored as in (Fig. 3.10) but with 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 (**B**) 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).

Figure 3.13. Secondary structure diagram of RNAs in the SSU processome. Individual nucleotides of RNAs are indicated with their base pairing interactions. 5' ETS (grey), 18S (black) and U3 snoRNA (light-blue) are shown. Regions of the 18S rRNA that have been remodeled in the SSU processome are highlighted in red.


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

Coordination of the 5’ ETS

The improved resolution of the SSU processome has allowed us to re-analyze the architecture of UtpA and UtpB, and their role in stabilizing the 5’ ETS (Fig. 3.11). UtpA forms the base of the SSU processome, where it acts as central scaffold that recognizes the first three helices (helix I-III) of the 5’ ETS (Fig. 3.14). Helix I is bound by a set of loops and helical elements on top of the tandem β-propeller of Utp17 (Fig. 3.14C). 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 C-terminal domain (CTD) and WD40 domain to position helix II and to stabilize the junction between helices II and III (Fig. 3.14A, E). The WD40 domain of Utp15 and helix II further provide a binding platform for Noc4, which acts as the foundation of a lateral extension of the UtpA complex where the 3’ domain is placed (Fig. 3.14E, F). This extension is additionally stabilized by Bud21, which connects Noc4 with Utp4 (UtpA), Nop1 (U3 snoRNP) and helix III that rests on top of Utp4 (Fig. 3.14).

A short single stranded RNA region between helices II and III of the 5’ ETS, is coordinated by two other β-propeller-containing subunits located next to Utp4 and Utp17. We could unambiguously assign these WD40 domains as Utp5 (UtpA) and Utp18 (UtpB) (Fig. 3.14F).

Utp5 is integrated within UtpA through its CTD and a linker peptide, which runs along a conserved groove of Utp17 (Fig. 3.14D). Akin to this interaction, a C-terminal peptide expansion of Utp17 contacts the β-propeller of Utp5 before binding to Utp10, the only subunit of UtpA composed solely of helical repeats. Utp5, Utp10 (UtpA), Utp21 and Utp18 (UtpB) form the junction between the UtpA and UtpB complexes. Within this

junction Utp18 serves as a central nexus. The placement of the WD40 domain of Utp18 between three UtpA subunits (Utp4, Utp17 and Utp5) and near two UtpA linker regions (Utp5 and Utp15) interlocks the two largest subcomplexes of the SSU processome (Fig. 3.14 and 3.15).

Like Utp17, Utp18 employs extensive peptide-like motifs to facilitate protein-protein interactions (Fig. 3.15 and 3.16A). Three regions within the 230-residue N-terminus of Utp18 interact with the UtpB subunits Utp6, Utp21, the U3 snoRNP component Nop58 and Utp10 (UtpA) (Fig. 3.14A, B, 3.15A-D, 3.17A). 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 of Nop58 (Fig. 3.15C, 3.16A). Additionally, it features an Mtr4 arch interacting motif (AIM) (Thoms et al. 2015) which is located in a disordered region between the first and second N-terminal segment of Utp18 (Fig. 3.15D). The WD40 domain of Utp18 interacts with Snu13 (U3 snoRNP) and stabilizes a single-stranded region of the 5’ ETS immediately upstream of the 3’ hinge (nucleotides 275-280). Downstream of the 3’ hinge duplex (nucleotides 293-332), the 5’ ETS is mostly single stranded with a short stem loop (nucleotides 299-308) that is stabilized by Utp21 and Utp1 on one side and Utp18 and Utp7-Sof1-Utp14 on the other side (Fig. 3.15D, E. The following single stranded region between helices V and VI of the 5’ ETS (nucleotides 393-396) is bound by Imp3, which interacts with an N-terminal helix (residues 2-34) of Imp4. Together with the UtpB tetramer, Imp3 and Imp4 provide a binding surface for Mpp10.

Figure 3.14 Architecture of the UtpA complex and its interactions with the 5' ETS and Utp18. UtpA subunits are colored in different shades of blue as in (Fig. 3.10) The 5' ETS is shown in yellow with its helices labeled with roman numbers and the UtpB subunit Utp18 is depicted in red. All elements are shown in cartoon representation. Helical C-terminal domains (CTD), WD40-domains and N- and C-termini (NT, CT) of all subunits are labeled if applicable. (**A, B**) Two views of all subunits of UtpA and Utp18 bound to the first three helices of the 5' ETS. (**C**) Loops of Utp17 (light and dark pink) contact helix I of the 5' ETS and the CTD tetramer. Utp17 also interacts with Utp4 on its side surface. (**D**) A linker (light-pink) between the WD40-domain and the CTD of Utp5 runs along Utp17, forming a β-strand (dark-pink). (**E**) The N-terminal extension (dark-pink) and a linker of Utp15 (light-pink) place its WD40 domain and the CTD on opposite sides of helix II. (**F**) Helix III is coordinated by Utp4 and Bud21. The junction between helix II and III is stabilized by a linker of Utp15 and the WD40 domain of Utp18.

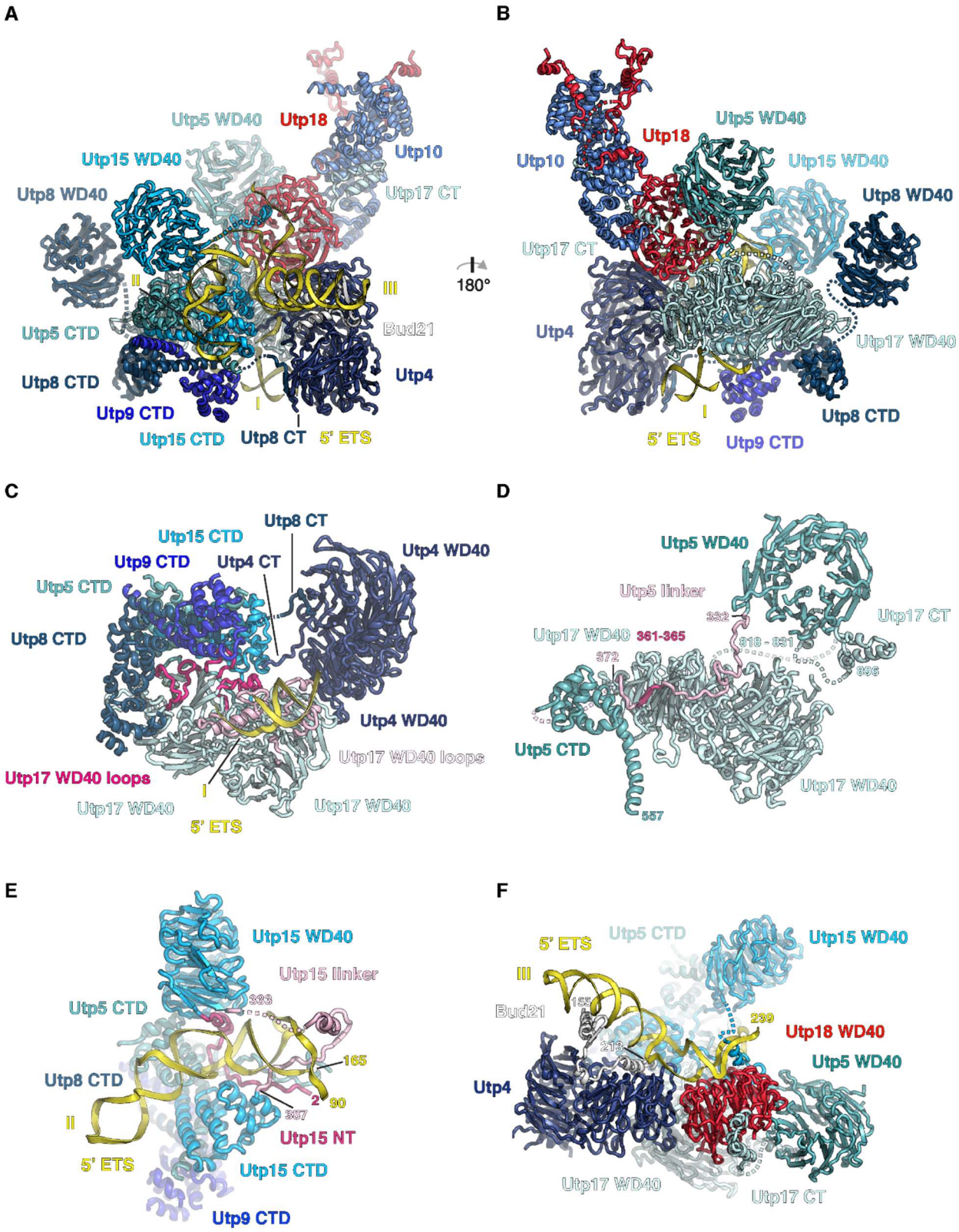
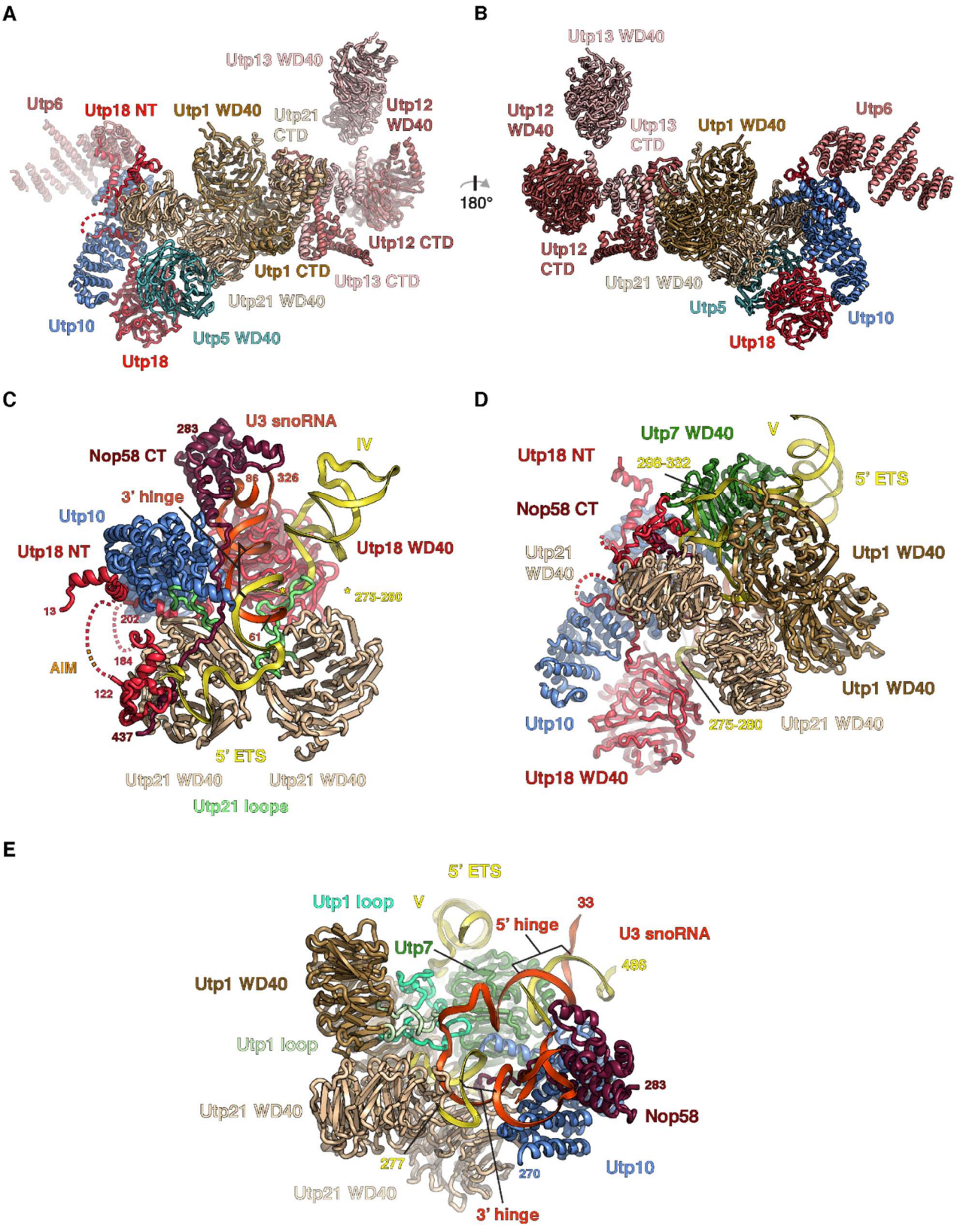

A
Utp5 WD40
Utp18
Utp10
Utp15 WD40
Utp8 WD40
Utp17 CT
II
III
180°
Utp5 CTD
Bud21
Utp8 CTD
Utp4
Utp9 CTD
I
Utp15 CTD
Utp8 CT
5' ETS
B
Utp5 WD40
Utp18
Utp10
Utp15 WD40
Utp8 WD40
Utp17 CT
Utp4
Utp17 WD40
Utp8 CTD
I
5' ETS
Utp9 CTD
C
Utp8 CT
Utp15 CTD
Utp4 WD40
Utp9 CTD
Utp4 CT
Utp5 CTD
Utp8 CTD
Utp4 WD40
Utp17 WD40 loops
Utp17 WD40 loops
I
Utp17 WD40
Utp17 WD40
5' ETS
D
Utp5 WD40
Utp5 linker
332
Utp17 CT
Utp17 WD40 361-365
372
818 - 831
896
Utp5 CTD
Utp17 WD40
557
E
Utp15 WD40
Utp15 linker
Utp5 CTD
333
165
Utp8 CTD
2
90
367
II
Utp15 NT
5' ETS
Utp15 CTD
Utp9 CTD
F
5' ETS
Utp5 CTD
Utp15 WD40
III
Bud21
155
239
Utp18 WD40
218
Utp5 WD40
Utp4
Utp17 WD40
Utp17 CT

Figure 3.15 Architecture of the UtpB complex and its interactions with RNA, the UtpA complex, and Utp7. UtpB subunits are colored in different shades of red and brown. The 5' ETS is shown in yellow with its helices labeled with roman numbers and the U3 snoRNA is shown in red with important sequence elements (3' hinge, 5' hinge) indicated. The UtpA subunits Utp10 and Utp5 are depicted in blue and Utp7 and Nop58 are illustrated in dark-green and burgundy respectively. All elements are shown in cartoon representation. Helical C-terminal domains (CTD), WD40-domains and N- and C-termini (NT, CT) of all subunits are labeled if applicable. (**A**, **B**) Two views of all subunits of UtpB, the Utp5 WD40 domain and Utp10. (**C**) A single stranded region of the 5' ETS (275-280) leading into the 3' hinge duplex is stabilized by Utp18 and a long loop (pale-green) of Utp21. A second loop (pale-green) of Utp21 binds Utp10. The exosome recruiting arch interacting motif (AIM) present in a linker of the N-terminal region of Utp18 is depicted in orange. (**D**) A predominantly single stranded region of the 5' ETS (298-332) is chaperoned by the N-terminal linker of Utp18 and the WD40 domains of Utp21, Utp1 and Utp7. (**E**) Two loops of Utp1 separate the 5' ETS and the U3 snoRNA between the 3' – and 5' hinges.


A
Utp13 WD40
Utp6
Utp18 NT
Utp1 WD40
Utp21 CTD
Utp12 WD40
180°
Utp12 CTD
Utp1 CTD
Utp13 CTD
Utp21 WD40
Utp10
Utp5 WD40
Utp18
B
Utp13 WD40
Utp12 WD40
Utp13 CTD
Utp1 WD40
Utp6
Utp12 CTD
Utp21 WD40
Utp5
Utp10
Utp18
C
U3 snoRNA
283
Nop58 CT
IV
86
326
3' hinge
Utp10
Utp18 NT
Utp18 WD40
275-280
13
202
61
AIM
184
122
437
5' ETS
Utp21 WD40
Utp21 WD40
Utp21 loops
D
Utp7 WD40
V
Utp18 NT
296-332
5' ETS
Nop58 CT
Utp21 WD40
Utp1 WD40
Utp1 WD40
Utp10
275-280
Utp21 WD40
Utp18 WD40
E
5' ETS
Utp1 loop
V
33
5' hinge
U3 snoRNA
Utp7
486
Utp1 WD40
Utp1 loop
283
Utp21 WD40
Nop58
277
270
Utp10
Utp21 WD40
3' hinge

Roles of extended proteins in the SSU processome

Many of the newly identified proteins share striking structural properties as they all contain long linkers that are used to weave through the structure and connect distant parts within the SSU processome (Fig. 3.10C, 3.16). At the base of the structure, Bud21 acts as a connector between Utp4 and Noc4 while a conserved C-terminal proline-glycine rich sequence of Rrt14 stabilizes the interaction of the L1-domain containing Utp30 with helix IV of the 5' ETS (Fig. 3.16B, F). In addition, Rrt14 contacts Utp11 located on the other side of helix IV where Utp11 interacts with Nop1 and Bud21 and connects this region with the core of the SSU processome (Fig. 3.16C). Sas10, Utp11 and Fcf2 each span at least 100 Ångstroms and are used to interconnect the U3 snoRNP with other important regions of the SSU processome (Fig. 3.16A-C). Strikingly, Nop1 (fibrillarin) is used as a binding platform for 5 proteins (Fcf2, Sas10, Utp24, Utp11 and Bud21) (Fig. 3.16G). The surfaces of the two Nop1 subunits - one located at the lower and the other at the upper end of U3 snoRNA - are used by these proteins. Here peptide backbone elements are used to form shared secondary structure elements within a β-barrel (Fcf2) or an extended β-sheet within Nop1 (Utp11, Bud21). Peptides from Sas10 and Utp24 interact similarly with Nop1 (Fig. 3.16G).

Utp14, Faf1, Lcp5 and Mpp10 are located near the top of the core of the SSU processome (Fig. 3.16). Here, Utp14 forms a highly unusual split structure in a functionally important region near the Utp7-Sof1 dimer and the A1 cleavage site (Fig. 3.16A, D). 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. Several hundred residues of Utp14 connect these two fragments. Close to the C-terminal segment, this sequence

contains the binding site for Dhr1, the essential DEAH box helicase that is responsible for displacing U3 snoRNA from early ribosome assembly intermediates (Zhu et al. 2016; Sardana et al. 2015). Faf1 is positioned in proximity to Utp14 and Utp7 and directly interacts with U3 snoRNA and pre-18S rRNA near the A1 cleavage site (Fig. 3.10B, C). Lcp5 interacts extensively with rRNA of the 5' domain and rpS9 whereas Mpp10 extends from Bms1 via Imp4 and Imp3 to the UtpB complex and helix 44.

In contrast to most of the other newly identified proteins, which have no visible globular domains, Nop14 contains a core helical repeat in addition to its long linker peptides (Fig. 3.10C). By directly binding to a second repeat protein, Noc4, the core domain of Nop14 is further expanded into a larger repeat structure. This composite helical repeat serves as a lateral structural extension of the UtpA complex and provides a scaffold for Enp1, Emg1 and parts of the 3' domain of the 18S rRNA (Fig. 3.16B), which later forms the beak structure in the mature small subunit. The repeat of Nop14 is flanked by N- and C-terminal extensions. A 75 amino-acid long C-terminal helix docks Nop14 into an opening between the Mpp10-Imp4 dimer and the Bms1-Rcl1 complex31 and points its C-terminal end towards the central cavity between the central and the 5' domain. The N-terminal extension of Nop14 binds Enp1, which caps the 3' domain of the 18S rRNA. Surprisingly, the N-terminal segment of Nop14 loops around this entire region and extends back to where the C-terminal long helix is located. Similar to Bms1, terminal extensions are used to fully integrate the Nop14-Noc4 complex within the center of the SSU processome (Fig. 3.16B). In addition to stabilizing Enp1 on top of the rRNA, this arrangement also positions the rRNA substrate in the active site of one of the Emg1 dimers. Peptides from Nop14 and Sas10 are used to provide structural support for the

dimeric Emg1 while also blocking the active site of one of its subunits. This enforces a structural asymmetry of the two Emg1 methyltransferase subunits so that only one active site is available for the methylation of base 1191 of the 3' domain (Fig. 3.16E).

Near the base of helix 41 of the 18S rRNA, we have identified rpS18, a ribosomal protein that is already positioned in a near-mature configuration with respect to the 18S rRNA (Fig. 3.16B). The C-terminus, which binds elements of the beak structure in the mature 40S subunit, adopts a different conformation in the context of the SSU processome, where this structure has not yet formed. Here, the C-terminus of rpS18 is stabilized by the long C-terminal helix of Nop14 as well as domain IV of Bms1 and a linker region of Mpp10 (Fig. 3.16B).

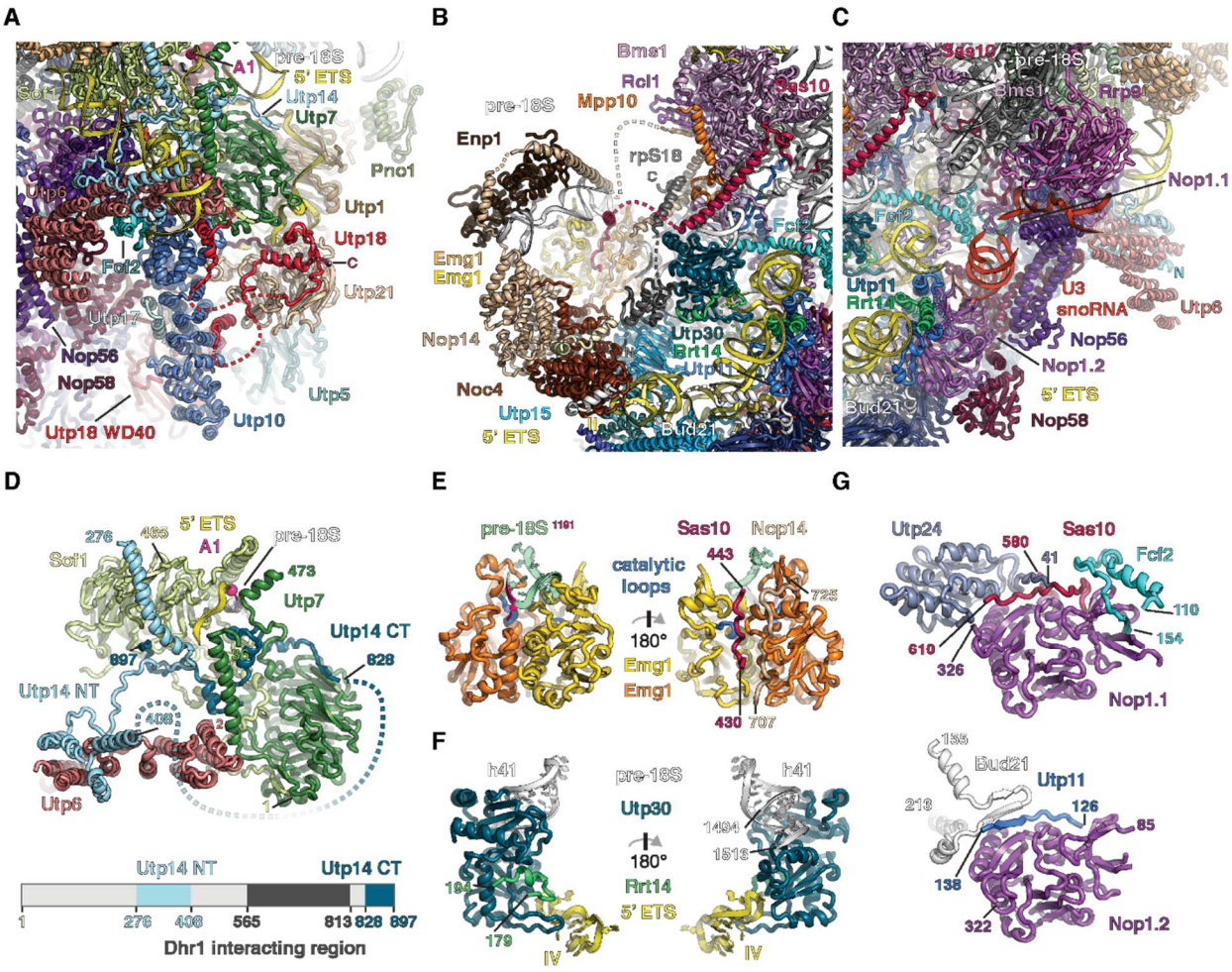


Figure 3.16. Diverse roles of peptides in the SSU processome. (A-C) Views of the network of peptides traversing through the SSU processome. (**A**) Interactions of Utp18 (UtpB) with Utp10 (UtpA), Nop58 (U3 snoRNP), Utp21 and Utp6 (UtpB). Sof1, Utp7 and Utp14 organize the A1 cleavage site (pink sphere). (**B**) Nop14, Noc4 and Enp1 cover the 3' domain of the pre-18S RNA and the methyltransferase Emg1. (**C**) Fcf2, Utp11, Bud21 and Sas10 bind to the U3 snoRNP (shades of purple) and extend through the SSU processome. (**D**) Accommodation of the A1 cleavage site (pink) by Utp7, Sof1, Utp14, and Utp6. N- and C-terminal parts of Utp14 are colored light-blue and dark-blue respectively. A schematic representation of modeled parts of Utp14 (shades of blue) is shown below. (**E**) Views of the Emg1 homodimer (orange, yellow) with catalytic loops (blue). Substrate pre-18S RNA (green) with target nucleotide (1191, pink) located in one active site while peptides of Sas10 (pink) and Nop14 (beige) occupy the other. (**F**) Utp30 and Rrt14 recognize the pre-18S and 5' ETS RNA. (**G**) Cartoon representation of the two copies of Nop1 (Nop1.1, Nop1.2) and interacting segments of Utp24, Sas10, Fcf2, Bud21 and Utp11.

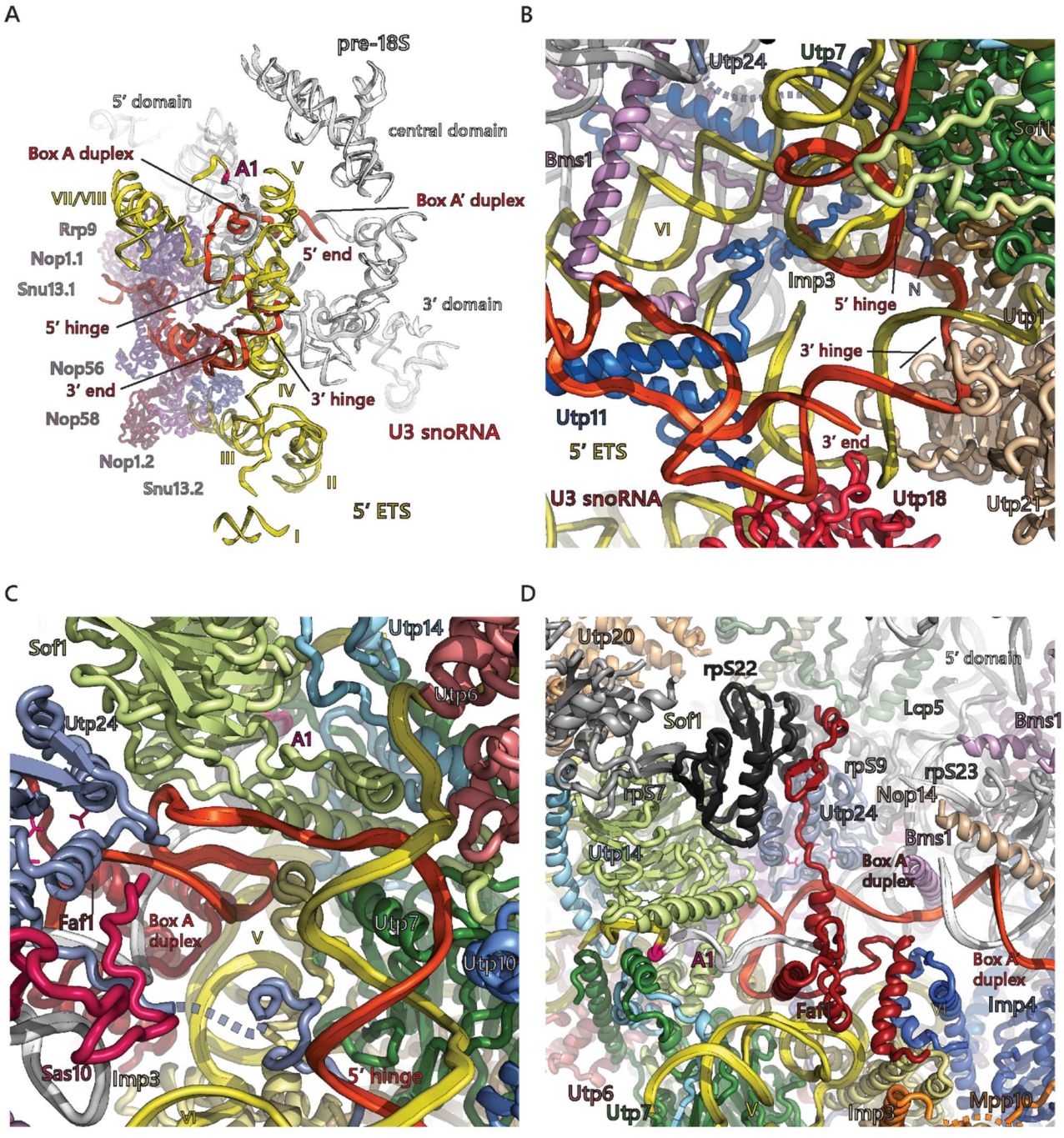
Protein-assisted RNA remodeling by U3 snoRNA

U3 snoRNA occupies a central position within the SSU processome and reaches from the outside into the core of the particle (Fig. 3.17A). By base pairing with its 5' and 3' hinges to nucleotides within the 5' ETS, it rigidifies the structural scaffold provided by the 5' ETS, which has been described previously. The 5' end of U3 snoRNA reaches further into the center of the SSU processome and base pairs with two regions of the pre-18S rRNA. A similar base pairing pattern was recently proposed for the SSU processome captured in a state upon Mtr4 depletion14. While Box A (U3 nucleotides 16-22) is base paired with nucleotides 9-15 of the pre-18S rRNA and organizes the pre-18S 5' end near the A1 cleavage site, Box A' (U3 nucleotides 3-13) is base paired with nucleotides 1111-1122 and re-organizes this region that is later in proximity to the central pseudoknot (Fig. 3.13 and 3.17A). A range of ribosome assembly factors is responsible for the stabilization of the four RNA duplexes that U3 snoRNA forms with the 5' ETS and the 18S precursor. Towards the outside of the particle, three proteins of the UtpB complex (Utp18, Utp21 and Utp1) are stabilizing the 3' hinge. While Utp18 is involved in rigidifying the junctions between helices II, III and IV of the 5' ETS, Utp21 and Utp10 (UtpA) form a clamp around the 3' hinge (Fig. 3.13E). Utp1 is used to bind the single stranded regions of both U3 snoRNA and 5' ETS with two long structured loops (residues 556-580 and 616-680). These loops act as a rudder, thereby separating the 5' ETS and the U3 snoRNA (Fig. 3.15E).

A short loop of U3 snoRNA, and the 5' hinge are coordinated by Imp3, Utp11, Bms1 and the N-terminus of Utp24 (Fig. 3.17B). Upstream of the 5' hinge, U3 snoRNA is stabilized by the Sof1-Utp7-Utp14 complex, which also provides a composite binding

site for a single stranded RNA that contains the A1 cleavage site (Fig. 3.16D, 3.17C, D). The Box A and Box A' duplexes are organized by the long C-terminal helices of Nop14 and Bms1 as well as Utp24, Faf1 and rpS23. Importantly, the captured state of the SSU processome contains an A0-cut precursor in a pre-A1 cleavage state where Utp24 is positioned close to its substrate but cleavage has not yet occurred. Faf1 is positioned between helix V of the 5' ETS and the Box A duplex near Imp3 and Imp4. A linker of Faf1 locks the Box A duplex in place and interacts with rpS22 on the opposite side, thereby occluding access to the active site of the nuclease Utp24 (Fig. 3.17D).

Figure 3.17. U3 snoRNA-mediated RNA remodeling. (**A**) Overview of the U3 snoRNP proteins (purple), U3 snoRNA (red) and their interactions with the 5' ETS (yellow) and pre-18S RNA (white). (**B**) View of protein elements stabilizing the 3' and 5' hinges. (**C**) View of proteins organizing the 5' hinge and the Box A duplex. (**D**) Overview of the Box A and Box A' duplexes of U3 snoRNA with pre-18S RNA. (**A-D**) Proteins and RNAs are color-coded. The N-terminus of Utp24 is indicated with an N. Helices of the 5' ETS are labeled with roman numbers. Important functional sequence elements of the U3 snoRNA (5' hinge, 3' hinge, Box A duplex, Box A' duplex) and visible domains of the 18S rRNA (3', 5' and central domain) are indicated. The cleavage site A1 is shown as pink sphere. Active site residues of Utp24 are highlighted as pink sticks.


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
Utp24
Utp7
Sof1
Bms1
VI
Imp3
N
5′ hinge
Utp1
3′ hinge
Utp11
5′ ETS
3′ end
U3 snoRNA
Utp18
Utp21
C
Sof1
Utp14
Utp6
Utp24
A1
Faf1
Box A duplex
V
Utp7
Utp10
Sas10
Imp3
5′ hinge
VI
D
Utp20
rpS22
5′ domain
Sof1
Lcp5
Bms1
rpS9
rpS23
rps7
Nop14
Utp24
Bms1
Utp14
Box A duplex
A1
Box A duplex
Faf1
VI
Imp4
Utp6
Utp7
V
Imp3
Mpp10

The GTPase Bms1 and is positioned in the core of the SSU processome.

Higher resolution information has allowed us to better understand the network of interactions that the GTPase Bms1 takes part in, within the core of the SSU processome (3.18A). Besides its translational GTPase architecture, with the four canonical domains I-IV, Bms1 interacts with Rcl1 through a dedicated linker domain, as previously described, and with Kre33 through a newly identified linker domain (Fig. 3.18B). Furthermore, Bms1 is anchored in the SSU processome through a long C-terminal extension which wraps around several factors and RNA, to end in a more than 70 residues alpha helix which spans the core of the particle (3.18B).

Domains I, II and III are lodged between Rcl1, Kre33 and helices 15 and 17 of the 5' domain, while domain IV reaches further into the core of the particle (Fig. 3.18, 3.19). These interactions with RNA are further stabilized by a network of peptides which span the SSU processome and interact extensively with Bms1. These peptides include Utp11, Mpp10, Sas10 and Nop14 (Fig. 18C). The central location of Bms1 in the SSU processome and its network of interactions with extended elements in the particle puts it at an ideal position to modulate the conformation of the SSU processome through in GTPase activity.

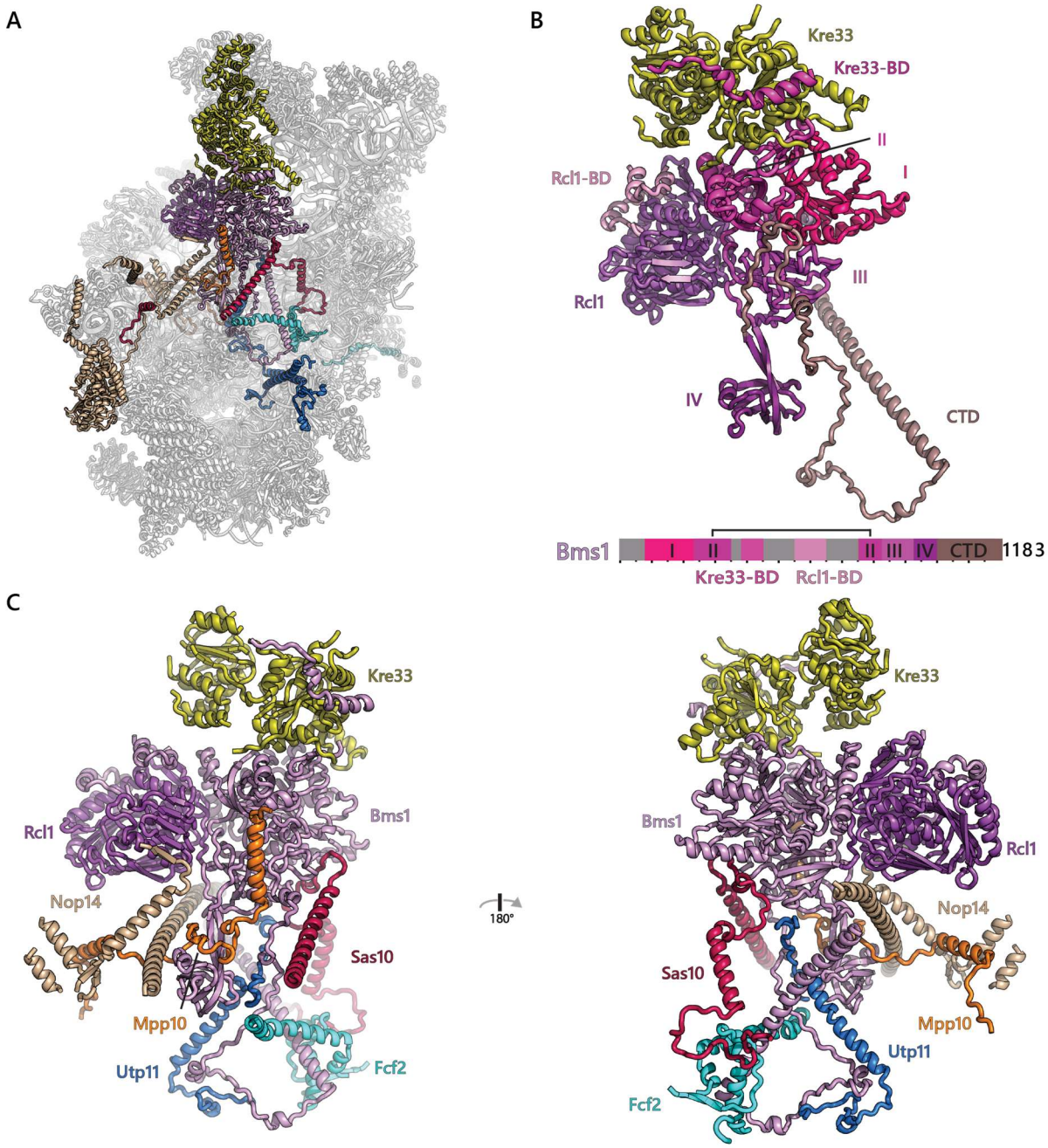


Figure 3.18. Structural analysis of Bms1 and its interaction partners. (**A**) Overview of Bms1 and its binding partners within the SSU processome. Proteins are colored as in (Fig. 3.10) with a transparent outline of the SSU processome in white. (**B**) Architecture of the Bms1-Rcl1 complex with Bms1 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.

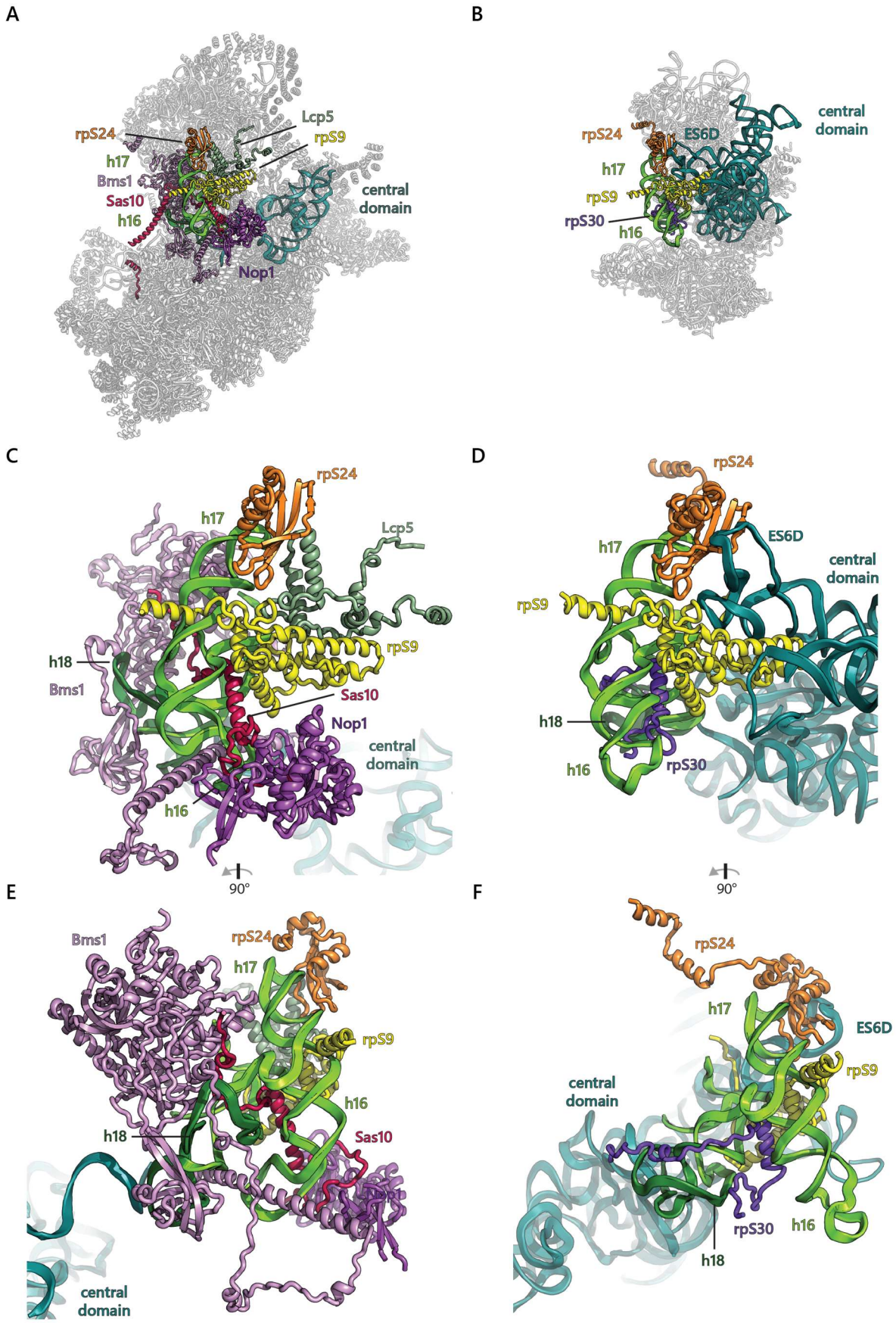
RNA remodeling prevents central pseudoknot formation

In the context of the SSU processome, Lcp5 and Sas10 are multi-functional proteins. In addition to Utp18, which contains a peptide that can interact with the exosome-associated helicase Mtr4 (Thoms et al. 2015), Lcp5 and Sas10 contain exosome interaction motifs (Sas10 domains) (Mitchell 2010). Lcp5 is positioned next to rpS9 and the 5' domain, a location (Fig. 3.19). This location is occupied by expansion segment ES6D in the mature small subunit, making the presence of Lcp5 mutually exclusion with the mature folding of the central domain relative to the 5' domain.

In contrast to Lcp5, the ordered parts of Sas10 are more elongated. Sas10 contains a solvent exposed N-terminal Sas10 domain, blocks the active site of Emg1 with a short peptide and bridges between the 5' and 3' domains with a long helix near Utp30 (Fig. 3.10B, C). In addition, it interacts with the U3 snoRNP through its C-terminus and is involved in RNA-protein remodeling by adopting a similar conformation, and occupying the same RNA binding site as rpS30 in the mature small subunit, close to rpS9 (Fig. 3.17B, C and 3.18).

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

Figure 3.19. Steric hindrance and molecular mimicry prevent premature ribosomal-RNA folding. (**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) (ref. 33). (**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
rpS24
Lcp5
rpS9
h17
Bms1
Sas10
h16
central domain
Nop1
B
central domain
rpS24
ES6D
h17
rpS9
rpS30
h16
C
rpS24
h17
Lcp5
rpS9
h18
Bms1
Sas10
Nop1
central domain
h16
90°
D
rpS24
h17
ES6D
central domain
rpS9
h18
h16
rpS30
90°
E
Bms1
rpS24
h17
rpS9
h16
h18
Sas10
central domain
F
rpS24
h17
ES6D
rpS9
central domain
rpS30
h16
h18

Within the SSU processome, the four structured domains of the 18S rRNA (5' central, 3' major and 3' minor domain) are segregated into different regions of the particle, thereby facilitating their separate maturation. RNA elements that are positioned in the vicinity of the central pseudoknot in the mature small subunit (Ben-Shem et al. 2011) are distinctly separated in the SSU processome (Fig. 3.20A, B). This is organized through multiple mechanisms. RNA mediated chaperoning of sequences close to the central pseudoknot in the 5' and central domains is accomplished through base pairing with U3 snoRNA boxes A and A' (Fig. 3.13, 3.17A, C, D). As a consequence of these interactions, sequences vicinal to the central pseudoknot, such as helix 27, adopt different conformations in the SSU processome, where a new stem loop forms between Rcl1 and Utp12 (Fig. 3.20A).

Mpp10 plays a central role in the remodeling of nucleotides close to helices 44 and 45. A partial unwinding of the region upstream of helix 44 results in an RNA loop (nucleotides 1628-1639) that is stabilized by Mpp10 (Fig. 3.20C). Due to this partial unwinding, 16 nucleotides of the opposite strand (nucleotides 1755-1769) are available to serve as a linker to helix 45, which is positioned 60 Ångstroms away on top of Pno1 that is held in place by Utp1 and Utp21 of the UtpB complex (Fig. 3.20C).

A second 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 (Fig. 3.20B). In the SSU processome, conserved elements of Bms1, Utp11 and Sas10 are employed in a concerted fashion to dramatically remodel helix 18 (h18, nucleotides 558-590) of the 18S rRNA (Fig. 3.19C-F, 3.20D). The C-terminal linker region and domain IV of Bms1 together with the

conserved N-terminal segment of Utp11 and a conserved linker region of Sas10 stabilize the remodeled RNA as well as rpS23, which is located in proximity to domains I-III of Bms1. Combined with the chaperone functions of Utp20 and Rrp5 in the 5' and central domains (Fig. 3.12), these examples highlight the precise and elaborate mechanisms that are employed to control tertiary interactions between ribosomal RNA domains.

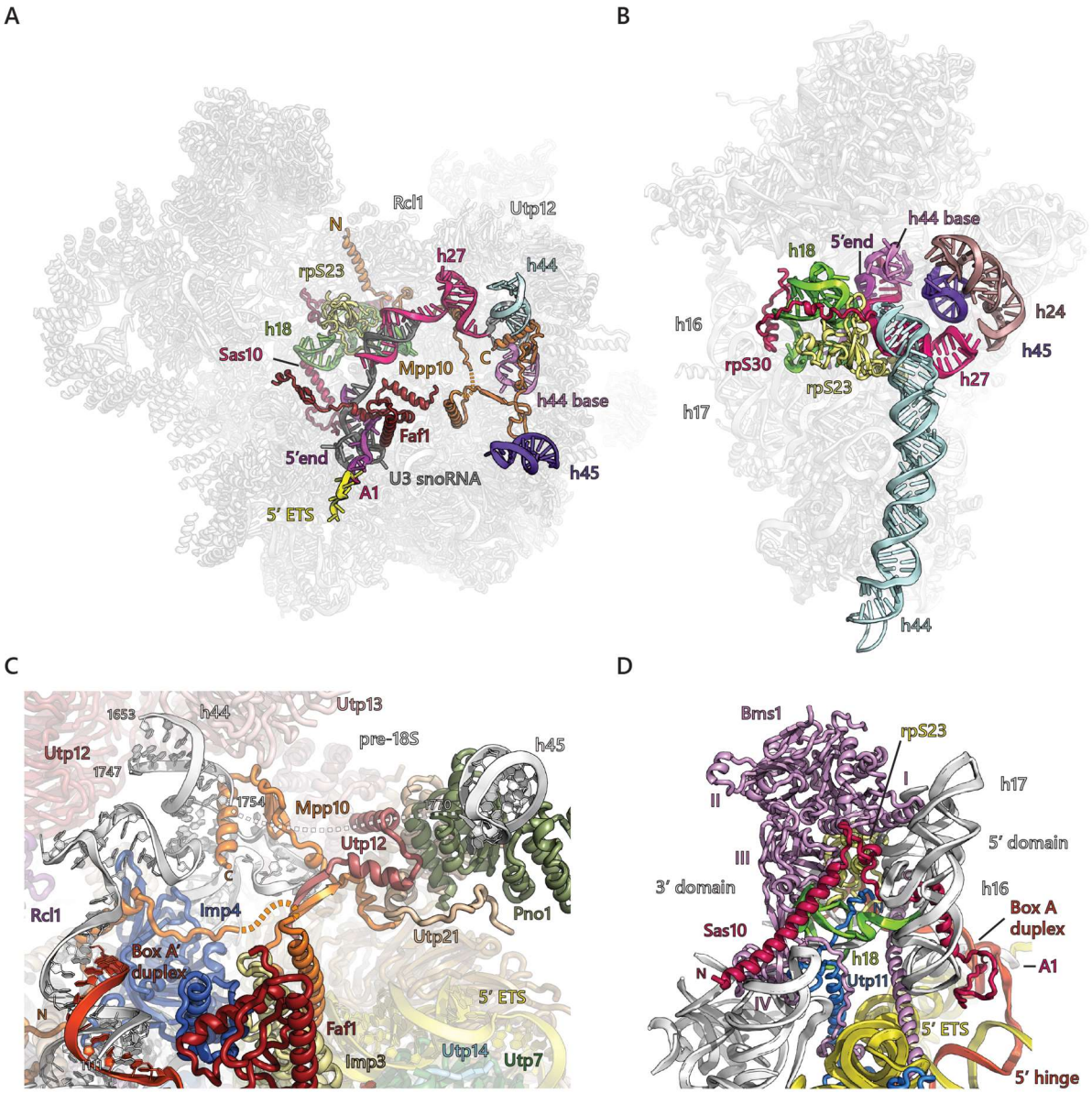


Figure 3.20. RNA remodeling prevents central pseudoknot formation. (**A**) The central pseudoknot and 18S rRNA elements in its vicinity are shown in color in their immature positions in the SSU processome (grey) and labeled with their corresponding mature 18S rRNA helix (h) number. Chaperoning RNA, ribosomal 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

Conclusions

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

A second function of the eukaryotic ribosome biogenesis machinery is the encapsulation and guided stabilization of a series of early ribosome assembly intermediates, which ultimately result in the mature small subunit (Lebaron et al. 2012).

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

Chapter IV - Large subunit processome assembly on pre-ribosomal RNA

The combined data from the cryo-EM structure of the small subunit processome and the study of its assembly has provided important insights into the role of many small subunit assembly factors. Early events in large subunit assembly, however, remain poorly understood. We therefore set out to use a similar approach to study the assembly of the large subunit processome.

An improved system to study large subunit processome assembly

Purification of pre-ribosomal complexes relies on the same approach as previously (Chaker-Margot et al. 2015). However, we decided to improve the purification strategy by inverting first and second capture step. In this version of the system, the expressed and truncated pre-rRNAs are purified in the first step and the resulting eluates are recaptured by protein affinity purification in the second step. For that purpose, we co-express a protease-cleavable MS2-GFP fusion and the truncated RNA, both under Gal1 promoters (Fig. 4.1A). The MS2 protein is fused to a nuclear localization signal (NLS) to promote its association with the truncated RNA in the cell.

The RNA constructs used here begin at the A3 site thereby mimicking A3 cleaved constructs. Previous studies have suggested that separation of the rDNA locus between the A2 and A3 sites, at nucleotides 2739 to 2747 of the 35S pre-rRNA, leads to viable ribosome production suggesting that there is no essential link between small and large subunit assembly factors (Liang and Fournier 1997). Target RNAs span the second half

of the 35S locus, ending at each of the structural domains of the 25S, I-VI (Fig. 4.1A). A shorter construct, which spanned the A3 site to the end of ITS2 was tested but was seemingly not amenable to purification (data not shown). The final and longest construct spans the entirety of the 25S and ends at the B0 site. A “full-length” construct that ended at the native end of the pre-rRNA underwent B0 cleavage and made isolation impossible (data not shown).

The nucleolar factor Cic1 was chosen as protein bait for the second capture. Cic1 is known to associate with ITS2 (Wu et al. 2016) and therefore represented an early factor likely to associate with all expressed constructs. By using Cic1 as protein bait, we ensured that ITS2 processing has not occurred yet, narrowing the type of pre-60S particles that can be isolated to earlier species.

Upon induction, truncated pre-rRNA are expressed and bound by MS2-GFP and the endogenous large subunit assembly machinery, including the sbp-tagged Cic1. During purification, the RNA molecules are captured by the anti-GFP resin (Fig. 4.1B). In the second step, the targeted pre-ribosomal complex is enriched by streptavidin capture, and subsequent biotin elution.

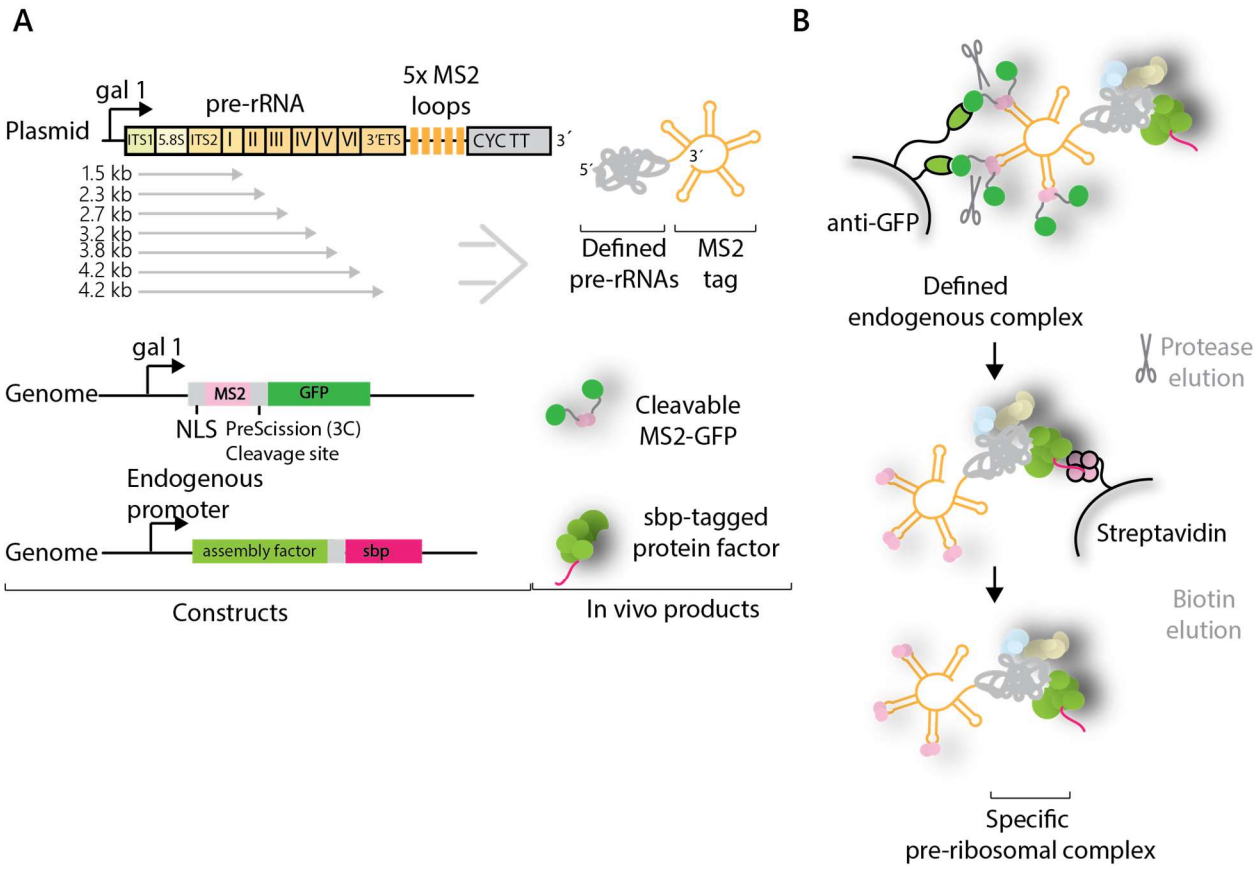


Figure 4.1 New purification technique of pre-ribosomal particles. (**A**) Plasmid overexpression of aptamer tagged pre-rRNAs of different lengths in genetically modified yeast strains containing a sbp-tagged ribosome assembly factor. Plasmids contain the gal1 promoter and CYC transcriptional terminator (CYC TT). Pre-rRNA domains, sbp-tagged ribosome assembly factor and MS2-GFP protein are shown. (**B**) Purification strategy using GFP and sbp as sequential tags, followed by biotin elution.

Stage specific assembly of the earliest large subunit assembly intermediates

As for the small subunit processome, the use of our purification system has allowed us to recapitulate the assembly of the earliest large subunit assembly intermediates. The negative control, wherein only the MS2 loops were expressed, yielded an accordingly clean sample, except for the presence of streptavidin (Fig. 4.2A). The addition of the first section of pre-rRNA, spanning from the A3 site to the end of domain I of the 25S, lead to the recruitment of approximately 10 factors. The subsequent addition of domains of the 25S induced the recruitment of more factors, although in a less dramatic fashion than what was observed for the small subunit processome (Chaker-Margot et al. 2015). Samples for all seven constructs, with termini at domain I-VI and B0, plus a negative control were purified in triplicates and analyzed by LC-MS. The median amounts (as quantified by areas) for each protein were then used (Fig. 4.2B, C). Only proteins with amounts within 100-fold of the protein bait Cic1, and with more than 5 unique peptides and were considered for the analysis. The amounts of Cic1, which should stay similar throughout, were used to normalize the samples. Once again, ribosomal proteins have been excluded from this analysis, as signals of highly abundant peptides from contaminating mature ribosomes prevent an unambiguous assignment of ribosomal proteins to a particular stage of ribosome assembly in our system.

This analysis has allowed us to determine the timing of recruitment of approximately 50 assembly factors on nascent pre-rRNA (Fig. 4.2B, C). The first construct (containing 5.8S rRNA, ITS2 and domain I) recruits 12 assembly factors, including several known binders of ITS2, Cic1, Nop7, Nop15 and Rlp7 (Wu et al. 2016). The helicase Has1, which is involved in small and large subunit processing (Emery et al. 2004) is also

incorporated in this particle. Strikingly, this first construct also induces the recruitment of the protein Pwp1, whose levels decrease drastically for longer constructs, indicating its transient role in early large subunit assembly. Extension of the pre-rRNA into domain II recruits another 11 factors. These include the large protein Rrp5 and its known interaction partners, Noc1 and Noc2. Rrp5 is required for the assembly of the small subunit while Noc1/Noc2 have been shown to co-precipitate with the SSU processome (Fig. 2.3, Appendix 7.2), likely through their interaction with Rrp5 (Hierlmeier et al. 2013). Nsa1, the nucleolar factor which defines the nucleolar pre-60S intermediate also binds at this stage (Kressler et al. 2008). While the addition of domain III recruits no specific assembly factors, domain IV induces the binding a single protein, Puf6. Domain V leads to the appearance of Nop13, a known nucleolar pre-60S factor and 3 members of the Npa/Urb subcomplex: Urb1, Nop8 and Rsa3. These proteins form a subcomplex, which also includes Urb2 and the helicase Dbp6, involved in the earliest steps of large subunit biogenesis (Rosado et al. 2007). Its role in large subunit assembly is not understood but this suggests association of this complex to domain V.

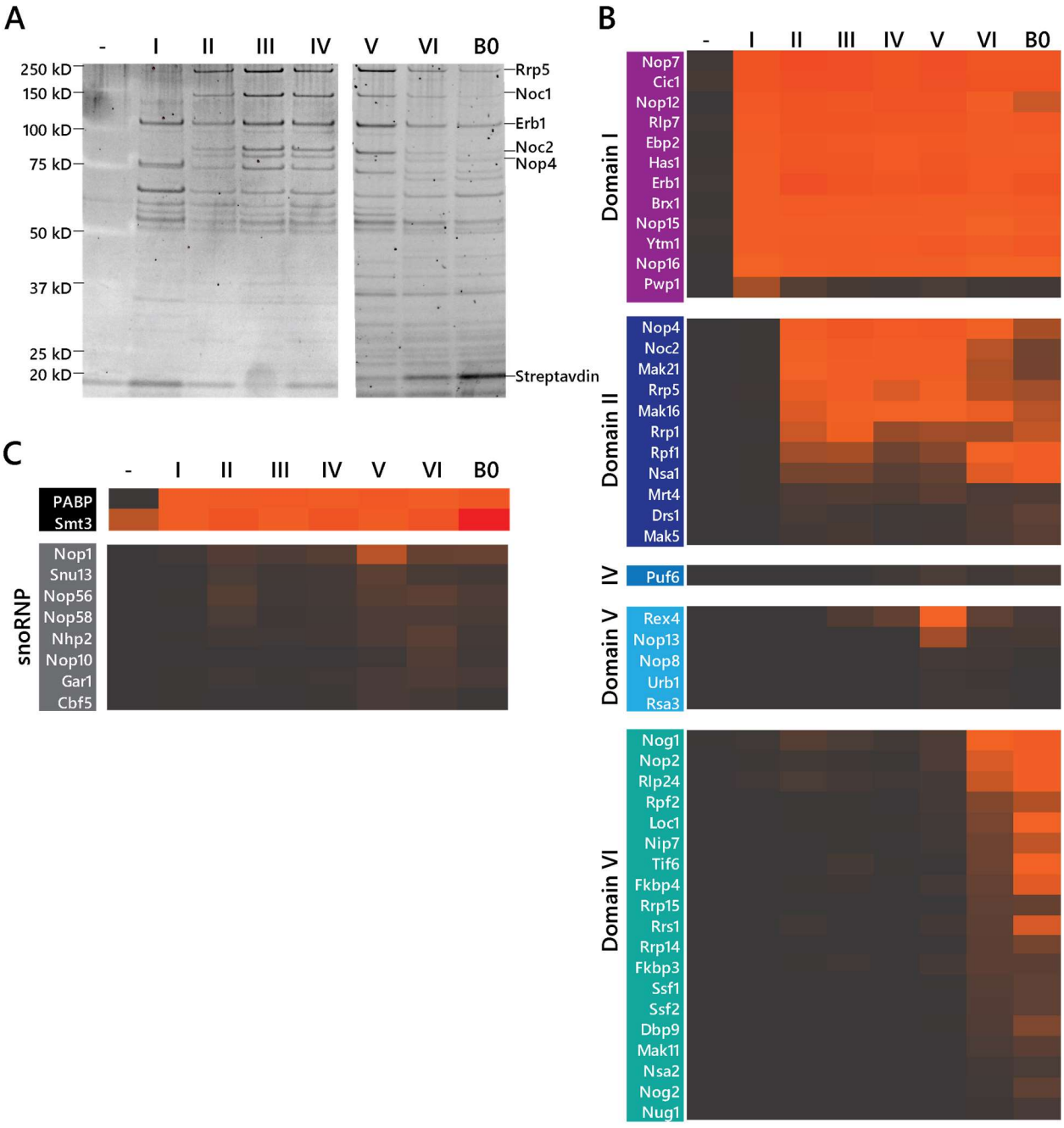
Completion of the 25S induced the recruitment of many factors: the GTPase Nog1 and the chaperones of the 5S rRNA, Rpf2 and Rrs1, among others. At this stage, Tif6 and Rlp24 are also recruited, and will stay associated with pre-60S particles until their removal in the cytoplasm. Northern Blotting analysis of the samples showed that the 5S rRNA and the box H/ACA snoRNA snR10 also appear with the addition of domain VI (Fig. 4.3A) snR10 is involved in the pseudouridylation of U2923, in domain V near domain VI (Ni et al. 1997). The truncated pre-rRNAs could not be observed by Northern Blot due to degradation resulting in low molecular weight smears (data not shown). This

prevented us from having a loading control. Differences in the levels of 5S rRNA and snR10 may therefore be due to differences in loading and should be interpreted carefully. Box C/D and box H/ACA snoRNP proteins were present throughout the samples. However, there was a moderate increase in signal in the levels of box H/ACA proteins (Cbf5, Nhp2, Nop10 and Gar1) at domain VI, consistent with the presence of snR10. The presence of other small RNAs was not tested.

Surprisingly, low levels of Nog2 were also detected in the A3-domain VI sample (Fig. 4.2B). Nog2 is a nuclear assembly factor whose binding site in the pre-60S particles overlaps with nucleolar factors such as Nip7, Sbp1 and Nop2 (Wu et al. 2016; Kater et al. 2017), which makes its association with nucleolar intermediates unlikely. This may be evidence that the A3-domain VI sample represents multiple pre-60S intermediates rather than a specific large subunit precursor, or that Nog2 can associate with the nucleolar pre-60S factors through peptide-like interactions. High levels of the ubiquitin like modifier SUMO (Smt3) were also detected in all samples (Fig. 4.2C)

The overexpressed truncated pre-rRNAs were analyzed in total RNA extracts by northern blotting using an MS2 probe (Fig. 4.3B). RNA bands corresponding to their expected size were visualized for each sample, but were weak for the domain VI and B0 truncated species. Surprisingly, we also observed the presence of a degradation product for all constructs that ended at domain II or beyond. Since the MS2 loops are located at the 3' end of the constructs, this allows mapping the approximate cleavage site to around nucleotide 1350 in domain II. The mechanism by which this cleavage happens is unknown.

Figure 4.2 Stage-specific assembly of the large subunit processome. (**A**) SYPRO-Ruby stained SDS-PAGE gel of pre-ribosomal complexes corresponding to increasing length of large subunit pre-RNA. Proteins indicated on the right were identified by corresponding molecular weight. (**B**) LC-MS analysis of ribosome assembly factors, based on a triplicate of samples, normalized against Cic1 and classified by recruitment stage. Proteins abundance are measured by total area under the curve and high abundance proteins are shown in orange/red while low abundance and absent proteins are shown in black/grey. (C) LC-MS analysis of the polyadenylate binding protein (PABP), Smt3 and the snoRNP proteins.


A
-
I
II
III
IV
V
VI
B0
250 kD
150 kD
100 kD
75 kD
50 kD
37 kD
25 kD
20 kD
Rrp5
Noc1
Erb1
Noc2
Nop4
Streptavdin
B
Domain I
Nop7
Cic1
Nop12
Rlp7
Ebp2
Has1
Erb1
Brx1
Nop15
Ytm1
Nop16
Pwp1
Domain II
Nop4
Noc2
Mak21
Rrp5
Mak16
Rrp1
Rpf1
Nsa1
Mrt4
Drs1
Mak5
IV
Puf6
Domain V
Rex4
Nop13
Nop8
Urb1
Rsa3
Domain VI
Nog1
Nop2
Rlp24
Rpf2
Loc1
Nip7
Tif6
Fkbp4
Rrp15
Rrs1
Rrp14
Fkbp3
Ssf1
Ssf2
Dbp9
Mak11
Nsa2
Nog2
Nug1
C
PABP
Smt3
snoRNP
Nop1
Snu13
Nop56
Nop58
Nhp2
Nop10
Gar1
Cbf5

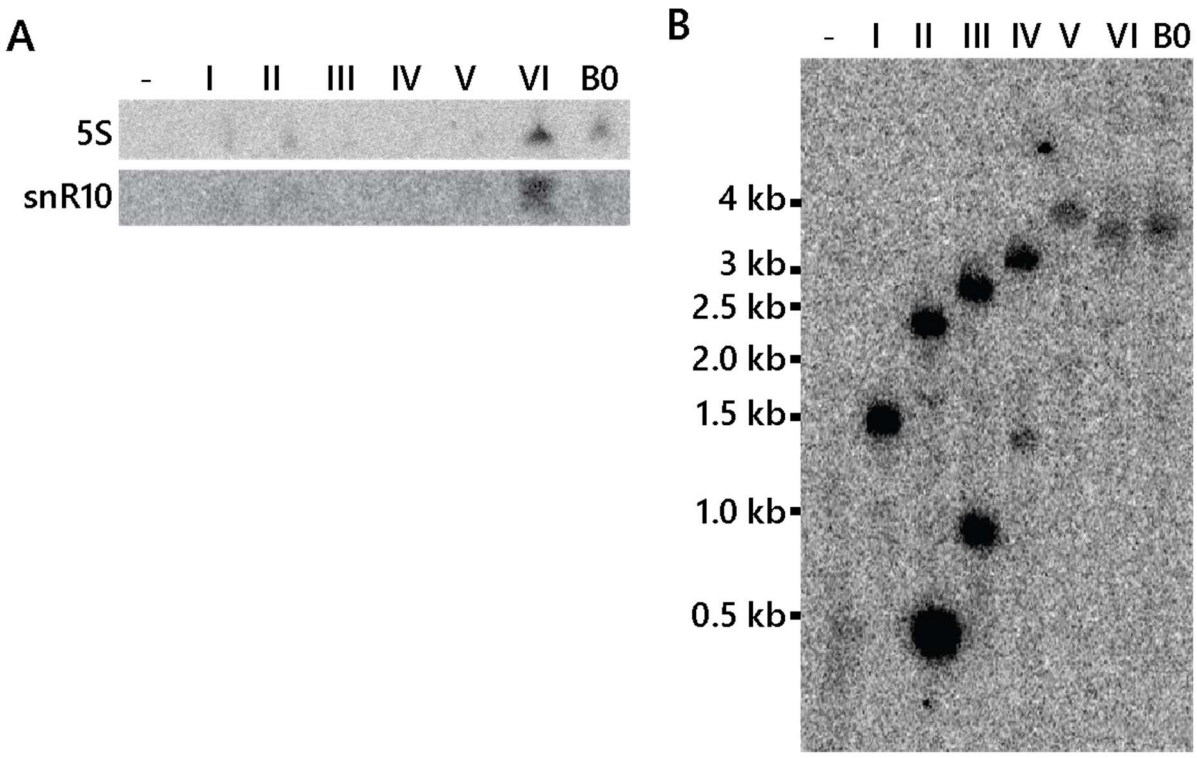


Figure 4.3 RNA analysis of pre-ribosomal complexes. (**A**) Northern blot with 5S and snR10 probes of RNA extracted from purified pre-ribosomal complexes. (**B**) Northern blot with MS2 probe of total RNA extracted from cells where truncated pre-rRNA were expressed.

Multiple binding sites of Rrp5 in ITS1.

Rrp5 is a multi-domain protein involved in both large and small subunit assembly (Fig. 4.4A). It is composed of 12 S1 RNA-binding domains followed by a C-terminal TPR repeat (Eppens et al. 1999). Noc1, also known as Mak21, and Noc2, two HEAT-repeat containing proteins, are known interactors of Rrp5 (Hierlmeier et al. 2013). The TPR domain of Rrp5 has been shown to bind to helix 24 in the small subunit processome (Barandun et al. 2017). Furthermore, Rrp5 is required for cleavage of the pre-rRNA at the A2 and A3 sites (Eppens et al. 1999; Vos et al. 2004). Within the early large subunit precursors, we see Rrp5 specifically recruited with the addition of domain II. This suggests it has another binding site in the 25S or with associated proteins. This data is consistent with previous cross-linking data *in vivo* (Lebaron et al. 2013).

Based on this data, we hypothesized that Rrp5 could be binding ITS1, and could have a role in bridging small and large subunit assembly complexes during transcription. To test this, we overexpressed truncated pre-rRNA ending at domain I but extending their start site further upstream in ITS1, to the A2 site and the D site, i.e. the 5' end of ITS1. We then performed tandem affinity purification on these particles to observe the absence or presence of Rrp5.

Extending the 5' end of our transcripts to both the A2 and D sites lead to the presence of Rrp5 (Fig. 4.4D). Levels of Rrp5 were somewhat lower than in the A3-II sample, and were not matched by the presence of Noc1. This suggests that Rrp5 weakly binds to ITS1, in a Noc1/Noc2 independent manner.

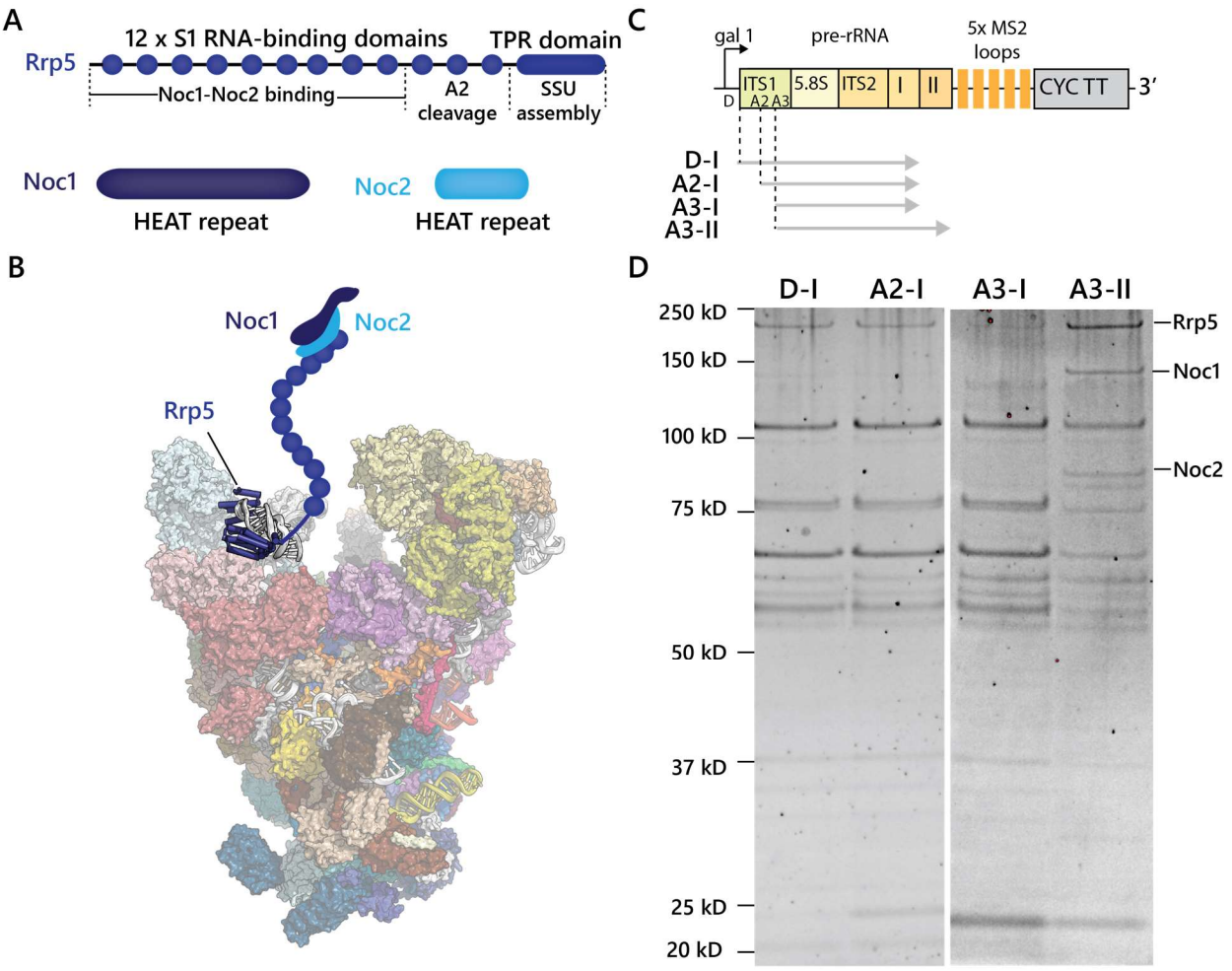


Figure 4.4 Multiple binding sites of Rrp5 on the pre-rRNA. (**A**) Domain architecture of Rrp5 and its interaction partners, Noc1 and Noc2. The functions of the different segments of Rrp5 are indicated (**B**) Location of the Rrp5 TPR repeat in the SSU processome and schematic representation of its N-terminal domains. (**C**) Constructs used to determine the existence of a binding site of Rrp5 in ITS1. (**D**). SYPRO-Ruby stained SDS-PAGE gel of purified pre-ribosomal complexes containing or not Rrp5.

Early processing of the 27SA RNA

While Rrp5 is involved in the very early steps of large subunit assembly, it is absent from later pre-60S particles, such as the Nsa1-containing nucleolar pre-60S. The latter was shown recently to contain a 27SB RNA species, that is, a large subunit precursor that has been processed at the 5' end in ITS1 and at the 3' end in the 3' ETS, but not in ITS2 (Sanghai et al. 2018). This led us to hypothesize that the transition from the earliest large subunit precursors, which contain Rrp5, to nucleolar pre-60S particles, which do not, might be brought about by 5' and 3' processing of the 27SA species. The 5' and 3' ends of the 27SB are located near each other in the nucleolar pre-60S structure and are therefore seemingly incompatible with the presence of 5' and 3' extensions (Kater et al. 2017; Sanghai et al. 2018). This would further suggest that 5' and 3' processing could lead to the emergence of the nucleolar pre-60S.

The presence of Rrp5 in the pre-ribosomal complexes we isolated suggested that those particles were unable to mature further. This, we postulated, could be caused by the presence of the MS2 loop at the 3' end of the constructs (Fig. 4.1A, 4.5A). To test this, we included a northern blot tag in the 25S (Fig. 4.5A), which allowed us to observe how efficiently overexpressed RNAs could be matured to 25S. Strikingly, the construct that began at the A3 site and ended at the native end of the rDNA locus, here called "full-length", is made much more efficiently into 25S than the constructs ending with domain VI or with the B0 site (Fig. 4.5B, C). This difference could be caused by the 3' MS2 loops interfering with normal 3' processing, or may be indicate that the 3' ETS is important for maturation of the RNA. Similarly, a "full-length" construct with 5' MS2 loops

does not form a 25S efficiently, which suggests that MS2 loops may interfere with processing.

We wanted to better understand what the requirements for efficient maturation of these precursor RNAs were. Particularly, we wanted to test whether cleavage of the 3' ETS was sufficient to lead to efficient 25S formation. If MS2 loops prevent normal cleavage of the 3' ETS, then restoring cleavage should result in normal 25S formation. For that purpose, we designed constructs truncated at the end of the 25S or B0, which were followed with the ribozyme hatchet (Fig. 4.5A), which leaves only three nucleotides after spontaneous cleavage, seconds after transcription (Li et al. 2015). The presence of the self-cleaving ribozyme at the 3' end of the RNA would mimic the cleavage by Rnt1.

Surprisingly, truncated constructs followed by hatchet were not efficiently processed into a 25S, unlike the "full-length" construct, in the absence or presence of the ribozyme (Fig. 4.5B, C). This suggests that 3' cleavage of the 27SA RNA is not sufficient for the early maturation of the large subunit precursor. Elements of the 3' ETS downstream of the B0 site may be necessary for efficient maturation of these precursors. Some other step, such as the action of a helicase may be required.

Lastly, we tested whether the removal of ITS1 from the constructs would be detrimental for the processing of large subunit precursors. An ITS1-less construct with a full-length 3' ETS (referred as FL-ΔITS) was not processed efficiently into a mature 25S (Fig. 4.5B, C). This suggests that ITS1 has an important role in the early assembly of the large subunit, consistent with previous genetic data (Liang and Fournier 1997)

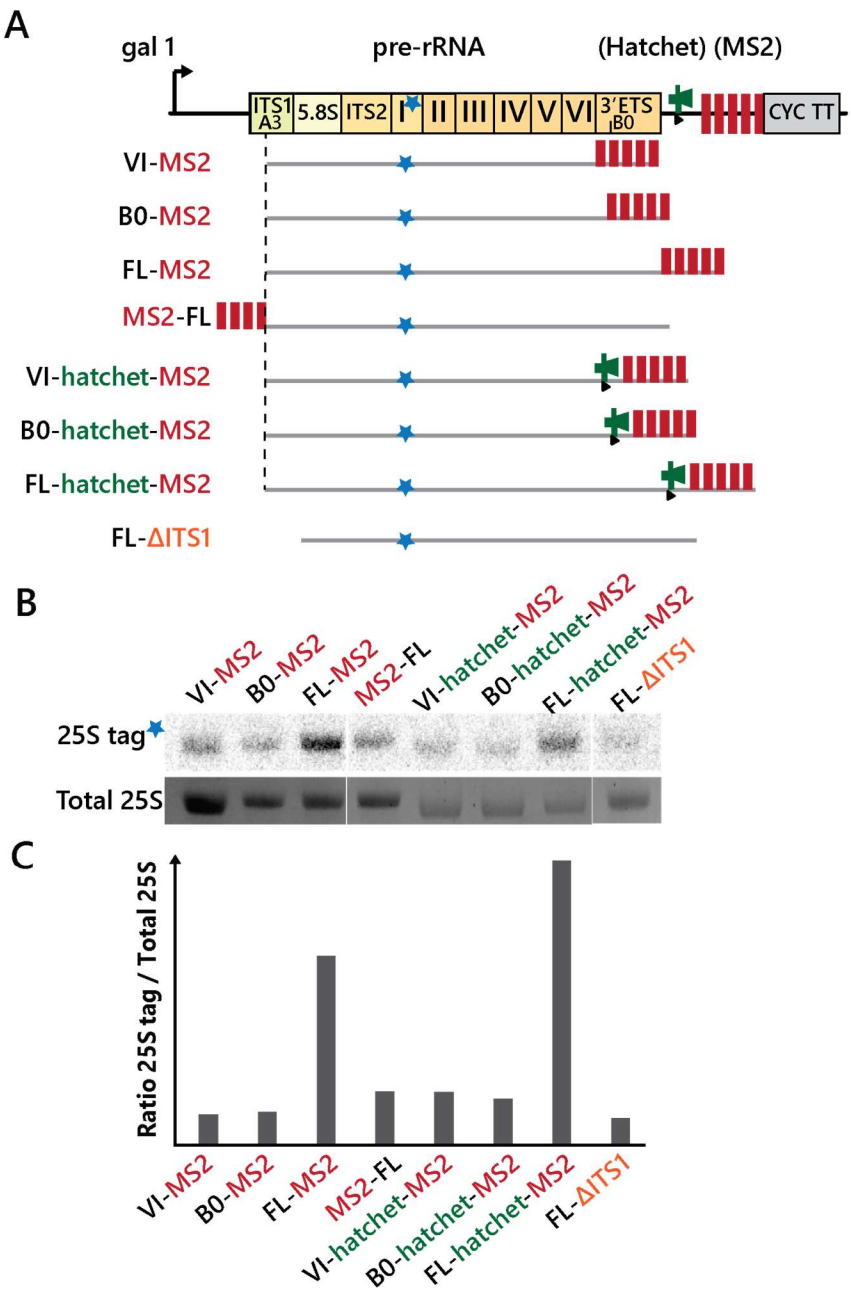


Figure 4.5 Processing of large subunit precursor RNAs. (**A**) Design of Northern blot tagged constructs used to determine processing efficiency of large subunit precursors. (**B**) Northern blotting analysis of mature tagged 25S coming from overexpressed constructs and corresponding levels of total 25S. (**C**) Relative amounts of tagged 25S to total 25S.

Conclusions

By overexpressing truncated pre-rRNA, we were able to recapitulate the assembly of ribosome biogenesis factors on the second half of the pre-rRNA locus. Large subunit assembly begins with the recruitment of many factors, which bind ITS2, 5.8S and domain I. With the addition of domain II, the Rrp5-Noc1-Noc2 complex, which is associated with the small subunit processome and ITS1, binds the forming large subunit, thereby bridging small and large subunit assembly. Further transcription does not lead to addition of many additional factors, since domains III-V are flexible in the early stages of pre-60S maturation (Kater et al. 2017; Sanghai et al. 2018). Completion of domain VI induces the incorporation of several factors, the 5S rRNA and association of snR10.

Using a similar approach, another group reported their results of the assembly of early large subunit precursors (Chen et al. 2017). While their results are mostly in good agreements with ours, there are a few differences. Their model has Rrp5-Noc1-Noc2 being recruited at domain I. They observe Nop53 recruited at domain VI, despite that the nucleolar pre-60S contain Erb1, whose presence is mutually exclusive with Nop53 (Kater et al. 2017; Sanghai et al. 2018). Several proteins that we classified as domain VI factors including Nip7, Nop2, Loc1, Rrs1 and Rpf2, were seen already at domain V. Lastly, a few proteins are absent from their model, such as Rrp14, Rrp15, or Rex4. Those differences may be partially due to the difference in the position of the MS2 tag, in ITS2 instead of at the 3' end, in their constructs.

Maturation of large subunit precursors requires processing at the 5' and 3' of the precursor RNA, in ITS1 and the 3' ETS. This processing involves cleavage and

resection of the spacer RNAs by Rnt1 and Rex1 for the 3’ end and Rrp17 or Rat1 for the 5’ end (Henry et al. 1994; van Hoof 2000; Oeffinger et al. 2009). We have determined here that cleavage is not sufficient for the maturation of the pre-60S into the nucleolar pre-60S. It will be interesting to determine the exact steps required for this transition and how it is coupled to Rrp5-Noc1-Noc2 removal.

Chapter V - Discussion

Early ribosome biogenesis illustrates the staggering amount of energy and control that eukaryotic cells dedicate to the assembly of the translation machinery. This process involves about 250 factors, whose combined mass exceed 10 Megadalton, more than 3-fold the mass of the mature ribosome. The study of the assembly of ribosome assembly factors onto the pre-rRNA and the structure of the small subunit processome has provided a new understanding of the role of more than 50 factors in early ribosome biogenesis. These results have also uncovered a new mechanism of the regulation of ribosome assembly, which illustrates how the cell performs extensive control of the formation of its translational machinery.

5.1 A new model for transcription coupled early ribosome biogenesis

The temporal assembly of ribosome assembly factors on the pre-rRNA has now been elucidated for both small and large subunit precursors. It is now possible to map this data onto the structures of the small subunit processome and nucleolar pre-60S particles, which we and others have obtained (Kornprobst et al. 2016; Chaker-Margot et al. 2017; Sun et al. 2017; Barandun et al. 2017; Cheng et al. 2017). This allows us to contextualize this dynamic data and gain new insights into the assembly of these particles (Fig. 5.1, 5.2, 5.3).

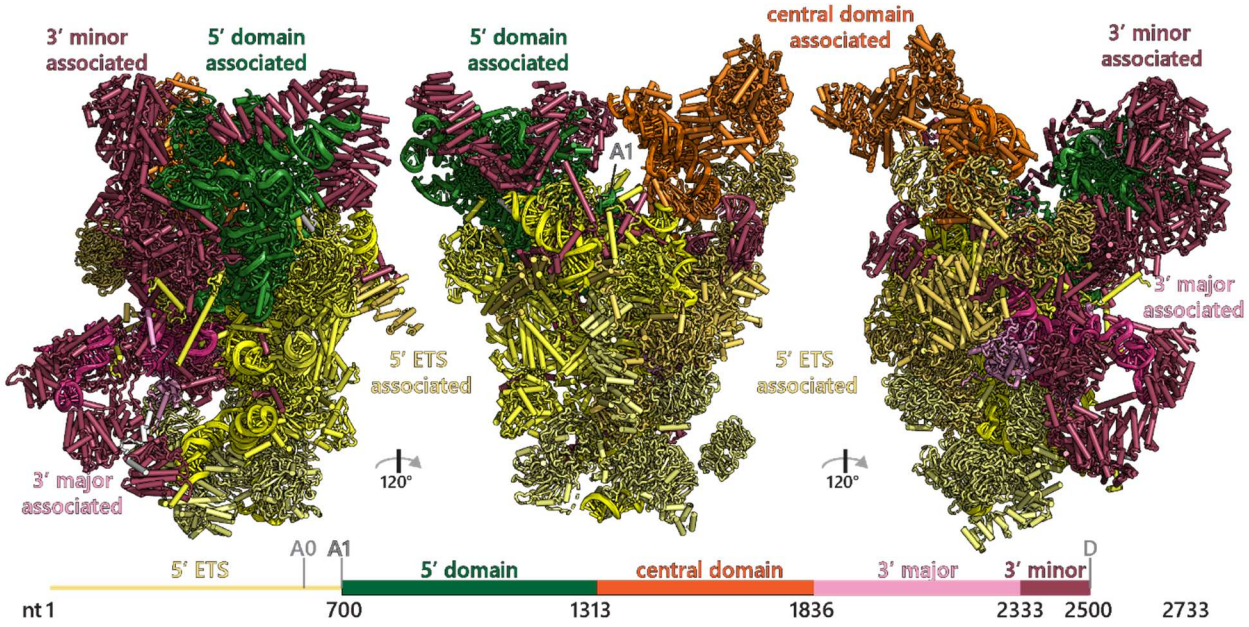


Figure 5.1 Mapping the assembly stages of the SSU processome. Three views of the structure of the SSU processome (PDB 5WLC) with factors colored according to their stage of assembly onto pre-rRNA, as defined in (Chaker-Margot et al. 2015). Diagram of the pre-rRNA up to the end of the 18S is shown below, with domains colored analogously to the model.

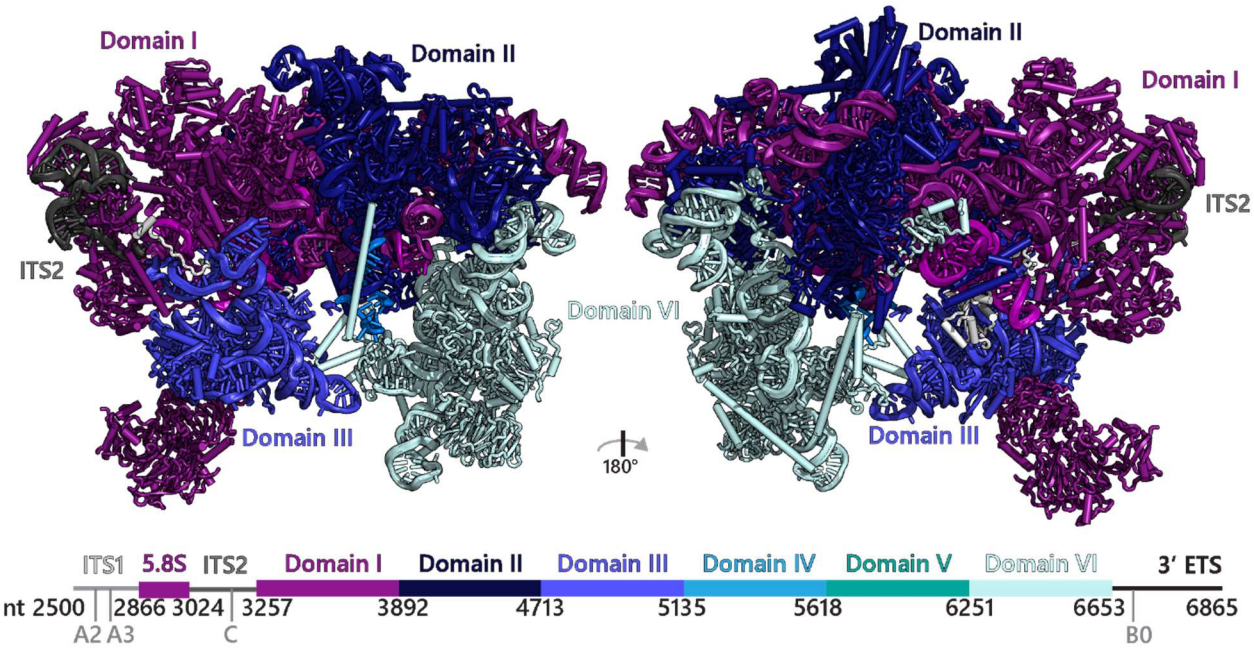


Figure 5.2 Mapping the assembly stages of the LSU processome. Two views of the combined structure of the nucleolar pre-60S (PDBs 6C0F, 6CB1) with factors colored according to their stage of assembly onto pre-rRNA, as defined in chapter IV. Diagram of the pre-rRNA from ITS1 to the end of the 35S is shown below, with domains colored analogously to the model.

Transcription of the pre-rRNA begins with the 5' ETS, which provides an architectural framework and facilitates the recruitment of the U3 snoRNP, UtpA and UtpB (Fig. 5.1, 5.3). The 5' ETS particle also induces the recruitment a several of individual proteins, which have many extended linkers, such as Mpp10, Bud21, Sas10, Utp11 or Fcf2. The formation of the SSU processome may be directed by an initially flexible set of peptide interactions. As SSU processome maturation progresses, the number of these peptide interactions likely increases and results in this stable nucleolar superstructure.

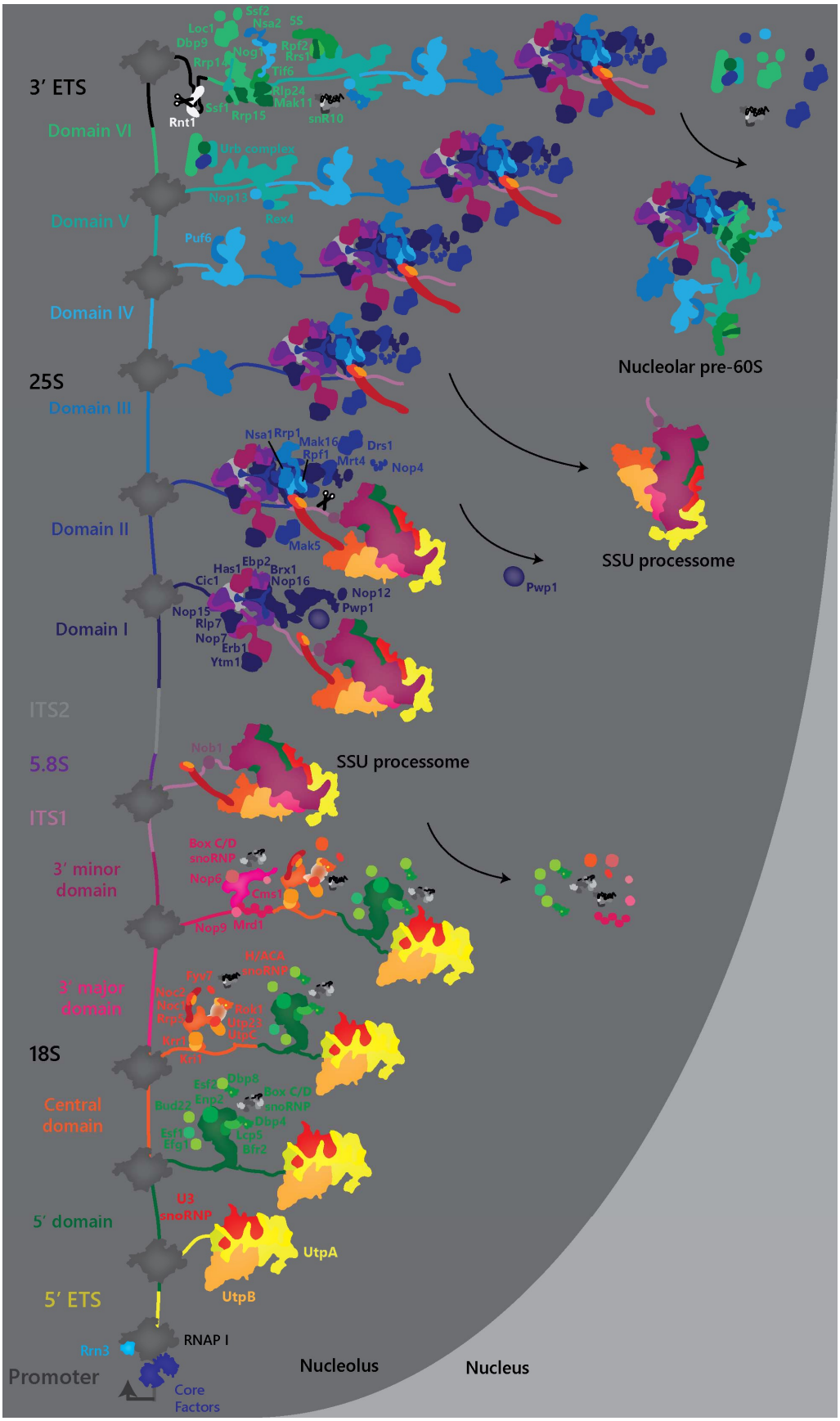
Interestingly, the folding of the 3' minor domain (Fig. 5.1) is essential for the recruitment of the outer shell of proteins. The placement of Bms1 and extensions of Nop14 at the core of the SSU processome suggests that domains of the 18S may be initially flexible relative to one another during co-transcriptional assembly of the SSU processome and only form a rigid structure with the completion of the 18S (Fig. 5.3). This may also explain the dissociation of many factors, such as Mrd1, which have a role in stabilizing important segments of the pre-rRNA, which are later buried in the SSU processome (Segerstolpe et al. 2013).

Addition of ITS1 leads to the recruitment of Nob1, which will stay bound to small subunit intermediates until its role in cleavage at the D site in the cytoplasm. Rrp5 also binds ITS1, between the A2 and A3 sites. Transcription up to domain I of the 25S induces the incorporation of factors forming one side of the emerging pre-60S (Fig. 5.2). The addition of domain II is stabilized by binding of Rrp5, Noc1 and Noc2, which join small and large subunit processomes (Fig. 5.3). This also leads to the incorporation of several factors, which completes the core of the nucleolar pre-60S (Fig. 5.2). This may further

trigger cleavage at the A2 site, which separates small and large subunit assembly pathways.

While the transcription of subsequent domains leads to few changes in the forming LSU processome, completion of the 25S induces the recruitment of many factors, which will form the interface of domain VI to the rest of the particle, as well as the 5S RNP (Fig. 5.2 and 5.3). However, the emergence of the nucleolar pre-60S requires first 5' and 3' processing in ITS1 and the 3' ETS respectively, and possibly other remodeling steps. This allows domain VI to fold near domains I and II (Fig. 5.3), leading to the conformation that was observed in the structure of the Nsa1-containing nucleolar pre-60S (Kater et al. 2017; Sanghai et al. 2018).

Figure 5.3 Co-transcriptional assembly of the large and small subunit processome. Diagram of the assembly of ribosome biogenesis factors onto the nascent pre-rRNA as it is being transcribed by RNA polymerase I. This leads to the emergence of the SSU processome and nucleolar pre-60S which have been structurally characterized.


3′ ETS
Domain VI
Domain V
Domain IV
25S
Domain III
Domain II
Domain I
ITS2
5.8S
ITS1
3′ minor domain
3′ major domain
18S
Central domain
5′ domain
5′ ETS
Promoter
Rnt1
Ssf1
Rrp15
Loc1
Ssf2
Nsa2
5S
Dbp9
Nog1
Rpf2
Rrs1
Rrp1
Tif6
Nip24
Mak11
snR10
Urb complex
Nop13
Rex4
Puf6
Nsa1
Rrp1
Mak16
Rpf1
Drs1
Mrt4
Nop4
Mak5
Pwp1
Has1
Ebp2
Brx1
Cic1
Nop16
Nop12
Pwp1
Nop15
Rlp7
Nop7
Erb1
Ytm1
Nob1
SSU processome
Box C/D snoRNP
Nop6
Cms1
Nop9
Mrd1
H/ACA snoRNP
Fyv7
Noc2
Noc1
Rrp5
Rok1
Utp23
UtpC
Krr1
Esf2
Dbp8
Enp2
Box C/D snoRNP
Bud22
Dbp4
Esf1
Lcp5
Efg1
Bfr2
U3 snoRNP
UtpA
UtpB
Rrn3
RNAP I
Core Factors
Nucleolar pre-60S
SSU processome
Nucleolus
Nucleus

5.2 Maturation of the small subunit processome

SSU processome particles must undergo several significant structural and compositional changes to mature into pre-40S particles, which adopt a near-mature state (Heuer et al. 2017; Scaiola et al. 2018). The domains of the 18S rRNA in the SSU processome are kept separated, in part by the concerted stabilization of non-mature RNA elements, which are later situated near the central pseudoknot (Fig. 5.4A-C). Maturation of the SSU processome will therefore require several steps including the formation of the central pseudoknot and RNA cleavage at site A1 (between the 5' ETS and the 18S), and site A2 (in ITS1), unwinding of the U3-18S duplexes, removal of many ribosome assembly factors, including the entire 5' ETS particle, and the inclusion of new ribosomal proteins and assembly factors, such as Tsr1, Rio2 and Ltv1, for nuclear export.

The most critical steps for SSU processome maturation are the cleavages of the A1 and A2 sites, formation of the central pseudoknot and removal of the 5' ETS particle. This requires the base-pairing of U3 snoRNA and the central pseudoknot sequences to be unwound. This is partially catalyzed by the DEAH-box helicase Dhr1 (also known as Ecm16) (Sardana et al. 2015). The structure of the SSU processome obtained from Dhr1-depleted cells (Sun et al. 2017) looks nearly identical to SSU processomes purified from starved cells (Barandun et al. 2017), where Dhr1 and its cofactor Utp14 (Sardana et al. 2014) are present (Fig. 5.4E). It is possible that conformational changes have to occur in order for the helicase to engage its substrate.

This transition may require the concerted activity of ATPases, such as Dhr1, Dhr2, Has1 and Rok1, and the GTPase Bms1. In structures of the SSU processome, U3-pre-rRNA

duplexes are heavily protected by a number of ribosome assembly factors. Any access to those RNA duplexes by helicases will require a significant conformational change. It is possible that conformational changes that lead to those processing events could be brought about by Bms1 in a GTP-dependent mechanism. Bms1 is strategically located at the core of the SSU processome between pre-18S domains and makes extensive contact with many factors that pierce through the particle. It is interesting that the GTPase site of Bms1 is solvent exposed in the SSU processome structures (Fig. 5.4D). GTPases typically require a GTPase activating protein (GAP) to stimulate GTPase activity. Bms1 would therefore be ideally positioned to be regulated by a trans-acting GAP. No such factor has been identified, but this would provide an ideal checkpoint for early small subunit maturation. Cleavage at the protected A1 site by Utp24, may follow U3-pre-rRNA duplex unwinding.

It should be noted that the role of GTP hydrolysis by Bms1 has not been characterized and may be required prior to the formation of the SSU processome. However, for several ATPases and GTPases of ribosome assembly, such as Rio1, Rio2, Nog2, Lsg1 and Efl1, nucleotidase activity results in the dissociation of these factors (Finch et al. 2011; Ferreira-Cerca et al. 2012, 2014; Matsuo et al. 2014; Malyutin et al. 2017)

Concomitant to the dramatic rearrangement of pre-18S domains, and the association of ribosomal proteins and late ribosome assembly factors, the entire 5'ETS particle has to be removed and degraded, a critical step to recycle ribosome assembly factors. This is thought be accomplished by the exosome, given that mutations to the exosome core or its associated cofactors lead to dramatic accumulation of the cleaved 5' ETS (Allmang et al. 2000). The mechanism by which this process is regulated is not fully understood,

but the SSU processome contains several domains that could recruit the RNA degradation machinery, providing some insights into how the exosome could be recruited.

Just like the large subunit assembly factor Nop53, UtpB-complex subunit Utp18 contains an Arch Interacting Motif (AIM), which can bind the noncatalytic domain of the exosome-associated helicase Mtr4 (Thoms et al. 2015) (Fig. 5.4E). This helicase binds the exosome and has a role in unwinding RNA elements and feeding these to the catalytic sites of the nuclear exosome (De la Cruz et al. 1998). Furthermore, two factors of the SSU processome, Sas10 and Lcp5, contain domains that have high homologies to the exosome cofactor Rrp47 (Costello et al. 2011) (Fig. 5.4E). However, unlike in the large subunit maturation pathway where recruitment of the exosome is used for precise processing of the 5.8S rRNA, here the recruitment is thought to lead to the degradation of the entire 5’ ETS.

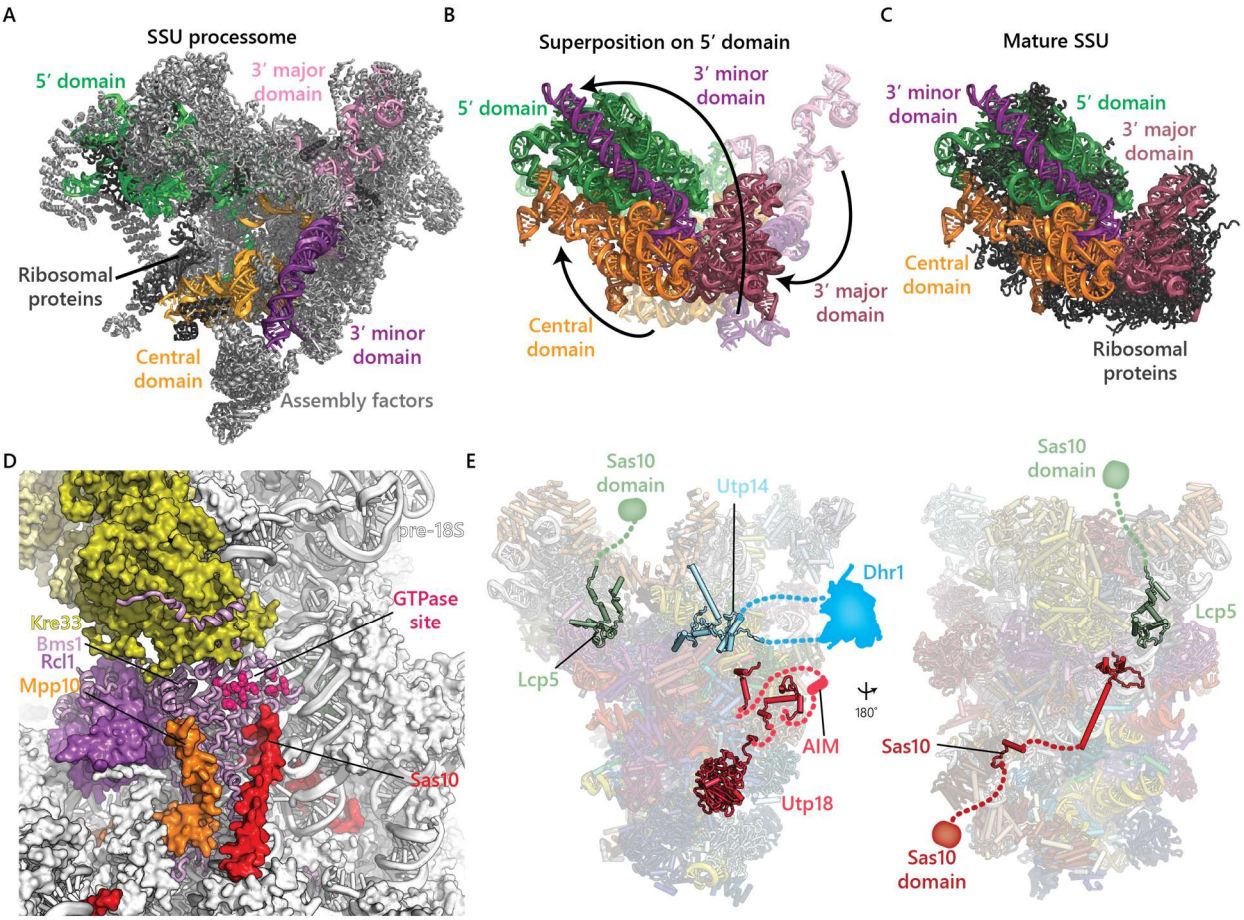


Figure 5.4 Maturation of the SSU processome into a pre-40S particle. (**A-C**) Comparative views of the domains of the 18S rRNA in the SSU processome (PDB: 5WLC, 5WYJ) and the mature small subunit (PDB: 4V88). Domains have to undergo large rotational movement in order to adopt their mature conformation as shown in (**B**). In (**B**), 18S domains in their SSU processome conformation are transparent. (**D**) View of Bms1 and its exposed GTPase site in the context of the SSU processome. (**E**) Cartoon representation of Utp18, Sas10, Lcp5 and Utp14 in the SSU processome. Utp18, Sas10 and Lcp5 are involved in recruiting the RNA degradation machinery through their AIM motif or the Sas10 domains respectively. Utp14 is bound to Dhr1, which is flexibly attached to the particle.

5.3 Maturation of the large subunit processome

Our study of large subunit processome assembly raises multiple questions concerning its further maturation. Our data suggests that 5’ and 3’ processing is required for further maturation of the particle into a nucleolar pre-60S particle (Fig. 4.5B, C), and that this process requires more than cleavage at the 3’ end, suggesting another step may be required for maturation of the large subunit processome.

The structure of the nucleolar pre-60S has shown that it is composed of a stable core composed of the 5.8S, ITS2 and domains I and II of the 25S, bound by several ribosome assembly factors as well as ribosomal proteins adopting their mature positions (Kater et al. 2017; Sanghai et al. 2018). A significant portion of the particles also have domain VI in its mature position. This conformation is mutually exclusive with the presence of the 3’ ETS or ITS1, which would clash in the observed conformation of the particle.

Similar to the small subunit processome, the transition from co-transcriptional assemblies of the large subunit to a *bonafide* pre-60S is still poorly understood. The presence in the earliest large subunit precursors of Rrp5, Noc1 and Noc2, which are absent from the nucleolar pre-60S, makes it tempting to speculate that 5’ and 3’ processing may be coupled to their removal. Early large subunit intermediates recruit several ATPases, including the DEAD-box proteins Dbp9, Has1, Drs1 and Mak5, all of which are essential. Their role in large subunit assembly is still uncharacterized but they could be involved in remodeling the early large subunit particles.

Lastly, structures of the nucleolar and nuclear pre-60S particles have shown that Erb1 and Nop53 employ a similar motif to bind to the surface of Nop7 (Wu et al. 2016; Kater et al. 2017; Sanghai et al. 2018). This is thought to prevent premature recruitment of the Mtr4-exosome by Nop53, which processes ITS2 in the nucleus (Thoms et al. 2015; Kater et al. 2017; Sanghai et al. 2018). Removal of Erb1 should be performed by the AAA+ protein Mdn1, which acts on its interactor Ytm1 through its ubiquitin-like fold (Baßler et al. 2010). It is currently unknown however how Mdn1 is prevented from acting on its substrate prematurely. Ytm1 is recruited to the pre-rRNA in the earliest steps of large subunit pre-rRNA transcription. Mdn1 was shown to be a SUMO-directed enzyme in mammalian cells (Raman et al. 2016). High levels of SUMOylation in early subunit precursors (Fig. 4.2C) might contribute to the recruitment of Mdn1 to these particles for their remodeling. It will be interesting to determine how Mdn1 engages the nucleolar pre-60S to understand this step in the early maturation of the large subunit.

5.4 Evolutionary origins factors of ribosome biogenesis

The structure of the small subunit processome has provided us with insights of the evolution of ribosome assembly factors. Ribosome assembly intermediates represent another example of eukaryotic superstructures, which have been described as large protein assembly specific to eukaryotic cells (Makarova et al. 2005). These include the nuclear pore complex (NPC), pre-initiation complexes of eukaryotic RNA polymerase, and the spliceosome, among others. The small subunit processome, like many of these structures, makes extensive use of similar folds, particularly β-propellers and helical repeats such as HEAT repeats, to perform an impressive diversity of functions. The

large multi-subunit complexes UtpA and UtpB represent a good example of this, using their combined 17 β-propeller domains to perform various stabilizing functions in the SSU processome, including mediating protein-protein interactions and stabilizing RNA elements of the SSU processome (Fig. 3.14, 3.15). Our discovery of structural and sequence homology between helical repeats present in both UtpA and UtpB points to their common origin (Fig. 3.5, 5.5A, B)

Structures of the early intermediates of ribosome biogenesis have also revealed that several ribosome assembly factors mimic components of the translation machinery to perform chaperoning functions. One example is the presence of Sas10 in the binding site of rpS30. While, Sas10 is multi-modular protein, its C-terminus mimics a loop-helix turn of rpS30 to bind helix 17 of the 5' domain of the 18S.

Ribosome assembly factors also use similar folds to proteins involved in translation to perform new functions. For instance, two factors that are necessary for both small and large subunit processing are homologs of the ribosomal protein Rpl1, which is conserved from bacteria to higher eukaryotes. In the bacteria *Thermus Thermophilus*, Rpl1 binds the 23S rRNA, in the large subunit. From this ancestral protein, paralogs have appeared in eukaryotes and use a similar binding surface to assist ribosome biogenesis (Fig. 5.5C-E). Utp30 and Cic1 use a similar fold to stabilize h41 of the pre-18S and ITS2 respectively. Interestingly, these two proteins have also developed a new binding surface, which is used to interact with Rrt14 in the case of Utp30, and with another region of ITS2, in Cic1. The peptide of Rrt14, which contacts Utp30, has a Pro-X-Pro-X-Pro motif which may be mimicking the ribose groups of RNA (Fig. 5.5C).

Our structure also revealed the architecture of the GTPase Bms1, which shares homology with translational GTPases and with another small subunit biogenesis factor, Tsr1 (McCaughan et al. 2016). This represents another example of proteins involved in normal translation function, such as eEF1A, being functionalized for the assembly of the ribosome (Fig. 5.5F-H). While a different class of GTPase, there are also similarities between the architectures of Nog1 and Bms1 (Wu et al. 2016), with a core GTPase domain flanked by linker regions anchoring the proteins into their respective particles.

More broadly, studies of the assembly of the small and large subunit processome have revealed somewhat different mechanisms in the biogenesis of the small and large ribosomal subunits. Small subunit biogenesis involves from its start the extensive coordination of all domains of the 18S, mostly organized by the 5' ETS and its associated factors. This provides a rigid scaffold on which domains of the 18S can be individually matured. Later stages in maturation will involve the controlled removal of this scaffold and the formation of a pseudo-mature state for the rRNA.

Conversely, structures of pre-60S intermediates and the assembly of the large subunit processome have suggested that large subunit biogenesis is more modular and does not involve a similarly central scaffold. While a similar number of factors are involved, large subunit assembly factors bind and dissociate the pre-60S and multiple waves, defining many distinct sites. Small subunit biogenesis involves the coordinated actions of most factors in its early biogenesis, after which only few new factors participate. This difference between large and small subunit maturations might reflect the greater size of the large subunit, which may not allow for the presence of several folds in mass of assembly factor bound to the precursors of the large subunit.

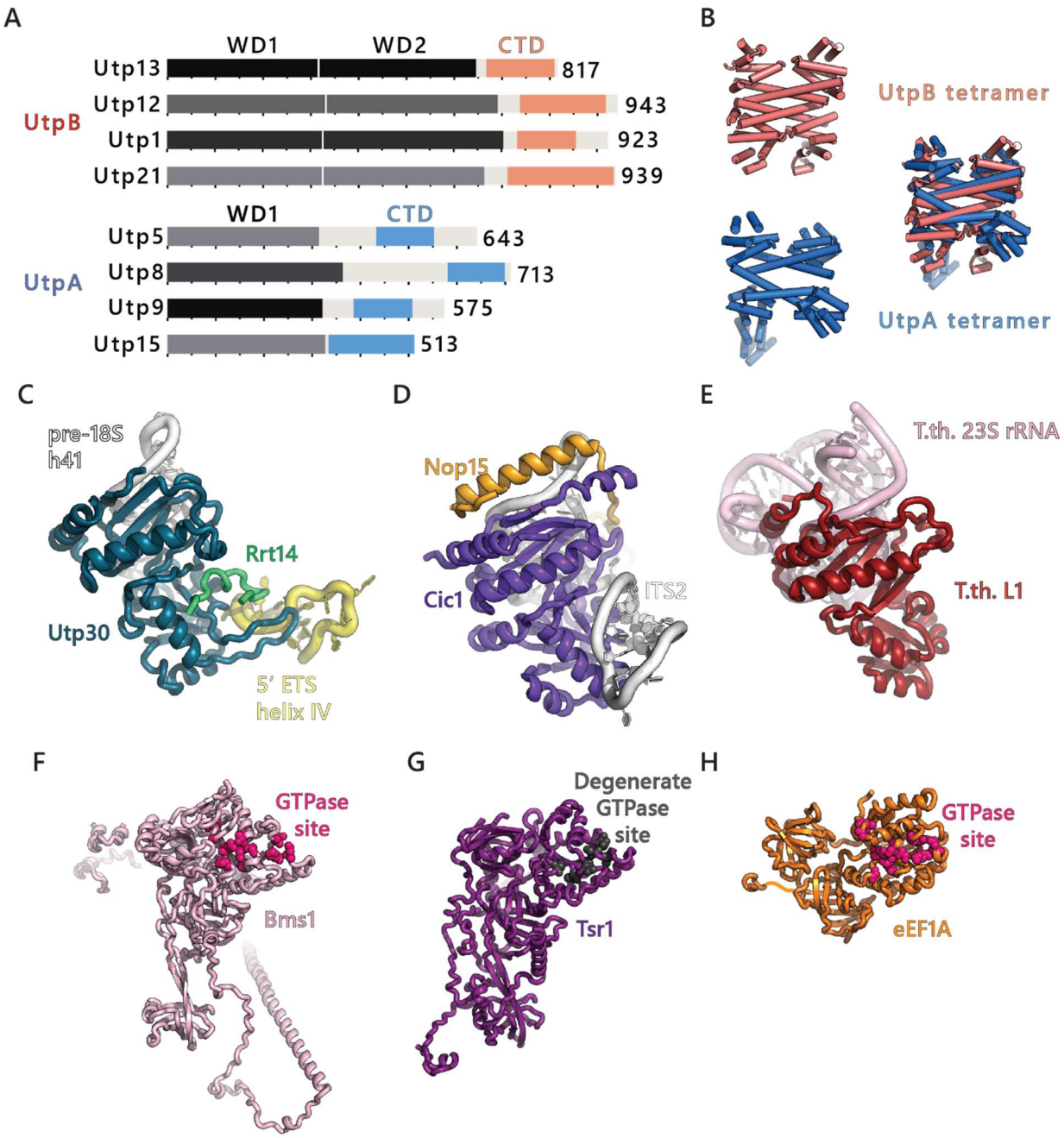


Figure 5.5 Homology between ribosome assembly factors. (**A**) Secondary structure diagram of the UtpB and UtpA tetramers, including the homologous C-terminal repeats (**B**) Cartoon representation of these homologous C-terminal helical bundles and superposition of their structure (**C**) Cartoon representation of Utp30 bound to the pre-18S and Rrt14 within the SSU processome (**D**) Cartoon representation of Cic1 bound to Nop15 and ITS2 within the Nog2 particle (**E**) Cartoon representation of bacteria Rpl1 bound to the 23S rRNA. (**F-H**) Structure of Bms1 (PDB 5WLC), Tsr1 (PDB 6FAI) and eEF1A (PDB 5O8W) with their GTPase site highlighted with either pink (active) or grey (inactive) spheres.

5.5 Emerging regulatory mechanisms of ribosome assembly

Translation and ribosome assembly have been long-known targets of the TOR complex, but recent insights have shown the extent to which all parts of the assembly are targeted by this pathway. With the goal of purifying small subunit processome particles, we have uncovered a mechanism by which small subunit biogenesis is halted under nutrient stress. This is consistent with similar data from the study of total pre-rRNA under various stress conditions, including nutrient starvation, heat and cold shock, or rapamycin treatment (Talkish et al. 2016; Kos-Braun et al. 2017). The accumulation of a nucleolar particle under stress conditions was also described in the large subunit pathway (Sanghai et al. 2018). Accumulation of these particles can be linked with the accumulation of 23 and 27S species that have been observed under similar conditions. The accumulation of such species was shown to not only be dependent on the TOR pathway, but also CK2, which happens to interact with the SSU processome factor UtpC to regulate ribosomal protein gene expression (Fig. 5.6).

The accumulation of nucleolar particles in ribosome biogenesis represents a new pathway by which TOR regulates ribosome assembly. However, the mechanism by which this process is modulated is still unclear. Sch9, a known downstream effector of TOR has been tied to ribosome biogenesis but the halting of ribosome assembly can still occur in its absence (Kos-Braun et al. 2017). It is tempting to speculate that the presence or absence of post-translational modifications on ribosome assembly factors may allow or prevent the normal maturation of the pre-ribosomal particles. No such factor or modification has been linked with this phenomenon.

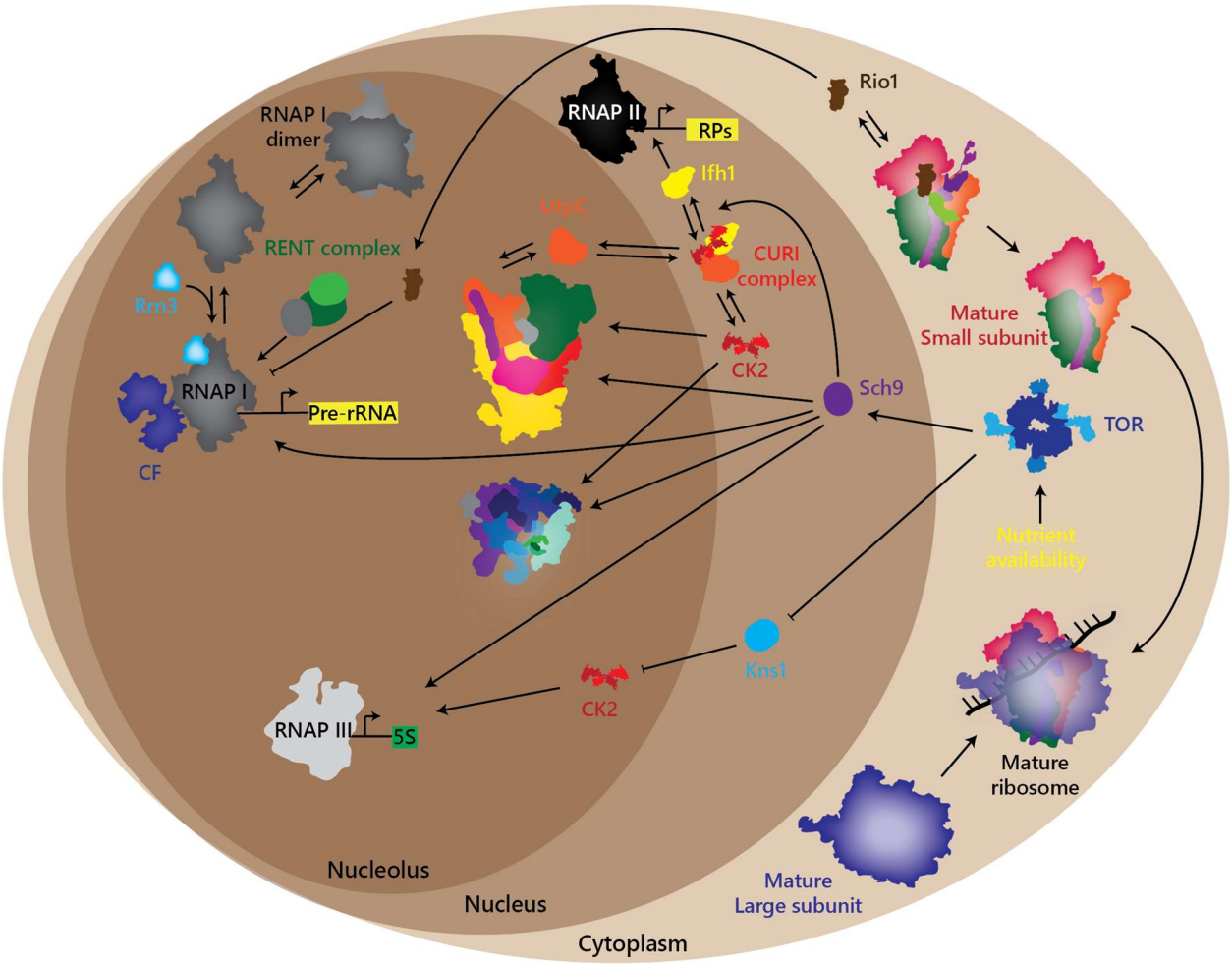


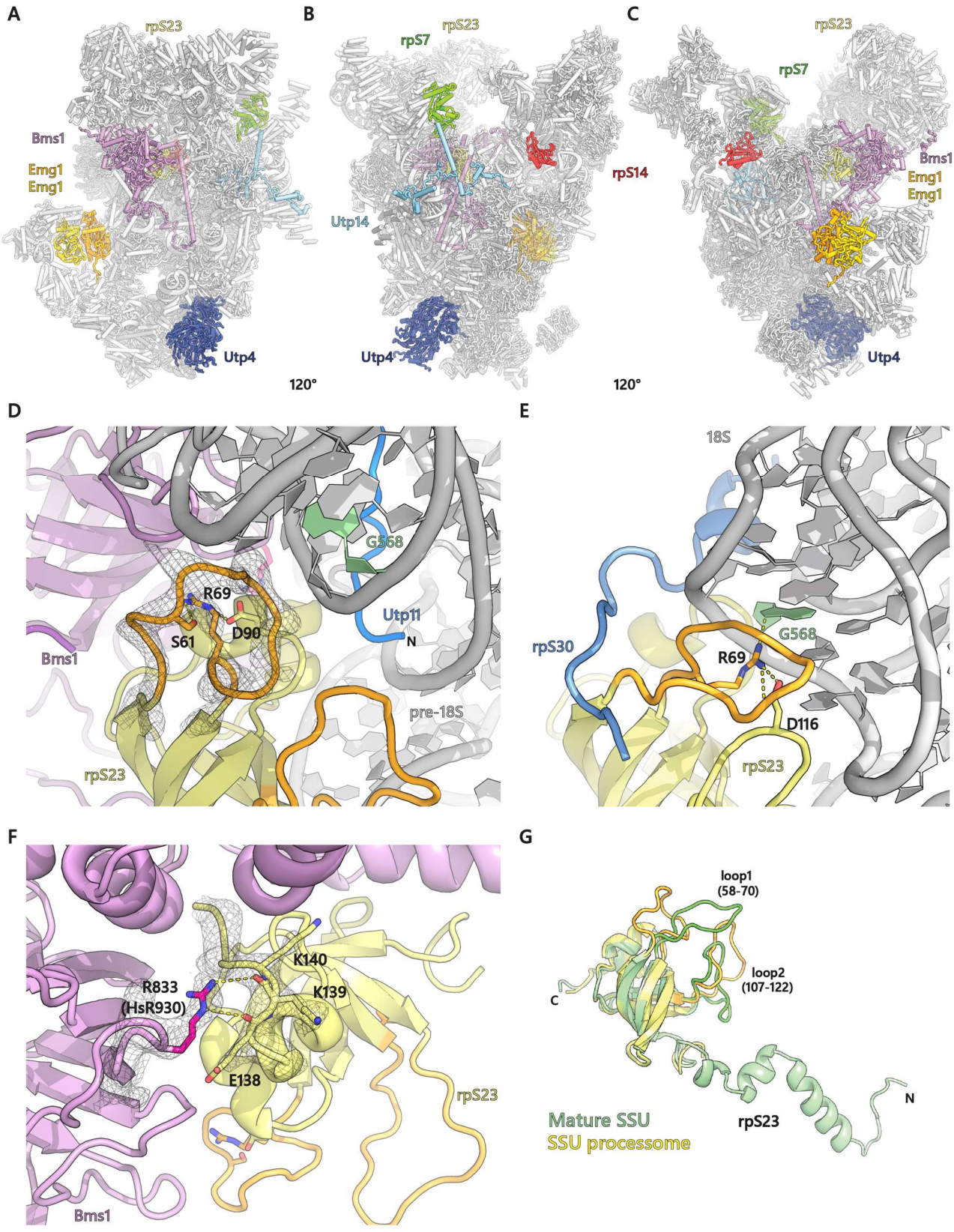
Figure 5.6 Updated view of regulation of ribosome biogenesis by the TOR and CK2 pathways. TOR is the master regulator of ribosome biogenesis. TOR affects Kns1 and Sch9, which regulate RNAP III transcription, RNAP II transcription of ribosomal protein genes, and RNAP I transcription of pre-rRNA. RNAP I can be silenced by dimerization which prevents Rrn3 association. Sch9/CK2 also regulate the further maturation of the SSU processome and the LSU processome.

5.6 Implications on human diseases

The diversity of diseases caused by defects in ribosome assembly has provided an additional need for a better mechanistic understanding of ribosome biogenesis. (Sondalle and Baserga 2014; Robson et al. 2016). The structure of the small subunit processome has now allowed us to map some mutations tied to specific ribosompathies in a structural context (Fig. 5.8A-C). In particular, two residues near the Bms1-rpS23 interface have been implicated in human diseases. While one residue is positioned in a dynamic loop of rpS23 (Paolini et al. 2017) (Fig. 5.8D, E,G), the other is located within Bms1 (Marneros 2013) and interacts with rpS23 (Fig. 5.8F). It is tempting to speculate that mutations in either of these residues may affect the structural transitions required for the resulting incorporation of rpS23 into the mature small ribosomal subunit.

A recent study has revealed the mechanism of pathogenesis for another related ribosomopathy, Diamond-Blackfan anemia, which is characterized by low white-blood cell counts (Khajuria et al. 2018). Reduced levels of ribosome production did not result in different ribosomes but lead to an overall decrease in translation levels. This affected certain transcripts preponderantly, including a transcription factor associated with the hematopoietic differentiation program, which resulted in the anemia. This theory however fails to explain why mutations in other ribosome assembly factors, or components of the ribosome, lead to such a wide variety of syndromes. Future work will need to detail how mutations causing other ribosomopathies lead to such different symptoms.

Figure 5.7 Structural context of proteins implicated in ribosomopathies. (**A-C**) Three views of the SSU processome (grey), related by 120-degree rotations. Proteins associated with ribosomopathies are highlighted in colors as follows: Emg1 (yellow, orange), Utp4 (blue), rpS23 (pale-yellow), rpS14 (red), rpS7 (green), Bms1 (violet), and Utp14 (cyan). (**D and E**) Comparative views of rpS23 (pale-yellow) in the context of the SSU processome (D) and the mature small subunit (E). (D) View of rpS23 (pale-yellow) in the context of the SSU processome. Utp11 (blue), Bms1 (violet) and pre-18S rRNA (light-grey) are shown. A loop of rpS23 (residues 58-70) is shown in orange. Arginine 69 (R69) of rpS23 and its interaction partners are depicted as sticks, G568 of pre-18S rRNA is highlighted in green. (**E**) View of rpS23 in the context of the mature small ribosomal subunit (PDB 4V88). RpS23 (pale-yellow), rpS30 (blue) and 18S rRNA (light-grey) are shown. A loop of rpS23 (residues 58-70) is colored orange. Arginine 69 (R69) and Aspartate 116 (D116) are depicted as sticks and nucleotide G568 is highlighted in green. (**F**) Structural context of Bms1 arginine 833 (human R930) in the SSU processome. Bms1 (violet) and rpS23 (pale-yellow) are shown. Arginine 833 and its interacting residues are shown as sticks. (**G**) Superposition of rpS23 in the context of the mature small ribosomal subunit (pale-green) and the SSU processome (pale-yellow). Loops that adopt different conformations (loop1 and loop2) are highlighted in darker shades.


A
rpS23
Bms1
Emg1
Emg1
Utp4
120°
B
rpS23
rpS7
rpS14
Utp14
Utp4
C
rpS23
rpS7
Bms1
Emg1
Emg1
Utp4
120°
D
G568
R69
Utp11
N
S61
D90
Bms1
pre-18S
rpS23
E
18S
rpS30
G568
R69
D116
rpS23
F
K140
R833
(HsR930)
K139
E138
rpS23
Bms1
G
loop1
(58-70)
loop2
(107-122)
C
N
rpS23
Mature SSU
SSU processome

5.7 Summary and outlook

While the factors needed for ribosome biogenesis have been known for decades, there had been a lack of knowledge on their function, and the timing of their involvement in ribosome assembly. This body of work, and that of others, has elucidated the association of more than 120 factors to the pre-rRNA. Combined with structures of nucleolar intermediates for both small and large subunit assembly, it offers a clearer view of early ribosome biogenesis. However, much remains to be learned.

Several factors, which are transiently associated with the nascent pre-rRNA, are not part of the later particle which have been characterized structurally. Furthermore, cryo-EM, which has been used to study all intermediates of ribosome assembly, is an ensemble approach, which leaves several flexible regions unresolved in the reconstruction. For instance, more than one megadalton of ribosome assembly factors could not be seen in the structures of the SSU processome. Other approaches will have to be used to characterize the more dynamic steps of the early assembly of nucleolar pre-ribosomal particles.

The wealth of structural information available for later intermediates of ribosome biogenesis, particularly in large subunit assembly, combined with genetic and biochemical data, presents us with a clear picture of the last cytoplasmic steps in the formation of the ribosome. Conversely, information concerning the transitions between nucleolar and nuclear particles is still lacking. The concerted action of several enzymes, RNA helicases, ATPases and GTPases, is required for this transition. The study of the activity of these enzymes will likely provide us with some clues as to the mechanisms that underlie later nucleolar maturation.

Finally, new modes of regulation of ribosome assembly have surfaced from recent results, including ours. The master regulator Tor and the kinase CK2 have been shown to be the main signaling controllers in this pathway but the mechanism in still unclear. Enzymes involved in the transition between nucleolar and later particles seem like ideal targets to halt or resume ribosome assembly, but extensive characterization of downstream targets of Tor will have to be done to elucidate the mechanism of this process. Given the importance of ribosome biogenesis in overall cellular health, and its involvement in human diseases, including cancer, a clearer mechanistic perspective of ribosome assembly and its regulation will be essential.

Chapter VI - Methods

6.1 Molecular cloning

All molecular cloning was done using either DH5α or Stbl3 cells. The full list of constructs used for the study of early ribosome biogenesis can be found in Table 6.1

Cloning of RNA sequences

Ribosomal RNA sequences were amplified directly for BY4741 yeast genomic DNA. Sequences corresponding to all constructs used in Chapter II and IV from *Saccharomyces cerevisiae* rDNA were cloned downstream of the Gal1 promoter of a derivative of pESC URA (Agilent Technologies) containing five copies of the MS2 aptamer tag at the 5' or 3´ end. Constructs containing sequences of the 25S carry a northern blot tag to track specifically the fate of those transcripts (Beltrame and Tollervey 1992). For some constructs, the hatchet ribozyme is located at the 3' end of the RNA (Li et al. 2015).

Cloning of FLAG-H14-sbp-3c-MS2

A mutant of the MS2 phage coat protein, which lacks two residues required for oligomerization (LeCuyer et al. 1995) was cloned into a derivative of pRSFDuet1 containing an N-terminal FLAG-H14-sbp tag followed by a 3C protease site.

Table 6.1 Plasmids used for this body of work. Antibiotic resistance for bacteria or yeast (if applicable) carried by each plasmid are indicated.

PLASMID NAME	PROMOTER - GENE OF INTEREST	BACT. RESIS.	YEAST RESIS.	SOURCE
PRSD70	pRSFDUET_T7::FLAG-H14-sbp-3c-MS2	Kan	-	Chapter II
PSKD3	pRSFDUET_T7::pelB-3K1K-H10	Kan	-	All
PSKD19	pRSFDUET_T7:: H14-SUMO-SAVSBPM18	Kan	-	All
PSKB186	pRDN-hygl	Amp	URA/LEU	Chernoff et al. 1994
PSKB187	pRDN4	Amp	LEU/TRP	Chernoff et al. 1994
PSKB202	pMS2x5-RDN4	Amp	LEU/TRP	Chernoff et al. 1994
PSKB42	pESC_Gal1::MS2x5	Amp	URA	Chapter II/IV
PSKB78	pESC_Gal1::MS2x5_5'ETS	Amp	URA	Chapter II
PSKB117	pESC_Gal1:: 5'ETS_ MS2x5	Amp	URA	Chapter II
PSKB221	pESC_Gal1:: 5'ETS-5' domain_ MS2x5	Amp	URA	Chapter II
PSKB222	pESC_Gal1:: 5'ETS-central domain_ MS2x5	Amp	URA	Chapter II
PSKB223	pESC_Gal1:: 5'ETS-3' major domain_ MS2x5	Amp	URA	Chapter II
PSKB224	pESC_Gal1:: 5'ETS-3' minor domain_ MS2x5	Amp	URA	Chapter II
PSKB259	pESC_Gal1:: 5'ETS-ITS1_ MS2x5	Amp	URA	Chapter II
PSKB347	pESC_Gal1:: A3-domain I_25Stag_ MS2x5	Amp	URA	Chapter IV
PSKB355	pESC_Gal1:: A3-domain II_ 25Stag _MS2x5	Amp	URA	Chapter IV
PSKB356	pESC_Gal1:: A3-domain III_ 25Stag _MS2x5	Amp	URA	Chapter IV
PSKB357	pESC_Gal1:: A3-domain IV_25Stag _MS2x5	Amp	URA	Chapter IV
PSKB358	pESC_Gal1:: A3-domain V_25Stag _MS2x5	Amp	URA	Chapter IV
PSKB349	pESC_Gal1:: A3-domain VI_25Stag _MS2x5	Amp	URA	Chapter IV
PSKB419	pESC_Gal1:: A3-B0_25Stag _MS2x5	Amp	URA	Chapter IV
PSKB364	pESC_Gal1:: A3-FL_25Stag _MS2x5	Amp	URA	Chapter IV
PSKB286	pESC_Gal1:: D-domain I_25Stag _MS2x5	Amp	URA	Chapter IV
PSKB346	pESC_Gal1:: A2-domain I_25Stag _MS2x5	Amp	URA	Chapter IV
PSKB471	pESC_Gal1:: MS2x5_A3-FL_25Stag	Amp	URA	Chapter IV
PSKB472	pESC_Gal1:: A3-domain VI_25Stag_hatchet _MS2x5	Amp	URA	Chapter IV
PSKB473	pESC_Gal1:: A3-B0_25Stag _hatchet_MS2x5	Amp	URA	Chapter IV
PSKB474	pESC_Gal1:: A3-FL_25Stag _hatchet_MS2x5	Amp	URA	Chapter IV
PSKB482	pESC_Gal1:: 5.8S-FL_25Stag _ MS2x5	Amp	URA	Chapter IV

6.2 Recombinant protein purification and affinity resin generation

Expression and purification of FLAG-H14-sbp-3c-MS2

RIL cells carrying pRSFDuet1-FLAG-H14-sbp-3c-MS2 were grown at 37 °C in 2xYT with 30 µg/ml kanamycin to OD600 of 2.0. Overexpression of the fusion protein was induced with 1 mM IPTG for 20 hours at 20 °C. Cells were pelleted and resuspended in lysis buffer (20 mM Tris/HCl, pH 7.7, 500 mM NaCl, 0.1 % Triton) and lysed using a cell disrupter at 30 kPsi and the lysate was cleared by centrifugation at 40'000xg for 30 minutes. The cleared lysate was incubated with IMAC affinity resin for 1 hr. The resin was washed with 60 column volumes of lysis buffer and 40 column volumes of washing buffer (20 mM Tris/HCl, pH 7.7, 500 mM NaCl). The sbp-MS2 fusion was eluted with 2 column volumes of elution buffer (20 mM Tris/HCl, pH 7.7, 500 mM NaCl, 300 mM imidazole) and dialyzed overnight at 4 °C against 20 mM Tris/HCl, pH 7.7, 500 mM NaCl, 1 mM EDTA, 10 % glycerol. The protein was aliquoted, frozen in liquid nitrogen, and stored at -80 °C until further use.

Expression and purification of anti-GFP nanobody

RIL cells carrying pSKD3 (pRSFDUET-pelB-3K1K-H10) were grown at 37 °C in 2xYT with 30 µg/ml kanamycin to OD600 of 2.0. Overexpression of the fusion protein was induced with 1 mM IPTG for 20 hours at 28 °C. Cells were pelleted and resuspended in lysis buffer (20 mM HEPES, pH 7.7, 500 mM NaCl, 0.1 % Triton) supplemented with 1X Pepstatin, 1X PMSF, 1X E-64, and DNase. Cells were subsequently lysed using a cell disrupter at 30 kPsi and the lysate was cleared by centrifugation at 40,000 x g for 30

minutes. The cleared lysate was supplemented with 30 mM imidazole and passed twice over a 5 mL IMAC His-Trap column. The column was washed with 30 mL of lysis buffer supplemented with 60 mM imidazole. The nanobody was eluted with 20 mL of lysis buffer supplemented with 300 mM imidazole. The 20 mL elution was then concentrated to 5 mL with a 10 kDa M.W cutoff concentrator (Pall), and injected over a HiLoad Superdex 75 gel filtration column that has been previously equilibrated in 1X PBS supplemented with 500 mM NaCl. The soluble fraction of the protein eluted at a retention volume of 75 mL (total column volume is 125 mL). The protein peak was finally concentrated to 5 mg/mL for resin coupling.

Expression and purification of Streptavidin variant: SAVSBPM18

RIL cells carrying pSKD19 (pRSFDUET-H14-SUMO-SAVSBPM18) were grown at 37 °C in 2xYT with 30 µg/ml kanamycin to OD600 of 2.0. Overexpression of the fusion protein was induced with 1 mM IPTG for 20 hours at 28 °C. Cells were pelleted and resuspended in lysis buffer (20 mM HEPES, pH 7.7, 500 mM NaCl, 0.1 % Triton) supplemented with 1X Pepstatin, 1X PMSF, 1X E-64, and DNase. Cells were subsequently lysed using a cell disrupter at 30 kPsi and the lysate was cleared by centrifugation at 40,000 x g for 30 minutes. The cleared lysate was supplemented with 30 mM imidazole and passed twice over a 5 mL IMAC His-Trap column. The column was washed with 30 mL of lysis buffer supplemented with 60 mM Imidazole. The nanobody was eluted with 20 mL of lysis buffer supplemented with 300 mM imidazole. The 20 mL elution was then concentrated to 5 mL with a 30 kDa M.W cutoff concentrator (Pall), and injected over a HiLoad Superdex 200 gel filtration column that

had been equilibrated in 1X PBS supplemented with 500 mM NaCl. The soluble fraction of the protein eluted at a retention volume of 70 mL (total column volume is 125 mL). The protein peak was finally concentrated to 10 mg/mL for resin coupling.

Coupling of nanobody or streptavidin to NHS-sepharose resin

To couple any protein, nanobody or streptavidin, to NHS-sepharose resin, the same protocol was used, only adjusting the desired concentration of protein. One volume of NHS-sepharose slurry was poured in a 50 mL conical tube. This usually corresponded to 75-80% beads and isopropanol. The slurry was spun down 1 min at 1000 rpm on a table centrifuge. The isopropanol was discarded and the resin was subsequently activated by 1 volume (relative to the beads) of freshly prepared, cold 1 mM HCl. The resin was resuspended in the acid and spun down again 1 min at 1000 rpm. The resin was then washed with 1 volume of 1X PBS supplemented with 500 mM NaCl and spun down one last time. After discarding the buffer, one half volume of protein was added to the resin at twice the target concentration (e.g. 10 mL of resin is incubated with 5 mL of nanobdy at a concentration of 5 mg/mL for a final concentration of 2.5 mg/mL of resin). The sepharose was then incubated with the protein overnight at 4 °C. The coupling reaction is quenched the next morning by spinning down the resin, removing the liquid, adding 1 volume of quenching buffer (100 mM Tris, 500 mM NaCl) and incubating 1-3 hours at 4 C. The sepharose was then washed 10 times with 4 volumes of washing buffer (100 mM Tris pH 7.7, 500 mM NaCl, 0.1% Triton X-100) and 3 times with 4 volumes of quenching buffer. The beads are stored in a 1:1 slurry of beads and quenching buffer at 4 °C in 2 mL screw cap tubes.

6.3 Yeast strains and yeast genetic methods

All strains used in this body of work, unless specified otherwise, are derivatives of BY4741. The full list of strains used can be found in Table 6.2.

Genetic transformation of yeast cells

Transformation of genetic material in yeast was done with the following procedure. Yeast cells were grown in 10 mL of YPD supplemented with ampicillin per transformation reaction. These cultures were inoculated to an OD of approximately 0.1 and grown 6-8 hours to an OD of 0.5, at which point they were spun down at 2000 x g for 2 minutes. The yeast was washed with water twice and once in 1X TE plus 0.1M LiOAc. The transformation-competent yeast was finally resuspended in 50 µL of 1X TE plus 0.1M LiOAc per transformation reaction. Transforming DNA, either for plasmid transformation, C-terminal gene tagging or gene integration, and 50 µg of salmon sperm DNA was added to the 50 µL of yeast. The transformation reaction was supplemented with 300 µL of 40% PEG + 1X TE + 0.1M LiOAC and incubated 30 min at 30°C and heat-shocked 15 min at 42 °C. Plasmid transformation reactions were then directly plated on the appropriate selection plate, including antibiotic resistance. Genomic integration or C-terminal tagging reactions were recovered, regardless of selection marker, 2 to 4 hours at 30 °C in YPD with ampicillin before being spun down, resuspended in 1X TE and plated on the appropriate selection plate.

For plasmid transformations, the yeast was simply incubated with approximately 1 µg of the pure plasmid.

For C-terminal gene tagging, the yeast was incubated with a PCR amplicon of the tag and selection marker flanked by, on the 5' end, the last 50 bp of the gene (excluding the STOP codon) and, on the 3' end, 50 bp of sequences approximately 100 bp downstream of the gene. 100-200 µL of PCR reaction was purified using the reaction clean-up kit, and used wholly for transformation. Primers used for genetic tagging of all strains used in this body work are listed in table 6.3.

For genetic integration, the integrant was cloned into a vector containing 2 sequences of a non-essential genetic locus (Pep4 or Prb1). Between those sites, the gene of interest was cloned under the relevant promoter. A selection marker is also present between the 2 homology regions. 5-10 µg of this vector was then cut using restriction sites flanking the homology regions, cleaned by gel band extraction or using the reaction clean-up kit, and used wholly for transformation.

To test the viability of rRNA modifications, the strain L-1491, containing only one copy of rDNA located on a plasmid, was used (Chernoff et al. 1994). Five copies of the MS2 aptamer were inserted at the 5´ end of rDNA in plasmid pRDN4 to generate pMS2x5-RDN4. Plasmid shuffling was used to generate strain YSK145 from strain L-1491 and pMS2x5-RDN4.

Table 6.2 List of yeast strains used in this body of work.

YSK NAME	ALT NAME	GENOTYPE	SOURCE
YSK77	Utp10-T3	MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 UTP10-3myc-tev-GFP-3FLAG::HphMX4	Chapter II
YSK76	Utp1-T3	MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 UTP1-3myc-tev-GFP-3FLAG::HphMX4	Chapter II
YSK89	rrp9-T3	MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 RRP9-3myc-tev-GFP-3FLAG::HphMX4	Chapter II
YSK111	L-1491	MATa adel-14 his7-1 leu2-3, lys2-L864 trpl-A1 ura3-52 bearing *pRDN-hygl [URA3 LEU2-d rdn-hyg1 rdn-ani1]*	Chernoff et al. 1994
YSK114	L-1496	MATa adel-14 his7-1 leu2-3, lys2-L864 trpl-A1 ura3-52 bearing *pRDN4 [TRP1 LEU2-d rdn4]*	Chernoff et al. 1994
YSK145	L-1496 pMS2	MATa adel-14 his7-1 leu2-3, lys2-L864 trpl-A1 ura3-52 bearing *pMS2x5-RDN4 [TRP1 LEU2-d rdn4]*	Chapter II
YSK232	UTP1-T3/ Kre33-sbp	MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 UTP1-3myc-TEV-GFP-3FLAG::HphMX4 Kre33-linker-sbp::kanMX6	Chapter III
YSK187	MS2-3c-GFP Cic1-sbp	MATa his3Δ leu2Δ0 met15Δ0 ura3Δ0 pep4::kanMX6-iMultiyeast-NLS-HA-MS2-3c-GFP Cic1-linker-sbp::kanMX6	Chapter IV

Table 6.3 List of primers used for genetic tagging.

PRIMER NUMBER	PRIMER NAME	SEQUENCE
489	Utp10 F	TCAAGGTTGTTGAAAACGTTTTAGGGGAACCTTTTGATAGGTATTTAGATACGCTGCAGGTCGACGGATCC
511	Utp10 R	CTATCACGAATGAATAACAATATTTTCAATGAAAGATTTCGTTCAACAGCGGCAGATCCGCGGCCGCATAGG
391	Utp1 F	CTAATGAAAATGATTCCAGTGATGAAGAAGAAAATGAGAAAGAGCTTCCTACGCTGCAGGTCGACGGATCC
392	Utp1 R	TTTGTTGATATTAAATACAAATGTAATGCTTTATATAGTTCTATACTTGGGGCAGATCCGCGGCCGCATAGG
494	Rrp9 F	CAGGTGCCAGAAATGGTATATATTCAGCTGTCATTGACCAAACAGGCTTTACGCTGCAGGTCGACGGATCC
495	Rrp9 R	TTATTTCTATATACTTGGGAGCCGGAGCCGCTTTGCGTTCTGAGAGAGTGGGCAGATCCGCGGCCGCATAGG
1069	Kre33 F	AAGAGATGAAAGCTATGAAAAAACCAAGAAAGTCTAAAAAGGCTGCAAATACGCTGCAGGTCGACGGATCC
1070	Kre33 R	TGTAAAGGTTCAAACATCAACTATGTTTCTATTCTATATTATTGTACAAAGGCAGATCCGCGGCCGCATAGG
1244	Cic1 F	AGCTTGAAAAAGAATCTAGCGAGTCAGAAGCTGTCAAGAAGGCTAAAAGTACGCTGCAGGTCGACGGATCC
1245	Cic1 R	TATATAGATATATACATATGCACCGCACTCTATGAAATTCAAATTTTTTTGGCAGATCCGCGGCCGCATAGG

6.4 Expression and purification of yeast ribosome biogenesis complexes

Expression of truncated pre-rRNA fragment

Partially reconstituted small subuit processome intermediates were purified from yeast strains bearing one GFP tag on either Utp10, Utp1 or Rrp9. These strains were grown to saturation in YPD from glycerol stock and inoculated at an OD of 0.1 for plasmid-based transformation (see above). Plasmid used were based on the pESC_URA system and contained pre-ribosomal RNA fragments starting at the 5' end of the 5' ETS and stopping at several sites in the first half of the pre-ribosomal RNA. These RNAs also contained either 5' or 3' tags, consisting of 5 repeats of the MS2 aptamer. The transcript was placed under a galactose promoter and upstream of the CYC terminator.

After 3-4 days, or when yeast colonies were approximately 1 mm in diameter or more, several yeast colonies were picked and grown in fully synthetic medium containing 2% glucose, 100 µg/mL of ampicillin, no URA (but supplemented with LEU, HIS, and TRP) and 400 µg/mL hygromycin B, to saturation. Small cultures were subsequently used to inoculate 100-300 mL of fully synthetic medium containing 2% raffinose, ampicillin, no URA and 200 µg/mL hygromycin B. These 100-300 mL cultures were grown to saturation at 30 °C and were used to inoculate 1.5-6L of the same medium. Large cultures were grown to approximately OD=1.0 (between OD 0.5 and 1.5) and 2% galactose were added to them. Cultures were left 16 hrs at 30 °C to induce after which they were harvested. Cultures were spun down at 4000 x g for 4 min, then washed with cold water, spun again, and washed one last time with 1 volume (approx. 25 mL) of cold water supplemented with 1X Pepstatin, 1X PMSF and 1X E-64. Pellets were then frozen as "noodles" and stored at -80 °C until cryo-milling.

For large subunit processome intermediates, a different approach was used. Plasmids were transformed in a strain baring a C-terminal sbp-tag on Cic1 and expressing NLS-MS2-3c-GFP under a galactose promoter. Growth, induction and harvesting procedures were the same.

Cryo-milling of harvested yeast pellets.

All purifications of pre-ribosomal complexes were performed using cryo-milling as lysis method for the yeast cells. Noodles of yeast pellets were placed in the appropriate grinding jar, that was preemptively chilled with liquid nitrogen. For 5 to 15 g of noodles, a 50 mL jar was used with 3 grinding balls. For 16 to 50 g of noodles, the 125 mL jar was used with 10 grinding balls. The noodles were ground with a Retsch cryo-mill PM100, 4 times 3 minutes at 300 rpm, changing direction every 1 minute. Between cycles, the jar was placed back into liquid nitrogen and quickly opened to resuspend all solidified powder at the bottom and on the sides of the jar with a liquid nitrogen-cooled spatula. The powder was store at -80 °C and used for purification within 2-3 days of grinding.

Purification of pre-ribosomal complexes assembled on truncated pre-rRNA

For small subunit processome intermediates, cryo-ground yeast powder from 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) at a ratio of 10 mL per gram of powder, and cleared by centrifugation at 4 °C, 40,000 x g for 10 min. The cleared lysate was incubated with anti-GFP beads at a ratio of 100 µL of beads per gram of powder for 2 hours at 4 °C on a nutator. Beads were pelleted by centrifugation at 4 °C for 1 min at 130 x g. After 3 washes in buffer A, the immobilized SSU processome was incubated with TEV protease at 4 °C for 1 hour. After protease cleavage, the supernatant was applied to five times less streptavidin beads than anti-GFP beads, and incubated in buffer B (50 mM Tris-HCl pH 7.7 (20 °C), 150 mM NaCl, 1 mM EDTA) for 1 hour at 4 °C. Beads were subsequently washed four times in buffer B and the complex was eluted in the same buffer, supplemented with 5 mM D-Biotin. Composition of the SSU processome was analyzed on a 4-12% SDS-PAGE gradient gel and by LC-MS.

For large subunit processome intermediates, the procedure was the same except for the ratio of streptavidin beads to anti-GFP beads used was of 1 to 3, instead of 1 to 5.

Purification of small subunit processome particles

The SSU processome was endogenously purified from a genetically modified Saccharomyces cerevisiae BY4741 strain containing a C-terminal cleavable GFP tagged Utp1 (Utp1-3myc-TEV-GFP-3FLAG) and a streptavidin binding peptide tagged Kre33 (Kre33-sbp).

The described strain was grown at 30 °C in complete minimal medium containing 2% (w/v) raffinose until an optical density of 0.6-1 was reached. Subsequently 2% (w/v) galactose was added and the culture was grown to an optical density at 600 nm of 7.5 absorbance units. Cells were harvested at 4000 x g for 5 min, washed once with 1 L of ice cold ddH2O and once with a volume of ddH2O supplemented with protease inhibitors (E64, Pepstatin, PMSF) equal to the mass of the cell pellet. The final pellet was flash frozen in liquid nitrogen and lysed by cryogenic grinding using a Retsch Planetary Ball Mill PM100. Cryo-ground powder was stored at -80 °C or used immediately.

Cryo-ground 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) and cleared by centrifugation at 4 °C, 40,000 x g for 10 min. Buffer A was identified in a previous study, where suitable buffers were screened for the isolation of pre-ribosomal particles. The cleared lysate was incubated with anti-GFP beads for 2 hours at 4 °C on a nutator. Beads were pelleted by centrifugation at 4 °C for 1 min at 130 x g. After 3 washes in buffer A, the immobilized SSU processome was incubated with TEV protease at 4 °C for 1 hour. After protease cleavage, the supernatant was applied to streptavidin beads and incubated in buffer B (50 mM Tris-HCl pH 7.7 (20 °C), 150 mM NaCl, 1 mM EDTA) for 1 hour at 4 °C. Beads were subsequently washed four times in buffer B and the complex was eluted in the same buffer, supplemented with 5 mM D-Biotin. Composition of the SSU processome was analyzed on a 4-12% SDS-PAGE gradient gel and by LC-MS.

6.5 Protein analysis

SDS-PAGE

For compositional analysis of purified protein or protein complex, pure samples were run on 4-12% or 4-20% ExpressPlus SDS-PAGE gel from Genscript, ran in MOPS-SDS running buffer. The gels were subsequently stained in SYPRO-Ruby or Coomassie Brilliant Blue.

Mass spectrometry analysis of partially reconstituted assembly intermediates

All mass spectrometry analyses were conducted on a minimum of two separate large-scale experiments. These included 7 purifications from 7 different yeast strains as shown in Fig. 2.3. The heat map shown in Fig. 2.3B corresponds to the sample that was used for SDS-PAGE analysis shown in Fig. 2.3A.

Following isolation of approximately 5 µg of total protein, proteins were denatured in urea and reduced prior to alkylation of cysteines with iodoacetamide. Proteins were digested with LysC (Wako Chemicals) followed by trypsin (Promega) and desalted (Rappsilber et al. 2003). Approximately 2 µg of each sample was injected for LC-MS/MS analysis (QExactive Plus, Thermo Scientific). Peptides were separated using a 12 cm x 75 µm C18 column (Nikkyo Technos Co., Ltd. Japan) at a flow rate of 200 nL/min, with a 5-40 % gradient over 100 minutes (buffer A 0.1 % formic acid, buffer B 0.1 % formic acid in acetonitrile).

LC-MS/MS data were analyzed using MaxQuant (version 1.5.0.30) and Perseus software (version 1.5.0.9) (Cox and Mann 2008) searching against a Uniprot Saccharomyces cerevisiae database (July 2014), allowing oxidation of methionine and protein N-terminal acetylation, and filtering at a 1 % false discovery rate at the peptide and protein level. The “match between runs” option was enabled. Protein abundances are expressed as iBAQ (intensity-Based Absolute Quantitation) values (Schwanhüusser et al. 2011) and used to generate heat maps representing protein abundance across the constructs. For validation of gel band identities (Fig. 2.3), bands were excised, and proteins digested in-gel with trypsin. Mass spectrometry was conducted as above, except on an Orbitrap-XL mass spectrometer (Thermo Scientific). Protein identities were determined by maximal area-under-the-curve per band using Proteome Discoverer 1.4 (Thermo Scientific) in conjunction with Mascot 2.5.1 (Matrix Science) searching and filtering with Percolator at 1 % FDR.

Cross-linking and mass spectrometry analysis

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

To each aliquot, 1.5 µL of disuccinimidylsuberate (DSS; 50 mM in DMSO, Creative Molecules Inc.) was added to yield a final DSS concentration of 0.5 mM and samples were cross-linked for 30 minutes at 25 °C with 450 rpm 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 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. After digestion, the peptide mixture was acidified and extracted from the gel as previously described (Shi et al. 2014, 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 × 105; maximum injection time: 800 ms; isolation width: 1.4 m/z; normalized collision energy: 24%; charge: 3–7; intensity threshold: 2.5 × 103; peptide match: off; dynamic exclusion tolerance: 1,500 mmu. Cross-linked peptides were identified from mass spectra by pLink51. All spectra were manually verified as previously (Shi et al. 2014). Cross-links were visualized using xiNET (Combe et al. 2015), are available online (Barandun et al. 2017)

6.6 RNA Analysis

Northern Blotting of Total RNA

All analysis of cellular total RNA was done using the following approach. Approximately 0.1 g of yeast powder, or pellets corresponding to 30 mL of saturated culture, were resuspended in 1 mL of TRIzol (Life Technologies) and extracted according to the manufacturer's instructions, as described below.

Yeast material was homogenized by extensive vortexing. If the yeast was already lysed by cryo-milling, no further lysis was required. If the yeast had not been lysed, approximately 100 µL of small acid-washed glass beads were added. Those samples were subsequently vortexed 5 minutes and incubated 5 min at 65 °C. After the yeast was completely resuspended and lysed, 200 µL of chloroform was added to the tubes. Tubes were incubated 2 min at room temperature, vortexed quickly and spun down at 12,000 x g for 5 min. After centrifugation the aqueous phase containing the RNA (superior phase) was collected and placed in a new siliconized tube. RNA was precipitated from this phase by adding 1 volume, approximately 600 µL, of cold isopropanol. Samples were incubated 5 min in the -20 °C freezer and were then spun down 10 min at 12,000 x g. Supernatants after centrifugation was discarded and the RNA pellets were washed by adding 500 µL of RNase-free 75% ethanol and spinning 5 min at 7,500 x g. The supernatants were then discarded by pipetting it out carefully, as to not disturb the RNA pellets. The pellets were left to air-dry in the tubes at 65 °C for 5 min. The pellets were then resuspended in a minimum of 200 µL and up to 1 mL of RNase-free water, to resuspend the RNA fully.

Total RNA samples are separated on a denaturing 3.7% Formaldehyde - 1.2% Agarose gel (SeaKem LE, Lonza). RNA can be visualized by adding a few drops of 0.2 mg/mL ethidium bromide when casting the gel. The gel can also be stained after running 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 min. The large RNA were partially hydrolyzed by applying 50mM NaOH + 10 mM NaCl for 30 min. The NaOh solution was subsequently washed out with water.

The RNA was subsequently transferred onto a cationized nylon membrane (Zeta-Probe GT, Bio-Rad) using downward capillary transfer (Fig. 6.1). After rinsing the gel and membrane in water and then 20X SSC, the gel was laid on the nylon membrane, making sure that no bubble is left between. The gel and membrane was then stacked, with the membrane facing down, on two chromatography papers that have been soaked in 20X SSC, two dry chromatography paper, and a short stack of paper towels, in that order. The gel was then covered by two other soaked chromatography paper, a two long strips of paper bridging the stack to reservoirs of 20X SSC sitting on either side of the setup. Finally, a weight of approximately was applied to glass plate resting over the entire stack. The gel was left to transfer 90 min.

Figure 6.1 Setup for downward capillary transfer of RNA molecules on nylon membrane.

RNA was cross-linked to the membrane for northern blot analysis by UV irradiation at 254 nm with a total exposure of 120 millijoules/cm2 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 min prior to addition of γ-^{32}P-end-labeled DNA oligo nucleotide probes.

To generate γ-^{32}P-end-labeled DNA oligo nucleotide probes, 1.5-2 µL of DNA oligonucleotide was incubated with 4 µL of γ-32P-ATP and 1 µL T4 PNK in 20 µL total T4 PNK buffer (NEB). The reaction was incubated at 37 °C for 30 min and subsequently cleaned-up by using the illustra MicroSpin G-25 columns (GE). All probes used in this study are listed in table 6.4.

Probes were allowed to hybridize 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 min 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).

Northern blot of purified particles

The S. cerevisiae SSU processome was purified as described above and RNA was extracted from the final D-Biotin elution with 1 mL TRIzol (Life Technologies), as described for total RNA extraction. During the isopropanol precipitation step, 10 mg of GlycoBlue Coprecipitant (Thermo Fisher) was added.

0.5 µg of isolated SSU processome RNA and 0.4 µg of purified 5' ETS-18S[1-194nt] control RNA were 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 min, RNA species were visualized with a Gel Doc EZ Imager (Bio-Rad) (Fig. 3.2) 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/cm2 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 min prior to addition of γ-^{32}P-end-labeled DNA oligo nucleotide probes. Probes were prepared as previously. Probes were

allowed to hybridize 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 min 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 6.4. Northern blotting probes used in this body of work.

PROBE	SEQUENCE	SOURCE
MS2	ACATGGGTGATCCTCATG	Chapter II/IV
5' ETS	GGAATGGTACGTTTGATATCGCTGATTTGAGA	Chapter II/III
A0-A1	CCCACCTATTCCCTCTTGC	Chapter III
18S	GGCTCCCTCTCCGGAATCGAACCCTT	Chapter III
3' MINOR	GGAAACCTTGTTACGACTTTTAG	Chapter III
D-A2	TCTTGCCCAGTAAAAGCTCTCATGC	Chapter III
A2-A3	TGTTACCTCTGGGCCCCGATTG	Chapter III
A3-5.8S	CCAGTTACGAAAATTCTTG	Chapter III
U3	TAGATTCAATTTCGGTTTCTC	Chapter III
5S	CTACTCGGTCAGGCTC	Chapter IV
SNR10	GTGTTACGAATGGCTGTTA	Chapter IV
18S TAG	CGAGGATCGAGGCTTT	Chapter IV
25S TAG	TGAAGCTCTCGAGTGTACCT	Chapter IV

6.7 Electron microscopy

Negative stain electron microscopy

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). The vaccum pump was applied on the glow discharger 5 min, after which the grids were glow-discharged 30 s at 50 mA. 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. Good classes, containing a total of 11,700 particles, were selected and used to generate an initial model of the SSU processome using e2initialmodel.py.

Cryo-electron microscopy 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.

Cryo-EM data collection

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.). 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 Å2. Data collection parameters can be found in Table 3.1.

Cryo-EM data processing

All 32 movie frames were gain corrected, aligned and dose weighted using MotionCor2 (Zheng et al. 2017). The first 2 and the last 3 frames were discarded. Other parameters for MotionCor2 were: 5x5 patches, 2x binning, 10 iterations and a tolerance of 0.5. CTFFIND 4.1.5 was used for estimating the contrast transfer function (CTF) (Rohou and 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. 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 (Fig. 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 (Supplementary Figs. 2, 3b). By using a mask encompassing UtpA, the Nop14-Noc4 heterodimer and the 3’ domain, we were able to obtain continuous density for the RNA and the Enp1 repeat protein and also improved the density for UtpA significantly (Fig. 3.4, 3.5C).

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 (Fig. 3.4, 3.5D).

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 (Fig. 3.4, 3.5E, 3.10). 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.

6.8 Model building and refinement

Protein modeling

SSU processome proteins were initially modeled using available crystal structures as well as homology models. Homology models were generated using Phyre2 (Kelley et al. 2015). Alternatively, one of the most similar structures identified by HHPRED was docked into the cryo-EM density map using Chimera (Pettersen et al. 2004).. The following crystal structures from S. cerevisiae were docked: Snu13 (PDB 2ALE), Rrp9 (PDB 4J0X), the Rcl1/Bms1 complex (PDB 4CLQ), Emg1 (PDB 3OIJ), Enp1 (PDB 5WWO), Utp22 and Rrp7 (PDB 4M5D), Rrp5 (PDB 5WWM), Krr1 (PDB 4QMF) and parts from the mature 18S rRNA as well as ribosomal proteins (PDB 4V88). Crystal structures and homology models were fitted into the density in COOT (Emsley and Cowtan 2004). A list of all used structural templates can be found in Appendix 7.3.

For the initial unassigned RNA, ideal A-form RNA duplex models were initially placed into the electron density maps, followed by manual adjustments in COOT. The secondary structure model together with the positions of the 3’ and 5’ hinges provided sufficient reference points for the complete tracing of the 5’ ETS RNA. The model for U3 snoRNA is based on the coordinates present in PDB code 5JPQ and was manually adjusted. After all identified protein folds and RNAs were fitted, unaccounted protein density was filled with ideal helices, strands and linkers in COOT (Emsley and Cowtan 2004). Assignments of these proteins elements was done using experimental cross-links (Fig. 3.9), secondary structure predictions (Drozdetskiy et al. 2015; Yang et al. 2017) and visible side-chain information.

Bioinformatic analysis of the 5' ETS RNA to generate secondary structure model

Sequences corresponding to nucleotides 1-700 of the S. cerevisiae 5’ ETS were fused to the first 30 nucleotides of the 18S rRNA, a 12 nucleotide linker (N12), the first 22 nucleotides of U3 snoRNA, a second 12 nucleotide linker (N12), the U3 snoRNA 5’ hinge, a third 12 nucleotide linker and the U3 snoRNA 3’ hinge. The combined sequence was analyzed using LocARNA-P (Will et al. 2012).

Structure refinement and validation

The model was refined against overall map 1 (3.8 Å) in PHENIX (Adams et al. 2010) with phenix.real_space_refine and secondary structure restraints for proteins and RNAs. The final refined model was resampled and scaled to the final cryo-EM map using Chimera (Pettersen et al. 2004). Model and maps were masked and model vs. map FSC curves were calculated using EMAN2 (Tang et al. 2007). To test for overfitting, atoms of the final model were randomly displaced by 0.5 Å using PHENIX (Adams et al. 2010) and refined against half-map 2 using the refinement strategy outlined above. The refined model was resampled, scaled and low-pass filtered to the nominal resolution of the final map using RELION-2.0 (Kimanius et al. 2016). FSC curves were calculated as described above with half-map 1 or half-map 2.

Structural analysis

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

Chapter VII. Appendix

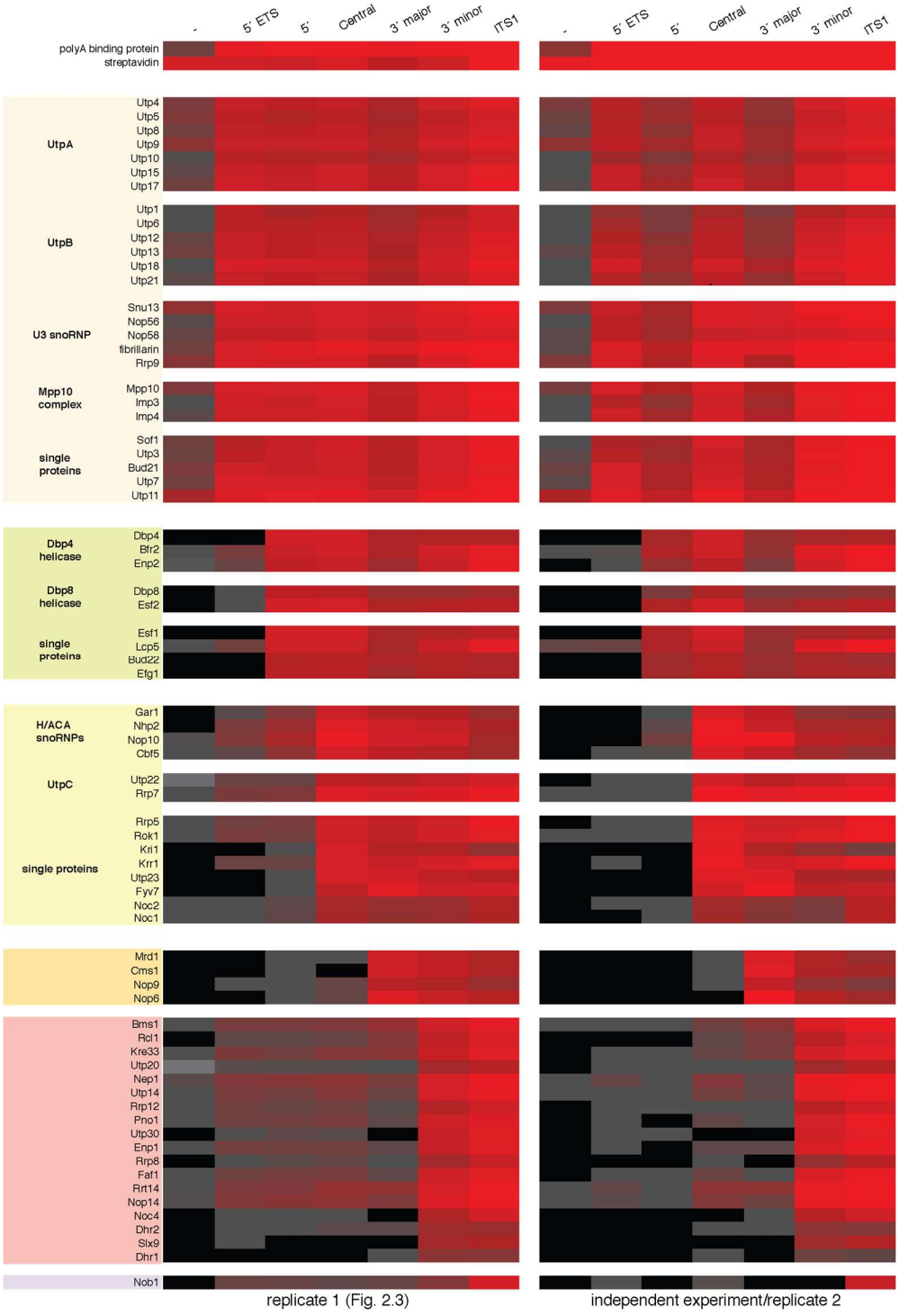


Figure 7.1 Heat map of protein abundance as a function of transcript length for SSU processome assembly factors for both replicates.

Table 7.1. List of proteins identified by mass spectrometry analysis of the yeast SSU processome sample. Proteins involved in small subunit assembly are in bold and those assigned in the structure, in red.

Protein	Unique Peptides	Area	Protein	Unique Peptides	Area	Protein	Unique Peptides	Area
RRP5	116	8.163E9	**BFR2**	29	6.280E9	**NOC2**	15	4.571E8
UTP20	100	2.601E9	**RS4A**	28	1.389E10	**RS5**	14	6.937E9
KRE33	91	1.650E10	**SAS10**	27	4.370E9	**RS6**	13	7.497E9
UTP10	70	4.786E9	**IMP4**	26	9.996E9	**RS16**	13	4.127E9
UTP22	60	5.502E9	**UTP9**	25	7.078E9	**RS24**	13	2.613E9
BMS1	59	6.483E9	**KRR1**	24	1.006E10	**FAF1**	12	1.981E9
UTP12	58	6.115E9	**RL1D1**	24	4.948E9	**BUD21**	12	1.823E9
UTP4	54	6.856E9	**RCL1**	22	6.826E9	RL3	12	4.042E8
UTP14	54	6.772E9	**UTP18**	22	5.470E9	**RS14**	11	7.429E9
UTP21	53	5.575E9	**NOC4**	22	1.801E9	**RS22**	11	3.872E9
RRP12	49	3.118E9	**FRBL**	21	6.512E9	**HAS1**	11	2.523E8
PWP2	45	5.991E9	**EMG1**	21	3.785E9	**RS7**	10	5.464E9
UTP17	42	8.543E9	**RS8A**	20	3.121E9	**RS17**	10	1.487E9
MPP10	39	7.806E9	**UTP5**	20	2.486E9	**RS23**	9	3.772E9
UTP13	37	9.317E9	**LCP5**	20	7.151E8	**RRT14**	8	3.679E9
ROK1	37	5.420E9	**RS9**	19	6.310E9	**RS13**	8	2.281E9
ENP2	36	6.068E9	**FCF2**	19	3.750E9	**FCF1**	8	2.122E9
DCA13	35	1.057E10	**IMP3**	18	5.821E9	RL3	8	1.287E9
UTP7	35	8.780E9	**MAK21**	18	3.166E8	RL19	8	1.126E9
RRP9	35	7.370E9	**UTP11**	16	9.813E9	RL7	8	6.282E8
UTP6	35	4.842E9	**PNO1**	16	7.774E9	RL10	8	5.728E8
ENP1	34	6.013E9	**RRP7**	16	3.293E9	RL13	8	5.327E8
UTP8	34	5.426E9	**RS18**	16	2.542E9	RL15	8	2.864E8
NOP56	33	8.045E9	**RRP8**	16	9.373E8	G3P3	8	2.233E8
UTP15	31	8.553E9	**DHR1**	16	2.632E8	HSP75	8	5.255E7
NOP58	31	6.283E9	**RS11**	15	8.466E9	RL20	7	5.852E8
NOP14	30	1.507E10	**NOB1**	15	2.373E9	SGD1	7	1.179E8

Protein	Unique Peptides	Area	Protein	Unique Peptides	Area	Protein	Unique Peptides	Area
SNU13	6	3.186E9	H2B1	3	1.290E8	MAK11	2	6.789E6
RS28	6	7.194E8	RL33	3	1.261E8	PYC1	2	2.712E6
RL18	6	4.032E8	RL1	3	1.197E8	NUG1	2	2.057E6
RL14	6	1.621E8	PDC1	3	6.209E7	RRS1	2	1.236E6
NOP4	6	1.254E8	RL26	3	4.727E7	**RS1**	1	1.498E9
ERB1	6	1.078E8	RL27	3	3.244E7	RL17	1	9.717E8
RL28	6	1.040E8	HSP71	3	3.213E7	RL17	1	9.664E8
PESC	6	5.018E7	**RS3**	3	1.984E7	RL21	1	6.181E8
PMA1	6	4.616E7	NOP2	3	1.833E7	RL21	1	6.181E8
RS12	5	1.292E9	KPYK1	3	1.777E7	RL34	1	4.999E8
RS15	5	3.777E8	RRP1	3	1.727E7	RL6	1	2.588E8
YRA1	5	1.141E8	**RS1A**	2	1.777E9	RL4	1	1.846E8
CIC1	5	7.650E7	SMT3	2	1.359E9	**ESF2**	1	1.663E8
SLX9	5	7.620E7	**RS25**	2	3.545E8	RL24	1	1.595E8
RL8	5	5.197E7	YD13A	2	2.661E8	ALF	1	1.195E8
TAN1	5	4.040E7	RL42	2	2.069E8	RL29	1	9.317E7
CBF5	5	1.625E7	RL4	2	1.846E8	RLP24	1	6.982E7
RS7	4	4.248E9	RL43	2	1.794E8	**NHP2**	1	6.810E7
RL35	4	1.083E9	RL6B	2	1.618E8	DBP8	1	6.756E7
RL16	4	5.178E8	RLP7	2	1.592E8	GRP78	1	5.175E7
RS26	4	2.657E8	H3	2	1.195E8	NOP16	1	4.745E7
H2A1	4	1.900E8	BRX1	2	1.063E8	EF1A	1	4.181E7
RL25	4	1.598E8	**RS27**	2	8.504E7	RM11	1	2.025E7
RL11	4	1.223E8	H4	2	6.228E7	PMG1	1	1.997E7
RS2	4	1.090E8	PGK	2	4.475E7	RL23	1	1.955E7
NOP12	4	3.238E7	LOC1	2	3.008E7	RPF2	1	1.732E7
RS31	3	7.788E8	**UTP25**	2	2.377E7	RL38	1	1.712E7
RL39	3	6.638E8	DBP9	2	1.952E7	**GAR1**	1	1.318E7
RL36	3	1.589E8	NOG1	2	1.117E7	RRP15	1	1.061E7
RL16	3	1.533E8	YTM1	2	9.343E6	RPF1	1	9.951E6
						MAK16	1	8.799E6

Table 7.2 Organization of the PDB model of the SSU processome (PDB 5WLC)

	Chain ID	Molecule name	Total residues	Domains	Modeling	Initial PDB template	Maps used for building
RNA	L0	ETS rRNA	700		atomic (498 residues)	Manual building	Core map
	L1	pre-18S rRNA	1800		atomic (1025 residues)	4V88	All
	L2	U3 snoRNA	333		atomic (169 residues)	5GIN	Core map and overall map 2
ribosomal proteins	L3	rpS18_US13	146		atomic (2-7, 11-87, 99-120), poly-Alanine (121-128)	4V88	Core map
	L4	rpS4_ES4	261		atomic (14-241)	4V88	Overall map 2
	L5	rpS5_US7	225		atomic (13-225)	4V88	Core map
	L6	rpS6_ES6	236		atomic (1-78, 91-125)	4V88	Overall map 2
	L7	rpS7_ES7	190		atomic (4-97, 117-186)	4V88	Overall map 2
	L8	rpS8_ES8	200		atomic (2-123, 150-197)	4V88	Overall map 2
	L9	rpS9_US4	197		atomic (10-184)	4V88	Overall map 2
	LC	rpS16_US9	143		atomic (3-127)	4V88	Core map
	LD	rpS11_US17	156		atomic (6-19, 31-143)	4V88	Overall map 2
	LE	rpS22_US8	130		atomic (4-130)	4V88	Overall map 2
	LF	rpS24_ES24	135		atomic (4-93)	4V88	Overall map 2
	LG	rpS28_ES28	67		atomic (5-67)	4V88	Core map
	SR	rpS23_US12	145		atomic (41-144)	4V88	Core map
	NF	rpS13_US15	151		side-chain trimmed crystal structure (28-151)	4V88	Central domain map
	NG	rpS14_US11	137		side-chain trimmed crystal structure (10-120)	4V88	Central domain map
UtpA	LH	Utp17	896	tandem-WD40	atomic (9-101, 104-159, 166-730, 734-768)	4NSX	
				CTD	poly-Alanine (818-831), atomic (846-896)	Manual building	Core map
	LI	Utp8	713	WD40	poly-Alanine (27-358)	5IXD.O	
				CTD	poly-Alanine (457-548), atomic (549-617, 631-681)	Manual building	Core map and 3' domain map
				Strand shared with Utp4	atomic (711-723)	Manual building	
	LJ	Utp15	513	WD40	atomic (2-334)	Phyre model based on 3DM0	
				Linker	atomic (353-385)	Manual building	Core map and 3' domain map
				CTD	atomic (386-512)	Manual building	
	LK	Utp9	575	CTD	poly-Alanine (391-426), atomic (428-515)	Manual building	Core map and 3' domain map
	LL	Utp5	643	WD40	atomic (5-36,46-67, 72-262, 269-332)	Phyre model based on 3DM0	
				Linker	atomic (333-372)	Manual building	Core map and 3' domain map
				CTD	atomic (434-557)	Manual building	
	LN	Utp4	776	tandem-WD40	atomic (29-405, 413-500, 508-600, 608-666, 731-776), poly-Alanine (676-689)	2YMJ	3' domain map
	LM	Utp10 N-term	1769	Helical repeat	atomic (2-425), poly-Alanine (426-432)	Manual building	Core map

	Chain ID	Molecule name	Total residues	Domains	Modeling	Initial PDB template	Maps used for building
UtpB	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	Utp6	440	Helical repeat	atomic (2-207, 264-290, 296-326), poly-Alanine (327-431)	Manual building	Core map
	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	Utp13	817	tandem-WD40	poly-Alanine (7-650)	1NR0	Core map
				CTD	atomic (649-707, 717-817)	5ICA	
	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	Utp21	939	tandem-WD40	atomic (19-673)	4NSX	Core map
				Linker	atomic (702-766)	Manual building	
				CTD	atomic (810-939)	5ICA	
U3 snoRNP	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	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	Nop1	327		atomic (80-128, 134-326)	Phyre model based on 2IPX	Core map
	SD	Nop1	327		atomic (85-125,138-324)	Phyre model based on 2IPX	Core map
	SE	Snu13	126		atomic (5-125)	2ALE	Core map
	SF	Snu13	126		atomic (5-125)	2ALE	Core map
	SG	Rrp9	573	WD40	atomic (107-166, 184-375, 394-570)	4J0X	Overall map 2

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

Bibliography

Adams PD, Afonine P V., Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, et al. 2010. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr Sect D Biol Crystallogr* **66**: 213–221.

Albert B, Knight B, Merwin J, Martin V, Ottoz D, Gloor Y, Bruzzone MJ, Rudner A, Shore D. 2016. A Molecular Titration System Coordinates Ribosomal Protein Gene Transcription with Ribosomal RNA Synthesis. *Mol Cell* **64**: 720–733.

Albert V, Hall MN. 2015. mTOR signaling in cellular and organismal energetics. *Curr Opin Cell Biol* **33**: 55–66.

Allmang C, Mitchell P, Petfalski E, Tollervey D. 2000. Degradation of ribosomal RNA precursors by the exosome. *Nucleic Acids Res* **28**: 1684–91.

Anger AM, Armache JP, Berninghausen O, Habeck M, Subklewe M, Wilson DN, Beckmann R. 2013. Structures of the human and Drosophila 80S ribosome. *Nature* **497**: 80–85.

Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. 2000. The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 A Resolution. *Science (80-)* **289**: 905–920.

Barandun J, Chaker-Margot M, Hunziker M, Molloy KR, Chait BT, Klinge S. 2017. The complete structure of the small-subunit processome. *Nat Struct Mol Biol* **24**: 944–953.

Barrio-Garcia C, Thoms M, Flemming D, Kater L, Berninghausen O, Baßler J, Beckmann R, Hurt E. 2016. Architecture of the Rix1–Rea1 checkpoint machinery during pre-60S-ribosome remodeling. *Nat Struct Mol Biol* **23**: 37–44.

Baßler J, Kallas M, Pertschy B, Ulbrich C, Thoms M, Hurt E. 2010. The AAA-ATPase Rea1 Drives Removal of Biogenesis Factors during Multiple Stages of 60S Ribosome Assembly. *Mol Cell* **38**: 712–721.

Beltrame M, Tollervey D. 1995. Base pairing between U3 and the pre-ribosomal RNA is required for 18S rRNA synthesis. *EMBO J* **14**: 4350–6.

Beltrame M, Tollervey D. 1992. Identification and functional analysis of two U3 binding sites on yeast pre-ribosomal RNA. *EMBO J* **11**: 1531–42.

Ben-Shem A, De Loubresse NG, Melnikov S, Jenner L, Yusupova G, Yusupov M. 2011. The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science (80-)* **334**: 1524–1529.

Bleichen F, Gagnon KT, Brown BA, Stuart Maxwell E, Leschziner AE, Unger VM, Baserga SJ. 2009. A dimeric structure for archaeal box C/D small ribonucleoproteins. *Science (80-)* **325**: 1384–1387.

Bleichert F, Granneman S, Osheim YN, Beyer AL, Baserga SJ. 2006. The PINc domain

protein Utp24, a putative nuclease, is required for the early cleavage steps in 18S rRNA maturation. *Proc Natl Acad Sci* **103**: 9464–9469.

Bolze A, Mahlaoui N, Byun M, Turner B, Trede N, Ellis SR, Abhyankar A, Itan Y, Patin E, Brebner S, et al. 2013. Ribosomal protein SA haploinsufficiency in humans with isolated congenital asplenia. *Science (80-)* **340**: 976–978.

Brilot AF, Chen JZ, Cheng A, Pan J, Harrison SC, Potter CS, Carragher B, Henderson R, Grigorieff N. 2012. Beam-induced motion of vitrified specimen on holey carbon film. *J Struct Biol* **177**: 630–637.

Chagnon P, Michaud J, Mitchell G, Mercier J, Marion J-F, Drouin E, Rasquin-Weber A, Hudson TJ, Richter A. 2002. A missense mutation (R565W) in cirhin (FLJ14728) in North American Indian childhood cirrhosis. *Am J Hum Genet* **71**: 1443–9.

Chaker-Margot M, Barandun J, Hunziker M, Klinge S. 2017. Architecture of the yeast small subunit processome. *Science* **355**: eaal1880.

Chaker-Margot M, Hunziker M, Barandun J, Dill BD, Klinge S. 2015. Stage-specific assembly events of the 6-MDa small-subunit processome initiate eukaryotic ribosome biogenesis. *Nat Struct Mol Biol* **22**: 920–923.

Chen W, Xie Z, Yang F, Ye K. 2017. Stepwise assembly of the earliest precursors of large ribosomal subunits in yeast. *Nucleic Acids Res* **45**: 6837–6847.

Cheng J, Kellner N, Berninghausen O, Hurt E, Beckmann R. 2017. 3.2-Å-resolution structure of the 90S preribosome before A1 pre-rRNA cleavage. *Nat Struct Mol Biol* **24**: 954–964.

Chernoff YO, Vincent a, Liebman SW. 1994. Mutations in eukaryotic 18S ribosomal RNA affect translational fidelity and resistance to aminoglycoside antibiotics. *EMBO J* **13**: 906–913.

Choesmel V, Bacqueville D, Rouquette J, Noaillac-Depeyre J, Fribourg S, Crétien A, Leblanc T, Tchernia G, Da Costa L, Gleizes PE. 2007. Impaired ribosome biogenesis in Diamond-Blackfan anemia. *Blood* **109**: 1275–1283.

Clemons WM, May JLC, Wimberly BT, McCutcheon JP, Capel MS, Ramakrishnan V. 1999. Structure of a bacterial 30S ribosomal subunit at 5.5 Å resolution. *Nature* **400**: 833–840.

Combe CW, Fischer L, Rappsilber J. 2015. xiNET: Cross-link Network Maps With Residue Resolution. *Mol Cell Proteomics* **14**: 1137–1147.

Cox J, Mann M. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**: 1367–1372.

Delprato A, Al Kadri Y, Pérébaskine N, Monfoulet C, Henry Y, Henras AK, Fribourg S. 2014. Crucial role of the Rcl1p-Bms1p interaction for yeast pre-ribosomal RNA processing. *Nucleic Acids Res* **42**: 10161–10172.

Dever TE, Kinzy TG, Pavitt GD. 2016. Mechanism and regulation of protein synthesis in Saccharomyces cerevisiae. *Genetics* **203**: 65–107.

Dragon F, Compagnone-Post PA, Mitchell BM, Porwancher KA, Wehner KA, Wormsley S, Settlage RE, Shabanowitz J, Osheim Y, Beyer AL, et al. 2002. A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature* **417**: 967–970.

Drozdetskiy A, Cole C, Procter J, Barton GJ. 2015. JPred4: A protein secondary structure prediction server. *Nucleic Acids Res* **43**: W389–W394.

Dutca LM, Gallagher JEG, Baserga SJ. 2011. The initial U3 snoRNA:pre-rRNA base pairing interaction required for pre-18S rRNA folding revealed by in vivo chemical probing. *Nucleic Acids Res* **39**: 5164–5180.

Emery B, De La Cruz J, Rocak S, Deloche O, Linder P. 2004. Has1p, a member of the DEAD-box family, is required for 40S ribosomal subunit biogenesis in Saccharomyces cerevisiae. *Mol Microbiol* **52**: 141–158.

Emsley P, Cowtan K. 2004. Coot: Model-building tools for molecular graphics. *Acta Crystallogr Sect D Biol Crystallogr* **60**: 2126–2132.

Engel C, Gubbey T, Neyer S, Sainsbury S, Oberthuer C, Baejen C, Bernecky C, Cramer P. 2017. Structural Basis of RNA Polymerase I Transcription Initiation. *Cell* **169**: 120–131.e22.

Engel C, Plitzko J, Cramer P. 2016. RNA polymerase I-Rrn3 complex at 4.8 Å resolution. *Nat Commun* **7**: 3–7.

Engel C, Sainsbury S, Cheung AC, Kostrewa D, Cramer P. 2013. RNA polymerase i structure and transcription regulation. *Nature* **502**: 650–655.

Eppens NA, Rensen S, Granneman S, Raué HA, Venema J. 1999. The roles of Rrp5p in the synthesis of yeast 18S and 5.8S rRNA can be functionally and physically separated. *RNA* **5**: 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**: 1698–701.

Fayet-Lebaron E, Atzorn V, Henry Y, Kiss T. 2009. 18S rRNA processing requires base pairings of snR30 H/ACA snoRNA to eukaryote-specific 18S sequences. *EMBO J* **28**: 1260–1270.

Faza MB, Chang Y, Occhipinti L, Kemmler S, Panse VG. 2012. Role of Mex67-Mtr2 in the Nuclear Export of 40S Pre-Ribosomes. *PLoS Genet* **8**: 16–21.

Ferreira-Cerca S, Kiburu I, Thomson E, Laronde N, Hurt E. 2014. Dominant Rio1 kinase/ATPase catalytic mutant induces trapping of late pre-40S biogenesis factors in 80S-like ribosomes. *Nucleic Acids Res* **42**: 8635–8647.

Ferreira-Cerca S, Pöll G, Kühn H, Neueder A, Jakob S, Tschochner H, Milkereit P. 2007. Analysis of the In Vivo Assembly Pathway of Eukaryotic 40S Ribosomal

Proteins. *Mol Cell* **28**: 446–457.

Ferreira-Cerca S, Sagar V, Schäfer T, Diop M, Wesseling AM, Lu H, Chai E, Hurt E, Laronde-Leblanc N. 2012. ATPase-dependent role of the atypical kinase Rio2 on the evolving pre-40S ribosomal subunit. *Nat Struct Mol Biol* **19**: 1316–1323.

Finch AJ, Hilcenko C, Basse N, Drynan LF, Goyenechea B, Menne TF, Fernández ÁG, Simpson P, D'Santos CS, Arends MJ, et al. 2011. Uncoupling of GTP hydrolysis from eIF6 release on the ribosome causes shwachman-diamond syndrome. *Genes Dev* **25**: 917–929.

Fung HYJ, Fu SC, Brautigam CA, Chook YM. 2015. Structural determinants of nuclear export signal orientation in binding to exportin CRM1. *eLife* **4**: 1–19.

Gasse L, Flemming D, Hurt E. 2015. Coordinated Ribosomal ITS2 RNA Processing by the Las1 Complex Integrating Endonuclease, Polynucleotide Kinase, and Exonuclease Activities. *Mol Cell* **60**: 808–815.

Gelperin D, Horton L, Beckman J, Hensold J, Lemmon SK. 2001. Bms1p, a novel GTP-binding protein, and the related Tsr1p are required for distinct steps of 40S ribosome biogenesis in yeast. *RNA* **7**: 1268–1283.

Gerbi S. 1996. Expansion segments: regions of variable size that interrupt the universal core secondary structure of ribosomal RNA. In *In Ribosomal RNA structure, evolution, processing, and function in protein biosynthesis (ed. Zimmermann RA, Dahlberg AE)* (eds. R. Zimmermann and A. Dahlberg), pp. 71–87, Telford-CRC Press.

Ghalei H, Schaub FX, Doherty JR, Noguchi Y, Roush WR, Cleveland JL, Elizabeth Stroupe M, Karbstein K. 2015. Hrr25/CK1δ-directed release of Ltv1 from pre-40S ribosomes is necessary for ribosome assembly and cell growth. *J Cell Biol* **208**: 745–759.

Ghalei H, Trepreau J, Collins JC, Bhaskaran H, Strunk BS, Karbstein K. 2017. Erratum: The ATPase Fap7 Tests the Ability to Carry Out Translocation-like Conformational Changes and Releases Dim1 during 40S Ribosome Maturation (Molecular Cell (2017) 67(6) (990–1000.e3) (S1097276517306044) (10.1016/j.molcel.2017.08.007)). *Mol Cell* **68**: 1134–1146.e6.

Grandi P, Rybin V, Baßler J, Petfalski E, Strauß D, Marzioch M, Schäfer T, Kuster B, Tschochner H, Tollervey D, 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. *Mol Cell* **10**: 105–115.

Granneman S, Lin CY, Champion EA, Nandineni MR, Zorca C, Baserga SJ. 2006. The nucleolar protein Esf2 interacts directly with the DExD/H box RNA helicase, Dbp8, to stimulate ATP hydrolysis. *Nucleic Acids Res* **34**: 3189–3199.

Granneman S, Petfalski E, Swiatkowska A, Tollervey D. 2010. Cracking pre-40S ribosomal subunit structure by systematic analyses of RNA-protein cross-linking. *EMBO J* **29**: 2026–2036.

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 J* **30**: 4006–4019.

Greber BJ, Gerhardy S, Leitner A, Leibundgut M, Salem M, Boehringer D, Leulliot N, Aebersold R, Panse VG, Ban N. 2016. Insertion of the Biogenesis Factor Rei1 Probes the Ribosomal Tunnel during 60S Maturation. *Cell* **164**: 91–102.

Gupta N, Culver GM. 2014. Multiple in vivo pathways for Escherichia coli small ribosomal subunit assembly occur on one pre-rRNA. *Nat Struct Mol Biol* **21**: 937–943.

Han Y, Yan C, Nguyen THD, Jackobel AJ, Ivanov I, Knutson BA, He Y. 2017. Structural mechanism of ATP-independent transcription initiation by RNA polymerase I. *eLife* **6**: 1–24.

Harms JM, Schlünzen F, Zarivach R, Bashan A, Gat S, Agmon I, Bartels H, Franceschi F, Yonath A. 2001. High Resolution Structure of the Large Ribosomal Subunit from a Mesophilic Eubacterium. *Cell* **107**: 679–688.

Hayama R, Rout MP, Fernandez-Martinez J. 2017. The nuclear pore complex core scaffold and permeability barrier: variations of a common theme. *Curr Opin Cell Biol* **46**: 110–118.

Hedges J, West M, Johnson AW. 2005. Release of the export adapter, Nmd3p, from the 60S ribosomal subunit requires Rpl10p and the cytoplasmic GTPase Lsg1p. *EMBO J* **24**: 567–579.

Henry Y, Wood H, Morrissey JP, Petfalski E, Kearsey S, Tollervey D. 1994. The 5’ end of yeast 5.8S rRNA is generated by exonucleases from an upstream cleavage site. *EMBO J* **13**: 2452–63.

Heuer A, Thomson E, Schmidt C, Berninghausen O, Becker T, Hurt E, Beckmann R. 2017. Cryo-EM structure of a late pre-40S ribosomal subunit from Saccharomyces cerevisiae. *eLife* **6**: e30189.

Hierlmeier T, Merl J, Sauert M, Perez-Fernandez J, Schultz P, Bruckmann A, Hamperl S, Ohmayer U, Rachel R, Jacob A, et al. 2013. Rrp5p, Noc1p and Noc2p form a protein module which is part of early large ribosomal subunit precursors in S. cerevisiae. *Nucleic Acids Res* **41**: 1191–1210.

Hoffmann NA, Jakobi AJ, Moreno-Morcillo M, Glatt S, Kosinski J, Hagen WJH, Sachse C, Müller CW. 2015. Molecular structures of unbound and transcribing RNA polymerase III. *Nature* **528**: 231–236.

Hughes JM, 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. *EMBO J* **10**: 4231–9.

Hunziker M, Barandun J, Petfalski E, Tan D, Delan-Forino C, Molloy KR, Kim KH, Dunn-Davies H, Shi Y, Chaker-Margot M, et al. 2016. UtpA and UtpB chaperone nascent pre-ribosomal RNA and U3 snoRNA to initiate eukaryotic ribosome assembly. *Nat*

Commun **7**: 12090.

Iacovella MG, Golfieri C, Massari LF, Busnelli S, Pagliuca C, Dal Maschio M, Infantino V, Visintin R, Mechtler K, Ferreira-Cerca S, et al. 2015. Rio1 promotes rDNA stability and downregulates RNA polymerase I to ensure rDNA segregation. *Nat Commun* **6**: 6643.

Jansen R, Tollervey D, Hurt EC. 1993. A U3 snoRNP protein with homology to splicing factor PRP4 and G beta domains is required for ribosomal RNA processing. *EMBO J* **12**: 2549–58.

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. *Mol Cell Biol* **23**: 4344–55.

Kappel L, Loibl M, Zisser G, Klein I, Fruhmann G, Gruber C, Unterweger S, Rechberger G, Pertschy B, Bergler H. 2012. Rlp24 activates the AAA-ATPase Drg1 to initiate cytoplasmic pre-60S maturation. *J Cell Biol* **199**: 771–782.

Karbstein K, Jonas S, Doudna JA. 2005. An essential GTPase promotes assembly of preribosomal RNA processing complexes. *Mol Cell* **20**: 633–643.

Kater L, Thoms M, Barrio-Garcia C, Cheng J, Ismail S, Ahmed YL, Bange G, Kressler D, Berninghausen O, Sinning I, et al. 2017. Visualizing the Assembly Pathway of Nucleolar Pre-60S Ribosomes. *Cell* **171**: 1599–1610.e13.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**: 845–858.

Kempers-Veenstra AE, Oliemans J, Offenberg H, Dekker AF, Piper PW, Planta RJ, Klootwijk J. 1986. 3'-End formation of transcripts from the yeast rRNA operon. *EMBO J* **5**: 2703–2710.

Khajuria RK, Munschauer M, Ulirsch JC, Fiorini C, Ludwig LS, McFarland SK, Abdulhay NJ, Specht H, Keshishian H, Mani DR, et al. 2018. Ribosome Levels Selectively Regulate Translation and Lineage Commitment in Human Hematopoiesis. *Cell* **173**: 90–103.e19.

Khatter H, Myasnikov AG, Natchiar SK, Klaholz BP. 2015. Structure of the human 80S ribosome. *Nature* **520**: 640–645.

Khatter H, Vorländer MK, Müller CW. 2017. RNA polymerase I and III: similar yet unique. *Curr Opin Struct Biol* **47**: 88–94.

Khoshnevis S, Askenasy I, Johnson MC, Dattolo MD, Young-Erdos CL, Stroupe ME, Karbstein K. 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 Biol* **14**: 1–25.

Kimanius D, Forsberg BO, Scheres SHW, Lindahl E. 2016. Accelerated cryo-EM structure determination with parallelisation using GPUS in RELION-2. *eLife* **5**.

Klinge S, Voigts-Hoffmann F, Leibundgut M, Arpagaus S, Ban N. 2011. Crystal structure of the eukaryotic 60S ribosomal subunit in complex with initiation factor 6. *Science (80-)* **334**: 941–948.

Kornprobst M, Turk M, Kellner N, Cheng J, Flemming D, Koš-Braun I, Koš M, Thoms M, Berninghausen O, Beckmann R, et al. 2016. Architecture of the 90S Pre-ribosome: A Structural View on the Birth of the Eukaryotic Ribosome. *Cell* **166**: 380–393.

Kos-Braun IC, Jung I, Koš M. 2017. Tor1 and CK2 kinases control a switch between alternative ribosome biogenesis pathways in a growth-dependent manner. *PLoS Biol* **15**.

Koš M, Tollervey D. 2005. The putative RNA helicase Dbp4p is required for release of the U14 snoRNA from preribosomes in Saccharomyces cerevisiae. *Mol Cell* **20**: 53–64.

Kressler D, Roser D, Pertschy B, Hurt E. 2008. The AAA ATPase Rix7 powers progression of ribosome biogenesis by stripping Nsa1 from pre-60S particles. *J Cell Biol* **181**: 935–944.

Krogan NJ, Peng WT, Cagney G, Robinson MD, Haw R, Zhong G, Guo X, Zhang X, Canadien V, Richards DP, et al. 2004. High-Definition Macromolecular Composition of Yeast RNA-Processing Complexes. *Mol Cell* **13**: 225–239.

Kucukelbir A, Sigworth FJ, Tagare HD. 2014. Quantifying the local resolution of cryo-EM density maps. *Nat Methods* **11**: 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* **5**: 909–917.

Kühn H, Hierlmeier T, Merl J, Jakob S, Aguissa-Touré AH, Milkereit P, Tschochner H. 2009. The Noc-domain containing C-terminus of Noc4p mediates both formation of the Noc4p-Nop14p submodule and its incorporation into the SSU processome. *PLoS One* **4**.

Kutay U, Guttinger S. 2005. Leucine-rich nuclear export signals: born to be weak. *Trends Cell Bio* **15**: 121–124.

Lafontaine D, Delcour J, Glasser A-L, Desgres J, Vandenhaute J. 1994. The DIM1 Gene Responsible for the Conserved m62Am62A Dimethylation in the 3′-Terminal Loop of 18 S rRNA is Essential in Yeast. *J Mol Biol* **241**: 492–497.

Lafontaine DLJ, Tollervey D. 2000. Synthesis and Assembly of the Box C+D Small Nucleolar RNPs. *Mol Cell Biol* **20**: 2650–2659.

Lebaron S, Schneider C, Van Nues RW, Swiatkowska A, Walsh D, Böttcher B, Granneman S, Watkins NJ, Tollervey D. 2012. Proofreading of pre-40S ribosome maturation by a translation initiation factor and 60S subunits. *Nat Struct Mol Biol* **19**: 744–753.

Lebaron S, Segerstolpe Å, French SL, Dudnakova T, de lima Alves F, Granneman S, Rappsilber J, Beyer AL, Wieslander L, Tollervey D. 2013. Rrp5 binding at multiple

sites coordinates Pre-rRNA processing and assembly. *Mol Cell* **52**: 707–719.

LeCuyer KA, Behlen LS, Uhlenbeck OC. 1995. Mutants of the Bacteriophage MS2 Coat Protein That Alter Its Cooperative Binding to RNA. *Biochemistry* **34**: 10600–10606.

Lee SJ, Baserga SJ. 1999. Imp3p and Imp4p, two specific components of the U3 small nucleolar ribonucleoprotein that are essential for pre-18S rRNA processing. *Mol Cell Biol* **19**: 5441–52.

Leidig C, Thoms M, Holdermann I, Bradatsch B, Berninghausen O, Bange G, Sinning I, Hurt E, Beckmann R. 2014. 60S ribosome biogenesis requires rotation of the 5S ribonucleoprotein particle. *Nat Commun* **5**.

Leulliot N, Bohnsack MT, Graille M, Tollervey D, Van tilbeurgh H. 2008. The yeast ribosome synthesis factor Emg1 is a novel member of the superfamily of alpha/beta knot fold methyltransferases. *Nucleic Acids Res* **36**: 629–639.

Li S, Lünse CE, Harris K a, Breaker RR. 2015. Biochemical analysis of hatchet self-cleaving ribozymes. *RNA* 1–7.

Liang W-Q, Fournier MJ. 1997. Synthesis of functional eukaryotic ribosomal RNAs in trans: Development of a novel in vivo rDNA system for dissecting ribosome biogenesis. *Proc Natl Acad Sci* **94**: 2864–2868.

Liang WQ, Fournier MJ. 1995. U14 base-pairs with 18S rRNA: A novel snoRNA interaction required for rRNA processing. *Genes Dev* **9**: 2433–2443.

Lin J, Lai S, Jia R, Xu A, Zhang L, Lu J, Ye K. 2011. Structural basis for site-specific ribose methylation by box C/D RNA protein complexes. *Nature* **469**: 559–564.

Lin J, Lu J, Feng Y, Sun M, Ye K. 2013. An RNA-Binding Complex Involved in Ribosome Biogenesis Contains a Protein with Homology to tRNA CCA-Adding Enzyme. *PLoS Biol* **11**.

Lo KY, Li Z, Bussiere C, Bresson S, Marcotte EM, Johnson AW. 2010. Defining the pathway of cytoplasmic maturation of the 60S ribosomal subunit. *Mol Cell* **39**: 196–208.

Loc'h J, Blaud M, Réty S, Lebaron S, Deschamps P, Bareille J, Jombart J, Robert-Paganin J, Delbos L, Chardon F, et al. 2014. RNA Mimicry by the Fap7 Adenylate Kinase in Ribosome Biogenesis. *PLoS Biol* **12**.

Lygerou Z, Mitchell P, Petfalski E, Séraphin B, Tollervey D. 1994. The POP1 gene encodes a protein component common to the RNase MRP and RNase P ribonucleoproteins. *Genes Dev* **8**: 1423–1433.

Ma C, Wu S, Li N, Chen Y, Yan K, Li Z, Zheng L, Lei J, Woolford JL, Gao N. 2017. Structural snapshot of cytoplasmic pre-60S ribosomal particles bound by Nmd3, Lsg1, Tif6 and Reh1. *Nat Struct Mol Biol* **24**: 214–220.

Makarova KS, Wolf YI, Mekhedov SL, Mirkin BG, Koonin E V. 2005. Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucleic*

Acids Res **33**: 4626–4638.

Malyutin AG, Musalgaonkar S, Patchett S, Frank J, Johnson AW. 2017. Nmd3 is a structural mimic of eIF5A, and activates the cpGTPase Lsg1 during 60S ribosome biogenesis. *EMBO J* **36**: 854–868.

Marmier-Gourrier N, Cléry A, Schlotter F, Senty-Ségault V, Branlant C. 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 Res* **39**: 9731–9745.

Marneros AG. 2013. BMS1 Is Mutated in Aplasia Cutis Congenita. *PLoS Genet* **9**.

Martin R, Hackert P, Ruprecht M, Simm S, Brüning L, Mirus O, Sloan KE, Kudla G, Schleiff E, Bohnsack MT. 2014. A pre-ribosomal RNA interaction network involving snoRNAs and the Rok1 helicase. RNA **20**: 1173–1182.

Mastronarde DN. 2005. Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* **152**: 36–51.

Matsuo Y, Granneman S, Thoms M, Manikas RG, Tollervey D, Hurt E. 2014. Coupled GTPase and remodelling ATPase activities form a checkpoint for ribosome export. *Nature* **505**: 112–116.

Mayer C, Zhao J, Yuan X, Grummt I. 2004. mTOR-dependent activation of the transcription factor TIF-IA links rRNA synthesis to nutrient availability. *Genes Dev* **18**: 423–434.

McCann KL, Baserga SJ. 2013. Mysterious ribosomopathies. *Science (80-)* **341**: 849–850.

McCaughan UM, Jayachandran U, Shchepachev V, Chen ZA, Rappsilber J, Tollervey D, Cook AG. 2016. Pre-40S ribosome biogenesis factor Tsr1 is an inactive structural mimic of translational GTPases. *Nat Commun* **7**: 11789.

Menne TF, Goyenechea B, Sánchez-Puig N, Wong CC, Tonkin LM, Ancliff PJ, Brost RL, Costanzo M, Boone C, Warren AJ. 2007. The Shwachman-Bodian-Diamond syndrome protein mediates translational activation of ribosomes in yeast. *Nat Genet* **39**: 486–495.

Meyer AE, Hung N-J, Yang P, Johnson AW, Craig EA. 2007. The specialized cytosolic J-protein, Jjj1, functions in 60S ribosomal subunit biogenesis. *Proc Natl Acad Sci U S A* **104**: 1558–63.

Meyer B, Wurm JP, Kötter P, Leisegang MS, Schilling V, Buchhaupt M, Held M, Bahr U, Karas M, Heckel A, et al. 2011. The Bowen-Conradi syndrome protein Nep1 (Emg1) has a dual role in eukaryotic ribosome biogenesis, as an essential assembly factor and in the methylation of Ψ1191 in yeast 18S rRNA. *Nucleic Acids Res* **39**: 1526–1537.

Miller OL, Beatty BR. 1969. Visualization of Nucleolar Genes. *Science (80-)* **164**: 955–957.

Mitchell P. 2010. Rrp47 and the function of the Sas10/C1D domain. *Biochem Soc Trans* **38**: 1088–92.

Mitchell P, Petfalski E, Shevchenko A, Mann M, Tollervey D. 1997. The exosome: A conserved eukaryotic RNA processing complex containing multiple 3'→5' exoribonucleases. *Cell* **91**: 457–466.

Morrissey JP, Tollervey D. 1997. U14 small nucleolar RNA makes multiple contacts with the pre-ribosomal RNA. *Chromosoma* **105**: 515–522.

Mougey EB, O'Reilly M, Osheim Y, Miller OL, Beyer A, Sollner-Webb B. 1993. The terminal balls characteristic of eukaryotic rRNA transcription units in chromatin spreads are rRNA processing complexes. *Genes Dev* **7**: 1609–1619.

Neyer S, Kunz M, Geiss C, Hantsche M, Hodirnau VV, Seybert A, Engel C, Scheffer MP, Cramer P, Frangakis AS. 2016. Structure of RNA polymerase I transcribing ribosomal DNA genes. *Nature* **540**: 607–610.

Ni J, Tien AL, Fournier MJ. 1997. Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell* **89**: 565–573.

Nogi Y, Yano R, Nomura M. 1991. Synthesis of large rRNAs by RNA polymerase II in mutants of Saccharomyces cerevisiae defective in RNA polymerase I. *Proc Natl Acad Sci U S A* **88**: 3962–6.

Noller H, Hoffarth V, Zimniak L. 1992. Unusual resistance of peptidyl transferase to protein extraction procedures. *Science (80-)* **256**: 1416–1419.

Oeffinger M, Dlakić M, Tollervey D. 2004. A pre-ribosome-associated HEAT-repeat protein is required for export of both ribosomal subunits. *Genes Dev* **18**: 196–209.

Oeffinger M, Zenklusen D, Ferguson A, Wei KE, El Hage A, Tollervey D, Chait BT, Singer RH, Rout MP. 2009. Rrp17p Is a Eukaryotic Exonuclease Required for 5′ End Processing of Pre-60S Ribosomal RNA. *Mol Cell* **36**: 768–781.

Palade GE. 1955. A small particulate component of the cytoplasm. *J Cell Biol* **1**: 59–68.

Paolini NA, Attwood M, Sondalle SB, Vieira CM dos S, van Adrichem AM, di Summa FM, O'Donohue MF, Gleizes PE, Rachuri S, Briggs JW, et al. 2017. A Ribosomopathy Reveals Decoding Defective Ribosomes Driving Human Dysmorphism. *Am J Hum Genet* **100**: 506–522.

Pelletier J, Thomas G, Volarevi S. 2017. Ribosome biogenesis in cancer: New players and therapeutic avenues. *Nat Rev Cancer* **18**: 51–63.

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 Res* **39**: 8105–8121.

Perez-Fernandez J, Roman A, De Las Rivas J, Bustelo XR, Dosil M. 2007. The 90S Preribosome Is a Multimodular Structure That Is Assembled through a Hierarchical Mechanism. *Mol Cell Biol* **27**: 5414–5429.

Pertschy B, Saveanu C, Zisser G, Lebreton A, Tengg M, Jacquier A, Liebminger E, Nobis B, Kappel L, van der Klei I, et al. 2007. Cytoplasmic Recycling of 60S Preribosomal Factors Depends on the AAA Protein Drg1. *Mol Cell Biol* **27**: 6581–6592.

Petrov AS, Bernier CR, Gulen B, Waterbury CC, Hershkovits E, Hsiao C, Harvey SC, Hud N V., Fox GE, Wartell RM, et al. 2014. Secondary structures of rRNAs from all three domains of life. *PLoS One* **9**: 1–6.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera - A visualization system for exploratory research and analysis. *J Comput Chem* **25**: 1605–1612.

Philippi A, Steinbauer R, Reiter A, Fath S, Leger-Silvestre I, Milkereit P, Griesenbeck J, Tschochner H. 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 Res* **38**: 5315–5326.

Pilsl M, Crucifix C, Papai G, Krupp F, Steinbauer R, Griesenbeck J, Milkereit P, Tschochner H, Schultz P. 2016. Structure of the initiation-competent RNA polymerase i and its implication for transcription. *Nat Commun* **7**: 1–12.

Rabl J, Leibundgut M, Ataide SF, Haag A, Ban N. 2011. Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. *Science (80-)* **331**: 730–736.

Raman N, Weir E, Müller S. 2016. The AAA ATPase MDN1 Acts as a SUMO-Targeted Regulator in Mammalian Pre-ribosome Remodeling. *Mol Cell* **64**: 607–615.

Ramesh M, Woolford JL. 2016. Eukaryote-specific rRNA expansion segments function in ribosome biogenesis. RNA **22**: 1153–1162.

Rappsilber J, Ishihama Y, Mann M. 2003. Stop And Go Extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* **75**: 663–670.

Robson A, Owens NDL, Baserga SJ, Khokha MK, Griffin JN. 2016. Expression of ribosomopathy genes during Xenopus tropicalis embryogenesis. *BMC Dev Biol* **16**: 1–13.

Rohou A, Grigorieff N. 2015. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**: 216–221.

Rosado I V., Dez C, Lebaron S, Caizergues-Ferrer M, Henry Y, de la Cruz J. 2007. Characterization of Saccharomyces cerevisiae Npa2p (Urb2p) Reveals a Low-Molecular-Mass Complex Containing Dbp6p, Npa1p (Urb1p), Nop8p, and Rsa3p Involved in Early Steps of 60S Ribosomal Subunit Biogenesis. *Mol Cell Biol* **27**: 1207–1221.

Sadian Y, Tafur L, Kosinski J, Jakobi AJ, Wetzel R, Buczak K, Hagen WJ, Beck M, Sachse C, Müller CW. 2017. Structural insights into transcription initiation by yeast RNA polymerase I. *EMBO J* **36**: e201796958.

Sanchez-Casalongue ME, Lee J, Diamond A, Shuldiner S, Moir RD, Willis IM. 2015. Differential Phosphorylation of a Regulatory Subunit of Protein Kinase CK2 by Target of Rapamycin Complex 1 Signaling and the Cdc-like Kinase Kns1. *J Biol Chem* **290**: 7221–7233.

Sanghai ZA, Miller L, Molloy KR, Barandun J, Hunziker M, Chaker-Margot M, Wang J, Chait BT, Klinge S. 2018. Modular assembly of the nucleolar pre-60S ribosomal subunit. *Nature*.

Sardana R, Liu X, Granneman S, Zhu J, Gill M, Papoulas O, Marcotte EM, Tollervey D, Correll CC, Johnson AW. 2015. The DEAH-box Helicase Dhr1 Dissociates U3 from the Pre-rRNA to Promote Formation of the Central Pseudoknot. *PLoS Biol* **13**: 1–25.

Sarkar A, Thoms M, Barrio-Garcia C, Thomson E, Flemming D, Beckmann R, Hurt E. 2017. Preribosomes escaping from the nucleus are caught during translation by cytoplasmic quality control. *Nat Struct Mol Biol* **24**: 1107–1115.

Saveanu C, Bienvenu D, Namane A, Gleizes PE, Gas N, Jacquier A, Fromont-Racine M. 2001. Nog2p, a putative GTPase associated with pre-60S subunits and required for late 60S maturation steps. *EMBO J* **20**: 6475–6484.

Scaiola A, Peña C, Weisser M, Böhringer D, Leibundgut M, Klingauf-Nerurkar P, Gerhardy S, Panse VG, Ban N. 2018. Structure of a eukaryotic cytoplasmic pre-40S ribosomal subunit. *EMBO J* e98499.

Schäfer T, Maco B, Petfalski E, Tollervey D, Böttcher B, Aebi U, Hurt E. 2006. Hrr25-dependent phosphorylation state regulates organization of the pre-40S subunit. *Nature* **441**: 651–655.

Schäfer T, Strauß D, Petfalski E, Tollervey D, Hurt E. 2003. The path from nucleolar 90S to cytoplasmic 40S pre-ribosomes. *EMBO J* **22**: 1370–1380.

Schluenzen F, Tocilj A, Zarivach R, Harms J, Gluehmann M, Janell D, Bashan A, Bartels H, Agmon I, Franceschi F, et al. 2000. Structure of Functionally Activated Small Ribosomal Subunit at 3.3 Å Resolution. *Cell* **102**: 615–623.

Schneider DA. 2012. RNA polymerase I activity is regulated at multiple steps in the transcription cycle: Recent insights into factors that influence transcription elongation. *Gene* **493**: 176–184.

Schuch B, Feigenbutz M, Makino DL, Falk S, Basquin C, Mitchell P, Conti E. 2014. The exosome-binding factors Rrp6 and Rrp47 form a composite surface for recruiting the Mtr4 helicase. *EMBO J* **33**: 2829–2846.

Schuller JM, Falk S, Fromm L, Hurt E, Conti E. 2018. Structure of the nuclear exosome captured on a maturing preribosome. *Science (80-)* eaar5428.

Schwanhüusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M. 2011. Global quantification of mammalian gene expression control. *Nature* **473**: 337–342.

Segerstolpe Å, Granneman S, Björk P, De Lima Alves F, Rappsilber J, Andersson C, Högbom M, Tollervey D, Wieslander L. 2013. Multiple RNA interactions position Mrd1 at the site of the small subunit pseudoknot within the 90S pre-ribosome. *Nucleic Acids Res* **41**: 1178–1190.

Shajani Z, Sykes MT, Williamson JR. 2011. Assembly of Bacterial Ribosomes. *Annu Rev Biochem* **80**: 501–526.

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. *Mol Cell Biol* **19**: 6012–9.

Sharma S, Langhendries JL, Watzinger P, Kotter P, Entian KD, Lafontaine DLJ. 2015. Yeast Kre33 and human NAT10 are conserved 18S rRNA cytosine acetyltransferases that modify tRNAs assisted by the adaptor Tan1/THUMPD1. *Nucleic Acids Res* **43**: 2242–2258.

Shi Y, Fernandez-Martinez J, Tjioe E, Pellarin R, Kim SJ, Williams R, Schneidman-Duhovny D, Sali A, Rout MP, Chait BT. 2014. Structural Characterization by Cross-linking Reveals the Detailed Architecture of a Coatomer-related Heptameric Module from the Nuclear Pore Complex. *Mol Cell Proteomics* **13**: 2927–2943.

Shi Y, Pellarin R, Fridy PC, Fernandez-Martinez J, Thompson MK, Li Y, Wang QJ, Sali A, Rout MP, Chait BT. 2015. A strategy for dissecting the architectures of native macromolecular assemblies. *Nat Methods* **12**: 1135–1138.

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 Res* **42**: 3194–3206.

Soltanieh S, Osheim YN, Spasov K, Trahan C, Beyer AL, Dragon F. 2015. DEAD-Box RNA Helicase Dbp4 Is Required for Small-Subunit Processome Formation and Function. *Mol Cell Biol* **35**: 816–830.

Sondalle SB, Baserga SJ. 2014. Human diseases of the SSU processome. *Biochim Biophys Acta - Mol Basis Dis* **1842**: 758–764.

Strunk BS, Novak MN, Young CL, Karbstein K. 2012. A translation-like cycle is a quality control checkpoint for maturing 40S ribosome subunits. *Cell* **150**: 111–121.

Sun Q, Zhu X, Qi J, An W, Lan P, Tan D, Chen R, Wang B, Zheng S, Zhang C, et al. 2017. Molecular architecture of the 90S small subunit pre-ribosome. *eLife* **6**.

Tackett AJ, DeGrasse JA, Sekedat MD, Oeffinger M, Rout MP, Chait BT. 2005. I-DIRT, a general method for distinguishing between specific and nonspecific protein interactions. *J Proteome Res* **4**: 1752–1756.

Talkish J, Biedka S, Jakovljevic J, Zhang J, Tang L, Strahler JR, Andrews PC, Maddock JR, Woolford JL. 2016. Disruption of ribosome assembly in yeast blocks cotranscriptional pre-rRNA processing and affects the global hierarchy of ribosome biogenesis. RNA **22**: 852–866.

Talkish J, Campbell IW, Sahasranaman A, Jakovljevic J, Woolford JL. 2014. Ribosome

Assembly Factors Pwp1 and Nop12 Are Important for Folding of 5.8S rRNA during Ribosome Biogenesis in Saccharomyces cerevisiae. *Mol Cell Biol* **34**: 1863–1877.

Tang G, Peng L, Baldwin PR, Mann DS, Jiang W, Rees I, Ludtke SJ. 2007. EMAN2: An extensible image processing suite for electron microscopy. *J Struct Biol* **157**: 38–46.

Thoms M, Thomson E, Baßler J, Gnädig M, Griesel S, Hurt E. 2015. The Exosome Is Recruited to RNA Substrates through Specific Adaptor Proteins. *Cell* **162**: 1029–1038.

Thomson E, Rappsilber J, Tollervey D. 2007. Nop9 is an RNA binding protein present in pre-40S ribosomes and required for 18S rRNA synthesis in yeast. *RNA* **13**: 2165–74.

Thomson E, Tollervey D. 2010. The Final Step in 5.8S rRNA Processing Is Cytoplasmic in Saccharomyces cerevisiae. *Mol Cell Biol* **30**: 976–984.

Tomecki R, Labno A, Drazkowska K, Cysewski D, Dziembowski A. 2015. hUTP24 is essential for processing of the human rRNA precursor at site A1but not at site A0. *RNA Biol* **12**: 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* **4**: 1636–1652.

Torreira E, Louro JA, Pazos I, González-Polo N, Gil-Carton D, Duran AG, Tosi S, Gallego O, Calvo O, Fernández-Tornero C. 2017. The dynamic assembly of distinct RNA polymerase i complexes modulates rDNA transcription. *eLife* **6**: 1–23.

Trapman J, Retèl J, Planta RJ. 1975. Ribosomal precursor particles from yeast. *Exp Cell Res* **90**: 95–104.

Turowski TW, Lebaron S, Zhang E, Peil L, Dudnakova T, Petfalski E, Granneman S, Rappsilber J, Tollervey D. 2014. Rio1 mediates ATP-dependent final maturation of 40S ribosomal subunits. *Nucleic Acids Res* **42**: 12189–12199.

Ulbrich C, Diepholz M, Baßler J, Kressler D, Pertschy B, Galani K, Böttcher B, Hurt E. 2009. Mechanochemical Removal of Ribosome Biogenesis Factors from Nascent 60S Ribosomal Subunits. *Cell* **138**: 911–922.

van Hoof A. 2000. Three conserved members of the RNase D family have unique and overlapping functions in the processing of 5S, 5.8S, U4, U5, RNase MRP and RNase P RNAs in yeast. *EMBO J* **19**: 1357–1365.

Vannini A, Cramer P. 2012. Conservation between the RNA Polymerase I, II, and III Transcription Initiation Machineries. *Mol Cell* **45**: 439–446.

Venema J, Vos HR, Faber AW, Van Venrooij WJ, Raué HA. 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 **6**: 1660–1671.

Vorländer MK, Khatter H, Wetzel R, Hagen WJH, Müller CW. 2018. Molecular mechanism of promoter opening by RNA polymerase III. *Nature* **553**: 295–300.

Vos HR, Faber AW, De Gier MD, Vos JC, Raué HA. 2004. Deletion of the three distal S1 motifs of Saccharomyces cerevisiae Rrp5p abolishes pre-rRNA processing at site A2without reducing the production of functional 40S subunits. *Eukaryot Cell* **3**: 1504–1512.

Warner JR. 1999. The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci* **24**: 437–440.

Wasmuth E V., Lima CD. 2012. Exo- and Endoribonucleolytic Activities of Yeast Cytoplasmic and Nuclear RNA Exosomes Are Dependent on the Noncatalytic Core and Central Channel. *Mol Cell* **48**: 133–144.

Watkins NJ, Bohnsack MT. 2012. The box C/D and H/ACA snoRNPs: Key players in the modification, processing and the dynamic folding of ribosomal RNA. *Wiley Interdiscip Rev RNA* **3**: 397–414.

Wegierski T, Billy E, Nasr F, Filipowicz W. 2001. Bms1p, a G-domain-containing protein, associates with Rcl1p and is required for 18S rRNA biogenesis in yeast. *RNA* **7**: S1355838201012079.

Weis F, Giudice E, Churcher M, Jin L, Hilcenko C, Wong CC, Traynor D, Kay RR, Warren AJ. 2015. Mechanism of eIF6 release from the nascent 60S ribosomal subunit. *Nat Struct Mol Biol* **22**: 914–919.

Wells GR, Weichmann F, Colvin D, Sloan KE, Kudla G, Tollervey D, Watkins NJ, Schneider C. 2016. The PIN domain endonuclease Utp24 cleaves pre-ribosomal RNA at two coupled sites in yeast and humans. *Nucleic Acids Res* **44**: 5399–5409.

Wiederkehr T, Prétôt RF, Minvielle-Sebastia L. 1998. Synthetic lethal interactions with conditional poly(A) polymerase alleles identify LCP5, a gene involved in 18S rRNA maturation. *RNA* **4**: 1357–1372.

Will S, Joshi T, Hofacker IL, Stadler PF, Backofen R. 2012. LocARNA-P: Accurate boundary prediction and improved detection of structural RNAs. *RNA* **18**: 900–914.

Wimberly BT, Brodersen DE, Clemons Jr WM, Morgan-Warren RJ, Carter AP, Vonrhein C, Hartsch T, Ramakrishnan V. 2000. Structure of the 30S ribosomal subunit. *Nature* **407**: 327–339.

Woolford JL, Baserga SJ. 2013. Ribosome biogenesis in the yeast Saccharomyces cerevisiae. *Genetics* **195**: 643–681.

Wu S, Tutuncuoglu B, Yan K, Brown H, Zhang Y, Tan D, Gamalinda M, Yuan Y, Li Z, Jakovljevic J, et al. 2016. Diverse roles of assembly factors revealed by structures of late nuclear pre-60S ribosomes. *Nature* **534**: 133–137.

Yang Y, Heffernan R, Paliwal K, Lyons J, Dehzangi A, Sharma A, Wang J, Sattar A, Zhou Y. 2017. Spider2: A package to predict secondary structure, accessible surface area, and main-chain torsional angles by deep neural networks. In *Methods*

in Molecular Biology, Vol. 1484 of, pp. 55–63.

Yao W, Roser D, Köhler A, Bradatsch B, Baßler J, Hurt E. 2007. Nuclear Export of Ribosomal 60S Subunits by the General mRNA Export Receptor Mex67-Mtr2. *Mol Cell* **26**: 51–62.

Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, Cate JHD, Noller HF. 2001. Crystal structure of the ribosome at 5.5 Å resolution. *Science (80-)* **292**: 883–896.

Zhang L, Wu C, Cai G, Chen S, Ye K. 2016. Stepwise and dynamic assembly of the earliest precursors of small ribosomal subunits in yeast. *Genes Dev* **30**: 718–732.

Zheng S, Lan P, Liu X, Ye K. 2014. Interaction between ribosome assembly factors Krr1 and Faf1 is essential for formation of small ribosomal subunit in yeast. *J Biol Chem* **289**: 22692–22703.

Zheng S, Ye K. 2014. Purification, crystallization and preliminary X-ray diffraction analysis of Imp3 in complex with an Mpp10 peptide involved in yeast ribosome biogenesis. *Acta Crystallogr Sect FStructural Biol Commun* **70**: 918–921.

Zheng SQ, Palovcak E, Armache JP, Verba KA, Cheng Y, Agard DA. 2017. MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**: 331–332.

Zhu J, Liu X, Anjos M, Correll CC, Johnson AW. 2016. Utp14 Recruits and Activates the RNA Helicase Dhr1 To Undock U3 snoRNA from the Preribosome. *Mol Cell Biol* **36**: 965–978.

Zinder JC, Lima CD. 2017. Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. *Genes Dev* **31**: 88–100.