Stepwise and dynamic assembly of the earliest precursors of small ribosomal subunits in yeast

Liman Zhang,[1,5] Chen Wu,[1,2,3,4,5] Gaihong Cai,[1] She Chen,[1] and Keqiong Ye[3,4]

[1]National Institute of Biological Sciences, Beijing, Beijing 102206, China; [2]College of Biological Sciences, China Agricultural University, Beijing 100193, China; [3]Key Laboratory of RNA Biology, [4]Beijing Key Laboratory of Noncoding RNA, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China

The eukaryotic ribosomal RNA (rRNA) is associated cotranscriptionally with numerous factors into an enormous 90S preribosomal particle that conducts early processing of small ribosomal subunits. The assembly pathway and structure of the 90S particle is poorly understood. Here, we affinity-purified and analyzed the constituents of yeast 90S particles that were assembled on a series of plasmid-encoded 3′-truncated pre-18S RNAs. We determined the assembly point of 65 proteins and the U3, U14, and snR30 small nucleolar RNAs (snoRNAs), revealing a stepwise and dynamic assembly map. The 5′ external transcribed spacer (ETS) alone can nucleate a large complex. When the 18S rRNA is nearly complete, the 90S structure undergoes a dramatic reorganization, releasing U14, snR30, and 14 protein factors that bind earlier. We also identified a reference state of 90S that is fully assembled yet has not undergone 5′ETS processing. The assembly map present here provides a new framework to understand small subunit biogenesis.

The ribosome that translates mRNA into protein is a large RNA–protein complex (RNP) composed of a small 40S subunit (SSU) and a large 60S subunit (LSU) in eukaryotes. Decades of studies in the yeast *Saccharomyces cerevisiae* have shown that ribosome synthesis is highly complicated and requires not only ribosomal RNAs (rRNAs) and ribosomal proteins (r-proteins) but also >200 *trans*-acting assembly factors (AFs) and many small nucleolar RNAs (snoRNAs) (Henras et al. 2008; Kressler et al. 2010; Woolford and Baserga 2013; de la Cruz et al. 2015). The eukaryotic ribosome synthesis machinery is much more complex than that of bacteria and dates back to the origin of eukaryotes (Ebersberger et al. 2014). Defects in ribosome assembly cause various human diseases (Freed et al. 2010).

The assembly of both ribosomal subunits starts with the transcription of a 35S precursor rRNA (pre-rRNA) in the nucleolus by RNA polymerase I (Pol I). The primary transcript contains SSU (18S) and LSU (5.8S and 25S) rRNA as well as the 5′ external transcribed spacer (ETS) and 3′ETS and internal transcribed spacer 1 (ITS1) and ITS2. The pre-rRNA is rapidly modified by H/ACA and C/D snoRNPs (Kos and Tollervey 2010) and then processed to remove spacers in the context of various preribosomal particles. The 5′ pre-18S region is assembled cotranscriptionally into the 90S preribosomal particle or the SSU processome (Dragon et al. 2002; Grandi et al. 2002; Osheim et al. 2004). Upon cleavage of 5′ETS at sites A0 and A1 and of ITS1 at site A2, a pre-40S particle containing the 20S pre-RNA intermediate is released and exported to the cytoplasm for final maturation to the SSU. The 3′ LSU part of pre-rRNA is associated with many AFs and r-proteins and a separately transcribed 5S rRNA into a pre-60S particle that eventually develops into the LSU.

More than 70 AFs, four processing snoRNAs (U3, U14, snR30, and snR10) and a subset of r-proteins have been genetically or biochemically linked with the formation of the 90S particle (Phipps et al. 2011). The absence of 90S factors commonly leads to defects in 20S pre-RNA production. Some 90S factors form stable subcomplexes, including the U3 snoRNP, UTPA/tUTP, UTPB, UTPC, and MPP10 subcomplexes (Lee and Baserga 1999; Watkins

---

et al. 2000; Grandi et al. 2002; Gallagher et al. 2004; Krogan et al. 2004).

The assembly pathway and structure of the 90S particle is poorly understood. The electron microscopy images for Miller chromatin spreads visualize two major phases during the cotranscriptional formation of 90S particles (Miller and Beatty 1969; Osheim et al. 2004). In the characteristic "Christmas tree" view, nascent rRNA transcripts of increasing lengths emanate from rRNA genes and are decorated with a knob at the 5′ end. Short transcripts with an incomplete 18S region contain a small terminal knob (~15 nm). When the pre-18S is fully transcribed, it is dramatically compacted to a large particle (~40 nm) that may correspond to the mature 90S particle (Osheim et al. 2004). However, this chromatin spreading technique cannot resolve finer assembly steps, and the composition of observed particles is unknown.

The relative assembly order has been deduced for a few AFs by analyzing their interdependencies for 90S incorporation (Perez-Fernandez et al. 2007, 2011), but this approach is indirect and low-throughput and cannot reveal the actual assembly times and assembly sites of AFs. Here, we construct a comprehensive spatiotemporal assembly map of 90S particle by analyzing a series of progressively assembled 90S particles. This map rationalizes many functional interactions among 90S AFs and reveals novel intermediates during early assembly of the SSU.

Results

Purification and analysis of progressively assembled 90S particles

To reveal the order in which AFs and processing snoRNAs assemble on the transcribing pre-rRNA, we expressed a series of 3′-truncated pre-18S RNAs in yeast and purified the in vivo assembled RNPs (Fig. 1A,B). We reasoned that the composition of these particles would provide biochemical snapshots for assembling 90S particles in temporal order. The pre-18S RNAs were expressed under the control of an RNA Pol II-driven *GAL7* promoter on a multiple-copy 2µ plasmid. Such plasmid-encoded 35S, pre-18S, and pre-5.8S/25S RNAs have been shown to produce functional ribosomal subunits (Nogi et al. 1991; Liang and Fournier 1997), indicating that they undergo productive ribosome assembly.

We affinity-purified the plasmid-derived 90S particles by using an MS2 coat protein-binding RNA motif (Hoareau-Aveilla et al. 2011) attached to the 5′ end of pre-18S RNA and then via a tandem affinity purification (TAP) tag fused to a known 90S protein. The bait protein Utp9-TAP was most frequently used in the second purification step because it is a component of the UTPA complex and likely assembles at a very early stage (Gallagher et al. 2004; Perez-Fernandez et al. 2007, 2011). Pwp2-TAP and Noc4-TAP were also used in some purifications.

The purified proteins were identified by mass spectrometry in a comprehensive manner. To estimate the abundance of each protein, the total spectral count (SpC) of all peptides belonging to a protein was first normalized
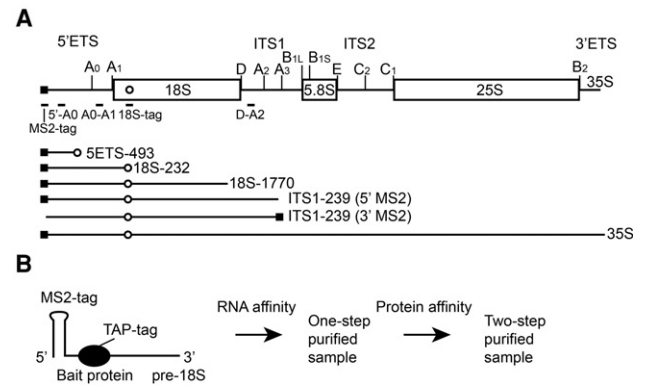

Figure 1. Purification of progressively assembled 90S particles. (*A*) Structural diagram and processing sites of 35S pre-rRNA. A few representative pre-18S fragments are shown. The plasmid-derived pre-18S RNAs contain a MS2 tag (square) at the 5′ end and an unique 18S tag sequence (circle) after the 3′ end or nucleotide 232 of 18S. The ITS1-239 RNA with a 3′ MS2 tag was also constructed. The locations of hybridization probes are indicated. (*B*) Two-step affinity purification of plasmid-derived 90S particles. The particles were affinity-purified first via an MS2 tag fused to pre-18S RNA and then via a tandem affinity purification (TAP)-tagged bait protein.

by one-hundredth of its residue number, yielding SpCs per 100 residues (SCPHR). The SCPHR is proportional to the absolute molar amount of protein and allows comparison of protein abundance within a sample. The SCPHR was further normalized against certain reference proteins, yielding the relative spectral abundance factor (RSAF). The RSAF provided a convenient, semiquantitative means to compare the relative stoichiometry of proteins within and across samples. In addition, the total SpC of UTPA proteins was used as an indicator of yield, and the total SCPHR of 90S AFs over that of all detected proteins was calculated as a purity estimation.

The second purification step enriched the 90S proteins by 3.4-fold on average and yielded more constant RASF values across samples and more specific binding patterns (Fig. 2; Supplemental Fig. S1). In particular, many RNA-binding proteins (H/ACA RNP proteins, Rrp5, Nsr1, Nop6, Pno1, and Rrp43) displayed nonspecific binding patterns in one-step-purified samples (Supplemental Fig. S1). Thus, our discussion is primarily based on two-step-purified samples. Nevertheless, the one-step-purified samples could occasionally preserve weakly associated proteins that failed to survive the second purification step.

The rRNA truncation sites were initially designed in reference to known functional and structural sites of pre-18S, including pre-rRNA-processing sites, snoRNA-binding sites, and boundaries of four 18S domains (Fig. 2). When many new proteins were found to be recruited between two consecutive RNA fragments, additional intermediate fragments were created to resolve the assembly order in finer steps. Each pre-18S RNA was named after its 3′ end position in 5′ETS, 18S, or ITS1. For example, 18S-5 refers to pre-18S that terminates at nucleotide 5 of 18S. We analyzed a total of 27 pre-rRNAs and identified 65 proteins that
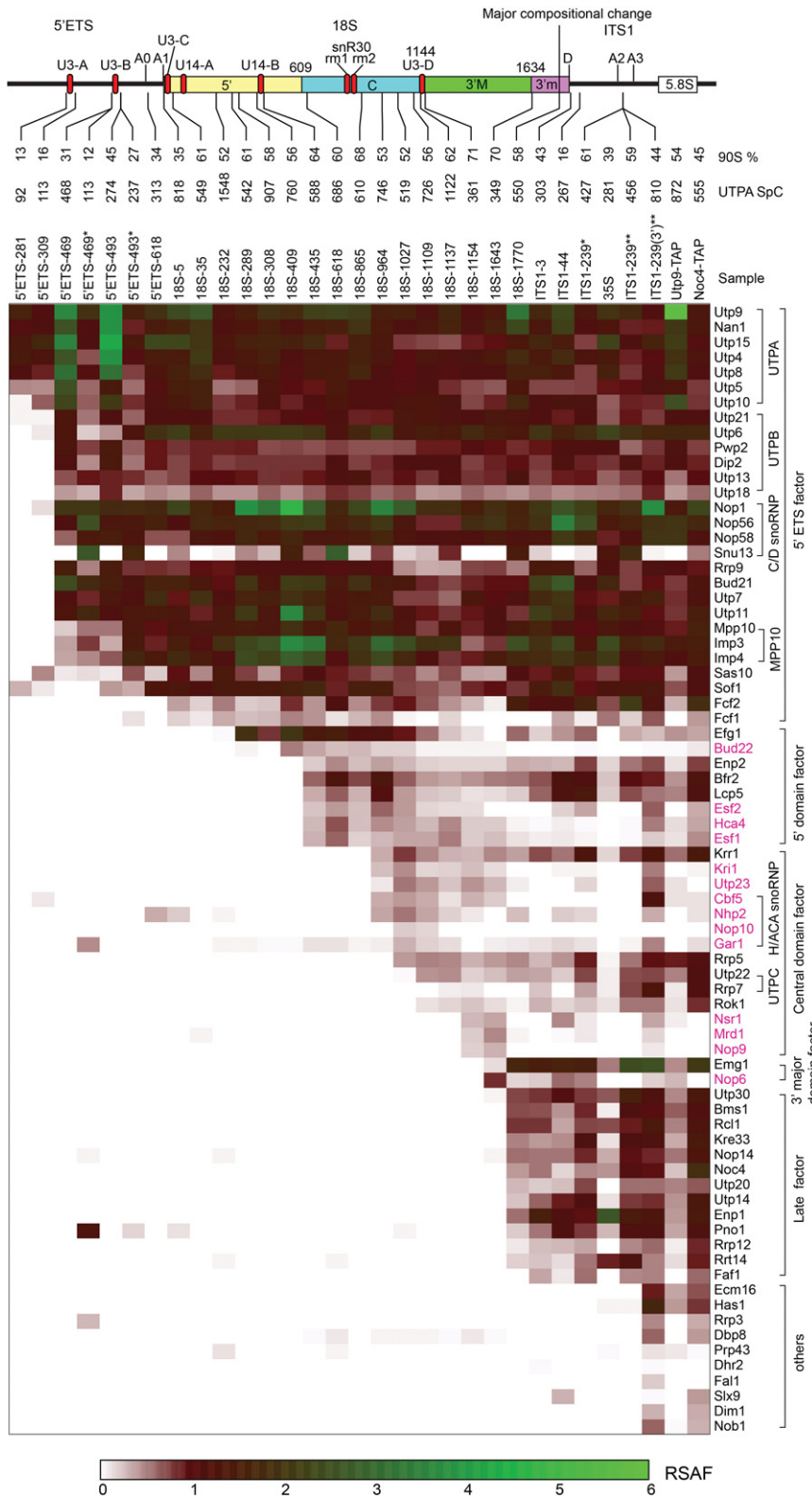
Figure 2. Heat map of 90S proteins associated with elongating pre-18S RNAs. The proteins identified from two-step-purified samples are color-coded according to their RSAF values, normalized against UTPB by default and against UTPA for the 5′ETS-281, 5′ETS-309, 5′ETS-469*, 5′ETS-493*, and ITS1-239* samples. The second purification step used Utp9-TAP, Pwp2-TAP (*), or Noc4-TAP (**). The plasmid-derived pre-rRNAs contain an MS2 tag at the 5′ end by default. ITS1-239(3′) has an MS2 tag at the 3′ end. Pwp2-TAP and Noc2-TAP are chromosomal particles. The subcomplexes and functional groups of AFs are indicated. The labile factors that dissociate in mature 90S particles are colored magenta. The total SpC of seven UTPA proteins and the total RSAF (or SCPHR) of 90S proteins over that of all detected proteins (90S%) are shown for each sample at the *top* of the figure. A cartoon of pre-18S is displayed with annotations of functional sites and domain boundaries.

show a clear pattern of binding to pre-18S RNAs (Fig. 2). Additionally, identified AFs with a low abundance and unrecognized binding pattern are in Supplemental Data Set 1. Overall, increasing numbers of 90S proteins were detected with longer pre-18S RNAs, indicating that the pre-18S rRNA is assembled progressively as it grows.

We checked the pre-rRNAs and processing snoRNAs in two-step-purified samples by Northern blot analysis. Most

particles contained intact pre-18S RNAs as major species (Fig. 3C,D), indicating that RNA degradation was not a major problem during purification. The U3, U14, snR30, and snR10 snoRNAs are required for early 18S rRNA pro-

cessing and should associate with 90S particles at some time points. Except for snR10 that had extremely weak signals, U3, U14, and snR30 displayed clear binding patterns (Fig. 4).
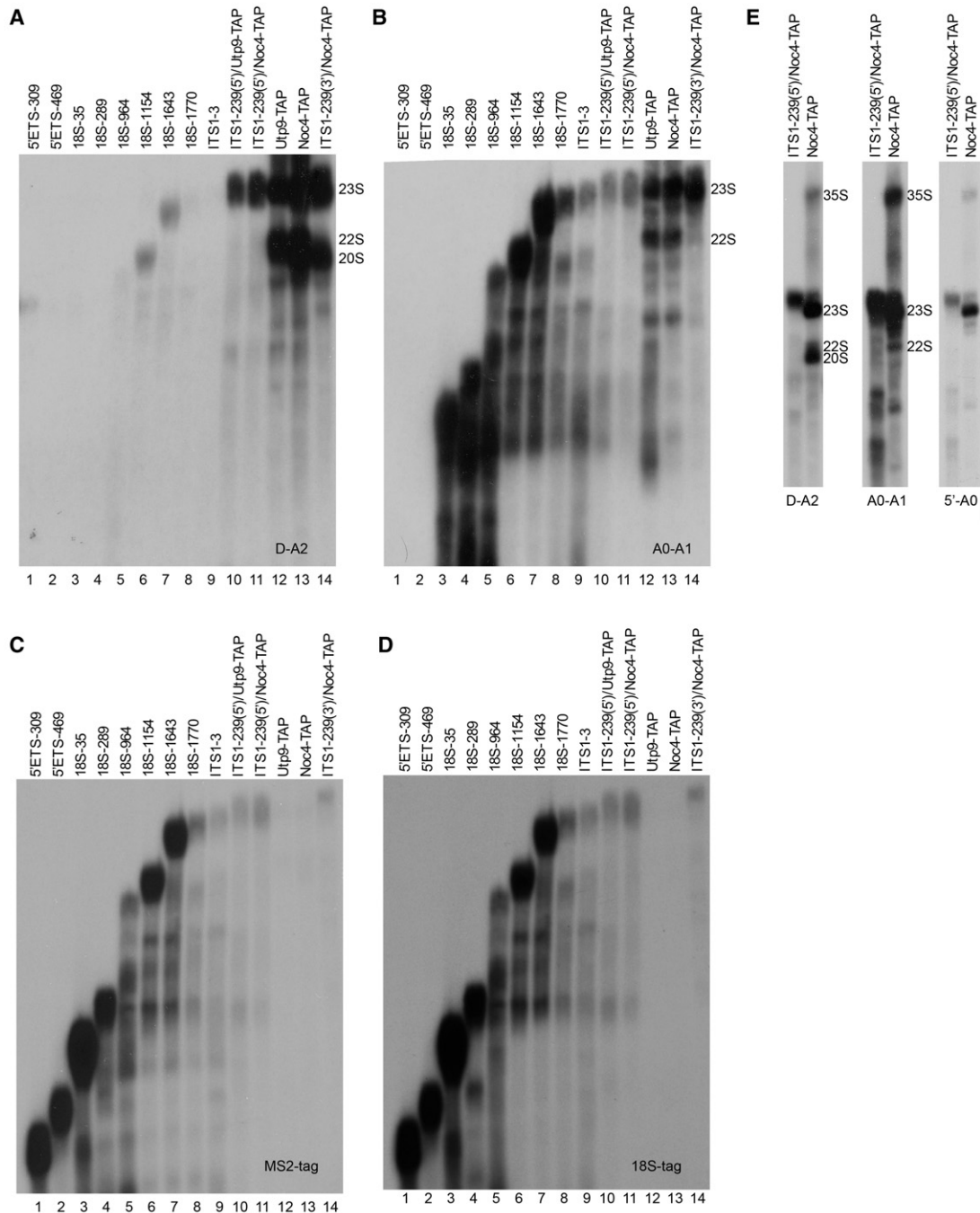

Figure 3. Pre-rRNAs in purified 90S particles. (*A–D*) Northern blot of pre-rRNAs in purified 90S particles. The 90S particles assembled on pre-18S RNAs were affinity-purified via a 5′ MS2 tag and Utp9-TAP by default. The ITS1-239 particle was also purified via a 3′ MS2 tag (lane *14*) and Noc4-TAP (lanes *11,14*). (Lanes *12,13*) Utp9-TAP and Noc4-TAP are chromosomal particles. The RNAs were separated in an agarose–formaldehyde gel and hybridized to the D-A2 (*A*), A0-A1 (*B*), MS2 tag (*C*), and 18S tag (*D*) probes. The hybridization sites on pre-rRNA are indicated in Figure 1A. The top region of this gel containing 35S pre-rRNA was accidentally lost. (*E*) Duplication of Northern blot for the 5′ MS2-tagged ITS1-293/Noc4-TAP particle and the chromosomal Noc4-TAP particle. The RNAs were hybridized to the D-A2, A0-A1, and 5′-A0 probes.
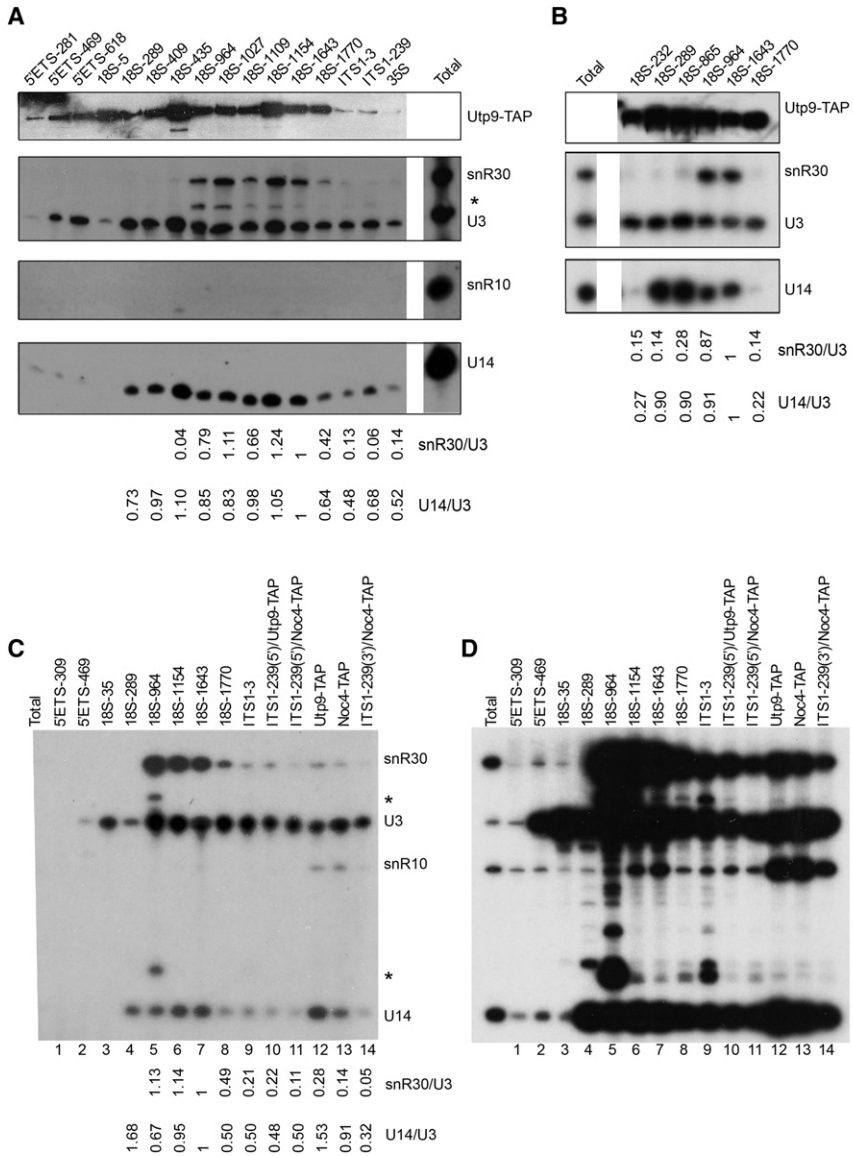
Figure 4. Processing snoRNAs in purified 90S particles. (*A*) Northern blot of snoRNAs in purified 90S particles. The 90S particles assembled on pre-18S RNAs were affinity-purified via a 5′ MS2 tag and Utp9-TAP. The Utp9-TAP protein was detected by Western blot with peroxidase–anti-peroxidase. The U3, U14, snR10, and snR30 RNAs were separated by a denaturing PAGE and detected by hybridization with specific ^{32}P-labeled DNA probes. Total RNA isolated from Utp9-TAP was used as a control. The snR30/U3 and U14/U3 ratios normalized to the 18S-1643 sample are shown at the *bottom*. (*B*) A replicate of *A* with additional samples from 18S-232 and 18S-865 particles. (*C*) Northern blot of snoRNAs in purified 90S particles. The same RNA samples were used as in Figure 3, A–D. The 90S particles assembled on pre-18S RNAs were affinity-purified via a 5′ MS2 tag and Utp9-TAP by default. The ITS1-239 particle was also purified via a 3′-MS2 tag (lane *14*) and Noc4-TAP (lanes *11,14*). (Lanes *12,13*) Utp9-TAP and Noc4-TAP are chromosomal particles. The four snoRNA probes were hybridized together to this gel. Asterisks mark major degradation products that appeared since the 18S-964 sample and were probably from snR30. (*D*) Longer exposure of *C*.

Assembly of the 5′ETS

The 5′ETS is a noncoding region preceding the 18S and contains two essential binding sites, termed A and B here, for U3 snoRNA (Beltrame and Tollervey 1995; Dutca et al. 2011; Marmier-Gourrier et al. 2011). U3 also binds at two additional sites, called C and D here, in the 5′ end and the central region of 18S. The C and D U3-binding sites form the central pseudoknot in the 40S ribosome (Fig. 5; Hughes 1996). The interaction of U3 with site D remains putative. The 5′ETS is processed at sites A0 and A1 by unidentified endonucleases to generate the 5′ end of 18S (Supplemental Fig. S2).

The shortest RNA analyzed was 5′ETS-281, which ends just before U3-binding site A. Seven UTPA proteins (Utp4, Utp5, Utp8, Utp9, Utp10, Utp15, and Nan1/Utp17) were copurified with 5′ETS-281. This finding provides direct evidence that UTPA is the first factor to be assembled on pre-rRNA (Gallagher et al. 2004; Perez-Fernandez et

al. 2007). Although Sof1 was detected in small amounts, it was not stably associated until the 3′ end of the expressed pre-rRNA reached 5′ETS-618.

The UTPA proteins were still the primary associated components for 5′ETS-309 RNA that includes U3-binding site A. Sas10, Sof1, Nop1 (a box C/D snoRNP protein bound to U3), and Utp6 (a UTPB protein) were present in low abundance. Other U3-associated proteins and UTPB proteins were not detected, indicating that the U3 snoRNP and the UTPB complex are not stably associated. Therefore, the base-pairing interaction with U3-binding site A alone is insufficient for recruiting the U3 snoRNP. It is noteworthy that the yield and percentage (13%–15%) of 90S proteins were repeatedly low for the two short RNAs, suggesting that UTPA binds weakly with the 5′-terminal region of the 5′ETS.

When the sequence linking U3-binding sites A and B was included in the 5′ETS-469 RNA, a large group of proteins were incorporated. These include six UTPB
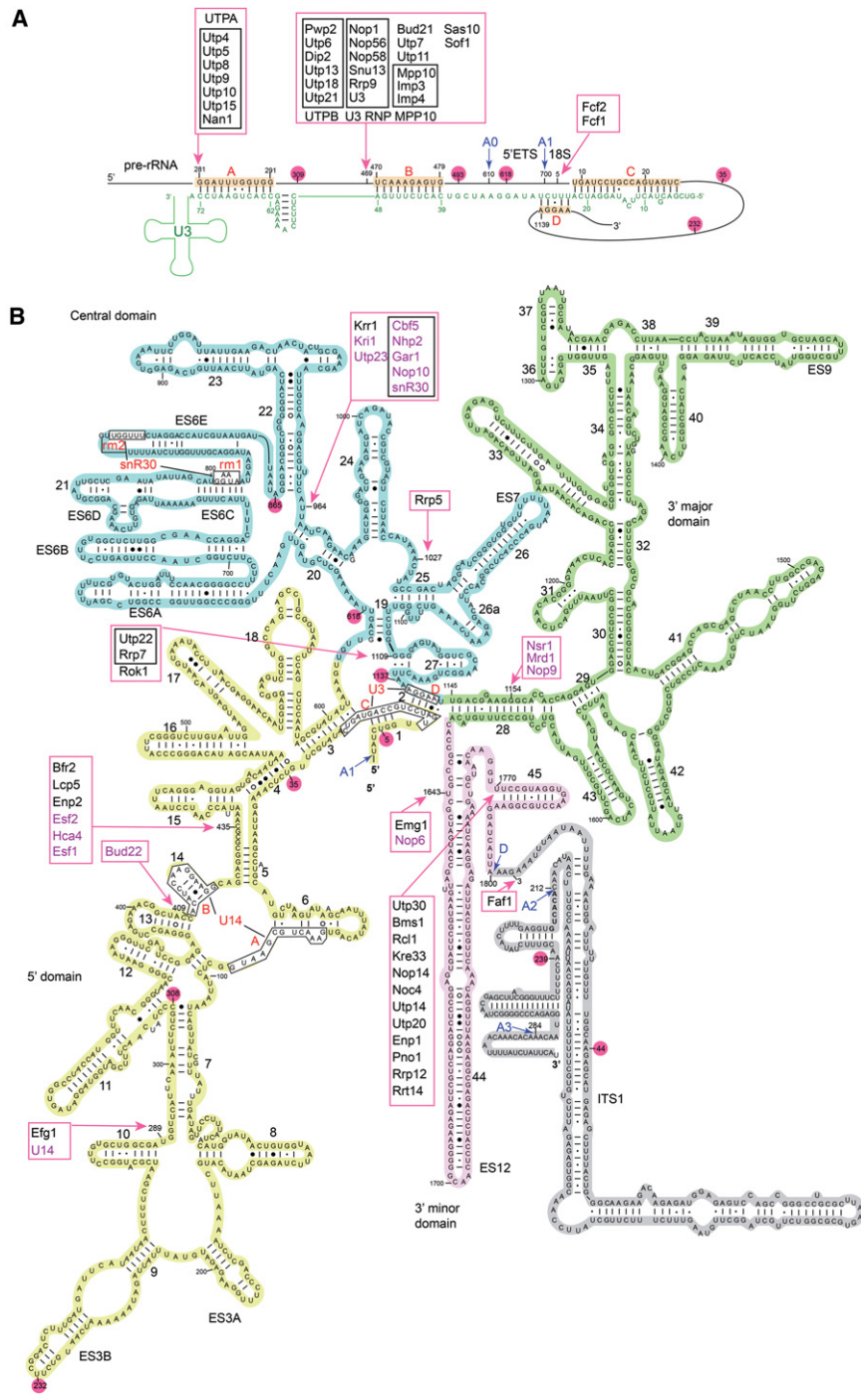
Figure 5. Ordered assembly of 90S factors on pre-18S RNAs. (*A*) Assembly of 90S factors on the 5′ETS region. The pre-18S RNA and U3 snoRNA are shown schematically. The U3-binding sites and pre-rRNA-processing sites are indicated. The 3′ end of each pre-18S RNA truncation is labeled with newly associated factors or with a magenta ball if no new factors are detected when compared with the last preceding RNA. Factors forming a subcomplex are grouped within a black box. Nucleotides are numbered independently for the 5′ETS and 18S regions. (*B*) Assembly of 90S factors on the 18S and ITS1 regions. The secondary structures of 18S rRNA are drawn according to the Comparative RNA Web Site Project (http://www.rna.ccbb.utexas.edu) and the yeast 40S structure. The pre-rRNA-processing sites; the binding sites of U3, U14, and snR30; and helix numbers are marked. Four 18S domains and ITS1 are shaded with different colors. Labile factors that dissociate in late assembly stages of 90S are colored magenta. Nucleotides are numbered independently for the 18S and ITS1 regions.

proteins (Pwp2/Utp1, Utp6, Dip21/Utp12, Utp13, Utp18, and Utp21) (Grandi et al. 2002; Krogan et al. 2004), five U3 snoRNP-associated proteins (Nop1, Nop56, Nop58, Snu13, and Rrp9) (Watkins et al. 2000), three proteins of the MPP10 complex (Mpp10, Imp3, and Imp4) (Lee and Baserga 1999), Bud21, Utp11, and Utp7. Nop1, Nop56, Nop58, and Snu13 are core proteins of C/D snoRNPs, yet Rrp9 is specific to the U3 snoRNP (Watkins et al. 2000). The presence of Rrp9 and its RSAF value being similar to that of other C/D proteins indicate

that U3 was specifically assembled. Importantly, U3 was also detected by Northern blot starting from 5′ETS-469 RNA (Fig. 4). We conclude that the sequence between U3-binding sites A and B, but not U3-binding site B, is required for initial recruitment of U3. Consistently, U3-binding site B has been shown to be less important than U3-binding site A for U3 association (Dutca et al. 2011).

The composition of the 90S particle did not change when the pre-rRNA was elongated to include U3-binding

site B (5′ETS-493) and site A0 (5′ETS-618). However, in the Utp9-TAP-purified particles of 5′ETS-493 and 5′ETS-469 RNAs, many proteins, particularly Mpp10, Imp3, Imp4, Sof1, and Sas10, were less abundant than UTPA proteins (Fig. 2). They were likely loosely bound in the two particles. In support of this idea, the abundance of these proteins was higher when the second purification step was omitted (Supplemental Fig. S1) or was conducted using Pwp2-TAP as a bait to ensure the binding of UTPB (Fig. 2). UTPB is required for assembly of U3 snoRNA and Imp4 into 90S particles (Dosil and Bustelo 2004; Perez-Fernandez et al. 2007, 2011). In contrast, most bound 90S proteins displayed more or less similar abundance in the 5′ETS-618 particle, indicating that the sequence between U3-binding site B and cleavage site A0 stabilizes the association of UTPB and MPP10 complexes.

Two new proteins, Fcf2 and Fcf1/Utp24 (Rempola et al. 2006), were recruited by 18S-5 RNA that includes the entire 5′ETS and processing site A1. Fcf1 is a putative PIN domain endonuclease important for A1 and/or A2 site cleavage (Bleichert et al. 2006). The appearance of Fcf1 near site A1 is consistent with a direct role in processing at A1. Fcf1 was substoichiometric in the 18S-5 and later particles, likely because of its weak association.

Assembly of the 5′ domain (nucleotides 1–608)

The 18S rRNA folds into four structurally distinct domains—the 5′ domain (nucleotides 1–608), the central domain (nucleotides 609–1144), the 3′ major domain (nucleotides 1145–1634), and the 3′ minor domain (nucleotides 1635–1800)—in the 40S ribosome structure (Fig. 5B; Ben-Shem et al. 2011). Inclusion of U3-binding site C in the 18S-35 RNA did not change the composition of copurified proteins. Efg1 was detected to associate with the 18S-289 RNA but not the shorter 18S-232 RNA, suggesting that helices h9 and h10 are important for its recruitment. In addition, Northern analysis shows that U14 is associated with the 18S-289 RNA but not the shorter 18S-232 RNA (Fig. 4). U14 has two binding sites, called A (nucleotides 83–95) and B (nucleotides 410–423), on the 5′ domain of 18S rRNA. Binding to site A is related to the essential function of U14 in 18S processing, whereas binding to site B directs the nonessential function of 2′-O-methylation of nucleotide C415 (Liang and Fournier 1995). The association of U14 with 18S-289 suggests that U14 is recruited to the 90S particle through interaction with site A.

No new protein was bound to 18S-308 RNA. Bud22 was stably associated with 18S-409 RNA but may bind earlier because it was detected in the one-step-purified 18S-289 particle (Supplemental Fig. S1). Further extension to 18S-435 resulted in association of six new proteins: Bfr2, Lcp5, Enp2, Esf2, Hca4/Dbp4, and Esf1. Among the factors recruited by the 5′ domain, physical interactions exist between Brf2 and Lcp5 (Uetz et al. 2000). In addition, the helicase Hca4/Dbp4 is involved in the release of U14 that binds nearby (Liang et al. 1997; Kos and Tollervey 2005).

Assembly of the central domain (nucleotides 609–1144)

No new protein was recruited, as the pre-18S extended to contain the entire 5′ domain (18S-618) and extension segment 6 (ES6) in the central domain (18S-865). The longer 18S-964 RNA recruited seven new proteins; namely, Krr1, Kri1, Utp23, Cbf5, Nhp2, Gar1, and Nop10. The latter four are core proteins of H/ACA snoRNPs and should primarily associate with the snR30 snoRNA because snR30 started to be detected in the 18S-964 particle (Fig. 4). The association of snR30 with 18S-964 is consistent with the previous observation that snR30 binds at the two short motifs rm1 (nucleotides 801–805) and rm2 (nucleotides 836–840) in ES6 (Fayet-Lebaron et al. 2009). Our data show that the binding of snR30 additionally requires helices h22 and h23 of 18S rRNA. The coassembly of snR30 snoRNP with Krr1, Kri1, and Utp23 is supported by the previously demonstrated interactions of snR30 with Utp23 and Kri1 (Hoareau-Aveilla et al. 2011; Lu et al. 2013) and the interaction between Krr1 and Kri1 (Sasaki et al. 2000).

Further growth of pre-18S to 18S-1027 led to the recruitment of Rrp5, a long protein with 12 S1 RNA-binding domains and seven TPR repeats (Venema and Tollervey 1996). Rrp5 has been shown to cross-link to multiple sites along 35S pre-rRNA, including a site in ES6 (Lebaron et al. 2013). Utp22, Rrp7, and Rok1 were recruited to the 18S-1109 RNA. Utp22 and Rrp7 form a stable UTPC complex (Krogan et al. 2004; Lin et al. 2013). The assembly point of Rrp7 is consistent with its cross-linking sites in helix E of ES6 and helix h26 (Lin et al. 2013). In addition, Rok1 interacts with Rrp5 that assembles one step earlier (Torchet et al. 1998; Young et al. 2013).

Inclusion of helix h27 in 18S-1137 caused no change to the copurified 90S proteins. When the pre-18S RNA extended from 18S-1137 to 18S-1154 by adding 17 nucleotides (nt), three RNA-binding proteins (Nsr1, Mrd1, and Nop9) were coassembled (Kondo and Inouye 1992; Lee et al. 1992; Thomson et al. 2007; Segerstolpe et al. 2012). Interestingly, the additional 17 nt contain U3-binding site D that forms the central pseudoknot with U3-binding site C in the mature 40S structure (Fig. 5B), suggesting that the three proteins may play a role in the formation of the central pseudoknot. The assembly point of Mrd1 is also consistent with its cross-linking sites at helices h27 and h28 (Segerstolpe et al. 2012).

Assembly of the 3′ major domain (nucleotides 1145–1634)

The 3′ major domain forms the head of the 40S structure. Somehow, surprisingly, upon inclusion of the entire 3′ major domain (~500 nt) in 18S-1643, only two new proteins Emg1/Nep1 and Nop6 were recruited (Eschrich et al. 2002; Garcia-Gomez et al. 2011). The 3′ major domain might be largely unfolded at this point. The assembly point of Emg1 is consistent with its function in methylation of pseudouridine 1191 in the 3′ major domain (Meyer et al. 2011). In addition, a genetic interaction exists between Emg1 and Nop6 (Buchhaupt et al. 2007).

Assembly of the 3′ minor domain (nucleotides 1635–1800)

Remarkably, inclusion of helix h44 in 18S-1770 RNA led to association of a group of 12 proteins, including Utp30, Bms1, Rcl1, Kre33, Nop14, Noc4, Utp14, Utp20, Enp1, Pno1/Dim2, Rrp12, and Rrt14. Some of them (Utp30, Bms1, Rcl1, Kre33, and Nop14) may have bound earlier, since they were detected in small amounts (RSAF ~0.1) in the one-step-purified sample of 18S-1643 (Supplemental Fig. S1). Among these coassembled factors, interactions have been found between Noc4 and Nop14 (Milkereit et al. 2003) and between Bms1 and Rcl1 (Wegierski et al. 2001). We identified Rrt14 as a putative new component of the 90S particle that associates with 18S-1770 and longer RNAs but not with shorter RNAs. Rrt14 was previously related to regulation of rDNA transcription (Hontz et al. 2009).

The ITS1-3 RNA that contains the complete 18S region recruited the last protein, Faf1, in our map (Karkusiewicz et al. 2004; Shirai et al. 2004; Zheng et al. 2014). The amount of Faf1 was quite variable in different purifications and particles. No new factors were found when the pre-18S RNA was further elongated to the ITS1-44, ITS1-239, and 35S pre-rRNAs, indicating that the ITS1 plays a minor role in the initial recruitment of 90S AFs. Nevertheless, ITS1 can provide binding sites for AFs, such as Rrp5 (Lebaron et al. 2013).

Processing of the 5′ETS

We examined by Northern blot analysis whether the plasmid-derived pre-18S fragments can be processed. All unprocessed pre-18S RNAs, except for the 35S pre-rRNA, were detected by a probe hybridizing to the MS2 tag (Fig. 6). The largest species from the plasmid-derived 35S pre-rRNA migrated at a position similar to the ITS1-44 and ITS1-239 RNAs, suggesting that the 35S pre-rRNA was rapidly cleaved at site A3. This would also explain why few 60S AFs were detected in the purified 35S com-

plex (Supplemental Data Set 1). The pre-18S RNAs that possess a complete 18S region (ITS1–3 and longer RNAs) were processed into an RNA species of 18S size, consistent with the previous study (Liang and Fournier 1997). Except for 18S-1770, all pre-18S RNAs with an incomplete 18S region cannot be processed. A low amount of 18S-like RNA was found occasionally in the 18S-1770 sample (Fig. 6). The processing product of 18S-1170 lacks helix h45 and site D and is likely to be rapidly degraded in yeast, giving rise to its variable amounts. The processing activity of the 18S-1770 particle could be accounted for by its nearly complete set of AFs. However, the Faf1 protein essential for 5′ETS processing was not found (Karkusiewicz et al. 2004; Shirai et al. 2004; Zheng et al. 2014). The association of Faf1 is quite variable, and we cannot exclude that a small amount of Faf1 still associated with 18S-1770 but escaped detection.

Dramatic compositional changes at late stage of 90S assembly

Interestingly, the abundance of a subset of assembled proteins was significantly reduced in the 90S particles assembled on the 18S-1770 and longer RNAs. They include Bud22, Esf2, Hca4, and Esf1 bound to the 5′ domain; Kri1, Utp23, Cbf5, Mrd1, Nop9, and Nsr1 bound at the central domain; and Nop6 bound to the 3′ major domain. The quantity of Bud22 was already decreased in the 18S-1109 particle. Dbp8 showed a weak (RSAF~0.1) yet repeatable association with 18S-435 to 18S-1643 RNAs (Fig. 2; Supplemental Data Set 1) and may be a labile 5′ domain factor. These labile proteins appear to be released or their interaction with pre-18S RNA is significantly weakened when the 90S assembly is about to complete. In contrast, all proteins bound to the 5′ETS; Efg1, Bfr2, Lcp5, and Enp2 bound to the 5′ domain; Krr1, Rrp5, Utp22, Rrp7, and Rok1 bound to the central domain; and Emg1 bound to the 3′ major domain remained more or less stably associated in the nearly complete 90S particles.

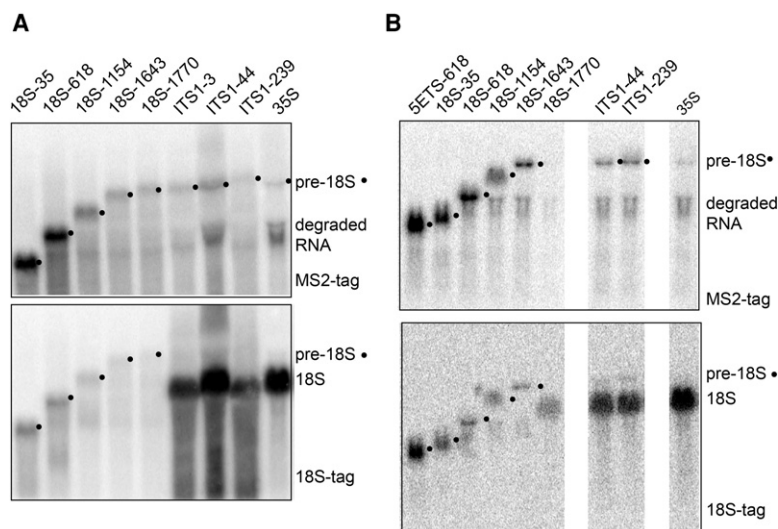**Figure 6.** Processing of pre-18S RNAs. (*A,B*) Total RNAs (15 μg) from UTP9-TAP/BY4741 yeast strains that expressed an indicated pre-rRNA from plasmid were separated in agarose–formaldehyde gels. Northern blot was conducted with ^{32}P-labeled DNA oligos that hybridize to the MS2 tag and the 18S tag present only in plasmid-encoded pre-18S RNAs. An 18S-like processing product was observed for 18S-1770 in *B* but not in *A*.

The labile and stable AFs display a dramatic contrast of abundance in the 5′ MS2-tagged ITS1-239 particle purified via Noc4-TAP (Fig. 2): The labile AFs are mostly absent, while most stable AFs are nearly stoichiometric. The ITS1-239/Noc4-TAP particle appears to represent a "clean" fully assembled 90S particle. The bait protein Noc4 is assembled at a late stage and is more likely to enrich mature 90S particles than early AFs. The pre-18S particles purified via early AFs (Utp9-TAP and Utp1-TAP) and the chromosome-derived 90S particles appear to include a fraction of early assembly intermediates and hence residual amounts of labile proteins.

Concomitant with the protein changes, the U14 and snR30 snoRNAs are also substantially reduced in the 18S-1770 and later particles (Fig. 4). In contrast, the U3 snoRNA remains stably associated. These data indicate that the U14 and snR30 snoRNAs are released at late stages of 90S assembly. The snR30 snoRNA appears to dissociate together with its binding proteins, Kri1 and Utp23 (Hoareau-Aveilla et al. 2011; Lu et al. 2013).

Comparison between plasmid- and chromosome-derived 90S particles

We purified the chromosome-derived 90S particles with the TAP approach. Two bait proteins—Utp9 and Noc4—that are assembled at early and late stages, respectively, were used. Compared with the Noc4-TAP particle, the Utp9-TAP particle contained smaller amounts of AFs that depend on 18S for assembly (Fig. 2) and larger amounts of U14 snoRNA (Fig. 4C, lanes 12–13), indicating that the early factor Utp9-TAP tends to enrich early assembly intermediates.

The protein composition of plasmid-derived 90S particles with a complete 18S region generally resembles that of the chromosome-derived 90S particles (Fig. 2). This provides biochemical evidence that Pol II transcribed pre-18S RNA is assembled normally. Compared with plasmid-derived pre-18S particles, the Noc4-TAP particle is associated with additional factors, including a few RNA helicases (Ecm16/Dhr1, Has1, Rrp3, and Dbp8), pre-40S factors (Slx9, Dim1, and Nob1), the exosome, and its cofactor (Mtr4), and a small amount of RNase MPR that cleaves the A3 site (Fig. 2; Supplemental Data Set 1; Lygerou et al. 1994).

Some of the extra factors present in the Noc4-TAP particle appear to associate with the 90S after the A0 and A1 sites of pre-rRNA are cleaved. For example, the exosome and Mtr4 are recruited to degrade the cleavage product of the 5′ETS (de la Cruz et al. 1998), and Slx9, Dim1, and Nob1 likely associate during the 90S-to-pre-40S transition. The Noc4-TAP particle is associated with many 60S AFs that should bind at the 3′ LSU region of 35S pre-rRNA (Supplemental Data Set 1). The pre-18S particles are totally absent of 60S AFs, as expected for a lack of LSU rRNA sequences.

Our plasmid-derived 90S particles were purified via the MS2 tag placed at the 5′ end of pre-18S. Such purification strategy would enrich 90S particles containing an unprocessed 5′ETS. Indeed, Northern blot analysis of pre-rRNA in purified 90S particles clearly showed that the ITS1-239/Noc4-TAP particle contained only the unprocessed transcript and no 20S pre-rRNA (Fig. 3A, lanes 10–11). In contrast, the chromosomal Utp9-TAP and Noc4-TAP particles contained the 35S and 23S pre-rRNAs with an intact 5′ETS and the 22S and 20S pre-rRNAs with the 5′ETS cleaved off (Fig. 3A,B, lanes 12–13, Fig. 3E). The 23S pre-rRNA resulted from cleavage of 35S at site A3 in the absence of prior cleavage at sites A0, A1, and A2 (Supplemental Fig. S2). If the MS2 tag was placed at the 3′ end of ITS1-239, the purified particle contained both the unprocessed transcript and 20S pre-rRNA, similar to the chromosomal particles (Fig. 3A, lane 14). Therefore, the use of the 5′ MS2 tag selectively purified the 90S particle prior to A0 and A1 cleavage.

The 3′ MS2-tagged ITS1-239/Noc4-TAP particle is also highly similar to the chromosomal Noc4-TAP particle in protein contents. Both contain the helicases, pre-40S factors, the exosome, and RNase MRP (Fig. 2; Supplemental Data Set 1). Interestingly, the 3′ MS2-tagged ITS1-239/Noc4-TAP particle contains abundant 60S AFs despite a lack of LSU rRNA sequences. The association of pre-60S with 90S could be mediated by protein–protein interactions, such as that between Rrp5 and Noc1(Hierlmeier et al. 2013), or via LSU r-protein contamination (see below).

r-proteins in the mature 90S particle

r-proteins are major contaminations in our purifications, which would preclude reliable analysis of their composition in 90S particles. Nevertheless, the 5′ MS2-tagged ITS1-239/Noc4-TAP particle is exceptionally pure: 90S proteins account for 59% of all detected molecules, SSU r-proteins account for 23.6%, and LSU r-proteins (contamination) account for 13.4%. The profile of SSU r-proteins in this particular sample should be informative (Fig. 7). By using an RSAF cutoff of 0.4, we defined a set of 18 r-proteins that are most likely present in mature 90S particles: S1, S4, S5, S6, S7, S8, S9, S11, S13, S14, S16, S17, S18, S19, S22, S24, and S28. Previous genetic studies have found 11 r-proteins to be essential for 20S pre-rRNA production (Ferreira-Cerca et al. 2005; de la Cruz et al. 2015). They should assemble into 90S particles that cleave the pre-rRNA at the A0, A1, and A2 sites. Except for S23 and S27, which have low RSAF values, all of them are present in our defined set, supporting the validity of the set. In addition, the set includes S5, S7, S18, S19, and S28, whose deletion does not block the formation of 20S pre-rRNA or 18S rRNA, and S4, S12, S17, and S22, whose deletion phenotype is not known at present. The set may still miss some r-proteins that are difficult to detect by mass spectrometry.

In vitro reconstitution of *Escherichia coli* 30S ribosome has revealed a hierarchical assembly order of r-proteins (Mizushima and Nomura 1970). Among 15 yeast SSU r-proteins that have a bacterial homolog, the 90S r-protein set includes all five primary binders, two out of three secondary binders, and only one out of seven tertiary binders. Remarkably, the SSU r-proteins that assemble early to the
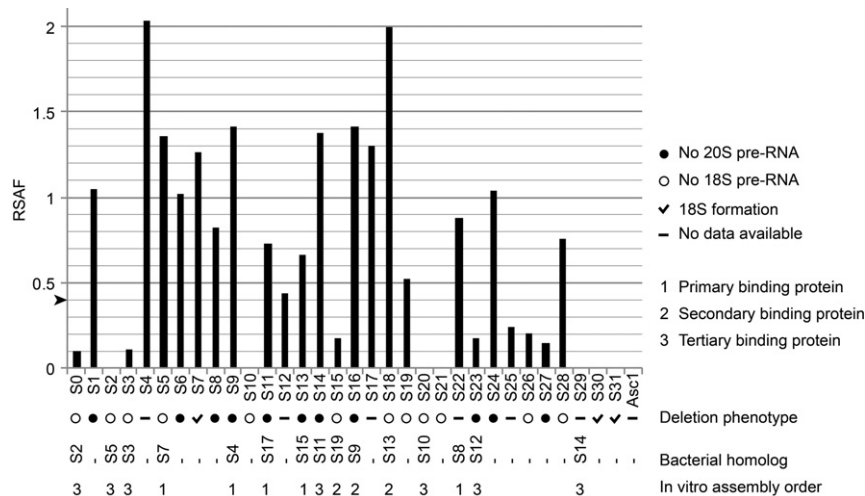
Figure 7. The r-proteins in the 5′ MS2-tagged ITS1-239/Noc4-TAP particle. The RSAF values of 33 SSU r-proteins were normalized against UTPB and are displayed. The cutoff of 0.4 to select 90S r-proteins is marked with an arrowhead. The phenotypes of r-protein deletion on 18S processing and the bacterial homologs of yeast r-proteins as well as their in vitro assembly order are shown at the *bottom* of the figure. A dash indicates no bacterial homolog.

Legend:
- ● No 20S pre-RNA
- ○ No 18S pre-RNA
- ✓ 18S formation
- − No data available
- 1 Primary binding protein
- 2 Secondary binding protein
- 3 Tertiary binding protein

bacterial rRNA in vitro appear to also associate early to the yeast pre-rRNA in vivo.

Assembly of snR10

snR10 is a nonessential yeast-specific H/ACA snoRNA that is involved in processing of 18S rRNA and pseudouridylation of U2923 in 25S rRNA (Tollervey 1987; Liang et al. 2010). Only background levels of snR10 were detected in the two-step-purified pre-18S particles (Fig. 4). Nevertheless, we noticed that snR10 was slightly enriched in the 18S-1154 and 18S-1643 particles (Fig. 4D, lanes 6–7), suggesting that snR10 may associate with the central domain and dissociate at 18S-1770, similar to snR30. In support of this notion, Rrp5, which makes a genetic interaction with snR10 (Venema and Tollervey 1996), is also recruited by the central domain. The level of snR10 was much higher in the chromosomal Utp9-TAP and Noc4-TAP particles that, in contrast to particles on the plasmid-encoded pre-rRNAs, will contain LSU sequences in the pre-rRNA. This observation suggests that association of snR10 is stabilized by its 25S target sequence, present in the 35S or associated pre-60S AFs. The 3′ MS2-tagged ITS1-239/Noc4-TAP particle is associated with pre-60S proteins and was also slightly enriched in snR10.

Discussion

We derived a spatiotemporal assembly map for the 90S particle by analyzing the protein and snoRNA components bound to a series of pre-18S RNAs of increasing lengths. Our data strongly suggest that the 90S particle is assembled progressively in 5′ to 3′ order on the transcribing pre-rRNA. The 90S particle should also incorporate a subset of r-proteins (Ferreira-Cerca et al. 2005), but the assembly point of r-proteins cannot be resolved because of their significant cross-contamination (Supplemental Data Set 1). Our map is fully compatible with previously identified assembly interdependencies between 90S factors (Dosil and Bustelo 2004; Perez-Fernandez et al. 2007, 2011; Hoareau-Aveilla et al. 2011;

Lin et al. 2013). If a factor is required for the assembly of another factor, the former always associates earlier than or simultaneously with the latter in our map. For example, UTPA is the first assembled factor in our map, and its component, Nan1, has been shown to be required for assembly of Pwp2 (UTPB), U3 snoRNA , Imp4, Rrp7 (UTPC), Bms1, and Utp20 (Perez-Fernandez et al. 2007, 2011).

The assembly map provides important insights into the binding sites and functional contexts of 90S factors. According to their RNA sequence requirements for assembly, we operationally classified AFs into the 5′ETS, 5′ domain, central domain, 3′ major domain, and late factors (Fig. 2). The classification only suggests, but cannot directly reveal, a factor's physical binding site, since an added RNA segment that leads to recruitment of a factor may directly bind the factor or induce a conformational change in preceding structures that enables recruitment of that factor. The 13 proteins induced by the 3′ minor domain are referred to as late factors without specifying their potential binding domains. Their actual binding sites are more uncertain because their association is accompanied by the release of many early factors. The expected large conformational changes may induce new binding sites in preceding domains. For example, the late factor Enp1 is known to bind at the 3′ major domain rather than the 3′ minor domain (Granneman et al. 2010; Strunk et al. 2011). In addition, a factor may have multiple binding sites on different domains, and the classification only suggests its initial binding site. As the pre-rRNA is elongated and new binding sites become available, the binding of an early associated factor may be stabilized. For example, Rrp5 is initially recruited by the central domain of 18S and would bind the ITS1 when it is available (Lebaron et al. 2013). Many 5′ETS factors are not stably associated until the transcript includes the A0 site. Despite these concerns, the classification provides an initial clue regarding the binding site and functional context of 90S AFs. Factors that assemble at similar times and places are more likely to be related in function.

We show that the 5′ETS alone can nucleate a large particle (molecular weight 2.1 MDa) that includes U3 snoRNA and 28 proteins. The U3 snoRNA and most of the 5′ETS factors are evolutionarily conserved in eukaryotes, suggesting that the 5′ETS particle is also functionally conserved for SSU biogenesis. The 5′ETS particle likely corresponds to the small knob on early rRNA transcripts that is universally detected in eukaryotes (Osheim et al. 2004). We notice that the composition of the 5′ETS particle closely matches that of the originally purified 28-protein "SSU processome" (Dragon et al. 2002). The 5′ETS particle may be a stable entity that resists dissociation and degradation during purification. In this regard, the 5′ETS factors are commonly more abundant than other factors in partially assembled 90S particles (Fig. 2). Alternatively, the 5′ETS particle might represent a physiologic A0/A1 cleavage product of 90S particles.

The 5′ MS2-tagged ITS1-239/Noc4-TAP particle is remarkably homogenous. Most of the 90S proteins and the core SSU r-proteins are of similar abundance, whereas the 14 labile protein factors and the extra factors found in the chromosomal Noc4-TAP particle are totally absent. The ITS1-239/Noc4-TAP particle appears to define a distinct state of 90S that has been fully assembled yet has not undergone 5′ETS processing. This state can serve as a reference to understand other states of 90S during its cotranscriptional assembly and transition to the pre-40S. The particle is composed of ITS1-239 RNA (similar with 23S pre-RNA), U3 snoRNA, at least 18 SSU r-proteins, and 51 stably associated AFs, and its molecular mass amounts to 5.0 MDa.

We show that the 90S particle undergoes a dramatic compositional change following the transcription of helix h44. Many factors bound at the 5′ and central domain, including U14 and snR30 snoRNAs, are released, coupled with association of 12 new proteins. The released snoRNAs and proteins are often difficult to detect in mature 90S particles (Grandi et al. 2002; Peng et al. 2004; Thomson et al. 2007). They actually function prior to the formation of mature 90S particles and have a shorter dwelling time than other stably associated factors. Our results reveal that the assembly of 90S is a highly dynamic process and involves intermediates that were not previously appreciated.

The 18S-1770 particle has acquired the ability of 5′ETS processing, raising the question of whether the dynamic event at 18S-1770 is due to pre-rRNA processing. Due to the use of the 5′ MS2 tag in RNA affinity purification, our purified cleavage-competent 90S particles contain only the unprocessed transcript, as shown clearly for the ITS1-239 particle. Therefore, the compositional change at 18S-1770 is not a consequence of pre-rRNA processing but rather an assembly event triggered by completion of helix h44.

U14 and snR30 have been considered as RNA chaperons assisting rRNA folding. We identified the precise time points at which they associate with and dissociate from pre-18S rRNA. The association of U14 and snR30 would prevent their binding regions from adopting mature structures. It is expected that their release will trigger further folding of the 5′ and central domains and promote 90S maturation.

The structural reorganization of 90S very likely corresponds to the dramatic compaction event that leads to the formation of the large knob observed in Miller spreads (Osheim et al. 2004) because both events occur at similar time points when the 18S region is near completion. The correspondence between our biochemical assembly map and previous electron microscopy observations (small and large knobs) demonstrates that the partially assembled 90S particles serve as an excellent model for investigating the cotranscriptional assembly process of the 90S preribosome. The assembly map derived here should guide detailed structural and mechanistic studies of early events in small ribosomal subunit synthesis.

The assembly process of pre-rRNA produced by Pol II from plasmids may not fully mimic that of pre-rRNA transcribed by Pol I from rDNA repeats. Yeasts containing only plasmid-derived ribosomes grew significantly slower and lack a proper nucleolus (Nogi et al. 1991). It is also unclear whether Pol II transcripts undergo cotranscriptional processing and modification similar to those shown for Pol I transcripts (Kos and Tollervey 2010).

Very recently, the Klinge group (Chaker-Margot et al. 2015) reported the protein compositions of 90S particles assembled on six 3′ MS2-tagged pre-18S RNAs. The assigned proteins are generally similar for the equivalent RNAs in the two studies. Our study additionally assigned Fcf2 and Fcf1 to the 5′ETS domain and Nsr1 to the central domain. The Klinge group's study (Chaker-Margot et al. 2015) additionally assigned Dbp8 to the 5′ domain; Fyv7, Noc1/Mak21, and Noc2 to the central domain; Cms1 to the 3′ major domain; and Rrp8, Dhr2, Slx9, Dhr1, and Nob1 to the complete 18S. In our study, Dbp8, Fyv7, and Cms1 were unassigned due to low signals, and Noc1/Mak21 and Noc2, two 60S AFs interacting with Rrp5 (Hierlmeier et al. 2013), were not detected. Some of the differences may be due to different experimental designs. In the Klinge group's study (Chaker-Margot et al. 2015), protein affinity purification was conducted before RNA affinity purification, and the MS2 tag was placed at the 3′ end of RNA (also at the 5′ end for the 5′EST RNA). We showed that placement of the MS2 tag at the 5′ or 3′ end will greatly affect the protein and RNA compositions of the purified cleavage-competent particles. Our study also describes the assembly of processing snoRNAs and the dynamic events occurring at late stages of 90S assembly.

Materials and methods

Plasmids

The pWL plasmids that carry an rDNA gene between the *GAL*7 promoter and terminator were generously provided by Skip Fournier (Liang and Fournier 1997). An RNA affinity tag containing two copies of the MS2 coat protein-binding motif followed by a tobramycin aptamer sequence (MS2-TOB tag) was described previously (Hoareau-Aveilla et al. 2011). The MS2-TOB tag was assembled by overlapping oligos, cloned into a pEASY-T vector,

and confirmed by sequencing. The MS2-TOB tag was PCR-amplified and inserted into plasmids pWL184 (ITS1-239), pWL207 (ITS1-44), and pWL109 (35S) before the first nucleotide of the pre-rRNA gene by using the transfer PCR approach (Erijman et al. 2011). Other pre-18S truncations were constructed by removing unwanted 3′ regions from longer genes with the Quik-Change method and appropriate primers (Zhang 2013). The pWL plasmids contain a 24-nt sequence tag between nucleotides 232 and 233 of 18S, which can be used to differentiate plasmid- and genome-encoded 18S RNAs. The plamsids encoding RNAs shorter than 18S-232 were constructed from the 18S-232 plasmid by deleting the sequence between the truncation site and the 18S tag, retaining the 18S tag at the 3′ end.

Purification of 90S particles

A plasmid that expresses the fusion of maltose-binding protein (MBP) and MS2 coat protein (MBP-MS2) was a gift from Melissa Moore (University of Massachusetts) (Jurica et al. 2002). The MBP-MS2 fusion protein was purified through amylose, heparin, and Q chromatography. Amylose beads (New England Biolabs) were charged with ~5 mg of MBP-MS2 protein per 1 mL of beads and washed with MS200 buffer (100 mM HEPES at pH 7.9, 200 mM KCl, 1 mM EDTA, 1 mM DTT, 10 mM β-mercaptoethanol, 0.02% NP-40).

Yeast was manipulated according to standard protocols. Unless specifically mentioned, yeast cells were grown in YPDA medium (1% yeast extract, 2% peptone, 0.003% adenine, 2% glucose) at 30°C. The UTP9-TAP/BY4741 (*MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0, UTP9-TAP::HIS3MX6*), PWP2-TAP/BY4741, and NOC4-TAP/BY4741 strains (Open Biosystems) were transformed with an rDNA plasmid and selected on Ura-deficient Synthetic Complete (SC) medium. A single clone was propagated in synthetic medium containing 2% galactose as the sole carbon source and lacking Ura (SGal-Ura), and the cells were further cultured in 6 L of YPGA medium (1% yeast extract, 2% peptone, 0.003% adenine, 2% galactose) to an OD_{600} of 1. The yeast cells were collected, washed with water, and stored at −80°C before use.

All purification steps were conducted at 4°C or on ice. Yeast cells were resuspended in 40 mL of MS200 buffer and broken with a high-pressure JN-3000 cell disruptor (JNBio). After centrifugation at 20,000*g* for 40 min, the supernatant was incubated with 500 µL of MBP-MS2 protein-loaded amylose beads for 1 h. The beads were washed with 100 mL of MS200 buffer and eluted with 1 mL of MSE buffer (100 mM HEPES at pH 7.9, 200 mM KCl, 1 mM EDTA, 2 mM DTT, 10 mM maltose) to obtain one-step-purified samples. About one-quarter was saved for mass spectrometric analysis, and the rest was incubated with 250 µL of calmodulin Sepharose beads (GE Healthcare) in 30 mL of CA100 buffer (25 mM Tris-HCl at pH 8.0, 100 mM NaCl, 1 mM MgCl_2, 1 mM imidazole, 2 mM CaCl_2, 10 mM β-mercaptoethanol, 0.02% NP-40) for 1 h. The beads were washed with 50 mL of CA100 buffer and eluted with 500 µL of CAE buffer (25 mM Tris-HCl at pH 8.0, 100 mM NaCl, 1 mM MgCl_2, 10 mM EGTA) to obtain two-step-purified samples. Alternatively, the protein affinity step was conducted with IgG beads. The eluate from the first-step purification was bound to IgG beads in buffer containing 20 mM HEPES-Na (pH 7.4), 110 mM acetate potassium, 0.5% Triton, 0.1% Tween-20, and 40mM NaCl for 30 min. The beads were washed, and the bound proteins were released by TEV protease cleavage for 12 h. RNA was extracted directly from the beads with TRIzol. The chromosomal 90S particles were purified from the UTP9-TAP/BY4741 and NOC4-TAP/BY4741 strain as described (Puig et al. 2001).

Mass spectrometric analysis

Samples were precipitated with 10% trichloroacetic acid, mixed with SDS loading buffer, heated for 2 min at 95°C, and separated briefly on 4%–20% Tris-glycine gradient SDS-PAGE gels. The gels were visualized by silver staining, and each entire gel lane was excised into three to five slices. Each slice was processed separately for mass spectrometric analysis. In brief, the gel slices were destained and digested in-gel with sequencing-grade trypsin (10 ng/µL trypsin, 50 mM ammonium bicarbonate at pH 8.0) overnight at 37°C. The peptides were sequentially extracted with 5% formic acid/50% acetonitrile and 0.1% formic acid/75% acetonitrile and then concentrated to ~20 µL. The extracted peptides were separated by an analytical capillary column (50 µm × 10 cm) packed with 5 µm of spherical C18 reversed-phase material (YMC). An Agilent 1260 capillary pump high-performance liquid chromatography (HPLC) system (Agilent Technologies) was used to generate the following HPLC gradient: 0%–5% B for 5 min, 5%–40% B for 45 min, and 40%–100% B for 13 min (A = 0.2 M acetic acid in water, and B = 0.2 M acetic acid/70% acetonitrile). The eluted peptides were sprayed into an LTQ mass spectrometer (Thermo Fisher Scientific) equipped with a nano-ESI ion source. The mass spectrometer was operated in data-dependent mode with one mass spectrometry scan followed by five tandem mass spectrometry scans for each cycle. The data files derived from the same gel lane were combined and searched on an in-house Mascot server (version 2.4, Matrix Science Ltd.) against a Uniprot *S. cerevisiae* database. Proteins were reported by their names in the database. The other search parameters were as follows: parent mass tolerance, 3 Da; product ion tolerance, 0.8 Da; protein N-term acetylation and methionine oxidation set as variable modifications; two missed enzymatic cleavage sites allowed; and peptide cutoff score, 0.05.

Protein quantification by SpC

The SpC of a protein is proportional to its amount and has been adapted as label-free semiquantitative indicators to quantify protein abundance (Zhu et al. 2010). In our analysis, the total SpC of matched peptides identifying a protein was divided by the residue number of the protein and multiplied by 100, yielding SCPHR. The SCPHR value normalizes the difference in protein length and is proportional to the absolute molar amount of protein. To compare protein abundance across different samples, the SCPHR value was further normalized against a set of reference proteins, yielding the RSAF. Specially, the RSAF of a protein was calculated by its SCPHR value divided by the averaged SCPHR value of reference proteins. The ideal reference proteins were common and stoichiometric components of purified 90S particles, and an RSAF of near 1 indicates a stoichiometric component of 90S particles. As the bait protein and its tightly associated proteins are often enriched, they are not ideal reference proteins. Under these considerations, the reference proteins were set as six UTPB proteins for most samples and seven UTPA proteins for samples purified via Pwp2-TAP (a component of UTPB). For the 5′ETS 281 and 5′ETS 309 samples, the UTPA proteins were the only major 90S proteins and the only choice of reference proteins. Proteins that are known to be present in the same subcomplex—for example U3 snoRNP, UTPA, and UTPB—normally have similar RSAF values, validating the use of RSAF as a semiquantitative indicator of protein abundance. To estimate the molar fraction of a group of proteins of interest, their SCPHR values were summed and divided by the total SCPHR value of all detected proteins. When a particle was purified multiple times, sometimes using different protein tags, the result with the highest fraction of 90S proteins was reported.

Many r-proteins have two closely related versions, A and B, in *S. cerevisiae*. As the two versions share almost identical sequences, they commonly have similar peptides detected and SpCs in mass spectroscopy. However, if one version lacks a unique peptide, it becomes undetected because all common peptides will be assigned to the other version. To correctly estimate the abundance of r-proteins, only the version with higher SpCs was counted, and it was always called version A.

Northern and Western blot analyses

RNA extraction and Northern blotting were performed as described (Lin et al. 2013). For snoRNA analysis, RNA was resolved in 8% polyacrylamide–8 M urea gels. For large RNA analysis, RNA was separated in 1.2% agarose–formaldehyde gels. The following probes were used for hybridization: 18S tag (5′-CGC CGAGGATCCAACTAGGGGGCT-3′), MS2 tag (5′-CGTACCC TGATGGTGTACGCC-3′), U3 (5′-GGATTGCGGACCAAGC TAA-3′), U14 (5′-TCACTCAGACATCCTAGG-3′), snR30 (5′- ATGTCTGCAGTATGGTTTTAC-3′), snR10 (5′-GTGTTACGA ATGGCTGTTA-3′), D-A2 (5′-CGGTTTTAATTGTCCTA-3′), A0-A1 (5′-AAAGAAACCGAAATCTCTTT-3′), and 5′-A0 (5′- CGAACGACAAGCCTACTCG-3′). Western blotting was conducted as described (Lin et al. 2013). The TAP tag was detected with the peroxidase–anti-peroxidase antibody (Sigma).

Acknowledgments

References

Beltrame M, Tollervey D. 1995. Base pairing between U3 and the pre-ribosomal RNA is required for 18S rRNA synthesis. *EMBO J* **14:** 4350–4356.

Ben-Shem A, Garreau de Loubresse N, Melnikov S, Jenner L, Yusupova G, Yusupov M. 2011. The structure of the eukaryotic ribosome at 3.0 A resolution. *Science* **334:** 1524–1529.

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.

Buchhaupt M, Kotter P, Entian KD. 2007. Mutations in the nucleolar proteins Tma23 and Nop6 suppress the malfunction of the Nep1 protein. *FEMS Yeast Res* **7:** 771–781.

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.

de la Cruz J, Karbstein K, Woolford JL Jr. 2015. Functions of ribosomal proteins in assembly of eukaryotic ribosomes in vivo. *Annu Rev Biochem* **84:** 93–129.

de la Cruz J, Kressler D, Tollervey D, Linder P. 1998. Dob1p (Mtr4p) is a putative ATP-dependent RNA helicase required for the 3′ end formation of 5.8S rRNA in *Saccharomyces cerevisiae*. *EMBO J* **17:** 1128–1140.

Dosil M, Bustelo XR. 2004. Functional characterization of Pwp2, a WD family protein essential for the assembly of the 90 S pre-ribosomal particle. *J Biol Chem* **279:** 37385–37397.

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

Dutca LM, Gallagher JE, 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.

Ebersberger I, Simm S, Leisegang MS, Schmitzberger P, Mirus O, von Haeseler A, Bohnsack MT, Schleiff E. 2014. The evolution of the ribosome biogenesis pathway from a yeast perspective. *Nucleic Acids Res* **42:** 1509–1523.

Erijman A, Dantes A, Bernheim R, Shifman JM, Peleg Y. 2011. Transfer-PCR (TPCR): a highway for DNA cloning and protein engineering. *J Struct Biol* **175:** 171–177.

Eschrich D, Buchhaupt M, Kotter P, Entian KD. 2002. Nep1p (Emg1p), a novel protein conserved in eukaryotes and archaea, is involved in ribosome biogenesis. *Curr Genet* **40:** 326–338.

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.

Ferreira-Cerca S, Poll G, Gleizes PE, Tschochner H, Milkereit P. 2005. Roles of eukaryotic ribosomal proteins in maturation and transport of pre-18S rRNA and ribosome function. *Mol Cell* **20:** 263–275.

Freed EF, Bleichert F, Dutca LM, Baserga SJ. 2010. When ribosomes go bad: diseases of ribosome biogenesis. *Mol Biosyst* **6:** 481–493.

Gallagher JE, Dunbar DA, Granneman S, Mitchell BM, Osheim Y, Beyer AL, Baserga SJ. 2004. RNA polymerase I transcription and pre-rRNA processing are linked by specific SSU processome components. *Genes Dev* **18:** 2506–2517.

Garcia-Gomez JJ, Babiano R, Lebaron S, Froment C, Monsarrat B, Henry Y, de la Cruz J. 2011. Nop6, a component of 90S pre-ribosomal particles, is required for 40S ribosomal subunit biogenesis in *Saccharomyces cerevisiae*. *RNA Biol* **8:** 112–124.

Grandi P, Rybin V, Bassler J, Petfalski E, Strauss D, Marzioch M, Schafer 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, 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.

Henras AK, Soudet J, Gerus M, Lebaron S, Caizergues-Ferrer M, Mougin A, Henry Y. 2008. The post-transcriptional steps of eukaryotic ribosome biogenesis. *Cell Mol Life Sci* **65:** 2334–2359.

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.

Hoareau-Aveilla C, Fayet-Lebaron E, Jady BE, Henras AK, Kiss T. 2011. Utp23p is required for dissociation of snR30 small nucleolar RNP from preribosomal particles. *Nucleic Acids Res* **40:** 3641–3652.

Hontz RD, Niederer RO, Johnson JM, Smith JS. 2009. Genetic identification of factors that modulate ribosomal DNA

transcription in *Saccharomyces cerevisiae*. *Genetics* **182:** 105–119.

Hughes JM. 1996. Functional base-pairing interaction between highly conserved elements of U3 small nucleolar RNA and the small ribosomal subunit RNA. *J Mol Biol* **259:** 645–654.

Jurica MS, Licklider LJ, Gygi SR, Grigorieff N, Moore MJ. 2002. Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA* **8:** 426–439.

Karkusiewicz I, Rempola B, Gromadka R, Grynberg M, Rytka J. 2004. Functional and physical interactions of Faf1p, a *Saccharomyces cerevisiae* nucleolar protein. *Biochem Biophys Res Commun* **319:** 349–357.

Kondo K, Inouye M. 1992. Yeast NSR1 protein that has structural similarity to mammalian nucleolin is involved in pre-rRNA processing. *J Biol Chem* **267:** 16252–16258.

Kos 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.

Kos M, Tollervey D. 2010. Yeast pre-rRNA processing and modification occur cotranscriptionally. *Mol Cell* **37:** 809–820.

Kressler D, Hurt E, Bassler J. 2010. Driving ribosome assembly. *Biochim Biophys Acta* **1803:** 673–683.

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.

Lebaron S, Segerstolpe A, 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.

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:** 5441–5452.

Lee WC, Zabetakis D, Melese T. 1992. NSR1 is required for pre-rRNA processing and for the proper maintenance of steady-state levels of ribosomal subunits. *Mol Cell Biol* **12:** 3865–3871.

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

Liang WQ, 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, Clark JA, Fournier MJ. 1997. The rRNA-processing function of the yeast U14 small nucleolar RNA can be rescued by a conserved RNA helicase-like protein. *Mol Cell Biol* **17:** 4124–4132.

Liang XH, Liu Q, King TH, Fournier MJ. 2010. Strong dependence between functional domains in a dual-function snoRNA infers coupling of rRNA processing and modification events. *Nucleic Acids Res* **38:** 3376–3387.

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:** e1001669.

Lu J, Sun M, Ye K. 2013. Structural and functional analysis of Utp23, a yeast ribosome synthesis factor with degenerate PIN domain. *RNA* **19:** 1815–1824.

Lygerou Z, Mitchell P, Petfalski E, Seraphin 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.

Marmier-Gourrier N, Clery A, Schlotter F, Senty-Segault 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.

Meyer B, Wurm JP, Kotter 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 Psi1191 in yeast 18S rRNA. *Nucleic Acids Res* **39:** 1526–1537.

Milkereit P, Strauss D, Bassler J, Gadal O, Kuhn H, Schutz S, Gas N, Lechner J, Hurt E, Tschochner H. 2003. A Noc complex specifically involved in the formation and nuclear export of ribosomal 40 S subunits. *J Biol Chem* **278:** 4072–4081.

Miller OL Jr, Beatty BR. 1969. Visualization of nucleolar genes. *Science* **164:** 955–957.

Mizushima S, Nomura M. 1970. Assembly mapping of 30S ribosomal proteins from *E. coli*. *Nature* **226:** 1214.

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* **88:** 3962–3966.

Osheim YN, French SL, Keck KM, Champion EA, Spasov K, Dragon F, Baserga SJ, Beyer AL. 2004. Pre-18S ribosomal RNA is structurally compacted into the SSU processome prior to being cleaved from nascent transcripts in *Saccharomyces cerevisiae*. *Mol Cell* **16:** 943–954.

Peng WT, Krogan NJ, Richards DP, Greenblatt JF, Hughes TR. 2004. ESF1 is required for 18S rRNA synthesis in *Saccharomyces cerevisiae*. *Nucleic Acids Res* **32:** 1993–1999.

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.

Perez-Fernandez J, Martin-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.

Phipps KR, Charette JM, Baserga SJ. 2011. The SSU processome in ribosome biogenesis—progress and prospects. *Wiley Interdiscip Rev RNA* **2:** 1–21.

Puig O, Caspary F, Rigaut G, Rutz B, Bouveret E, Bragado-Nilsson E, Wilm M, Seraphin B. 2001. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **24:** 218–229.

Rempola B, Karkusiewicz I, Piekarska I, Rytka J. 2006. Fcf1p and Fcf2p are novel nucleolar *Saccharomyces cerevisiae* proteins involved in pre-rRNA processing. *Biochem Biophys Res Commun* **346:** 546–554.

Sasaki T, Toh EA, Kikuchi Y. 2000. Yeast Krr1p physically and functionally interacts with a novel essential Kri1p, and both proteins are required for 40S ribosome biogenesis in the nucleolus. *Mol Cell Biol* **20:** 7971–7979.

Segerstolpe A, Granneman S, Bjork P, de Lima Alves F, Rappsilber J, Andersson C, Hogbom M, Tollervey D, Wieslander L. 2012. Multiple RNA interactions position Mrd1 at the site of the small subunit pseudoknot within the 90S pre-ribosome. *Nucleic Acids Res* **41:** 1178–1190.

Shirai C, Takai T, Nariai M, Horigome C, Mizuta K. 2004. Ebp2p, the yeast homolog of Epstein-Barr virus nuclear antigen 1-binding protein 2, interacts with factors of both the 60 S and the 40 s ribosomal subunit assembly. *J Biol Chem* **279:** 25353–25358.

Strunk BS, Loucks CR, Su M, Vashisth H, Cheng S, Schilling J, Brooks CL III, Karbstein K, Skiniotis G. 2011. Ribosome assembly factors prevent premature translation initiation by 40S assembly intermediates. *Science* **333:** 1449–1453.

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:** 2165–2174.

Tollervey D. 1987. A yeast small nuclear RNA is required for normal processing of pre-ribosomal RNA. *EMBO J* **6:** 4169–4175.

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.

Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, et al. 2000. A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403:** 623–627.

Venema J, Tollervey D. 1996. RRP5 is required for formation of both 18S and 5.8S rRNA in yeast. *EMBO J* **15:** 5701–5714.

Watkins NJ, Segault V, Charpentier B, Nottrott S, Fabrizio P, Bachi A, Wilm M, Rosbash M, Branlant C, Luhrmann R. 2000. A common core RNP structure shared between the small nucleoar box C/D RNPs and the spliceosomal U4 snRNP. *Cell* **103:** 457–466.

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:** 1254–1267.

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

Young CL, Khoshnevis S, Karbstein K. 2013. Cofactor-dependent specificity of a DEAD-box protein. *Proc Natl Acad Sci* **110:** E2668–E2676.

Zhang L. 2013. "Assembly of early precursors of yeast small ribosomal subunit and structural and functional analysis of Rrp9." PhD thesis, Graduate School of Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing.

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.

Zhu W, Smith JW, Huang CM. 2010. Mass spectrometry-based label-free quantitative proteomics. *J Biomed Biotechnol* **2010:** 840518.