Structure and RNA recognition of ribosome assembly factor Utp30

JIANFEI HU,[1,2,3,4] XING ZHU,[1] and KEQIONG YE[1,4]

[1]Key Laboratory of RNA Biology, CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
[2]Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China
[3]School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China
[4]University of Chinese Academy of Sciences, Beijing 100049, China

ABSTRACT

The 90S preribosomes are gigantic early assembly intermediates of small ribosomal subunits. Cryo-EM structures of 90S were recently determined, but many of its components have not been accurately modeled. Here we determine the crystal structure of yeast Utp30, a ribosomal L1 domain-containing protein in 90S, at 2.65 Å resolution, revealing a classic two-domain fold. The structure of Utp30 fits well into the cryo-EM density of 90S, confirming its previously assigned location. Utp30 binds to the rearranged helix 41 of 18S rRNA and helix 4 of 5′ external transcribed spacer in 90S. Comparison of RNA-binding modes of different L1 domains illustrates that they consistently recognize a short RNA duplex with the concaved surface of domain I, but are versatile in RNA recognition outside the core interface. Cic1 is a paralog of Utp30 associating with large subunit preribosomes. Utp30 and Cic1 share similar RNA-binding modes, suggesting that their distinct functions may be executed by a single protein in other organisms. Deletion of Utp30 does not affect the composition of 90S. The nonessential role of Utp30 could be ascribed to its peripheral localization and redundant interactions in 90S.

Keywords: ribosome assembly; 90S preribosome; crystal structure; RNA-binding protein

INTRODUCTION

Assembly of ribosome, the molecular machine for protein synthesis, is a fundamental and highly complex process that involves transcription, modification, and processing of ribosomal RNAs (rRNAs) and assembly of ribosomal proteins. More than 200 assembly factors (AFs) and many snoRNAs participate in ribosome assembly (Henras et al. 2008; Kressler et al. 2010; Woolford and Baserga 2013). A number of preribosomal particles are formed as the small 40S and large 60S subunit mature in the nucleolus, the nucleoplasm, and the cytoplasm.

Ribosome assembly in the budding yeast *S. cerevisiae* begins with the transcription of a long 35S precursor rRNA (pre-rRNA) that encodes sequences for 18S, 5.8S, and 25S rRNAs and four external and internal transcribed spacers (ETS and ITS). The 90S preribosome or small subunit processome is the earliest precursor of the small subunit that assembles cotranscriptionally on the 5′ part of the pre-rRNA in the nucleolus (Dragon et al. 2002; Grandi et al. 2002; Osheim et al. 2004). Approximately 70 AFs and U3, U14, and snR30 snoRNAs associate stepwise to the pre-rRNA to form the 90S

(Chaker-Margot et al. 2015; Zhang et al. 2016). A dozen AFs and snR30 and U14 snoRNAs are released upon maturation of 90S (Zhang et al. 2016). Following processing of pre-rRNA at the 5′ ETS and ITS1 regions and dramatic compositional changes, the 90S particle is converted into a pre-40S ribosome that is exported and develops into the small subunit in the cytoplasm. The early precursors of large subunits are assembled on the 3′ half of pre-rRNA (Chen et al. 2017) and evolve largely independently of small subunits.

Recent cryo-EM analyses of 90S particles have revealed its complex architecture (Kornprobst et al. 2016; Chaker-Margot et al. 2017; Sun et al. 2017). The 18S rRNA is assembled into several isolated substructures that are stabilized by abundant AFs. The 5′ ETS region, in complex with the UTPA, UTPB, and U3 snoRNP complexes and other factors, forms a large base organizing the structure of nascent ribosomes. Due to limited resolution of the cryo-EM density maps (4.5 Å, best), the quality of the current structural model of 90S critically depends on availability of high-resolution

---

structures of component proteins and RNAs. However, many AFs in 90S have no crystal structure determined.

One of the AFs associated with 90S is Utp30, which is homologous to ribosomal L1 proteins. Utp30 was first identified as a component of 90S preribosomes with the biochemical purification approach (Grandi et al. 2002). Deletion of Utp30 caused no obvious phenotype on ribosome biogenesis and yeast growth, but is synthetically lethal with a temperature-sensitive mutant of Emg1 (Nep1), which is an RNA modification enzyme in 90S (Schilling et al. 2012). Utp30 has a close paralog Cic1 (Nsa3) in the budding yeast that is involved in large subunit assembly (Bassler et al. 2001; Harnpicharnchai et al. 2001; Fatica et al. 2003). The human homolog of Utp30, cellular senescence-inhibited gene (CSIG) or ribosomal L1 domain-containing 1 (RSL1D1), is localized in the nucleolus and involved in regulation of cell proliferation and senescence (Ma et al. 2008; Cheng et al. 2015). In the ordered assembly pathway of 90S, Utp30 associates with the pre-rRNA when the 18S region becomes nearly complete (Chaker-Margot et al. 2015; Zhang et al. 2016). In the cryo-EM maps of 90S, a density in the mouth region was previously assigned to Utp30. The structure of the homologous L1 protein was used to fit the map, but discrepancy between the density and the model was evident.

In this work, we determine the first crystal structure of yeast Utp30. The structure confirms the previous assignment of the Utp30 location and provides an accurate template for modeling its structure in 90S. We describe in detail the RNA interaction of Utp30 in 90S and compare how different L1 domain proteins bind to their targets. We show that depletion of Utp30 causes little change of the protein composition of 90S. We also discuss the functional role and evolution of Utp30.

RESULTS

Crystal structure of Utp30

To facilitate crystallization, a few nonconserved residues were removed from both the N and C termini of Utp30 from *S. cerevisiae* (total of 274 residues), yielding a fragment with residues 11–256. The protein was expressed as a fusion to N-terminal six-His-tagged Smt3 in *E. coli*, which was cleaved during protein purification. In the absence of Smt3, Utp30 was soluble only under high salt conditions (>0.5 M NaCl). We crystallized native and selenium methionine (SeMet)-labeled Utp30 and determined its structure with de novo phasing. The structure was refined at 2.65 Å resolution to an R-work of 0.200 and an R-free of 0.266 with good geometry (Fig. 1A; Supplemental Table S1). The asymmetric unit contains two protomers, which are superimposable with a root mean standard deviation (rmsd) of 0.597 Å over 186 Cα pairs. The two protomers are nearly identical with conformational variations restricted to a few peripheral elements.

The Utp30 structure consists of two domains and shares a similar fold as ribosomal L1 proteins (Fig. 1B; Nikulin et al.

2003). Domain I is formed by both the N- and C-terminal sequences and is interrupted by domain II between strands β2 and β7. Domain I folds into a two-layer structure and contains a major four-stranded anti-parallel β-sheet (β8, β1, β9, and β10) and a minor two-stranded anti-parallel β-sheet (β2 and β7), which is separated from the major sheet. Two α-helices (α1 and α7) are packed at one side of these sheets. Domain II adopts a three-layer Rossmann fold and contains a four-stranded parallel β-sheet (β4, β3, β5, and β6) decorated by five α-helices (α2–α6) at both sides. Utp30 shares 13% sequence identity with the L1 protein from *Methanococcus jannaschii*. Their structures can be aligned with an rmsd of 3.6 Å over 189 Cα pairs (Fig. 1B). Domain I is better aligned than domain II in the two structures.

Evolution of Utp30

To examine the sequence conservation and evolutionary distribution of Utp30, we searched its homologs with PSI-Blast, aligned the sequences from representative organisms, and constructed a phylogenetic tree (Fig. 1C; Supplemental Figs. S1, S2). Both Utp30 and its paralog Cic1 can be clearly identified and distinguished in Saccharomycetales (Supplemental Figs. S1, S2). Many other organisms, for example, all species in Pesizomycotina, contain only a single homolog. Some organisms contain multiple homologs that are often closely related and cannot be classified as Utp30 or Cic1 (Supplemental Figs. S1, S2). Our analysis shows that Utp30 and Cic1 diverged in Saccharomycetales and most other organisms contain a single type of Utp30/Cic1 homolog. Our data revise the previous observation that both proteins are present in the larger group of fungi (Ebersberger et al. 2014).

Structure of Utp30 in 90S

In the cryo-EM maps of 90S (Kornprobst et al. 2016; Chaker-Margot et al. 2017; Sun et al. 2017), Utp30 was previously assigned to a density in the mouth region at the front face of 90S (Fig. 2A,B). The structure of Utp30 fits much better to the density than the previously used structural model of L1. Since some structural elements of Utp30 move notably in the 90S density, the structure was further adjusted and refined against the cryo-EM density (Fig. 2C; Supplemental Table S2). The cryo-EM densities of yeast 90S from different studies are well superimposable around Utp30 (Chaker-Margot et al. 2017; Sun et al. 2017). In the 90S, Utp30 primarily binds helix 41 of the 3′ major domain of 18S rRNA and also contacts the tip of helix 4 (H4) of 5′ ETS. A long α-helix (∼34 residues) from unassigned proteins contacts the tip loop of helix 41 and docks at domain I of Utp30 near helix α1 and strand β10. Otherwise, Utp30 apparently binds no other proteins.

Comparison of Utp30 structures in the crystal and in the 90S particle shows that the orientation between domains I and II is unchanged upon assembly into 90S, indicating
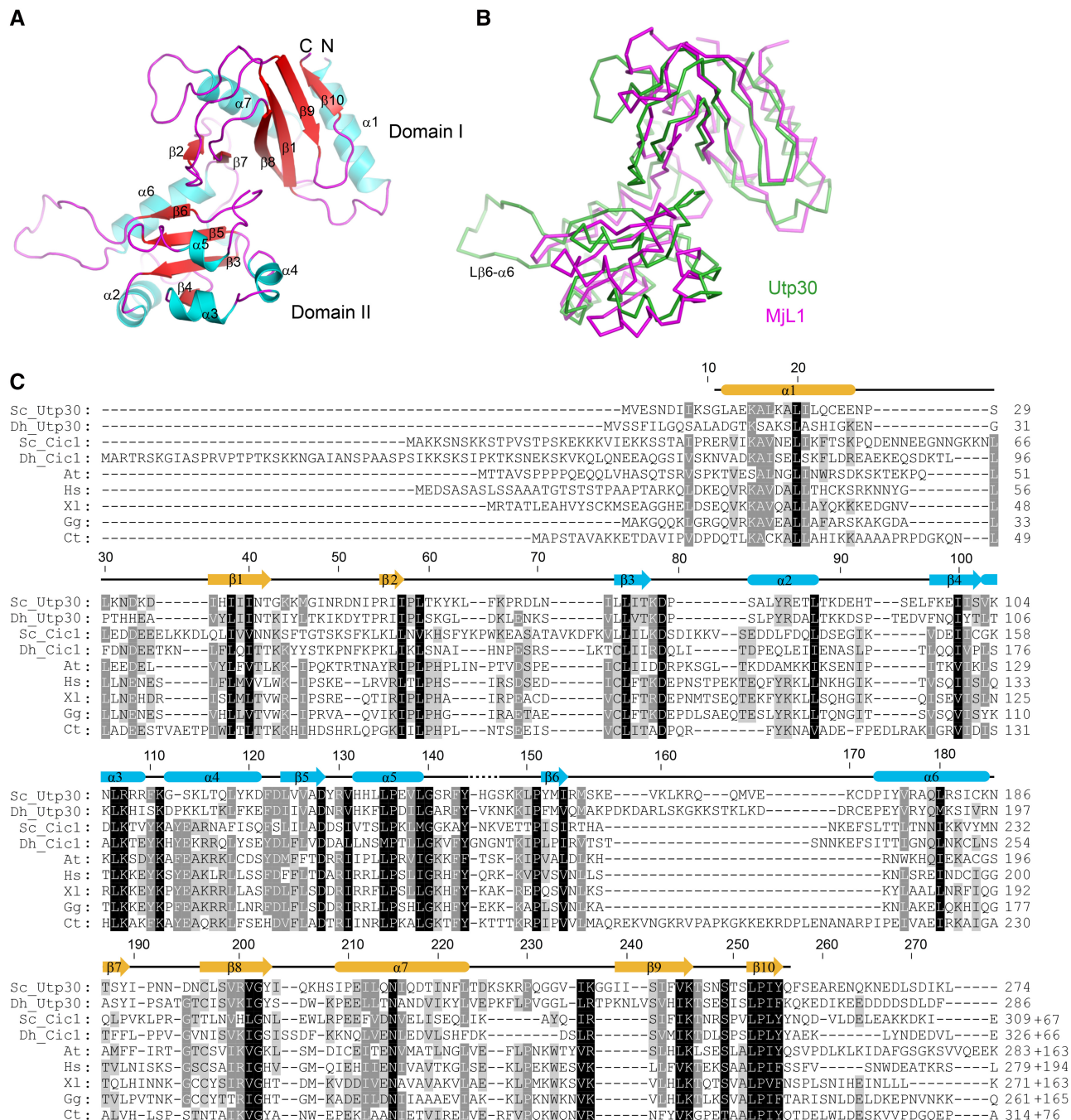
A

B

C

```
                                                          10        20
                                                         ━━━━━━━━━━━━━━
                                                              α1
Sc_Utp30: ------------------------------------------------MVESNDIIKSGLAEKAIKALILQCEENP------------S  29
Dh_Utp30: ------------------------------------------------MVSSFILGQSALADGTKSAKSLASHIGKEN----------G  31
Sc_Cic1:  --------------------MAKKSNSKKSTPVSTPSKEKKKVIEKKSSTAIPRERVIKAVNELIKFTSKPQDENNEEGNNGKKNL      66
Dh_Cic1:  MARTRSKGIASPRVPTPTKSKKNGAIANSPAASPSIKKSKSIPKTKSNEKSKVKQLQNEEAAQGSIVSKNVADKAISELSKFLDREAEKEQSDKTL---L  96
At:       ------------------------MTTAVSPPPPQEQQLVHASQTSRVSPKTVESAINGLINWRSDKSKTEKPQ-------L                 51
Hs:       ----------------MEDSASASLSSAAATGTSTSTPAAPTARKQIDKEQVRKAVDALITHCKSRKNNYG------L                      56
Xl:       -----------------MRTATLEAHVYSCKMSEAGGHEIDSEQVKKAVQALIAYQKKKEDGNV---------L                          48
Gg:       ------------------------------MAKGQQKLGRGQVRKAVEALIAFARSKAKGDA---------L                           33
Ct:       -------------MAPSTAVAKKETDAVIPVDPDQTLKACKALLAHIKKAAAAPRPDGKQN---L                                   49

           30         40         50         60         70         80         90        100
          ━━━━━━━━━►        ━━━━━━►              ━━━►          ━━━━━━━━━        ━━━━►
             β1              β2                   β3                α2               β4
Sc_Utp30: LKNLKD------IHIIINTGKKMGINRDNIPRIIPLTKYKL--FKPRDLN------ILLITKDP--------SALYRETLTKDEHT---SELFKEIISVK 104
Dh_Utp30: PTHHEA------VYLIINTKIYLTKIKDYTPRIIPLSKGL--DKLENKS------VLLVTKDP--------SLPYRDALTKKDSP---TEDVFNQIYTLT 106
Sc_Cic1:  LEDEEELKKDLQLIVNNKSFTGTSKSFKLKLINVKHSFYKPWKEASATAVKDFKVLLIKDSDIKKV--SEDDLFDQLDSEGIK------VDELICGK    158
Dh_Cic1:  FDNEETKN---IFLQLITKKYYSTKPNFKPKLIKLSNAI---HNPESRS----LKTCLIIRDQLI-----TDPEQLEILENASLP-----TLQQIVPLS 176
At:       LEEPEL------YLFVTLKK-IPQKTRTNAYRIPLPHPLIN-PTVLSPE------ICLIIDDRPKSGL--TKDDAMKKIIKSENIP------ITKVIKLS 129
Hs:       LLNPNES------IFLMVVLWK--IPSKE--LRVRITLPHS----IRSLSED------ICLFTKDEPNSTPEKTEQFYRKLLNKHGIK----TVSQIISLQ 133
Xl:       LLNPHDR------ISLMITVWR--IPSRE--QTIRIPLPHA----IRPEACD------VCLFTKDEPNMTSEQTEKFYKKLLSQHGIK-----QISEVIISLN 125
Gg:       LLNPNES------VHLLYTVWK--IPRVA--QVIKIPLPHG----IRAETAE------VCLFTKDEPDLSAEQTESLYRKLLTQNGIT-----SVSQYTISYK 110
Ct:       LADEESTVAETPIWLTITTKKHIHDSHRLQPGKIIILPHPL---NTSEEIS------VCLITADPQR-------FYKNAVADE-FPEDLRAKIGRVIDIS   131

           110        120        130        140        150        160        170        180
          ━━━━ ━━━━━━►     ━━►      ━━━━        ━━━━┄┄┄┄┄━►                              ━━━━━━━━━
           α3    α4        β5      α5             β6                                       α6
Sc_Utp30: NLRRREKG-SKLTQLYKDFDIVVADYRVHHLLPEVLGSRDY-HGSKKLPYMIRMSKE----VKLKRQ---QMVE---------KCDPIYVRAQLRSICKN 186
Dh_Utp30: KLKHISKDPKKLTKLFKEFDIIVADNRVHKFLPDLLGAREY-VKNNKIPFMVQMAKPDKDARLSKGKKSTKLKD-------DRCEPEYVRYQMKSIVRN  197
Sc_Cic1:  DLKTVYKAEYARNAFISQFSLILADDSIVTSLPKIMGGKAY-NKVETTPISIRTHA------------------NKEFSLTTITNNIKKVYMN       232
Dh_Cic1:  ALKTEYIKHYEKRRQLYSEYDLFLVDDALLNSMPTILGKVTGYGNGNTKITLPIRVTST--------------SNNKEFSITTIGNQLNKCLNS      254
At:       KLKSDYKAPEAKRKLCDSYDMFFTDRRIIPLLPRVIGKKEF-TSK-KLPVADLDKH--------------------RNWKHQIEKACGS           196
Hs:       TLKKEYKSYEAKLRLLSSFDFFLTDARIRRLLPSLIGRHFY-QRK-KVPVSVNLLS--------------------KNLSREINDCIGG           200
Xl:       RLKKEYKPYEAKRRLLASFDIFLSDDRIRRFLPSLLGKHFY-KAK-KEPQSVNLKS--------------------KYIAALLNRFIQG           192
Gg:       TLKKEYKPEAKRRLNRFDIFLSDDRIRRLLPSHLGKHFY-EKK-KAPLSVNLKA--------------------KNLAKELQKHIQG            177
Ct:       HLKAKEKAYEAQRKLFSEHDVFLAPTRIINRLPKALGKTIY-KTTTKRPIPVVLMAQREKVNGKRVPAPKGKKEKRDPLENANARPIPELVAEIRKAIGA 230

           190        200        210        220        230        240        250        260        270
          ━━━►     ━━━━━━►         ━━━━━━━━━              ━━►                ━━━━━━►     ━━►
           β7        β8              α7                   β9                 β10
Sc_Utp30: TSYI-PNN-DNCLSVRVGYI--QKHSIPELLQNIQDTINFITDKSKRLQGGV-IKGGII---SIFVKTSNSTSLPIYQFSEARENQKNEDLSDIKL----- 274
Dh_Utp30: ASYI-PSATGTCISVKIGYS--DW-KPEELLTNANDVIKYLVEPKFLEVGGL-LRTPKNLVSVHIKTSESISLPIFKQKEDIKEEDDDDSDLDF------ 286
Sc_Cic1:  QLPVKLPR-GTTLNVHLGNL--EWLRPEEFVDNVELISEQLIK------AYQ-IR------SIFIKTNRSPVLPLYYNQD-VLDELEAKKDKI------E  309 +67
Dh_Cic1:  TFFL-PPV-GVNISVKIGSISSDF-KKNQLVENLEDVLSHFDK--------DSLR------SVMIKTDLSPSLPLYYAEK-------LYNDEDVL----- 326 +66
At:       AMFF-IRT-GTCSVIRVGKL--SM-DICEITENVMATLNGIVE--FLENKWTYVR------SLHLKLSESLALPIYQSVPDLKLKIDAFGSGKSVVQEEK 283 +163
Hs:       TVLNISKS-GSCSAIRIGHV--GM-QIEHIIENLVAVTKGLSE--KLEEKWESVK------LLFVKTEKSAALPIFSSFV-----SNWDEATKRS----L 279 +194
Xl:       TQLHINNK-GCCYSIRVGHT--DM-KVDDIVENAVAVAKVLAE--KLEMKWKSVK------VLHLKTQTSVALPVENSPLSNIHEINLLL---------K 271 +163
Gg:       TVLPVTNK-GCCYTTIGHT--GM-KADELLDNIIAAAEVIAK--KLEKNWKNVK------VLHLKTLKSVALPIFTARISNLDELDKEPNVNKK----Q 261 +165
Ct:       ALVH-LSP-STNTAIKVGYA--NW-EPEKLAANIETVIRELVE-RFVEQKWQNVR------NFYVKGPETAALPIYQTDELWLDESKVVPDGQEP----A 314 +76
```

FIGURE 1. Crystal structure of Utp30. (*A*) Ribbon representation of Utp30 structure. The α-helices, β-strands, and loops are colored in cyan, red, and magenta, respectively. The secondary structures, N and C termini, and two domains are labeled. (*B*) Structural alignment between Utp30 and L1 from *Methanococcus jannaschii* (MjL1, PDB code 1U63). (*C*) Multiple sequence alignment of Utp30/Cic1 family proteins. The sequences are from *Saccharomyces cerevisiae* (Sc), *Debaryomyces hansenii* (Dh), *Arabidopsis thaliana* (At), *Homo sapiens* (Hs), *Xenopus laevis* (Xl), *Gallus gallus* (Gg), and *Chaetomium thermophilum* (Ct). Residues with 100%, 80%, and 60% conservation are shaded in black, gray, and light gray, respectively. The secondary structures observed in the crystal structure of Utp30 are shown on the *top* and colored by domain.

that the interdomain connection of Utp30 is rather rigid. Major structural changes occur at the RNA-binding interface of domain II (Fig. 2D). Helices α2, α3, and α4 and the loop between helix α5 and strand β6 experience considerable movements upon binding helix 41 of 18S rRNA. The confor-

mation of the latter three structural elements also differs between two protomers in the asymmetric unit of crystal, indicating that they are intrinsically flexible (Fig. 2E). The most dramatic movement occurs at the long loop linking strand β6 and helix α6 in domain II that flips by ∼90° to make contact
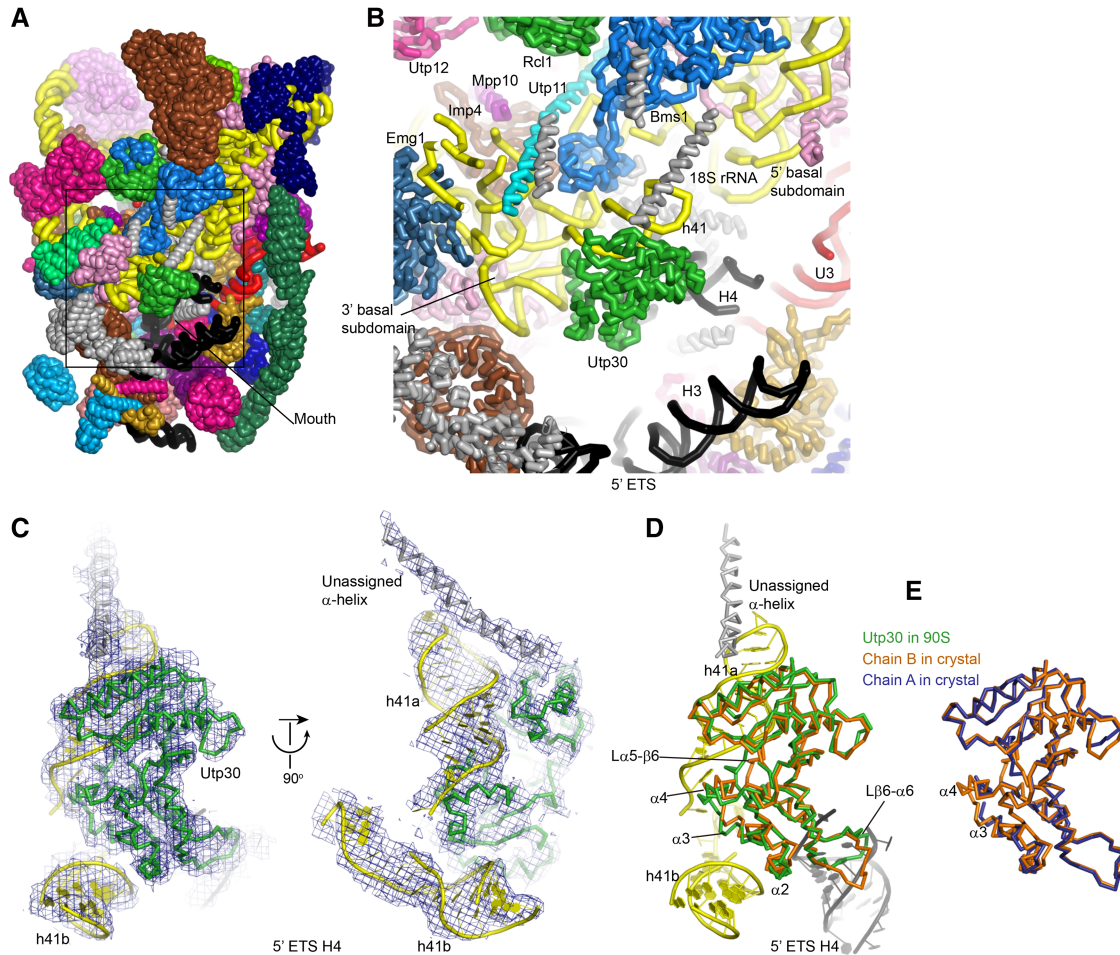
FIGURE 2. Structure of Utp30 in 90S. (*A*) Structure of 90S preribosome showing the front face. Individual proteins are color-coded and unassigned proteins are in gray. The 18S, 5′ ETS, and U3 RNAs are colored yellow, black, and red, respectively. (*B*) Zoomed-in view of the mouth. (*C*) Cryo-EM density map of the ΔMtr4 90S is fitted with Utp30 and its binding partners. Two perpendicular views are shown. (*D*) Alignment of Utp30 structures in crystal (orange) and in 90S (green). (*E*) Structural alignment of two Utp30 protomers in crystal.

with helix 4 of the 5′ ETS. Notably, the corresponding loop is of similar length in Utp30 proteins, but much shorter in L1 and Cic1 proteins (Fig. 1B,C). The loop appears to be elongated in Utp30 for binding the 5′ ETS RNA.

Interaction of Utp30 with helix 41

Helix 41 consists of a top stem (h41a), a base stem (h41b), and middle internal loops (Fig. 3A). In the structure of the mature 40S subunit (Ben-Shem et al. 2011), the two stems fold back into a compact parallel structure (Fig. 3B). In the 90S structure, the base stem is still packed into the 3′ basal subdomain of 18S rRNA as in the mature subunit, but the top stem is projected to be perpendicular to the base stem (Fig. 3C). Parts of the two internal loops linking the top and base stems were not modeled in 90S.

The elongated structure of Utp30 primarily binds the top stem of helix 41 and also docks at the base stem. The top stem is clamped by two domains of Utp30 (Fig. 3C,D). The

concaved surface of the major β-sheet of domain I wraps around the A-form duplex structure of the top stem at the minor groove. The residues at the RNA-binding interface of domain I are mainly of polar and hydrophobic nature (Fig. 3D,E). These indicate that domain I recognizes the top stem of helix 41 mainly by shape complementarity and collective interactions.

Domain 2 contacts the low part of h41a and the 3′ internal loop (nt 1514–1517) from the major groove side with helix α4 and the loop between helix α5 and strand β6. In addition, domain II contacts the base stem of helix 41 with helices α2 and α3. In contrast to the noncharged RNA-binding surface of domain I, the RNA-binding surface of domain II is highly basic (Fig. 3E). Domain II employs abundant electrostatic interactions for binding RNA, a strategy different from that used by domain I.

The majority of the residues conserved in Utp30/Cic1 family proteins are buried inside the structure and apparently play a structural role (Fig. 1C). In general, the exposed surface

FIGURE 3. Interaction of Utp30 with 90S. (*A*) Secondary structure of 18S rRNA helix 41. The nucleotides contacted by Utp30 are colored in red and unmodeled in 90S are in gray. The top and base stems of helix 41 are termed as 41a and 41b. (*B*) Structure of helix 41 in mature 40S ribosome. The h41b structure is aligned to its counterpart in the *right* panel of *C*. The nucleotides unmodeled in 90S are colored in gray. (*C*) Interactions of Utp30 with 90S. (*D*) Detailed interactions of Utp30 with helix 41. Residues at the RNA-binding interface are shown as sticks. (*E*) Electrostatic potential surface of Utp30. The surface is colored from blue to red for positively to negatively charged regions. (*F*) Conservation surface of Utp30. The residues that are conserved in 100% and 80% of nine aligned Utp30/Cic1 sequences in Figure 1C are colored in orange and yellow, respectively. (*G*) Association of mutant Utp30 to 90S. The utp30Δ/Noc4-TAP strain was transformed with a pRS415 plasmid expressing wild-type (WT) or mutant Utp30. The Noc4-TAP particle was immunoprecipitated (IP) with IgG-coated beads and analyzed with SDS-PAGE and Western blotting. (*H*) Quantification of Utp30 binding. The Utp30 to Noc4 volume ratio was calculated and normalized against WT Utp30. Averages of three experiments and standard deviations are shown.

of Utp30 is not extensively conserved, compared to ribosomal L1 proteins (Nikulin et al. 2003; Nevskaya et al. 2005). A few basic residues at the RNA-binding interface, K245 in domain I and R107, K111, and R130 in domain II, are highly conserved (Fig. 3F).

We mutated a few residues at the RNA-binding interface of Utp30 and assessed its association with 90S particles immunoprecipitated via tandem affinity purification (TAP)-tagged Noc4, a stable component of mature 90S (Fig. 3G,H). The double mutation H37E/I39G and single mutation K245E at the core RNA-binding interface of domain I reduced the amount of bound Utp30 by ~50%, validating the importance of the RNA-binding interface. The other analyzed mutations, S198E at domain I and R107E/K111E at domain II, have only minor effects.

Comparison with other L1 domain–RNA complex structures

Ribosomal L1 proteins are universally present in the large subunit of bacterial, archaeal, and eukaryotic ribosomes. L1 binds to helices 76–78 of large subunit rRNAs, forming the mobile L1 stalk. In bacteria and archaea, L1 also binds to the leader sequence of its own mRNA, thereby negatively reg-

ulating its expression at the translational level. Structures of L1 in complex with rRNA and mRNA have been determined (Fig. 4A–D; Nikulin et al. 2003; Nevskaya et al. 2005). The core regions of rRNA and mRNA recognized by L1 share significant structural similarity and consist of two perpendicular RNA helices (called P1 and P2 for simplicity) and two interconnecting loops (J1-2 and J2-1). Domain I of L1 binds P1 duplexes (helix 77 in rRNA) of rRNA and mRNA with the concaved surface of the major sheet. Domain I additionally contacts the J1-2 region and the basal part of P2 helices (helix 78 in rRNA) with the loop between helix α1 and strand β1. Domain II of L1 contacts the J2-1 region (Loop B in rRNA). The J2-1 region of rRNA is much longer than the corresponding region of mRNA and interacts more extensively with L1, which contributes to the higher binding affinity of L1 to rRNA.

In the Utp30–rRNA complex structure, domain I of Utp30 binds to a short duplex (h41a) of rRNA, which is equivalent to the P1 duplex in L1-RNA complexes (Fig. 4E,F). The binding mode of domain I with the P1 duplex is generally preserved for Utp30 and L1, although the orientation of the duplex is shifted by ~30° between the two complex structures and the identity of RNA-binding residues are not conserved. The other RNA interactions are totally different for Utp30
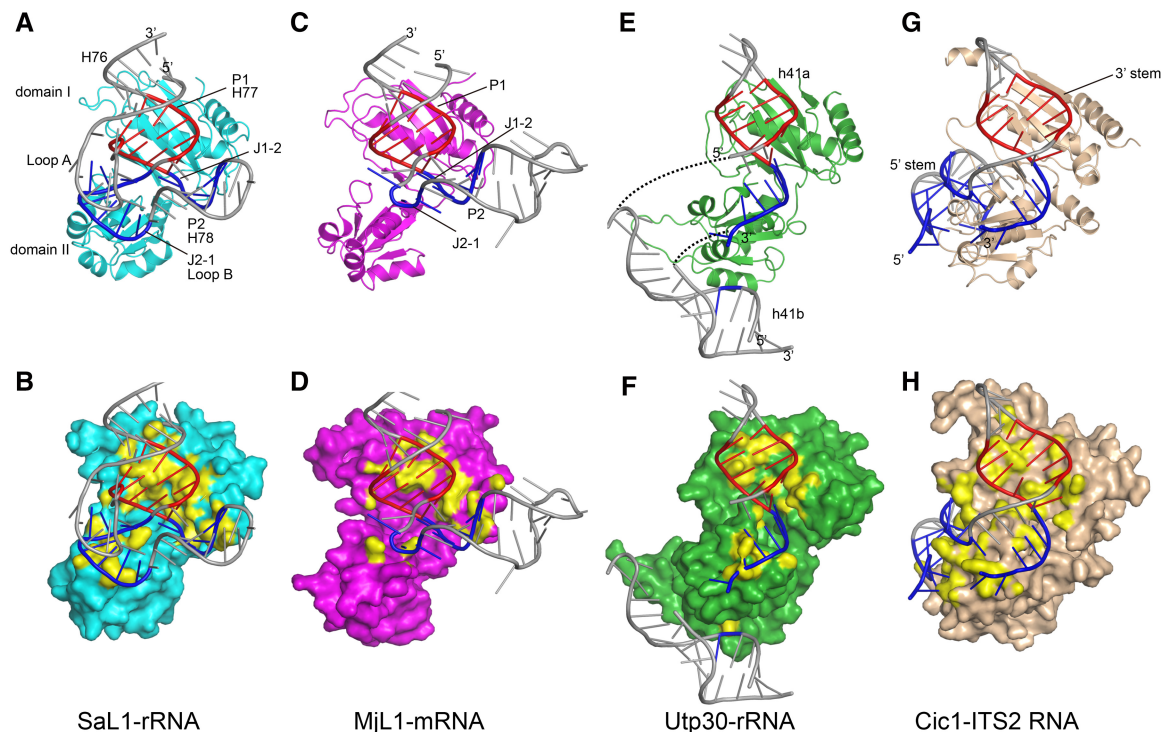
FIGURE 4. Comparison of RNA recognition by L1 domains. (*A,B*) Structure of L1 from *Sulfolobus acidocaldarius* in complex with a fragment of 23S rRNA from *T. thermophilus* (PDB code 1MZP). (*C,D*) Structure of L1 from *M. jannaschii* L1 in complex with a fragment of its mRNA (PDB code 1U63). (*E,F*) Structure of Utp30–rRNA complex in 90S preribosome. (*G,H*) Structure of Cic1–RNA complex in pre-60S preribosome (PDB code 3JCT). These structures are aligned by protein. The RNA duplexes contacted by domain I are colored in red; the other protein-binding regions of RNA are colored in blue. The proteins are shown in surface representation in *B*, *D*, *F*, and *H* and colored in yellow for the atoms that are within 4 Å of any RNA atoms.

and L1. There are no equivalent structures of J1-2 and P2 bound by domain I of Utp30. Instead, the h41a duplex is more extended and reaches into domain II of Utp30.

The structures of Cic1 and Utp30 are superimposable with an rmsd of 3.535 Å over 159 Cα atom pairs. Cic1 binds two stems of ITS2 RNA in the cryo-EM structure of a pre-60S particle (Fig. 4G,H; Wu et al. 2016). The 3′ stem is equivalent to the P1 duplex recognized by L1. The top part of the 3′ stem is recognized in a classic way by the concaved surface of domain I, and the low part of the 3′ stem extends to contact the front face of domain II. Both Cic1 and Utp30 recognize similarly an elongated duplex by domains I and II, consistent with the close evolutionary relationship between them. Cic1 also extensively contacts the 5′ stem with the back side of domain II; this interaction is unique to Cic1.

Comparison of the four RNA complex structures demonstrates that L1 domains maintain the most conserved interaction between domain I and an RNA duplex and are also highly adaptive to recognize various RNA targets.

Utp30-associated and depleted 90S particles

Each ribosome AF binds to the pre-rRNA for a certain period. To determine the association duration of Utp30, we engineered Utp30 with a C-terminal TAP-tag, affinity purified the Utp30-containing preribosomes, and identified the associated proteins with semi-quantitative mass spectrometry (Fig. 5; Supplemental Dataset 1). The Utp30-TAP particle has a very similar protein profile as the particle purified via Noc4-TAP (Zhang et al. 2016), indicating that the association periods of Utp30 and Noc4 are largely overlapped. Utp30 and Noc4 have been shown to assemble into 90S at the same late stage during the ordered assembly pathway of 90S (Chaker-Margot et al. 2015; Zhang et al. 2016). The RNA helicase Prp43 and its binding partner Pxr1 (Chen et al. 2014) are notably enriched in the Utp30-TAP particle. The functional implication of this observation is unclear at present. Utp30 also copurified with small amounts of Ltv1, Tsr1, and Rio2, which are AFs specifically bound to pre-40S particles. These suggest that Utp30 and Noc4 may not be released at the exact same time during the transition of 90S to pre-40S.

To understand the role of Utp30 on 90S assembly, the Noc4-TAP particle was purified in a yeast strain where the Utp30 gene was deleted. The Utp30-deleted yeast grew normally as the wild-type strain (data not shown), as reported previously (Schilling et al. 2012). Other than the absence of Utp30, no proteins display significant changes compared to the wild-type Noc4-TAP particle (Fig. 5). This provides biochemical evidence that Utp30 is dispensable for 90S assembly and function.

DISCUSSION

We have determined a high-resolution crystal structure of Utp30 and built a more accurate model for Utp30 in
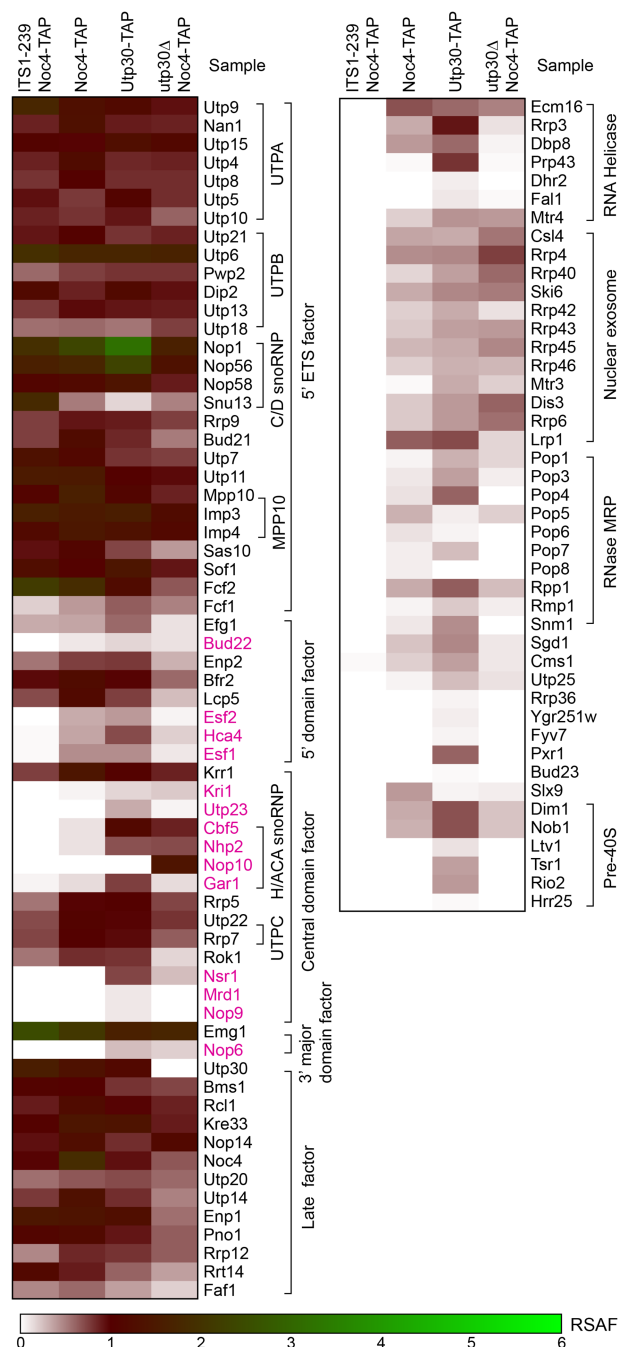
FIGURE 5. Heatmap of proteins in 90S preribosomes. The first sample (ITS1-239/Noc4-TAP) was purified via a plasmid-derived 5′-MS2-tagged pre-rRNA fragment ending at position 239 of ITS1 and Noc4-TAP and represents an unprocessed state of mature 90S with most labile factors released. The other samples were purified via Noc4-TAP, Utp30-TAP, and Noc4-TAP in utp30Δ strain. The first two samples from our previous study were included for comparison. Relative spectral abundance factors (RSAF) were normalized against six UTPB proteins and color-coded. The labile factors dissociated in mature 90S particles are colored magenta.

the 90S preribosome. The target RNA structure and RNA-binding mode of Utp30 are different from those of other characterized L1-domain proteins. Comparison of the RNA

complex structures of L1, Cic1, and Utp30 illustrates that recognition of a short RNA duplex by the concaved surface of domain I is the defining feature of L1 domains. Outside the core interaction interface, L1 domains are highly versatile in binding various RNA structures.

Utp30 and Cic1 play distinct functions in the budding yeast and are part of 90S and pre-60S preribosomes, respectively. The two paralogs were evolved only in Saccharomycetales. Most other eukaryotes contain only a single Utp30/Cic1 family protein or multiple closely related proteins that cannot be distinguished as Utp30 or Cic1. It is unclear whether the single Utp30/Cic1 family protein from other organisms is the functional counterpart of Utp30 or Cic1. At the sequence level, the single Utp30/Cic1 homologs are more related to the Cic1 clade of Saccharomycetales than the Utp30 clade (Supplemental Fig. S2). Moreover, Cic1 is an essential gene, but Utp30 is not. The function of Cic1 should be more likely to be preserved than Utp30 in the course of evolution. These considerations suggest that the single Utp30/Cic1 homolog functions as Cic1.

Alternatively, the Utp30/Cic1 homolog could play dual roles in both 90S and pre-60S preribosomes. As Utp30 and Cic1 bind similarly to the core structure of their target RNAs (h41a of 18S rRNA and 3′ stem of ITS1), a single protein may fulfill the function to bind both targets. To support this notion, the only Utp30/Cic1 protein of *Chaetomium thermophilum* has been shown to be present in 90S (Kornprobst et al. 2016).

We have shown that deletion of Utp30 did not affect the composition of 90S preribosome, indicating that no protein depends on Utp30 for association. Structurally, Utp30 is bound at the periphery of the 90S structure and makes primary interaction with the ejected helix 41 of 18S rRNA but little interactions with other proteins. The rather isolated interaction network of Utp30 may account for its nonessentiality to 90S assembly and function. Moreover, helix 41 is additionally contacted by an unassigned α-helix, which may compensate for the loss of Utp30.

Although Utp30 is dispensable for ribosome assembly, it makes a positive genetic interaction with Emg1. Deletion of Utp30 enhanced the growth defect of a temperature-sensitive mutant of Emg1 and overexpression of Utp30 suppressed the phenotype (Schilling et al. 2012). Structurally, Utp30 and Emg1 both bind to the 3′ major domain of 18S rRNA, which forms the head in the mature subunit (Fig. 2B). Combining the genetic and structural data, Utp30 appears to play an auxiliary role in the assembly of the head. The robust architecture and redundant interactions of the 90S preribosome could tolerate a certain degree of structural defects in the head assembly.

MATERIALS AND METHODS

Gene cloning and protein purification

The gene encoding residues 11–256 of Utp30 was PCR-amplified from yeast genomic DNA and cloned into a modified pET-28a vector with the In-Fusion method (TaKaRa). The expressed protein contains an N-terminal His_6-Smt3 tag. All constructs were verified by DNA sequencing.

The expression vector was transformed into *E. coli* BL21(DE3) strain. The bacteria were cultured at 37°C in LB medium containing 50 mg/mL kanamycin. When the OD600 value reached 0.8, the culture was cooled to 18°C and supplied with 1.0 mM isopropyl β-D-1-thiogalactopyranoside to induce protein expression for 16 h. The cells were harvested by centrifugation, resuspended in lysis buffer (50 mM Tris–HCl pH 8.0, 200 mM NaCl), and then broken in a high-pressure JN-3000 PLUS cell disruptor (JNBIO). After the lysate was clarified by centrifugation at 4°C, the supernatant was loaded onto a 5-mL HisTrap column (GE Healthcare) and the target protein was eluted with 0.5 M imidazole in lysis buffer. The concentration of NaCl in the pooled fractions was raised to 0.5 M to increase protein solubility. The target protein was then digested with Ulp1 for 4 h at 4°C to remove the His_6-Smt3 tag. The target protein was loaded onto a heparin column (GE Healthcare) pre-equilibrated in 0.5 M NaCl and Tris–HCl (pH 8.0) and eluted with a 0.5–1 M NaCl gradient. The eluted protein was concentrated with ultrafiltration devices of 10 kDa cutoff (Amicon) and further purified with a HiLoad 16/60 Superdex 200pg column (GE Healthcare) pre-equilibrated in 50 mM Tris–HCl (pH 8.0) and 1 M NaCl. The protein sample was concentrated to 10–40 mg/mL for crystallization or storage at −80°C.

The seleno-methionine (SeMet)-labeled protein was expressed in M9 medium as described previously (Zhang et al. 2014) and purified in the same way as the native protein except that all buffer solutions were supplied with 2 mM dithiothreitol.

Crystallization and structure determination

Initial crystallization screening was conducted at 18°C using the sitting-drop method by mixing 120 nL each of protein (50 mM Tris–HCl, pH 8.0, 1 M NaCl) and reservoir solutions. The identified crystallization conditions were optimized using the hanging-drop method with 1 µL each of protein and reservoir solutions equilibrated against 0.5 mL of reservoir solution. The native protein of Utp30 11–256 was crystalized from 1.45 M trisodium citrate dehydrate and 0.1 M 2-(*N*-morpholino)ethanesulfonic acid (MES) (pH 6.5). The SeMet-labeled protein was crystallized from 1.35 M trisodium citrate dehydrate and 0.1 M MES (pH 6.5).

The crystals were cryoprotected in 25% glycerol made from the reservoir solutions. Diffraction data were collected at BL17U and BL19U beamlines of the Shanghai Synchrotron Radiation Facility (SSRF) and processed with HKL3000 (Minor et al. 2006). A data set was collected for the native crystal at the wavelength of 0.97776 Å. Another data set was collected for the SeMet-labeled crystal at 0.97919 Å. Both crystals are isomorphic and belong to space group $C222_1$ with two molecules in the asymmetric unit. The structure was determined based on the native and derivative data sets with the single isomorphous replacement with anomalous scattering (SIRAS) method by using AutoSol of PHENIX (Adams et al. 2010). The model was initially built with AutoBuild in PHENIX and further adjusted in COOT (Emsley et al. 2010). The structure was refined against the native data set with PHENIX. The current model contains two Utp30 molecules with residues 11–110, 117–136, and 144–256 in chain A and residues 11–144 and 147–256 in chain B, 15 water molecules, and one phosphate

ion. Ramachandran analysis shows that 93.36% of the residues are in the favored regions and 6.64% in the allowed regions.

To model the Utp30 structure in 90S, the more complete chain B was docked into the cryo-EM density map of ΔMtr4 90S (Sun et al. 2017). The gap between residues 144 and 147 was filled and some loops were manually adjusted according to the density. The model was refined with molecular dynamic flexible fitting (Trabuco et al. 2008). The remodeled Utp30 structure and a few surrounding structural elements, including nt 1474–1537 of 18S rRNA, nt 247–265 of 5′ ETS RNA, an unassigned helix, were further refined against the ΔMtr4 90S map with phenix.real_space_refine (Supplemental Table S2; Adams et al. 2010).

Yeast studies

All strains were derived from BY4741 (Mat a, leu2Δ0, Met15Δ0, ura3Δ0). The Noc4-TAP strain (BY4741, Noc4-TAP::His3MX6) was purchased from Open Biosystems. The utp30Δ/Noc4-TAP strain (BY4741, utp30Δ::KanMX6, Noc4-TAP::His3MX6) was constructed from the Noc4-TAP strain by homologous recombination replacing the Utp30 gene with a PCR product generated from the pFA6A-kanMX6 cassette (Longtine et al. 1998). The Utp30-TAP strain (BY4741, Utp30-TAP::His3MX6) was constructed from BY4741 by inserting a PCR product of the TAP cassette (CaMBP-TEV-2xProtA-His3MX6) generated from the genomic DNA of Noc4-TAP strain. Yeast cells were transformed using the lithium acetate method. All strains were confirmed by PCR.

The Utp30 gene was cloned into plasmid pRS415 downstream from a GPD promoter and a 3HA-tag. Mutations were introduced by the QuikChange approach. The utp30Δ/Noc4-TAP strain was transformed with pRS415-Utp30 and grown in Leu-free Synthetic Complete medium or YPD (1% yeast extract, 2% peptone, 2% glucose) medium.

Preribosomes were purified as previously described (Sun et al. 2017). Anti-HA antibody (CST, 1:2000 dilution) and peroxidase-anti-peroxidase (PAP) antibody (Sigma-Aldrich, 1:5000 dilution) were used for Western blotting. The chemiluminescent signals were imaged with ChemiScope 5300 Pro (Clinx Science Instruments) and quantified with Quantity One (Bio-Rad). Mass spectrometry analysis was conducted as previously described (Chen et al. 2017). The total spectral counts per 100 residues (SCPHR) was calculated for each identified protein and further normalized against a group of reference proteins (UTPB proteins) to obtain the relative spectral abundance factor (RSAF) (Zhang et al. 2016).

Sequence analysis

About 1000 homologous sequences of Utp30 (*E*-value $<2 \times 10^{-7}$) were identified from the Reference Proteins Database of Eukaryota in the National Center for Biotechnology Information (NCBI) with two rounds of PSI-Blast search (Altschul et al. 1997). Few ribosomal L1 proteins were retrieved from the search. The taxonomy report was generated by NCBI. The sequences from a few selected organisms and the ribosomal L1 protein of *S. cerevisiae* were compiled and aligned with Muscle in Jalview (Waterhouse et al. 2009). Phylogenetic trees were calculated with neighbor joining using BLOSUM62 in Jalview.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

REFERENCES

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

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25:** 3389–3402.

Bassler J, Grandi P, Gadal O, Lessmann T, Petfalski E, Tollervey D, Lechner J, Hurt E. 2001. Identification of a 60S preribosomal particle that is closely linked to nuclear export. *Mol Cell* **8:** 517–529.

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 Å resolution. *Science* **334:** 1524–1529.

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.

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

Chen YL, Capeyrou R, Humbert O, Mouffok S, Kadri YA, Lebaron S, Henras AK, Henry Y. 2014. The telomerase inhibitor Gno1p/PINX1 activates the helicase Prp43p during ribosome biogenesis. *Nucleic Acids Res* **42:** 7330–7345.

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

Cheng Q, Yuan F, Lu F, Zhang B, Chen T, Chen X, Cheng Y, Li N, Ma L, Tong T. 2015. CSIG promotes hepatocellular carcinoma proliferation by activating c-MYC expression. *Oncotarget* **6:** 4733–4744.

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.

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.

Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of *Coot. Acta Crystallogr D Biol Crystallogr* **66:** 486–501.

Fatica A, Oeffinger M, Tollervey D, Bozzoni I. 2003. Cic1p/Nsa3p is required for synthesis and nuclear export of 60S ribosomal subunits. *RNA* **9:** 1431–1436.

Grandi P, Rybin V, Bassler J, Petfalski E, Strauss D, Marzioch M, Schäfer T, Kuster B, Tschochner H, Tollervey D, et al. 2002. 90S pre-ribosomes include the 35S pre-rRNA, the U3 snoRNP, and 40S subunit processing factors but predominantly lack 60S synthesis factors. *Mol Cell* **10:** 105–115.

Harnpicharnchai P, Jakovljevic J, Horsey E, Miles T, Roman J, Rout M, Meagher D, Imai B, Guo Y, Brame CJ, et al. 2001. Composition and functional characterization of yeast 66S ribosome assembly intermediates. *Mol Cell* **8:** 505–515.

Henras AK, Soudet J, Gérus 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.

Kornprobst M, Turk M, Kellner N, Cheng J, Flemming D, Koš-Braun I, Koš M, Thoms M, Berninghausen O, Beckmann R, et al. 2016. Architecture of the 90S pre-ribosome: a structural view on the birth of the eukaryotic ribosome. *Cell* **166:** 380–393.

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

Longtine MS, McKenzie A III, Demarini DJ, Shah NG, Wach A, Brachat A, Philippsen P, Pringle JR. 1998. Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae. Yeast* **14:** 953–961.

Ma L, Chang N, Guo S, Li Q, Zhang Z, Wang W, Tong T. 2008. CSIG inhibits PTEN translation in replicative senescence. *Mol Cell Biol* **28:** 6290–6301.

Minor W, Cymborowski M, Otwinowski Z, Chruszcz M. 2006. *HKL*-3000: the integration of data reduction and structure solution–from diffraction images to an initial model in minutes. *Acta Crystallogr D Biol Crystallogr* **62:** 859–866.

Nevskaya N, Tishchenko S, Gabdoulkhakov A, Nikonova E, Nikonov O, Nikulin A, Platonova O, Garber M, Nikonov S, Piendl W. 2005. Ribosomal protein L1 recognizes the same specific structural motif in its target sites on the autoregulatory mRNA and 23S rRNA. *Nucleic Acids Res* **33:** 478–485.

Nikulin A, Eliseikina I, Tishchenko S, Nevskaya N, Davydova N, Platonova O, Piendl W, Selmer M, Liljas A, Drygin D, et al. 2003. Structure of the L1 protuberance in the ribosome. *Nat Struct Biol* **10:** 104–108.

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.

Schilling V, Peifer C, Buchhaupt M, Lamberth S, Lioutikov A, Rietschel B, Kötter P, Entian KD. 2012. Genetic interactions of yeast NEP1 (EMG1), encoding an essential factor in ribosome biogenesis. *Yeast* **29:** 167–183.

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.

Trabuco LG, Villa E, Mitra K, Frank J, Schulten K. 2008. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16:** 673–683.

Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. 2009. Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25:** 1189–1191.

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

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

Zhang C, Lin J, Liu W, Chen X, Chen R, 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.

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