2096–2106 *Nucleic Acids Research, 2018, Vol. 46, No. 4* *Published online 18 January 2018*
doi: 10.1093/nar/gky011

Structural and functional analysis of ribosome assembly factor Efg1

Sheng Shu[1,2,3] and Keqiong Ye[1,3,4,*]

[1]Graduate School of Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100730, China, [2]National Institute of Biological Sciences, Beijing 102206, China, [3]Key Laboratory of RNA Biology, CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China and [4]University of Chinese Academy of Sciences, Beijing 100049, China

Received September 11, 2017; Revised December 07, 2017; Editorial Decision January 03, 2018; Accepted January 06, 2018

ABSTRACT

Ribosome biogenesis in eukaryotes is a complicated process that involves association and dissociation of numerous assembly factors and snoRNAs. The yeast small ribosomal subunit is first assembled into 90S pre-ribosomes in an ordered and dynamic manner. Efg1 is a protein with no recognizable domain that is associated with early 90S particles. Here, we determine the crystal structure of Efg1 from *Chaetomium thermophilum* at 3.3 Å resolution, revealing a novel elongated all-helical structure. Efg1 is not located in recently determined cryo-EM densities of 90S likely due to its low abundance in mature 90S. Genetic analysis in *Saccharomyces cerevisiae* shows that the functional core of Efg1 contains two helical hairpins composed of highly conserved residues. Depletion of Efg1 blocks 18S rRNA processing at sites A1 and A2, but not at site A0, and production of small ribosomal subunits. Efg1 is initially recruited by the 5′ domain of 18S rRNA. Its absence disturbs the assembly of the 5′ domain and inhibits release of U14 snoRNA from 90S. Our study shows that Efg1 is required for early assembly and reorganization of the 5′ domain of 18S rRNA.

INTRODUCTION

Ribosomes are essential and conserved nanomachines in all organisms and are responsible for proteins synthesis. Consisting of four ribosomal RNAs (rRNAs) and 79 ribosomal proteins, the ribosome in *Saccharomyces cerevisiae* is assembled in a highly complicated and dynamic process (1–3). This process begins with the transcription of 35S precursor rRNA (pre-rRNA) by RNA polymerase I (Pol I) and 5S rRNA by Pol III. The 35S pre-rRNA is processed into the 18S, 5.8S and 25S rRNAs following removal of four external and internal transcribed spacers (ETS and ITS). In addition to binding of ribosomal proteins, the pre-rRNA also associates with >200 *trans*-acting assembly factors (AFs) and small nucleolar RNAs (snoRNAs) during ribosome maturation. A series of pre-ribosomal particles of the small 40S and large 60S subunits are formed, coupled with pre-rRNA processing, structural reorganization and translocation from the nucleolus through the nucleoplasm to the cytoplasm.

Early assembly of both small and large subunits occurs in a stepwise manner during the transcription of pre-rRNA (4–7). The 5′ region of 35S pre-rRNA is assembled into the 90S pre-ribosome or the small subunit processome (8,9) and the 3′ region is packed into early pre-60S ribosomes. Nearly 70 AFs and U3, U14 and snR30 snoRNAs associate in an orderly manner with the pre-rRNA to form the 90S (4,5). Some AFs, U14 and snR30 snoRNAs that are recruited by the 18S rRNA sequence are released at late assembly stages of 90S (4). The 5′ ETS region nucleates the assembly of UTPA, UTPB and U3 snoRNP subcomplexes and other AFs. The recently determined cryo-EM structures of 90S have shown that the 5′ ETS and associated factors form a large base organizing several partially folded subdomains of 40S ribosomes (10–12). Within the 90S, the 35S pre-rRNA is cleaved at sites A0 and A1 of 5′ ETS and at site A2 of ITS1. Following dramatic structural reorganization, a pre-40S ribosome is released, exported to the cytoplasm and matures into 40S subunits. The 35S pre-rRNA can also be cleaved at site A3 of ITS1, yielding a 90S pre-ribosome containing 23S pre-rRNA.

Efg1 is a protein component of 90S pre-ribosomes (4,5) and contains an uncharacterized domain (DUF2361 or Efg1 domain). The *efg1* deletion strain is significantly delayed in exiting from the G_1 phase after α-factor arrest, from which the gene name EFG1 (Exit From G_1) is derived (13). Efg1 is required for efficient 18S rRNA processing and its depletion causes accumulation of 35S and 23S pre-rRNAs and reduction of 20S pre-rRNA (14). Efg1 deletion is synthetic lethal with mutants of Emg1, another 90S AF and leads to increased sensitivity to paromomycin (15). During the course of stepwise assembly of 90S, Efg1 is re-

*To whom correspondence should be addressed. Tel: +84 10 6488 7672; Email: yekeqiong@ibp.ac.cn

cruited by the 5′ domain sequence of 18S rRNA (4,5). In the cryo-EM densities of 90S, many AFs, including Efg1, were not located and many densities were unassigned (10–12). It is unclear whether and where Efg1 is located in the cryo-EM density of 90S in the absence of the Efg1 structure.

In this study, we determine the crystal structure of Efg1 from *Chaetomium thermophilum* (ctEfg1) and analyze the function of Efg1 in pre-rRNA processing and 90S assembly using the model organism *S. cerevisiae*. We show that Efg1 is required for proper assembly of the 5′ domain of 18S rRNA and release of U14 snoRNA.

MATERIALS AND METHODS

Gene cloning and protein purification

The ctEfg1 gene was amplified from *C. thermophilum* genomic DNA by joining two exon sequences with overlap polymerase chain reaction (PCR) and cloned into a modified pET28a vector. ctEfg1 was expressed as a fusion protein with an N-terminal His_6-Smt3 tag. The construct was confirmed by DNA sequencing. The vector was transformed into the *Escherichia coli* Rosetta(DE3) strain (Cwbiotech). Selenomethione (SeMet)-substituted ctEfg1 was expressed in M9 medium as described previously (16). After the cells were grown at 37°C to A600 of ∼0.8, the temperature was shifted to 18°C and the protein expression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside overnight. The cells were harvested with centrifugation, resuspended in buffer A (50 mM Tris–HCl, pH 8.0 and 500 mM NaCl) and lysed by sonication. After centrifugal clarification, the supernatant was filtered through 0.45 μm filters and loaded onto a HisTrap column (GE Healthcare). The column was washed with buffer A containing 25 mM imidazole and eluted with a linear gradient of 25–500 mM imidazole in buffer A. All buffers used in the following purification procedures included 1 mM dithiothreitol (DTT). The His_6-Smt3 tag was cleaved with Ulp1 proteinase for 1 h at 4°C. The cleavage reaction was diluted with 25 mM HEPES buffer, pH 7.6, to a NaCl concentration of 250 mM and loaded onto a heparin column (GE Healthcare). The column was eluted with a linear gradient of 250–1000 mM KCl in 25 mM of HEPES buffer, pH 7.6. The fractions containing the target protein were pooled and concentrated to ∼1 ml. The protein was further purified with a HiLoad 16/60 Superdex 200 column (GE Healthcare) running in buffer C (10 mM Tris–HCl, pH 8.0, 200 mM NaCl and 1 mM DTT).

Crystallization

Initial crystallization screens were performed with the sitting-drop vapor diffusion method at 18°C by mixing 100 nl of protein solution with 100 nl of reservoir solution using a Mosquito robot (TTP Labtech). The initial crystallization conditions were optimized with the hanging-drop vapor diffusion method by mixing 1 μl each of protein and reservoir solutions. Crystals of SeMet-labeled ctEfg1 (16 mg/ml in 10 mM Tris–HCl, pH 8.0, 200 mM NaCl and 1 mM DTT) were grown from 2.4 M sodium malonate (pH 7.0). The crystals were cryoprotected in 20% glycerol prepared in the reservoir solution and flash cooled in liquid nitrogen.

Table 1. Data collection and refinement statistics

	SeMet-ctEfg1
Data collection	
Space group	$P6_2$
Cell dimensions	
a, b, c (Å)	99.1, 99.1, 90.2
α, β, γ (°)	90, 90, 120
Wavelength (Å)	0.9792
Resolution range (Å)	50–3.30 (3.36–3.30)
Unique reflections	7701(385)
Redundancy	22 (20.1)
$\langle I \rangle / \langle \sigma(I) \rangle$	19.3 (2.6)
Completeness (%)	99.9 (100)
R-merge	0.362 (>1)
CC1/2	(0.952)
CC*	(0.988)
Refinement	
Resolution range (Å)	39–3.3 (3.55–3.3)
No. of reflections	7381 (1371)
R-work	0.198 (0.277)
R-free	0.252 (0.351)
No. of atoms	1493
Average B-factor	31.37
RMSD bond length (Å)	0.010
RMSD bond angles (°)	1.077
Ramachandran plot	
Favored (%)	90.29
Allowed (%)	9.14
Outliers (%)	0.57

Statistics for the highest-resolution shell are shown in parentheses.

Data collection and structure determination

Diffraction data were collected at beamline BL17U of the Shanghai Synchrotron Radiation Facility and processed by HKL2000 (17). The statistics for data processing and structural refinement are summarized in Table 1. The crystal belongs to space group $P6_2$ and contains one molecule per asymmetric unit. The structure was determined with the single-wavelength anomalous dispersion (SAD) method using Autosol and Autobuild in PHENIX (18). The model was further manually adjusted in COOT (19) and refined in PHENIX. The final model contains residues 47–173 and 183–234 of ctEfg1.

Yeast plasmids, strains and medium

Yeast experiments were performed according to standard protocols. Yeast cells were grown in YPD (1% yeast extract, 2% peptone, 2% glucose), YPG (1% yeast extract, 2% peptone, 2% galactose), Synthetic Complete (SC) medium or appropriate SC dropout medium (Clontech). The *EFG1* gene was cloned into plasmid pRS415 under the *GPD* promoter with an N-terminal FLAG tag, yielding pRS415-GPD-FLAG-EFG1. Plasmids were constructed by the In-Fusion approach and deletion mutations were introduced by the QuikChange method. All plasmids were confirmed by DNA sequencing.

All yeast strains were derived from BY4741 (Mat a, leu2Δ0, Met15Δ0, ura3Δ0). The strains NOC4-TAP (BY4741, NOC4-TAP::His3MX), ENP1-TAP (BY4741, ENP1-TAP::His3MX) and UTP9-TAP (BY4741, UTP9-TAP::His3MX) were purchased from Open Biosystems. To construct conditional expression strains (GAL::HA-

efg1), NOC4-TAP or ENP1-TAP strains were transformed with a PCR product generated from the natNT2-GALL cassette in plasmid pYM-N28 (20). The positive clones were selected by the antibiotic natNT2 and confirmed by PCR. The efg1Δ/UTP9-TAP strain (BY4741, UTP9-TAP::His3MX, efg1Δ::natNT2) was constructed from the UTP9-TAP strain by homologous recombination replacing the *EFG1* gene with a PCR product generated from plasmid pYM-N28.

Yeast functional assays

The GAL::HA-efg1/ENP1-TAP strain was transformed with a pRS415-GPD plasmid expressing full length or truncated Efg1. Single clones of the transformants of similar size were cultured, 5-fold serially diluted and spotted onto plates containing SC medium with galactose or glucose. The plates were incubated at 30 or 37°C for 2 days or at 18°C for 4 days. Strains UTP9-TAP, efg1Δ/UTP9-TAP and efg1Δ/UTP9-TAP complemented with pRS415-GPD-EFG1 were spotted on YPD plates.

Immunoprecipitation was conducted as described below. For western blot analysis, proteins were separated in 4–20% sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) gels and transferred to 0.45 μm polyvinylidene fluoride (PVDF) membranes (GE Healthcare) using a semi-dry electrophoretic transfer cell (Bio-Rad). HRP-conjugated anti-Flag (1:10000, Sigma) and HRP-conjugated anti-HA antibodies (1:3000, CST) were used with appropriate dilution ratios. Ribosome profile assays were performed as previously described (21).

Purification of pre-ribosomes

The GAL::efg1/NOC4-TAP strain was first grown in YPG and then shifted to YPD medium to grow for 14 h at 30°C. Cells were collected by centrifugation and resuspended in one cell volume of lysis buffer (20 mM HEPES-KOH, pH 7.6, 110 mM K-acetate, 40 mM NaCl, 0.5% Triton and 0.1% Tween-20) supplemented by ethylenediaminetetraacetic acid-free protease inhibitor cocktail (Roche). The cells were lysed using steel balls and clarified by centrifugation. The supernatant was incubated with rabbit IgG-coated magnetic Beaver beads (Beaverbio) for 30 min. The beads were washed five times with 1 ml of lysis buffer and submitted to mass spectrometric analysis or northern blot analysis. The incomplete 90S particles assembled on MS2-tagged pre-rRNA fragments were purified as described (4).

Mass spectrometric analysis

Samples bound to IgG-coated beads were boiled at 95°C for 5 min in SDS loading buffer, resolved in SDS-PAGE and visualized with silver staining. Mass spectrometric analysis was conducted as described (6). The total spectral counts per 100 residues (SCPHR) were calculated for each identified protein and further normalized against UTPB proteins, yielding the relative spectral abundance factor (RSAF) (Supplementary Dataset 1) (4).

Northern blot

RNA extraction and northern blotting were carried out as described (21). For small RNA analysis, RNAs were resolved in 8% polyacrylamide-8 M urea gels and transferred to Hybond N^+ membranes (GE Healthcare) with semi-dry blotting. For large RNA analysis, RNAs were separated in 1.2% agarose-formaldehyde gels and transferred to membranes with 60 mBar vacuum (GE healthcare). The following probes were used for hybridization: D-A2 (5′-CGGTTTTAATTGTCCTA-3′), A0-A1 (5′-AAAGA AACCGAAATCTCTTT-3′), A2-A3 (5′-ATGAAAACT CCACAGTG-3′), 5′-A0 (5′-GGAAATGCTCTCTGTTC AAAAAGCTTTTACACTCTTGACCAGCGCACTC C-3′), snR30 (5′-ATGTCTGCAGTATGGTTTTAC-3′), U3 (5′-GGATTGCGGACCAAGCTAA-3′), snR10 (5′-G TGTTACGAATGGCTGTTA-3′), U14 (5′-TCACTCA GACATCCTAGG-3′).

Electrophoretic mobility shift assay

U14 snoRNA and the helix 10 region of 18S rRNA (18S-h10, 5′-UACGCAUGGCCUUGUGCUGGCGA-3′) were prepared by in vitro transcription and dephosphorylated with alkaline phosphatase (Takara). An 11-nt RNA (5′-CCAUGAGUGUU-3′) was purchased from Takara. The RNAs were 5′-^{32}P-labeled with T4 polynucleotide kinase (NEB) and purified with MicroSpin G-25 columns (GE Healthcare). U14 and 18S-h10 were annealed by heating at 95°C and slowly cooling down to room temperature. RNAs (∼0.2 nM) were mixed with 2-fold serial dilutions (0.03–8000 nM) of Efg1 in 20 μl of binding buffer (25 mM HEPES pH 7.6, 200 mM KCl, 2 mM $MgCl_2$, 0.01% NP40, 10% glycerol). The reactions were incubated at room temperature for 30 min and resolved in a native polyacrylamide gel run in Tris-glycine buffer (pH 8.3) at room temperature. The gels were dried and visualized with a Typhoon PhosphorImager (GE Healthcare).

RESULTS

Crystal structure of ctEfg1

We initially attempted to crystallize Efg1 from *S. cerevisiae* (scEfg1) but were not successful. We then shifted to ctEfg1 from the thermophilic fungus *C. thermophilum*, whose proteins are frequently used for structural studies due to stability (22). The full-length protein (310 residues) of SeMet-substituted ctEfg1 was purified and crystalized. The structure was determined with the SAD method and refined at 3.3 Å resolution to an R_{work}/R_{free} of 0.198/0.252 with good geometry (Figure 1A and Table 1). The asymmetric unit of crystal contains one molecule with residues 47–173 and 183–234 resolved (Figure 2).

ctEfg1 folds into an elongated rod-shaped structure composed of eight α-helices with the longest dimension of 87 Å (Figure 1A). The structure is composed of two helical hairpins (or anti-parallel coiled-coils) arranged in nearly opposite directions and a string of C-terminal α-helices bound at the center. Helices α1 and α2 fold back into hairpin 1 and helices α3 and α4 form hairpin 2. Helix α2 and α3 are immediately joined at an angle of 140°. A string of short

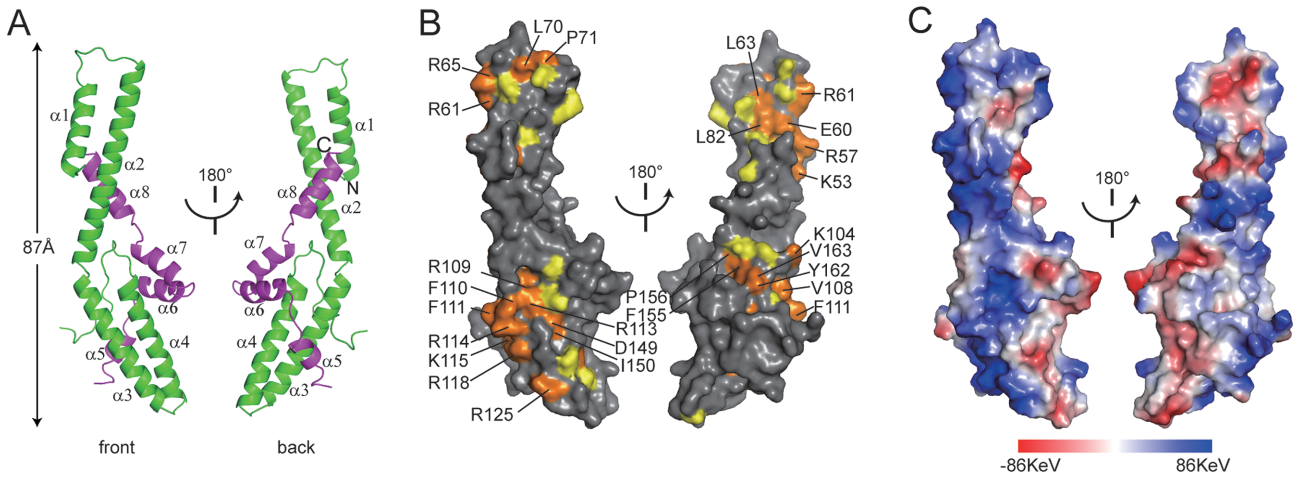

Figure 1. Crystal structure of ctEfg1. (**A**) Ribbon representation of ctEfg1 structure in front and back views. The core domain is colored green and the functionally dispensable region is colored magenta. Secondary structures and the N- and C-termini are labeled. (**B**) Conservation surface. Residues with 100 and 80% conservation, as defined in Figure 2, are colored in orange and yellow, respectively. Residues with 100% conservation are labeled. (**C**) Electrostatic potential surface. The surface is colored from blue to red for positively to negatively charged regions. The structural orientations are the same in A–C.

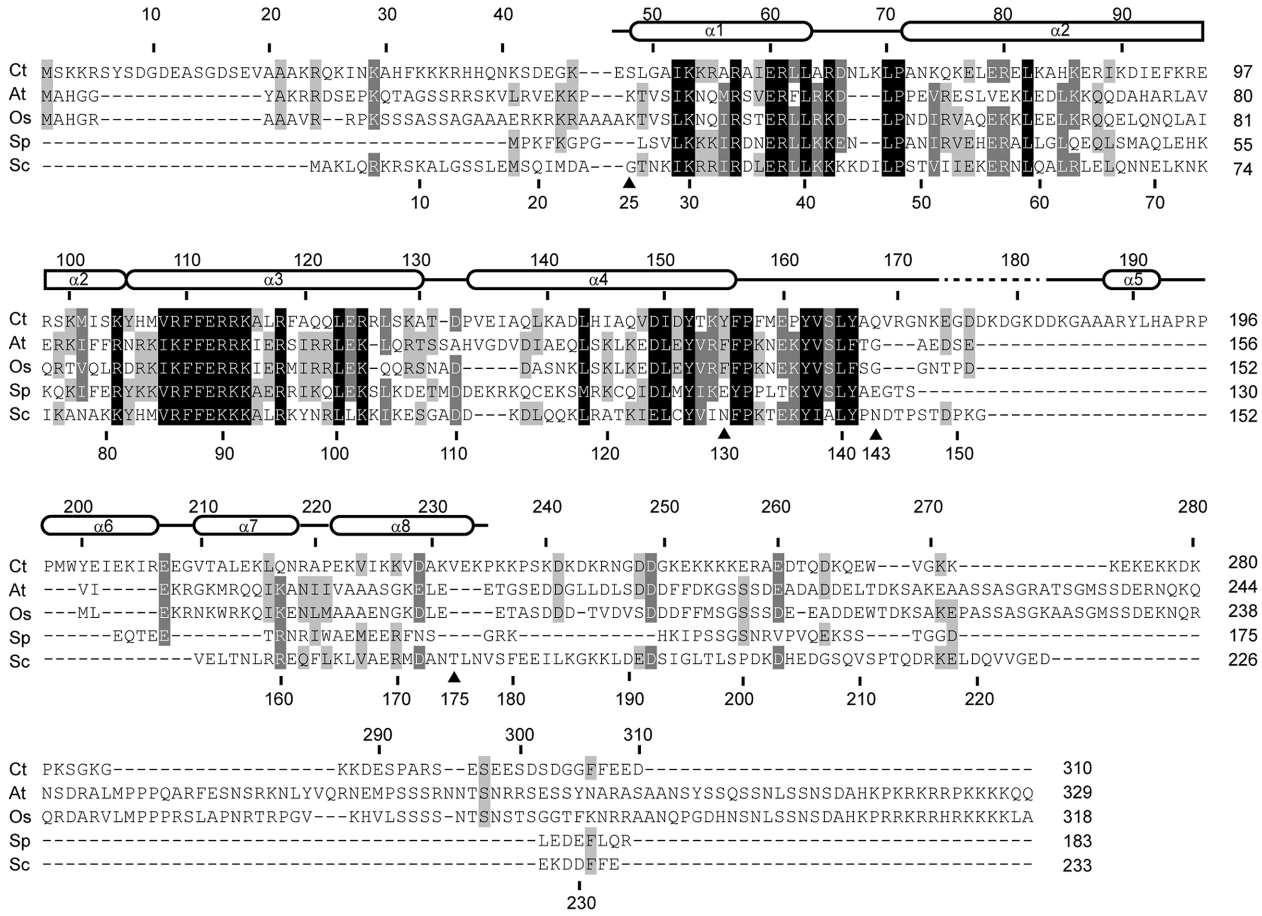

Figure 2. Multiple sequence alignment of Efg1. Homologous sequences of Efg1 from *Chaetomium thermophilum* (Ct), *Arabidopsis thaliana* (At), *Oryza sativa* (Os), *Schizosaccharomyces pombe* (Sp) and *Saccharomyces cerevisiae* (Sc) are aligned and displayed in GeneDoc with manual adjustment. The residues with 100, 80 and 60% conservation are shaded in black, gray and light gray, respectively. Similarity groups are defined as follows: D and E; N and Q; S and T; K and R; F, Y and W; and I, L, V and W. The secondary structures observed in the ctEfg1 structure are displayed on the top. The truncation sites of scEfg1 deletion mutants are marked with triangles.

helices, α5–α8, runs from the second to first hairpin along the back face of the structure. As a result, the N- and C-termini of the structure are in close proximity. We searched for structural homologs of ctEfg1 with the Dali server and found many hits (*Z*-score up to 7.2) (23). However, these structures contain only a single helical hairpin, which is a common structural motif, and no structure resembles the overall structure of ctEfg1.

A search of Efg1 homologs with PHI-BLAST retrieved hundreds of sequences from yeasts and plants but few from metazoans (14 hits, most of which were false hits with large *E*-values). Multiple sequence alignment of representative sequences showed that the residues around helices α1–α4 are highly conserved, whereas the residues in helices α5–α8 and the unstructured N- and C-terminal tails are poorly aligned (Figure 2). The most conserved residues are clustered on the tip of hairpin 1, the front face of hairpin 2 and the loop C-terminal of helix α4, suggesting that these regions are functionally important (Figure 1B). Efg1 may function as an RNA-binding protein as the structure contains over ten highly conserved basic residues (Figure 2) and the front face of the structure is highly positively charged (Figure 1C). Indeed, electrophonic mobility shift assay showed that Efg1 was capable of binding various RNAs, including U14 snoRNA, helix 10 of 18S rRNA and a short single-stranded RNA (Supplementary Figure S1). The binding was non-specific since U14 was completely competed off by excessive tRNA.

Functional core of Efg1

To assess the function of Efg1, we made conditional expression strains on the mother *S.* cerevisiae strain where Noc4 or Enp1, which are AFs in 90S, was fused to a C-terminal tandem affinity purification (TAP) tag. The genomic *EFG1* gene was placed under a galactose-dependent *GAL* promoter. The protein level of Efg1 was efficiently depleted after transfer to glucose medium for 9 h (Figure 3A). Upon depletion of Efg1, the yeast growth was significantly impaired, particularly at 37 and 18°C (Figure 3B), as previously reported (13). The *efg1Δ* strain on the UTP9-TAP background also showed a severe growth defect (Figure 3C).

To determine the functional domain of Efg1, several deletion mutants were expressed from plasmid and assessed for function (Figure 3B). The growth defect caused by Efg1 depletion was rescued by the wild-type (WT) Efg1 or its fragments that lack the residues equivalent to the unstructured tails and helices 5–8 in the ctEfg1 structure (fragments 1–175, 25–175 and 25–143). Although the 25–143 fragment was expressed at a low level (Figure 3D), it fully supported growth except at 37°C. The 25–143 fragment was co-immunoprecipitated by Noc4-TAP, indicating that it assembled into 90S pre-ribosomes (Figure 3D). A shorter fragment with residues 25–130 failed to rescue the growth defect. However, this fragment was not expressed in yeast (Figure 3D). The additionally deleted residues 131–143 in this fragment form the C-terminal loop of helix α4 that wraps around helix α3 in the ctEfg1 structure and appear to be important for protein stability. These data show that the functional core of Efg1 is composed of helices α1-α4 and an ensuing loop. The core domain of Efg1 is also the most evolutionarily conserved region (Figure 2).

It is notable that the *EFG1* conditional expression strains grew more slowly in galactose medium than the WT strains (Figure 3E). This indicates that overexpression of Efg1 has a dominant negative phenotype.

Efg1 is required for 40S formation and 18S rRNA processing

Ribosomal profile analysis showed that depletion of Efg1 led to disappearance of free 40S subunits and excessive free 60S subunits due to a shortage of 40S subunits for association (Figure 3F and G). This indicates that Efg1 is specifically required for production of 40S subunits, but not for 60S subunits.

To study the role of Efg1 in pre-rRNA processing (Supplementary Figure S2), we conducted northern blot analyses to examine pre-rRNA processing intermediates in total RNAs purified from WT and ΔEfg1 cells. Upon depletion or deletion of Efg1, 18S rRNA was significantly reduced, whereas the 25S rRNA levels were not changed (Figure 4A, lanes 1–3 and 7–8, EB staining), indicating that Efg1 specifically functions in 18S rRNA processing. The absence of Egf1 led to a strong accumulation of 35S and 23S pre-rRNAs and blocked production of 20S and 27SA2 pre-rRNAs (Figure 4A, lanes 1–3 and 7–8). The 22S pre-rRNA, an A0 cleavage product of 23S pre-rRNA, was detectable. These findings indicate that cleavage at the A1 and A2 sites is strongly inhibited by the absence of Efg1, and A0 cleavage still occurs. The 18S rRNA processing defect was largely rescued by expression of WT Efg1 or its functional core (Efg1 25–143) but was not rescued by the fragment Efg1 25–130 (Figure 4A, lanes 4–6). Overexpression of Efg1 in galactose also suppressed 18S rRNA processing, leading to reduction of 27SA2, 22S and 20S pre-rRNAs and accumulation of 23S pre-rRNA (Figure 4A, lane 2).

We also analyzed pre-rRNAs present in the purified 90S pre-ribosomes. The WT 90S pre-ribosomes purified via Noc4-TAP and Utp9-TAP contain various processing intermediates, including 35S, 33S, 23S, 22S and 20S pre-rRNAs (Figure 4B, lanes 1, 2 and 7). In the absence of Efg1, 35S and 23S pre-rRNAs were highly abundant, 33S and 22S pre-rRNAs produced by A0 cleavage were at similar levels as the WT yeast, and 20S pre-rRNA appeared to be absent. The A0 cleavage of pre-rRNA also produces a 5′-A0 fragment that is subsequently degraded by the exosome (24). The level of 5′-A0 fragment in ΔEfg1 90S was similar to that in WT 90S (Figure 5B), suggesting that A0 cleavage and subsequent degradation of 5′-A0 fragment are not significantly affected by Efg1 depletion.

Depletion of Efg1 inhibits release of U14 snoRNA from 90S

U3, U14, snR30 and snR10 snoRNAs are involved in 18S rRNA processing and bind the pre-rRNA for different periods. U3 is a prominent and stable component of 90S structure (10–12). By contrast, snR30 and U14 only associate briefly with early 90S particles and are released when the 90S is fully assembled (4). The assembly of snR10 has not been understood yet. We examined whether Efg1 depletion affects association of these snoRNAs with 90S.

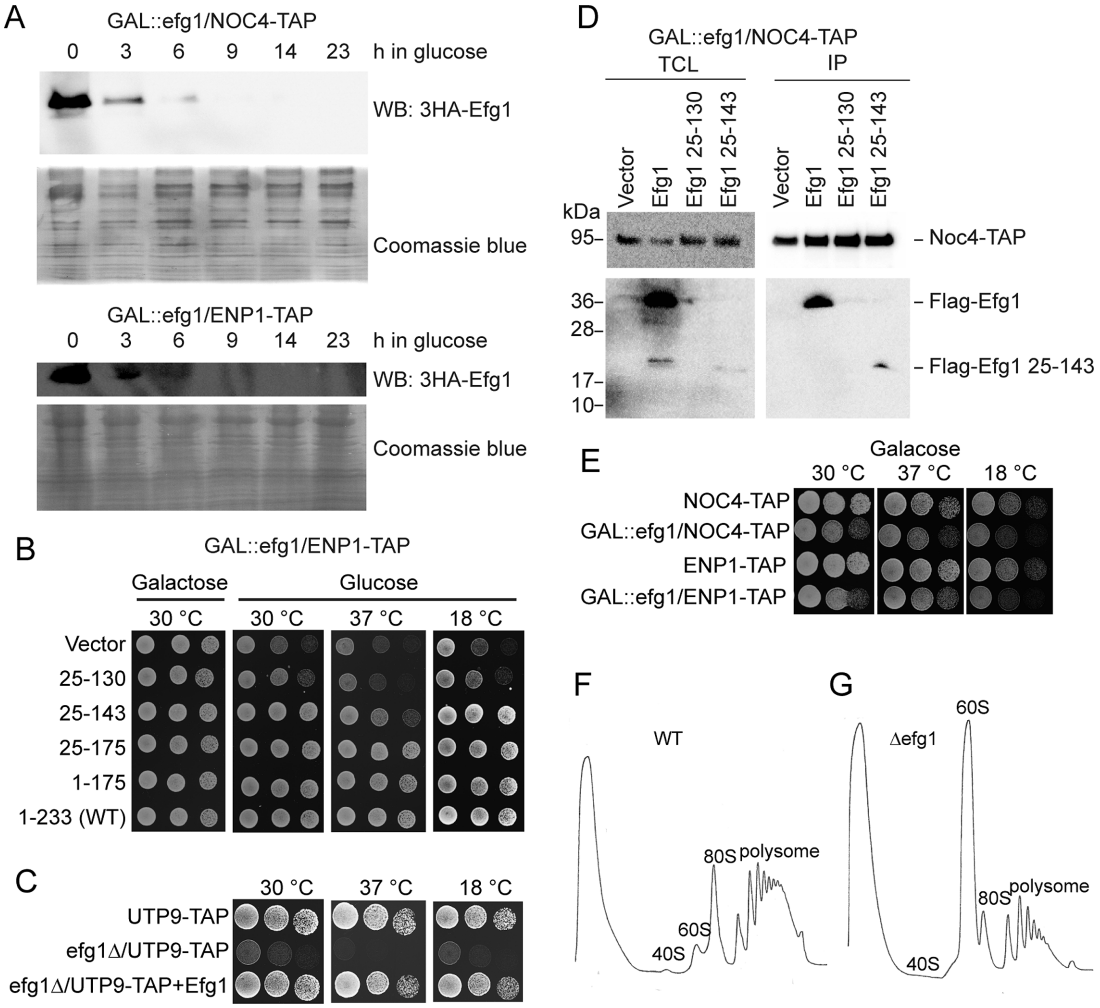

Figure 3. Functional assays of Efg1. (**A**) Expression of Efg1 in GAL::efg1/NOC4-TAP or GAL::efg1/ENP1-TAP strains after shift to glucose medium. The 3HA-tagged Efg1 was detected with western blot using an anti-HA antibody. The same membranes were stained with Coomassie blue to show equal loading of total proteins. (**B**) Growth assay. The GAL::efg1/ENP1-TAP strain was transformed with an empty pRS415-GPD vector or vectors expressing Efg1 WT or fragments containing indicated residues. Five-fold serial dilutions of transformants were grown on SC-Leu plates containing galactose or glucose. (**C**) Growth assay of strains UTP9-TAP, efg1Δ/UTP9-TAP and efg1Δ/UTP9-TAP complemented with a plasmid expressing WT Efg1. (**D**) Association of Efg1 mutants with 90S. The strains described in B were grown in YPD for 14 h. Total cell lysates (TCL) and immunoprecipitates (IP) of IgG-coated bead were analyzed with western blot using anti-HA and anti-Flag antibodies. (**E**) Overexpression of Efg1 has a dominate negative phenotype. Growth assay of WT and GAL::efg1 strains in YPG plates. (**F** and **G**) Ribosome profiles. Cell extracts from strains ENP1-TAP (WT) (F) and GAL::efg1/ENP1-TAP (G) after transfer to glucose medium for 14 h were fractionated on 7–50% sucrose gradients. UV absorbance was recorded at 254 nm.

These snoRNAs, except for U14, displayed similar expression levels in different samples (Figure 5A). Notably, the level of U14 was increased by ∼2-fold in the Efg1-depeleted strain (lane 7) and returned to normal levels when supplied with WT Efg1, but not with the Efg1 fragments 25–130 and 25–143 (lanes 8–10). The level of U14 seems to be inversely correlated with the efficiency of U14 release from 90S (see below, Figure 5B). However, the level of U14 was not changed much in the *efg1Δ* strain (Figure 5A, lanes 2 and 4). The reason underlying the change of U14 level is still unclear.

We quantified the amounts of U14, snR30 and snR10 relative to the stably associated U3 in the purified 90S particles (Figure 5B). The data provide little information about association of snR10, as it displays similar levels in all samples. The 90S particle assembled on a plasmid-derived pre-rRNA fragment ending at position 1643 of 18S (18S-1643) represents an immature 90S particle, which strongly associates with snR30 and U14 (4). Deletion of Efg1 did not affect the levels of snR30 and U14 in the 18S-1643 particle, suggesting that Efg1 is not required for their association (Figure 5B, lanes 1–2).

The Noc4-TAP particle contained very low levels of snR30 and U14 as they are largely released in mature 90S particles (Figure 5B, lane 5). Interestingly, depletion of Efg1 increased the level of U14 by ∼9-fold (cf. lanes 5 and 7) but did not affect the level of snR30. This indicates that Efg1 depletion specifically suppresses U14 release from 90S. Overexpression of Efg1 in galactose also increased the amount of retained U14 in 90S (Figure 5B, lane 6). The defect in U14 release was largely rescued by expression of WT Efg1 (Figure 5B, lane 8) but not the inactive fragment 25–130 (Fig-

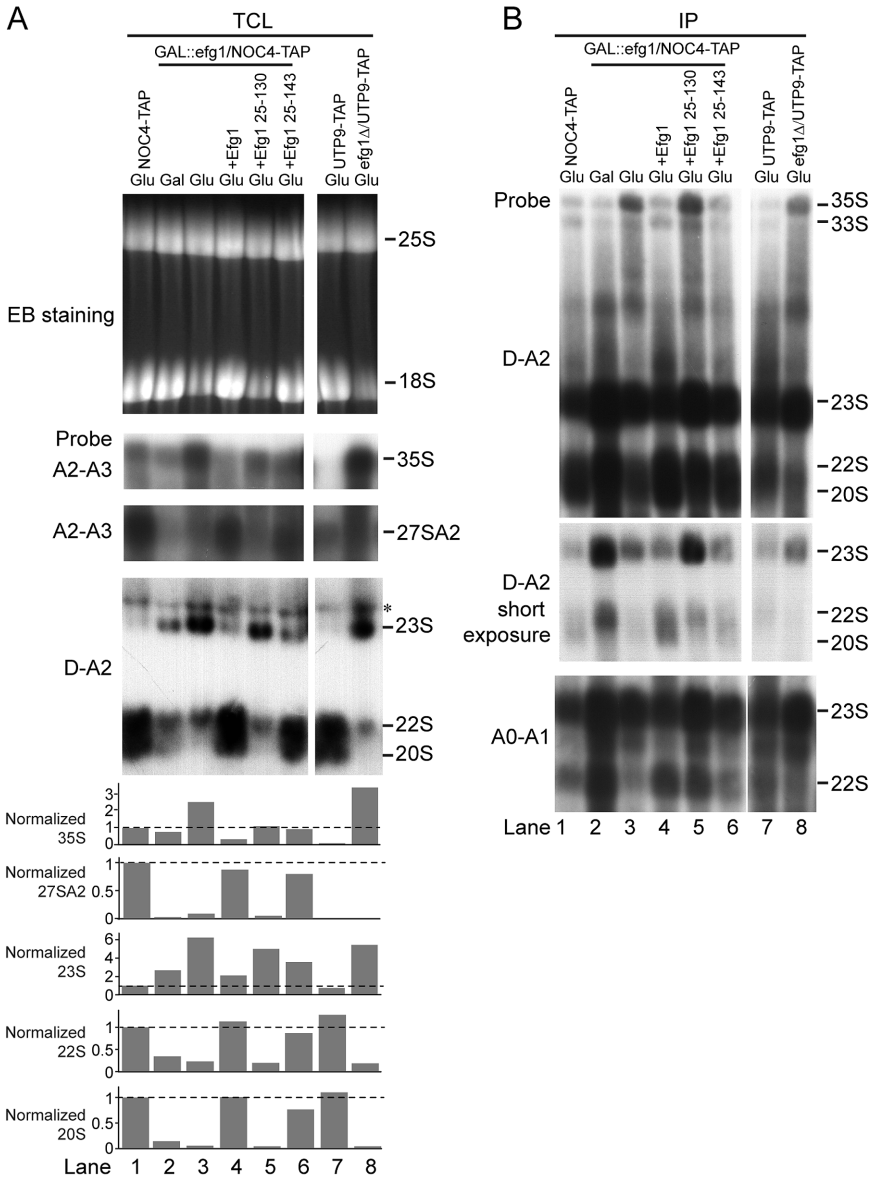

Figure 4. Northern blot analysis of pre-rRNA processing intermediates. (**A**) Pre-rRNAs in total RNAs. GAL::efg1/NOC4-TAP strains expressing no Efg1, full length Efg1 (Efg1) or its truncations (Efg1 25–130 and Efg1 25–143) were grown in YPG (Gal) and shifted to YPD (Glu) for 14 h to deplete Efg1. NOC4-TAP, UTP9-TAP and efg1Δ/UTP9-TAP strains were grown in YPD. RNAs (20 μg) extracted from TCL were resolved in 1.2% agarose-formaldehyde gels and stained with ethidium bromide (EB), transferred to membranes and hybridized against ^{32}P-labeled probes. Asterisk denotes non-specific bands of 25S rRNA. The volume of pre-rRNA was normalized against that from the Noc4-TAP strain (lane 1). The 27SA2 RNA cannot be distinguished in lane 8 due to strong background signals. (**B**) Pre-rRNAs in 90S particles. Growth of yeast cells and northern blotting were conducted as described above. RNAs were extracted from immunoprecipitates (IP) of IgG coated beads. A total of 4.5 μg of RNAs was loaded in each lane except for lane 2, which had 13.5 μg of RNA. The 22S and 20S pre-rRNAs were not well resolved in this gel.

ure 5B, lane 9). Although the fragment 25–143 is fully functional in growth assay and pre-rRNA processing (Figures 3B and 4), it did not restore the release of U14 from 90S (Figure 5B, lane 10).

The 90S particle purified via Utp9-TAP contained higher levels of snR30 and U14 compared to the Noc4-TAP particle (Figure 5, lane 4). As Utp9 binds the pre-rRNA at earlier stages than Noc4 (4,5), the purified Utp9-TAP particle should contain more early assembly intermediates of 90S that associate with snR30 and U14. Nevertheless, deletion of Efg1 also led to an increased accumulation of U14 (Figure 5B, lane 3). Together, our data show that Efg1 is required for efficient release of U14 from 90S.

Function of Efg1 on 90S assembly

To directly examine how the absence of Efg1 affects 90S assembly, we affinity purified Noc4-TAP particles in WT and Efg1-depleted strains and identified associated proteins by semi-quantitative mass spectrometry (Figure 6). The RSAF was calculated to compare the relative stoichiometry of proteins within and across samples (4,6). Complete depletion of

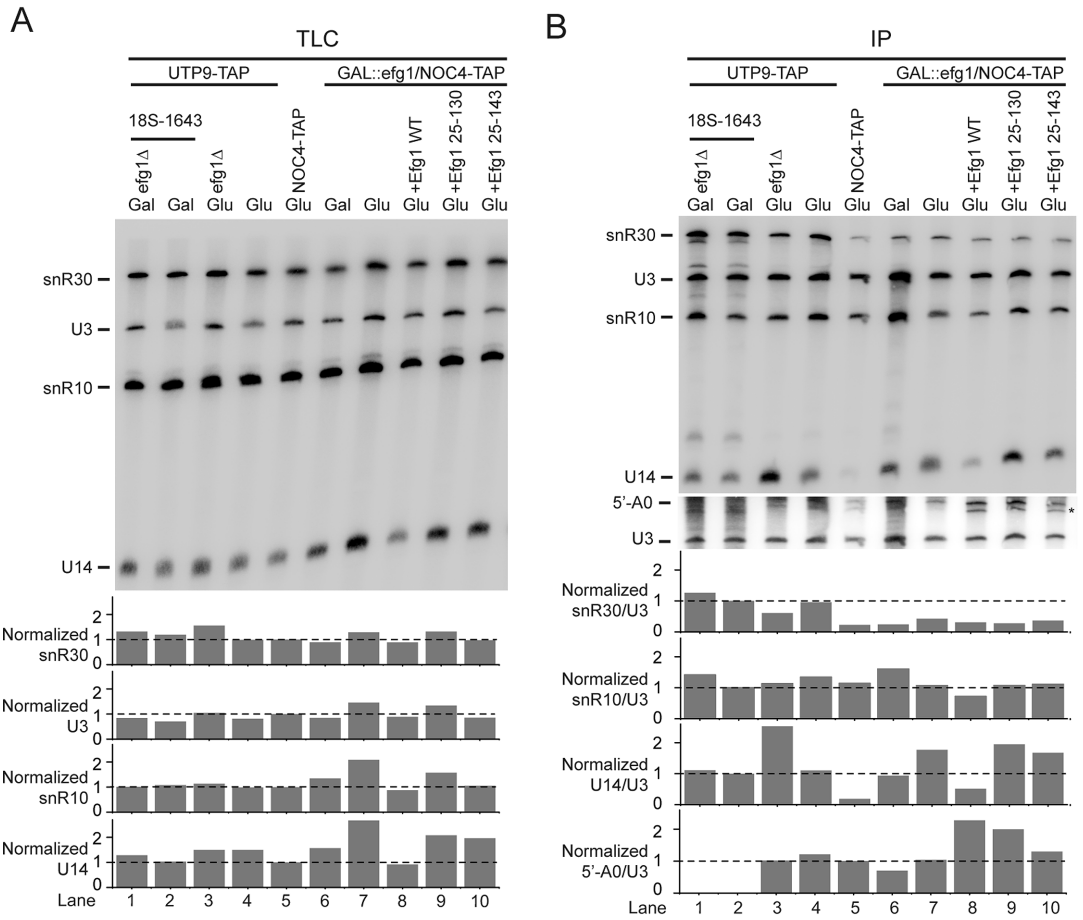

Figure 5. Northern blot analysis of snoRNAs and 5′-A0 fragments. (**A**) Expression levels of snoRNAs. A 5′-MS2-tagged pre-rRNA fragment ending at position 1643 of 18S (18S-1643) was expressed in UTP9-TAP or efg1Δ/UTP9-TAP strains. GAL::efg1/NOC4-TAP strains expressing no Efg1, full length Efg1 or its truncations (Efg1 25–130 and Efg1 25–143) were grown in YPG (Gal) and shifted to YPD (Glu) for 14 h to deplete Efg1. The NOC4-TAP, UTP9-TAP and efg1Δ/UTP9-TAP stains were grown in YPD. RNAs (7 μg) extracted from TCL were resolved in 8% polyacrylamide-8 M urea gels, transferred to membranes and hybridized against ^{32}P-labeled DNA probes. The volume of each snoRNA is normalized to that from NOC4-TAP strain (lane 5). (**B**) Analysis of snoRNAs and 5′-A0 fragments in 90S pre-ribosomes. RNAs were extracted from immunoprecipitates (IP) of IgG-coated beads. The incomplete 90S particle assembled on the 18S-1643 fragment pre-rRNA was affinity purified with MS2-tagged pre-rRNA and IgG immunoprecipitation (lanes 1 and 2). A total of 1.5 μg of RNAs was loaded in each lane except for lane 6, which had 4.5 μg of RNA. Asterisk indicates a degradation product of 5′-A0 fragment. The snR30/U3, snR10/U3 and U14/U3 volume ratios were normalized against those in the 18S-1643 particle (lane 2). The 5′-A0/U3 volume ratio was normalized against that in Noc4-TAP particle (lane 5) and was not determined for lanes 1–2 that contain only non-specific signals.

Efg1 was evidenced by the absence of Efg1 protein in ΔEfg1 90S samples. Most detected AFs had similar abundances within experimental variations in the WT and ΔEfg1 90S. Notably, in two samples of ΔEfg1 90S, the stably associated AFs Enp2, Bfr2 and Lcp5 displayed decreased abundance, whereas the labile factor Bud22 was enriched. All these affected AFs, like Efg1, are initially recruited by the 5′ domain of 18S rRNA (4,5). Therefore, Efg1 is important for assembly and release of its structural neighbors in the 5′ domain.

DISCUSSION

We characterized the structure and function of Efg1, an AF associated with 90S pre-ribosome. We showed that Efg1 adopts a novel all-helical fold, which is distinct from many helical solenoid structures present in 90S (10–12). We also showed that the core domain of Efg1 composed of two helical hairpins is required for pre-rRNA processing and yeast growth, whereas the terminal unstructured residues and the C-terminal helices of Efg1 are dispensable. Efg1 appears to be an RNA-binding protein since it has positively charged surface patches and binds RNAs *in vitro*. The *in vivo* binding target of Efg1 remains to be determined.

We searched the cryo-EM density of 90S for the presence of Efg1 but did not identify a density that fits the Efg1 structure. Efg1 is strongly reduced in mature 90S particles compared to the incomplete 90S intermediates assembled on pre-rRNA fragments (4,25). Mostly likely, the determined 90S cryo-EM structures represent mature states of 90S that contain little Efg1. In addition, the density of Efg1 may be averaged out due to mobility.

The assembly of 90S was recently found to be a highly dynamic process involving factor release and structural reorganization (4). A dozen labile AFs and U14 and snR30 snoRNAs that are initially recruited by 18S rRNA sequences are released when 90S is fully assembled. The cryo-EM structures of 90S have revealed major features of mature 90S particles (10–12), but the dynamic assembly pro-

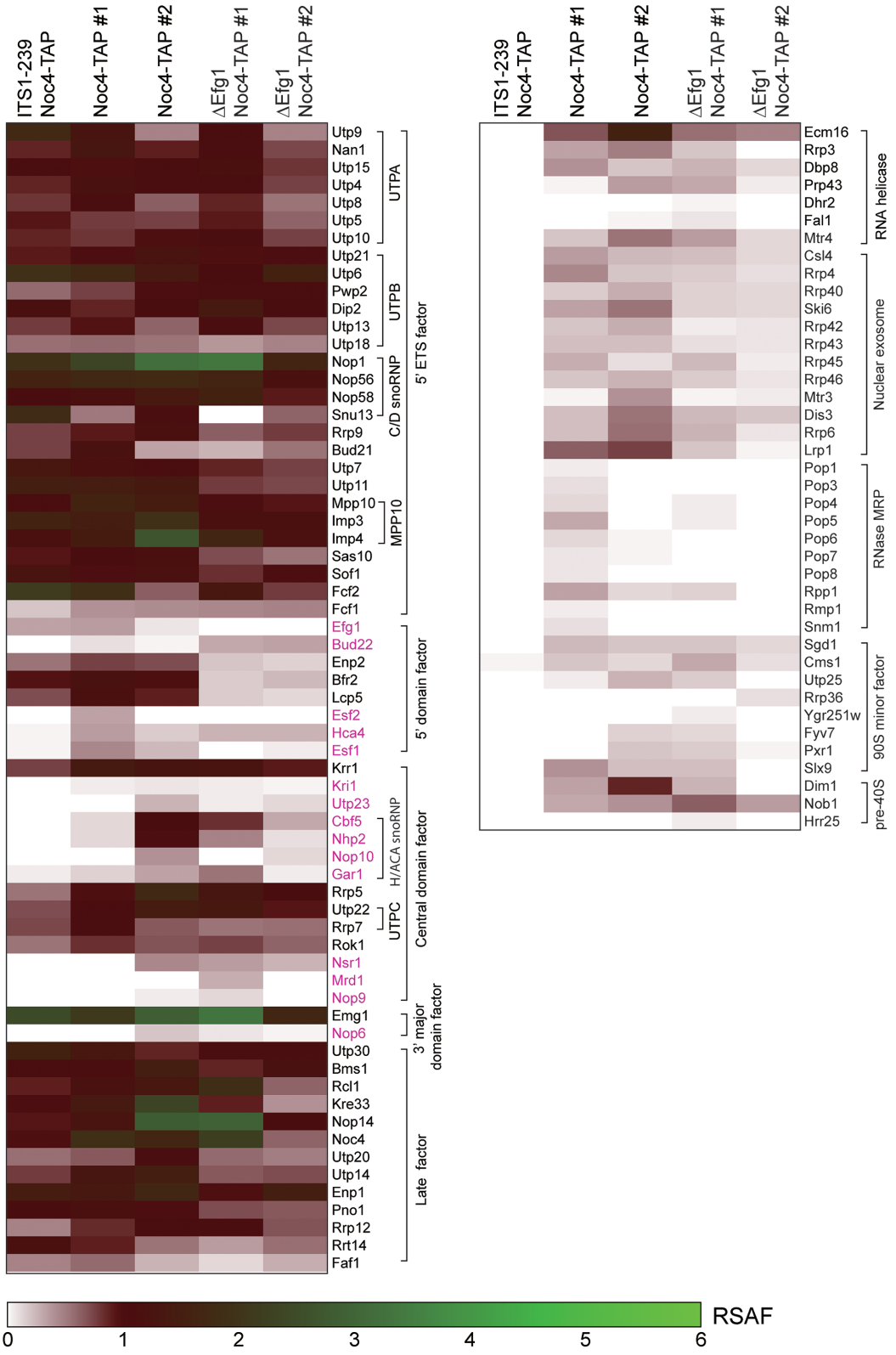

Figure 6. Heatmap of AFs in 90S pre-ribosomes. GAL::efg1/NOC4-TAP stain was grown in YPG and shifted to YPD for 14 h to deplete Efg1. The first two previously reported samples were included for comparison (4). The first sample was purified via a 5′-MS2-tagged pre-rRNA fragment ending at position 239 of ITS1 and Noc4-TAP and represents a fully assembled yet unprocessed 90S particle with most labile factors released. The Noc4-TAP #1 sample was purified with two affinity steps. The other three samples were purified with one step of IgG immunoprecipitation. Proteins are color-coded according to their RSAF values normalized against UTPB. Labile factors that are dissociated in mature 90S are colored in magenta.

cess leading to the formation of 90S is poorly understood. We recently identified the reason underlying release of a labile AF Nop9 (25). Nop9 directly binds a sequence of 18S rRNA that forms the central pseudoknot and helix 28. As helix 28 is already formed in the 90S structure, release of Nop9 is obligatory for 90S assembly to proceed. Our current study provides insight into the role of Efg1 in the assembly and dynamic reorganization of the 5′ domain of 18S rRNA.

During the ordered assembly of 90S, the 5′ domain sequence of 18S rRNA recruits a group of factors, including U14 snoRNA, Efg1, Bud22, Enp2, Bfr2, Lcp5, Esf2, Hca4/Dbp4 and Esf1 (4,5). These factors likely form a structural neighborhood in the nascent 5′ domain. U14 snoRNA makes an essential interaction with nucleotides 83–95 of 18S rRNA (26). Efg1 and U14 snoRNA are together associated with a pre-rRNA fragment ending at position 289 of 18S rRNA but not with a shorter fragment ending at position 232 (4), which suggests that assembly of Efg1 and U14 may be interdependent. However, our data show that the initial recruitment of U14 to pre-rRNA was not affected by Efg1 depletion.

In the fully assembled 90S, Enp2, Lcp5 and Bfr2 are still stably associated. Enp2 has been located in the cryo-EM structures of 90S and binds the 5′ domain of 18S rRNA (10,11). The other labile factors, including U14 and Efg1, would be released during 90S maturation. Our data demonstrate that Efg1 is important for the formation of the nascent 5′ domain and its reorganization during maturation of 90S. Compared to the WT mature 90S, the labile factors U14 and Bud22 show increased retention in ΔEfg1 90S, whereas the stable factors Enp2, Bfr2 and Lcp5 become weakly associated. The assembly of other 90S AFs that are distributed throughout the 90S structure are not affected. Therefore, Efg1 depletion mainly affects the association and dissociation of its structural neighbors in the 5′ domain. These structural defects in the 5′ domain appear to eventually inactivate 90S in processing the A1 and A2 sites and block production of 40S subunit.

We found that overexpression of Efg1 inhibits yeast growth, pre-rRNA processing and release of U14. As Efg1 needs to be released during 90S maturation, high concentration of Efg1 may suppress its release and hence arrest the 5′ domain in early assembly states.

Release of U14 requires unwinding of the base pairs formed between U14 and rRNA. The RNA helicases Hca4/Dbp4 and Has1 have been shown to be required for U14 release, although whether they directly unwind the U14–rRNA duplex is unknown (27–29). We showed that Efg1, which is apparently a structural protein without enzymatic activity, is also required for efficient release of U14. Efg1 likely functions together with these helicases in the structural reorganization of the 5′ domain.

DATA AVAILABILITY

The coordinates and structural factors have been deposited into the Protein Data Bank with accession code 5YMA.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to the staff from beamlines BL17U and BL19U of National Facility for Protein Sciences Shanghai (NFPS) and Shanghai Synchrotron Radiation Facility for assistance in data collection. We thank Zhensheng Xie and Fuquan Yang for mass spectrometry analysis and Hongjie Zhang for help in radioactive experiments.

FUNDING

Strategic Priority Research Program of Chinese Academy of Sciences [XDB08010203]; National Key R&D Program of China [2017YFA0504600]; National Natural Science Foundation of China [91540201, 31430024, 31325007]; 100 Talents Program of CAS. Funding for open access charge: National Natural Science Foundation of China [91540201].
Conflict of interest statement. None declared.

REFERENCES

1. Woolford,J.L. Jr and Baserga,S.J. (2013) Ribosome Biogenesis in the Yeast Saccharomyces cerevisiae. *Genetics*, **195**, 643–681.
2. Kressler,D., Hurt,E. and Bassler,J. (2010) Driving ribosome assembly. *Biochim. Biophys. Acta*, **1803**, 673–683.
3. Henras,A.K., Soudet,J., Gerus,M., Lebaron,S., Caizergues-Ferrer,M., Mougin,A. and Henry,Y. (2008) The post-transcriptional steps of eukaryotic ribosome biogenesis. *Cell Mol. Life Sci.*, **65**, 2334–2359.
4. Zhang,L., Wu,C., Cai,G., Chen,S. and Ye,K. (2016) Stepwise and dynamic assembly of the earliest precursors of small ribosomal subunits in yeast. *Genes Dev.*, **30**, 718–732.
5. Chaker-Margot,M., Hunziker,M., Barandun,J., Dill,B.D. and 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.
6. Chen,W., Xie,Z., Yang,F. and Ye,K. (2017) Stepwise assembly of the earliest precursors of large ribosomal subunits in yeast. *Nucleic Acids Res.*, **45**, 6837–6847.
7. Osheim,Y.N., French,S.L., Keck,K.M., Champion,E.A., Spasov,K., Dragon,F., Baserga,S.J. and Beyer,A.L. (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.
8. 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.
9. Dragon,F., Gallagher,J.E., Compagnone-Post,P.A., Mitchell,B.M., Porwancher,K.A., Wehner,K.A., Wormsley,S., Settlage,R.E., Shabanowitz,J., Osheim,Y. *et al.* (2002) A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature*, **417**, 967–970.
10. 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**, e22086.
11. Chaker-Margot,M., Barandun,J., Hunziker,M. and Klinge,S. (2017) Architecture of the yeast small subunit processome. *Science*, **355**, eaal1880.
12. Kornprobst,M., Turk,M., Kellner,N., Cheng,J., Flemming,D., Kos-Braun,I., Kos,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.
13. Kastenmayer,J.P., Ni,L., Chu,A., Kitchen,L.E., Au,W.C., Yang,H., Carter,C.D., Wheeler,D., Davis,R.W., Boeke,J.D. *et al.* (2006) Functional genomics of genes with small open reading frames (sORFs) in S. cerevisiae. *Genome Res.*, **16**, 365–373.
14. Peng,W.T., Robinson,M.D., Mnaimneh,S., Krogan,N.J., Cagney,G., Morris,Q., Davierwala,A.P., Grigull,J., Yang,X., Zhang,W. *et al.*

(2003) A panoramic view of yeast noncoding RNA processing. *Cell*, **113**, 919–933.
15. Schilling,V., Peifer,C., Buchhaupt,M., Lamberth,S., Lioutikov,A., Rietschel,B., Kotter,P. and Entian,K.D. (2012) Genetic interactions of yeast NEP1 (EMG1), encoding an essential factor in ribosome biogenesis. *Yeast*, **29**, 167–183.
16. Zhang,C., Lin,J., Liu,W., Chen,X., Chen,R. and Ye,K. (2014) Structure of Utp21 tandem wd domain provides insight into the organization of the UTPB complex involved in ribosome synthesis. *PLoS One*, **9**, e86540.
17. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
18. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.
19. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
20. Janke,C., Magiera,M.M., Rathfelder,N., Taxis,C., Reber,S., Maekawa,H., Moreno-Borchart,A., Doenges,G., Schwob,E., Schiebel,E. *et al.* (2004) A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. *Yeast*, **21**, 947–962.
21. Lin,J., Lu,J., Feng,Y., Sun,M. and 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.
22. Amlacher,S., Sarges,P., Flemming,D., van Noort,V., Kunze,R., Devos,D.P., Arumugam,M., Bork,P. and Hurt,E. (2011) Insight into structure and assembly of the nuclear pore complex by utilizing the genome of a eukaryotic thermophile. *Cell*, **146**, 277–289.
23. Holm,L. and Sander,C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.*, **233**, 123–138.
24. de la Cruz,J., Kressler,D., Tollervey,D. and 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.
25. Wang,B. and Ye,K. (2017) Nop9 binds the central pseudoknot region of 18S rRNA. *Nucleic Acids Res.*, **45**, 3559–3567.
26. Liang,W.Q. and Fournier,M.J. (1995) U14 base-pairs with 18S rRNA: a novel snoRNA interaction required for rRNA processing. *Genes Dev.*, **9**, 2433–2443.
27. Kos,M. and 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.
28. Liang,X.H. and Fournier,M.J. (2006) The helicase Has1p is required for snoRNA release from pre-rRNA. *Mol. Cell. Biol.*, **26**, 7437–7450.
29. Bohnsack,M.T., Kos,M. and Tollervey,D. (2008) Quantitative analysis of snoRNA association with pre-ribosomes and release of snR30 by Rok1 helicase. *EMBO Rep.*, **9**, 1230–1236.