Published online 4 January 2017

Nucleic Acids Research, 2017, Vol. 45, No. 6 ***3559–3567***
doi: 10.1093/nar/gkw1323

Nop9 binds the central pseudoknot region of 18S rRNA

Bing Wang[1,2,3] and Keqiong Ye[3,4,*]

[1]Department of Biochemistry and Molecular Biology, College of Life Sciences, Beijing Normal University, Beijing 100875, 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 October 05, 2016; Revised December 14, 2016; Editorial Decision December 16, 2016; Accepted December 19, 2016

ABSTRACT

The assembly of eukaryotic ribosomes requires numerous factors that transiently associate with evolving pre-ribosomal particles. The Pumilio repeat-containing protein Nop9 briefly associates with the 90S pre-ribosome during its co-transcriptional assembly. Here, we show that Nop9 specifically binds an 11-nucleotide sequence of 18S rRNA that forms the 3′ side of the central pseudoknot and helix 28 in the mature subunit. Crystal structures of Nop9 in the free and RNA-bound states reveal a new type of Pumilio repeat protein with a distinct structure, target sequence and RNA-binding mode. Nop9 contains 10 Pumilio repeats arranged into a U-shaped scaffold. The target RNA is recognized by two stretches of repeats in a bipartite manner, and three central bases are unrecognized as a result of the degeneracy of repeats 6 and 7. Our data suggest that Nop9 regulates the folding of 18S rRNA at early assembly stages of 90S.

INTRODUCTION

Eukaryotic ribosomes are synthesized in a complex and dynamic process in which four rRNAs and ∼80 ribosomal proteins (r-proteins) are assembled into small and large subunits (1,2). More than 200 *trans*-acting proteins function in ribosome assembly in the budding yeast *S. cerevisiae*. A large number of small nucleolar RNAs (snoRNAs) direct the modification of rRNA, and four special snoRNAs, U3, U14, snR30 and snR10, are involved in 18S rRNA processing. Assembly factors (AFs) and snoRNAs transiently associate with a series of pre-ribosomal particles formed during the assembly pathway of two ribosomal subunits. The evolvement of pre-ribosomes is coupled with pre-rRNA processing, the binding and release of AFs and snoRNAs and structural rearrangements.

Ribosome assembly starts in the nucleolus, where the 18S, 5.8S and 25S rRNAs and four spacer sequences are transcribed as a single precursor rRNA (pre-rRNA) from rDNA gene repeats. H/ACA and C/D snoRNPs rapidly modify rRNA sequences at specific sites. The growing 5′ end of the nascent transcript is co-transcriptionally packed into a large 90S pre-ribosomal particle, which is the earliest precursor of the small ribosomal subunit (3–6). Within the 90S particle, sites A0 and A1 in the 5′ external transcribed spacer (5′ ETS) are cleaved, thus generating the 5′ end of 18S rRNA, and the A2 site in the internal transcribed spacer 1 (ITS1) is cleaved, thereby splitting the precursors of the small and large subunits. The 90S particle is assembled in a stepwise and dynamic manner (7,8). Approximately 70 assembly factors, ∼20 r-proteins, and U3, U14 and snR30 snoRNAs are sequentially deposited on the growing transcript. After helix 44 near the 3′ end of 18S rRNA is transcribed, the particle undergoes a dramatic reorganization event including the association of 10 late AFs and the release of U14 and snR30 snoRNAs and ∼14 AFs that bind earlier (7). These association and dissociation events eventually lead to the formation of a mature 90S particle capable of 18S processing. The order and molecular mechanism of the late assembly events remain unknown.

One protein among the AFs released at the last step of 90S assembly is Nop9. Nop9 is a conserved nucleolar protein required for early 18S rRNA processing (9,10). APUM23, the ortholog of Nop9 in Arabidopsis, has also been shown to function in pre-rRNA processing (11). Nop9 contains multiple Pumilio (PUM) repeats that were first characterized in the Pumilio/*fem-3* mRNA-binding factor (PUF) family proteins. PUF proteins are sequence-specific RNA-binding proteins that commonly bind the 3′-untranslated region of mRNA and regulate mRNA translation and stability (12,13). Besides Nop9, another PUM repeat-containing protein, Puf6, is involved in the assembly of the ribosome, in this case the large subunit (10,14). Puf6 appears not to recognize specific RNA sequences, but it binds structured RNA (14).

*To whom correspondence should be addressed. Tel: +86-10-64887672; Email: yekeqiong@ibp.ac.cn

The RNA-binding domain of PUF proteins, known as the Pumilio homology domain (PUM-HD) or PUF domain, typically consists of eight PUM repeats that assemble into a crescent-shaped superhelical structure (15,16). Each PUM repeat contains ∼36 residues arranged into three α-helices. The RNA is bound at the inner concave surface of the PUF domain structure in a modular fashion, and one base is recognized by one repeat (17). Bases 1–8 are bound sequentially to repeats 8-1, resulting in an anti-parallel arrangement between the RNA and protein. Three key RNA-binding residues (constituting the tripartite recognition motif, TRM) are located at positions 1, 2 and 5 of a 5-residue motif (12XX5 motif) in the second helix of each PUM repeat (17–20). The sequence specificity is primarily determined by the residues at positions 1 and 5, which bind the edge of the base through hydrogen bonding and van der Waals interactions. The residue at position 2 frequently stacks over the bases.

In this study, we located the binding site of Nop9 on 18S rRNA and characterized the structural basis of the specific Nop9–RNA interaction. Nop9 binds to a structurally critical region of 18S rRNA and would block the folding of this region. This finding explains why Nop9 must be released at the final assembly stage of 90S.

MATERIALS AND METHODS

Gene cloning and protein purification

The Nop9 gene was amplified from yeast genomic DNA and cloned into a modified pETDuet-1 plasmid, resulting in the fusion of Nop9 protein to an N-terminal His_6-Smt3 tag and PreScission cleavage site. Deletion mutations were introduced with the QuikChange system and confirmed by DNA sequencing.

Escherichia coli Rosetta (DE3) (Novagen) cells were transformed with a plasmid expressing Nop9 or its fragments. Cells were grown in LB medium containing 30 μg/ml chloramphenicol at 37°C to an A_{600} of 0.5–0.8 and then induced with 0.2 mM isopropyl β-D-1-thiogalactopyranoside for 16 h at 16°C. The cells were pelleted, resuspended in lysis buffer (50 mM Tris–Cl, pH 8.0, 200 mM NaCl, 20 mM imidazole and 5% glycerol) and lysed by sonication. After clarification, the supernatant was lysed again by sonication to fragment nucleic acids. The supernatant was filtered through a 0.45 μm filter and then loaded onto a 5-ml HisTrap column (GE Healthcare). The column was washed with lysis buffer and eluted with a linear gradient of 20–500 mM imidazole in lysis buffer. The peak fractions were incubated with PreScission protease overnight at 4°C to remove the His_6-Smt3 tag and then loaded onto a 5-ml Q column (GE Healthcare) pre-equilibrated with buffer A (50 mM Tris–Cl, pH 8.0, 200 mM NaCl). The flow-through containing the target protein was collected and loaded onto a 5-ml heparin column (GF Healthcare) pre-equilibrated with buffer A. The column was washed with 0.2 M NaCl, 10 mM Tris–Cl, pH 8.0 and eluted with a 0.2–2 M NaCl gradient. The target protein eluted at approximately 400 mM NaCl was pooled and further purified with a HiLoad 16/60 Superdex 200 column (GE Healthcare) using a buffer of 10 mM Tris–Cl (pH 8.0) and 200 mM NaCl. The purified protein fractions were concentrated to 15 mg/ml for crystallization. The selenomethione (SeMet)-substituted protein was expressed in M9 medium as described (21). The SeMet-labeled protein was purified through a similar procedure, except that the lysis buffer contained 0.5 mM DTT, and all other buffers contained 1 mM DTT.

Crystallization

Initial crystallization screens were carried out using a Mosquito (TTP Labtech) by the sitting-drop vapor-diffusion method at 20°C with 100 nl each of the protein and reservoir solutions. Crystal optimization was conducted with the hanging-drop vapor-diffusion method by mixing 1 μl each of the protein and reservoir solutions. The crystal of SeMet-labeled full-length Nop9 protein (15 mg/ml in 10 mM Tris–Cl, pH 8.0 and 200 mM NaCl) was grown in 0.2 M sodium sulfate decahydrate and 19% PEG 3350 and cryoprotected in 20% glycerol prepared in the mother solution.

RNAs were chemically synthesized and purified by Takara. The Nop9–RNA complex was assembled from 60 μM of a Nop9 truncation containing residues 53–220 and 249–634 and 90 μM of RNA (5′-AAAGGAAUUGACGGAAGG-3′) in a buffer of 10 mM Tris–Cl, pH 8.0 and 200 mM NaCl. The crystal of the complex was grown from 0.2 M sodium citrate tribasic dihydrate and 18% PEG 3350 and cryoprotected in 20% glycerol prepared in the mother solution.

Data collection and structural determination

Diffraction data were collected at the Shanghai Synchrotron Radiation Facility beamline BL17U1 and BL18U1 and processed with HKL2000 (22). The apo structure of Nop9 was determined in Phenix with the single-wavelength anomalous dispersion method based on a selenium derivative dataset (23). The model was built in Coot (24) and refined in Phenix. The crystal of apo-Nop9 belongs to the space group $I2_12_12_1$ and contains one molecule per asymmetric unit. The final model contains residues 49–76, 80–164, 176–220 and 249–633.

The crystal of the Nop9–RNA complex belongs to space group $P2_1$ and contains two complexes per asymmetric unit. The complex structure was solved by molecular replacement using the apo structure of Nop9 as the search model in Phaser (25). The resultant map showed clear densities for the missing RNA. The current model contains two copies of the complexes, each containing Nop9 residues 48 (or 51)–164, 176–220 and 249–633, nucleotides (nt) 1140–1150 of 18S rRNA and 104 water molecules. Protein residues 48–52 were from the tag sequence.

Electrophoretic mobility shift assay

Synthetic RNAs (Takara) were labeled with [γ-^{32}P]ATP by T7 polynucleotide kinase and purified on G-25 spin columns. Approximately 0.1 nM labeled RNA was incubated with protein at 0.013–51.2 μM concentrations in a 10-μl volume reaction containing 25 mM HEPES-K (pH 7.9), 100 mM NaCl, 2 mM $MgCl_2$, 1 mM DTT, 0.01% NP-40,

and 10% glycerol at room temperature for 15 min. The reaction products were resolved in 5% native polyacrylamide gels running in Tris-glycine (pH 8.3) buffer at room temperature. The gels were dried and visualized using a Typhoon PhosphorImager (GE Healthcare).

Western blot analysis

The RNPs assembled on 5′-MS2-tagged pre-rRNAs were one-step purified via RNA affinity, as described previously (7). The bound proteins were resolved in SDS-PAGE and detected by western blotting. The TAP tag was detected with the peroxidase–anti-peroxidase (PAP) antibody (Sigma). Polyclonal antibodies were raised in mouse for recombinant proteins of Enp1 and Efg1 and in rat for recombinant proteins of Nop9 and Mrd1. These antibodies were used with appropriate dilution ratios: PAP (1:10 000), anti-Nop9 (1:2000) anti-Mrd1 (1:1000), anti-Enp1 (1:5000) and anti-Efg1 (1:1000).

Accession number

The coordinates and structural factors have been deposited into the Protein Data Bank with accession codes 5WTX and 5WTY.

RESULTS

Nop9 binds a central sequence of 18S rRNA

We have previously analyzed the AFs and snoRNAs associated with a series of 3′-truncated plasmid-encoded pre-rRNAs and found that three proteins, Nsr1, Mrd1 and Nop9, are recruited to an RNA fragment ending at position 1154 of 18S rRNA, but not to a slightly shorter fragment ending at position 1137 (Figure 1A–C) (7). When the transcript is elongated to position 1770, late factors associate, and about 12 proteins, including Nop9, Nsr1 and Mrd1 and snR30 and U14 snoRNAs, dissociate. Here, we checked the association of Utp9, Nop9, Mrd1, Enp1 and Efg1 with five representative transcripts by western blotting (Figure 1D). The data confirmed the assembly and release points of these AFs, as previously derived from mass spectrometry analysis. Nop9 and Mrd1 associated with only the 18S-1154 and 18S-1643 RNAs but not with the shorter and longer RNAs. Enp1, a late factor, began to bind when the 3′ end of the transcript reached position 1770 of 18S rRNA. Utp9 and Efg1, which are recruited by the 5′ ETS and the 5′ domain of 18S RNA, respectively, associated with the 18S-1137 and longer RNAs. The amount of Efg1 was significantly reduced in the ITS1-239 RNA, suggesting that Efg1 is released upon 90S maturation. The previous mass spectrometry data also indicated a reduction in Efg1 abundance with increasing RNA length (7).

These data indicate that the association of Nsr1, Mrd1 and Nop9 critically depends on nt 1137–1154 of 18S rRNA. Interestingly, these three proteins all contain single-stranded RNA-binding domains: Nsr1 and Mrd1 have two or four RNA recognition motif (RRM) domains (26–28), and Nop9 contains multiple PUM repeats (9). We wondered whether one of these proteins might directly recognize nt 1137–1154 of 18S rRNA. Because PUM repeats recognize RNA sequences by a recognition code (17–20), we attempted to predict the target sequence of Nop9. We identified ten PUM repeats, called R1 to R10, from the Nop9 sequence. Based on the amino acid residue identities at positions 1 and 5 of the 12XX5 motif, repeats 10 through 8 may recognize a GGG triplet, and repeats 6 to 1 may recognize a 6-mer sequence, CGACGG (Figure 1E, Table 1). Repeat 7 lacks a canonical recognition motif and appears to be degenerate. Interestingly, the two predicted sequences closely matched two discontinuous segments in the 18S sequence required for Nop9 binding (Figure 1E).

Encouraged by this prediction, we examined the binding of Nop9 to nt 1137–1154 of 18S rRNA by using electrophoretic mobility shift assay (Figure 1F). Indeed, Nop9 bound the RNA with an affinity of approximately 200 nM. Mutation of either the 5′ or 3′ part of the sequence significantly reduced the binding, indicating that Nop9 specifically recognizes the sequence (Figure 1E and F). Moreover, the same two mutations disrupted the binding of Nop9 to the 18S-1154 RNA *in vivo* (Figure 1D). As a control, the association of Utp9 and Efg1, which assemble earlier than Nop9, were not affected by the mutations. Notably, Mrd1 also failed to bind the mutated 18S-1154 RNAs, suggesting that Mrd1 depends on Nop9 or the 18S sequence for association. These *in vitro* and *in vivo* binding data indicate that Nop9 specifically recognizes nt 1137–1154 of 18S rRNA.

Structure of Nop9 in the free and RNA-bound states

To understand the structural basis of Nop9 recognition of its target RNA, we determined the structure of Nop9 in the free and RNA-bound states by X-ray crystallography. The structure of free Nop9 was determined by the Se-SAD method and refined to an R_{free} value of 0.266 at 3.1 Å (Table 2). The structure reveals 10 PUM repeats (R1 to R10) and two terminal pseudo-repeats (R1′ and R10′) that stack into a right-handed U-shaped superhelix (Figure 2A and B, Supplementary Figure S1). Each PUM repeat consists of three α-helices. The first and third α-helices are located at the outer convex surface, whereas the second α-helix is located at the inner concave surface. Repeat 1 is capped by two α-helices, termed R1′, and repeat 10 is capped by four α-helices, termed R10′.

We next determined the structure of Nop9 bound to the 1137–1154 sequence of 18S rRNA. To facilitate crystallization, the two terminal tails (residues 1–52 and 635–666) and an internal loop (residues 221–248) that are disordered in the free Nop9 structure were removed. The structure was refined to an R_{free} of 0.237 at 2.8 Å and includes two complexes in the asymmetric unit (Table 2). The two complexes are highly similar in local structure and RNA-protein interaction but have slightly different overall shape (Supplementary Figure S2). The electron density map is of sufficient resolution to distinguish guanine and adenine, allowing unambiguous assignment of the RNA sequence (Figure 2D).

The Nop9 structure generally resembles other PUF domain structures (15–17,29–32), but is longer (10 versus 8 repeats) and much more curved (U shape versus crescent shape) (Figure 2E and F). In this regard, Nop9 is similar to Puf6, which contains 11 repeats arranged in an L-shape (14).

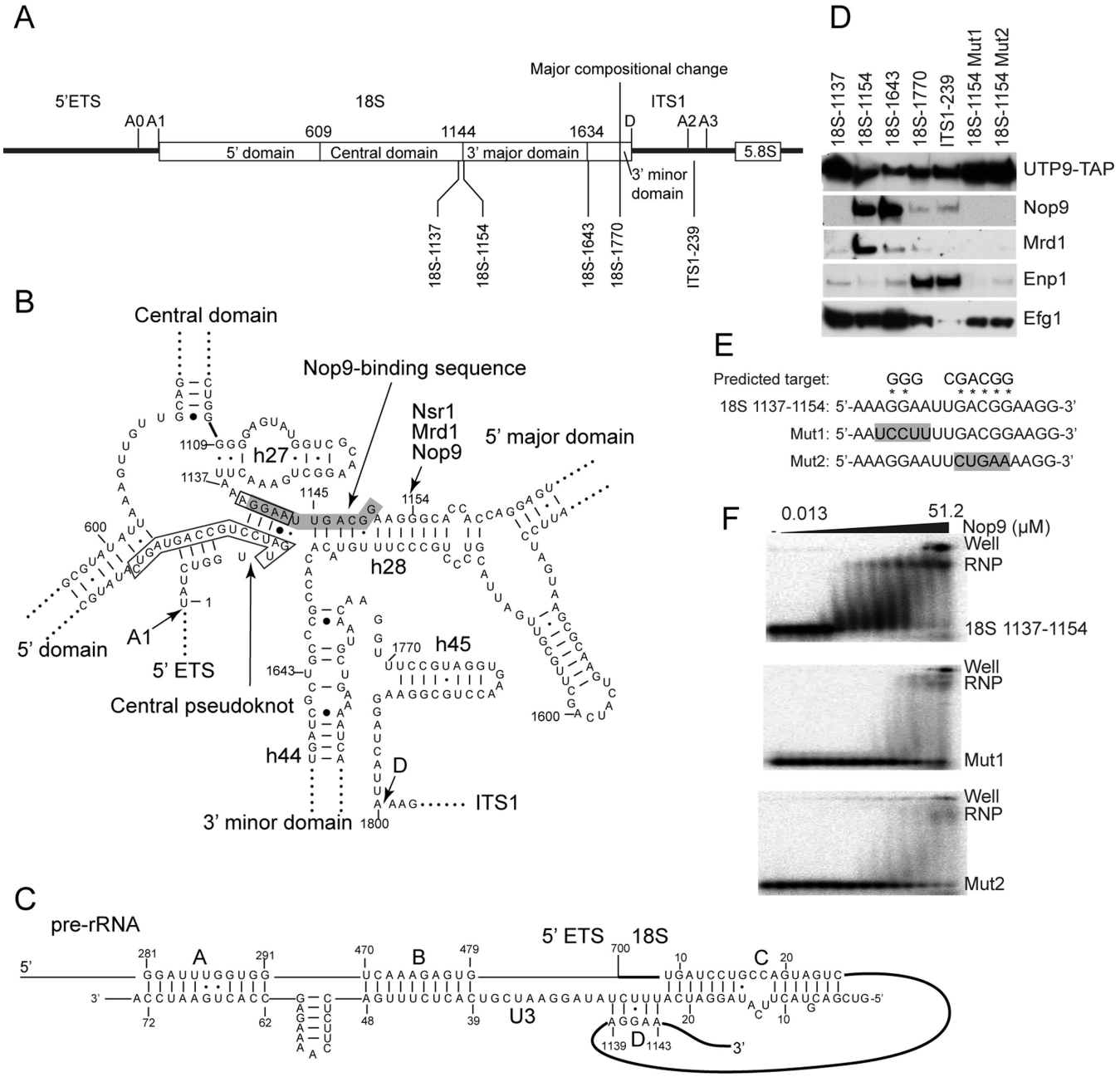

Figure 1. Nop9 specifically recognizes a central sequence of 18S rRNA. (**A**) Structure and processing sites of the 5′ portion of 35S pre-rRNA. The boundaries of four 18S domains and the ending positions of the analyzed RNA fragments are marked. The Nop9-binding sequence is shaded. (**B**) Secondary structure of 18S rRNA around the central pseudoknot. The two sites recognized by U3 snoRNA are boxed. (**C**) Secondary structural model showing interactions between the 5′ domain of U3 snoRNA and pre-rRNA. (**D**) Western blot analysis of AFs associated with pre-18S RNA fragments. The RNAs were expressed from plasmids in the Utp9-TAP strain and pulled down via a 5′ MS2-tag. The associated proteins were immunoblotted with PAP to detect the TAP tag in Utp9 and polyclonal antibodies raised against Nop9, Mrd1, Enp1 and Efg1. (**E**) Alignment of the two predicted target sequences of Nop9 with 18S rRNA. Also shown are two mutant sequences with the altered bases shaded. (**F**) RNA binding assay of Nop9. Sequence 1137–1154 of 18S rRNA and its mutants were 5′-^{32}P-labeled and assembled with Nop9 protein at concentrations of 0, 0.013, 0.025, 0.05, 0.1, 0.2, 0.4, 0.8, 1.6, 3.2, 6.4, 12.8 and 51.2 μM. The free and bound RNAs were resolved on native gels and visualized by autoradiograph.

Table 1. Prediction of target sequence of Nop9

Repeat	12XX5	Predicted base	Observed base
R1	SKXXE	G	G1150
R2	SHXXE	G	G1149
R3	SHXXR	C	C1148
R4	SPXXQ	A	A1147
R5	SHXXE	G	G1146
R6	AFXXR	C	—
R7	GTXXI	?	—
R8	SVXXE	G	A1142
R9	SHXXE	G	G1141
R10	SHXXD	G?	G1140

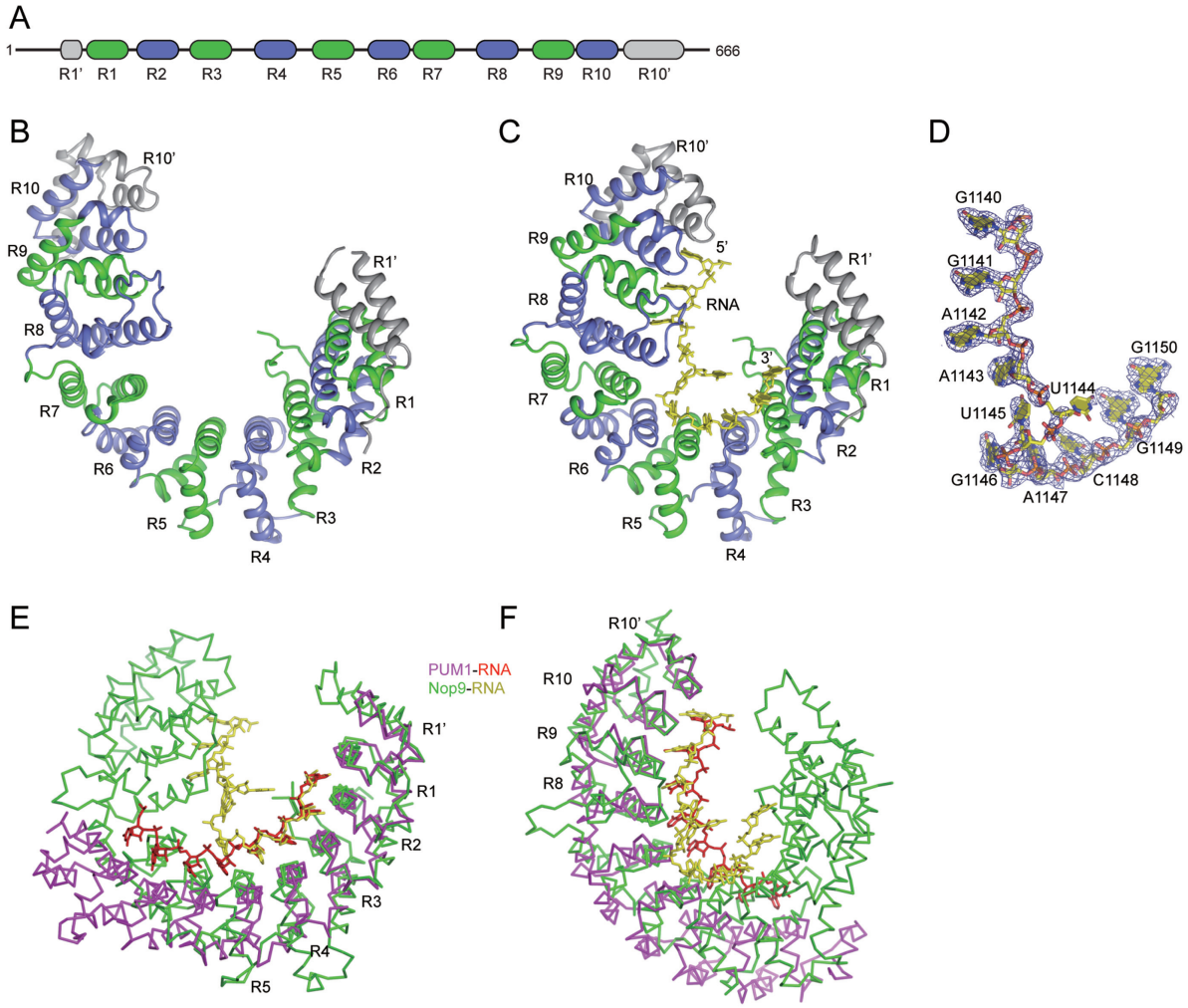

Figure 2. Crystal structure of Nop9 (**A**) Domain diagram of Nop9. (B and C) Ribbon representation of Nop9 structure in the free (**B**) and RNA-bound (**C**) states. The ten PUM repeats (R1–R10) are colored alternately in green and sky blue. R1′ and R10′ are colored gray. RNA is colored yellow. (**D**) The $2f_o - f_c$ electron density map, contoured at the 1.0σ level for the bound RNA. (**E** and **F**) Alignment of the RNA complex structures of Nop9 and human PUM1 (PDB code: 1M8Y). PUM1 repeats 1–5 and 6–8 were separately aligned to Nop9 repeats 1–5 and 8–10 in E and F, respectively. Nop9 and its bound RNA are colored in green and yellow. PUM1 and its bound RNA are colored in magenta and red.

The Nop9 structure becomes more compact after RNA binding (Figure 2B and C, Supplementary Figure S2B and D). The two Nop9–RNA complexes in the asymmetric unit also display slightly different curvature (Supplementary Figure S2A and C). The bending mainly occurs at the middle part of the structure around repeats 4–8. The plasticity of the Nop9 structure contrasts with the rigid structures of PUM1 and Puf4, which display no conformational change after RNA binding (15,17,32).

RNA sequence recognition by Nop9

Nucleotides 1140–1150, termed as the Nop9 binding sequence (NBS), are resolved in the complex structure and

Table 2. Data collection and refinement statistics

Crystal form	SeMet-Nop9	Nop9–RNA complex
Data collection		
Space group	$I2_12_12_1$	$P2_1$
Cell dimensions		
a, *b*, *c* (Å)	93.801, 134.846, 145.921	69.503, 102.744, 114.708
α, β, γ(°)	90.0, 90.0, 90.0	90.0, 95.5, 90.0
Wavelength (Å)	0.9793	0.9793
Resolution (Å)	50.0–3.1 (3.15–3.1)	50.0–2.8 (2.85–2.8)
R_{merge} (%)	19.5 (>100)	9.3 (83.0)
$<I>/<\sigma(I)>$	17.1 (1.7)	18.3 (3.0)
Completeness (%)	99.7 (100)	99.6 (100)
Redundancy	14.4 (14.7)	5.3 (5.7)
Refinement		
Resolution range (Å)	49–3.1 (3.27–3.1)	20–2.8 (2.85–2.8)
No. reflections	17 549 (2825)	39 356 (2255)
R_{work}	0.194 (0.271)	0.180 (0.269)
R_{free}	0.266 (0.345)	0.237 (0.360)
No. atoms	4460	9520
Protein	4460	8934
RNA	0	482
Water	0	104
Mean *B* (Å^2)	67.12	39.07
Rmsd bond lengths (Å)	0.010	0.008
Rmsd bond angles (°)	1.275	1.199
Ramachandran plot		
Favored (%)	90.65	95.74
Allowed (%)	8.04	3.99
Outliers (%)	1.31	0.28

Values in parentheses are for highest-resolution shell.

are bound at the inner surface of the Nop9 structure. Other nucleotides in the crystallized RNA are not visible in the density map and are apparently disordered. In contrast to the typical sequential recognition of an RNA sequence by PUM repeats (17), the 11 bases in the NBS are recognized in a bipartite manner. The five 3′ bases G1146–G1150 are bound consecutively by repeats 5 to 1, and the three 5′ bases G1140–A1142 are bound consecutively by repeats 10 to 8 (Figure 3A and C). The N- and C-terminal part of the Nop9–RNA complex can be aligned individually, but not as a whole, with the PUM1–RNA complex structure (Figure 2E and F) (17). As a result of the bipartite interaction, the 5′ and 3′ segments of the RNA are nearly perpendicular to each other.

The linker region between the two recognized RNA segments, namely nucleotides A1143, U1144 and U1145, is flipped out and is not recognized in a sequence-specific manner (Figure 3B). U1144 has little interaction with the protein and shows poor electron density (Figure 2D). The base of A1143 is stacked by the side chain of Arg485, but has no contact at the edge. The interruption of the consecutive recognition of the RNA sequence is caused by the degeneracy of repeats 6 and 7. Repeat 7 does not bind RNA at all, whereas repeat 6 contacts G1146 with a stacking interaction and a hydrogen bonding interaction to the 2′-hydroxyl group of G1146.

The Nop9–RNA complex structure illustrates the sequence-specific recognition mode for five guanines (G1140, G1141, G1146, G1149 and G1150), two adenines (A1142, A1147) and one cytosine (C1148). The basic binding mode is conserved, but the structure also reveals variations in RNA recognition by PUM repeats.

Except for G1140, the remaining four of the five guanines are recognized by an SE motif (residues at position 1 and 5), as previously observed (17). The hydroxyl group of the serine at position 1 forms a hydrogen bond with the N2 group of guanine. The carboxyl group of the glutamate at position 5 interacts with the N1 group and sometimes additionally with the N2 group of guanine. The side chains of the glutamate display conformational variations in the four repeats, resulting in slightly different interactions with the bases. Guanine recognition has previously been observed only once, by repeat 7 of the PUF domain. Although other recognition codes for guanine have been identified (18), the prevalence of the SE motif in Nop9 suggests that it is the most frequent code for guanine.

G1140 is the 3′ terminal nucleotide and is recognized by a novel SD motif in repeat 10 and extra residues in repeat 10′. In repeat 10, S565 binds the N2 group of guanine. Despite having a shorter side chain, Asp569 at position 5 contacts the N1 group of guanine like a glutamate. In repeat 10′, Asn608 further specifies the O6 atom of G1140 and Gln604 makes a stacking interaction with G1140.

A1147 is bound by an SQ motif in repeat R4 similarly to previous observations (17). Gln305 interacts with the N1 group of A1147 (Figure 3A). Ser301 makes a van der Waals interaction with the C2 group of A1142 and also stabilizes the side chain of Gln305. Unexpectedly, the other adenine, A1142, is recognized by an SE motif in repeat 8 that normally specifies guanine (Figure 3B). The C2 group of A1142 is contacted by a van der Waals interaction from Ser488, and the N6 group of A1142 is bound by the carboxylic group of Glu492. Val489 at position 2 of repeat 8 does not stack over A1142. Instead, the side chain of Arg485, located

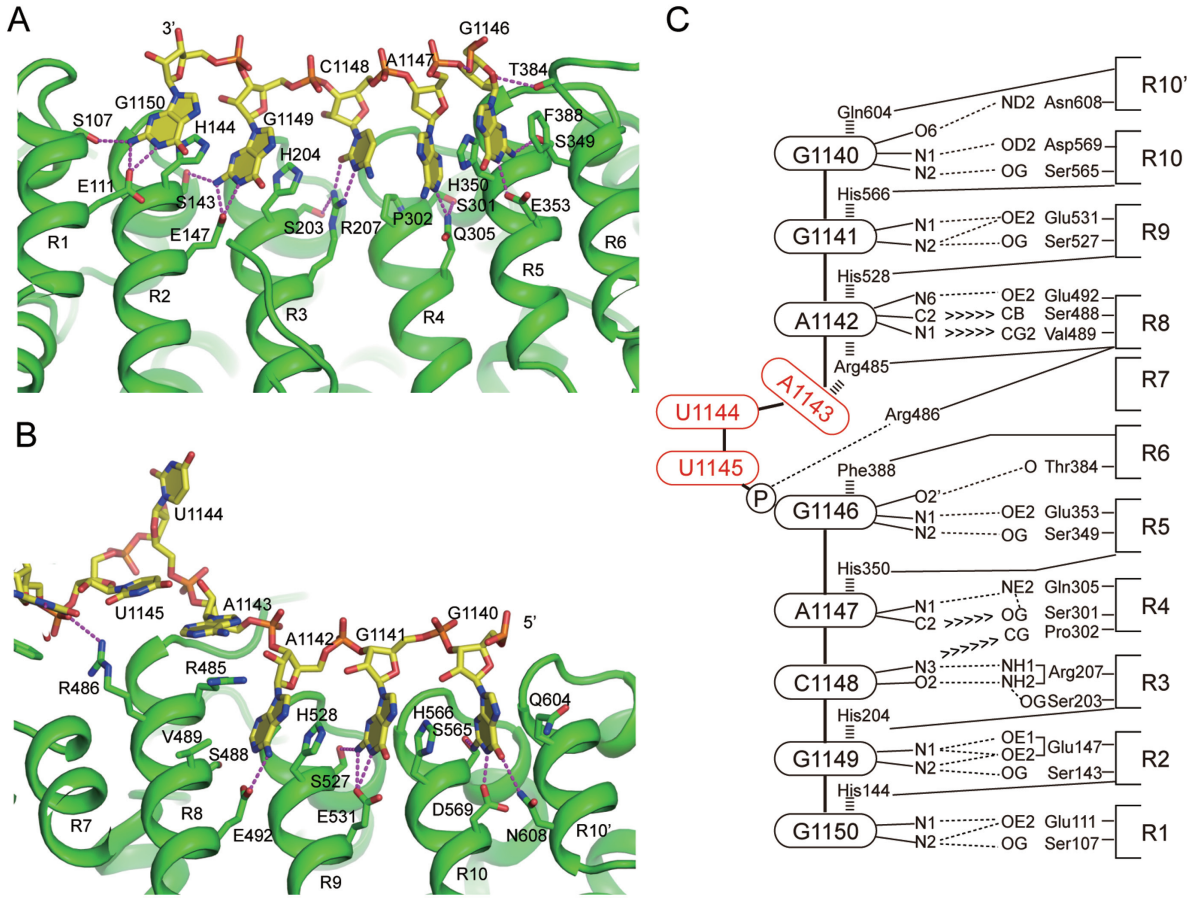

Figure 3. Interaction between Nop9 and target RNA. (**A** and **B**) Recognition of the 3′- (**A**) and 5′-half (**B**) of the target RNA. RNA and the side chains of the interacting protein residues are shown as sticks. Nitrogen is blue, oxygen is red, phosphorus is orange, and carbon is green for Nop9 and yellow for RNA. Dashed lines stand for hydrogen bonds. (**C**) Schematic of Nop9–RNA interaction. Dashed lines, '>>>' and stacked bars denote hydrogen bonds, van der Waals interactions and stacking interactions, respectively. The three unrecognized bases are colored in red.

at the extended N-terminus of the second helix of repeat 8, wedges between the bases of A1142 and A1143 and pushes A1143 out of the track of the 5′ RNA segment.

Our structure shows two divergent recognition codes for A1142 and G1140. Both bases are located at the termini of a consecutively recognized RNA sequence and adjacent to non-canonical repeats (R10′, R7). The different environment may alter the binding specificity and hence complicate the target prediction.

Naturally occurring PUM repeats that recognize cytosine have not been identified. However, directed evolution studies have shown that PUM repeats having a residue with a small side chain (Gly, Ala, Ser, Thr, Cys) at position 1 and an arginine residue at position 5 specifically bind to cytosine (19,20). Our structure illustrates for the first time how a cytosine is specifically recognized in a natural context. C1148 in the NBS is recognized by an SR motif in repeat R3, consistent with a previous prediction (19). Specifically, the side chain of Arg207 at position 5 makes two hydrogen bonds with the Watson-Crick edge of C1148. Ser203 at position 1 does not contact the base but stabilizes the conformation of Arg207. These interactions are very similar to those observed between cytosine and an engineered SYXXR motif (19). However, C1148 in our structure makes no stacking interaction with the residues at position 2 of repeats 3 and 4.

DISCUSSION

The final assembly stage of the 90S pre-ribosome consists of numerous association and dissociation events. The order and logic underling these complicated assembly events is unknown. Our biochemical and structural data reveal that Nop9 specifically recognizes nt 1140–1150 of 18S rRNA during co-transcriptional assembly of 90S, providing major insight into the function of Nop9 in 90S maturation. The binding site of Nop9 is located at a structurally critical region (Figure 1B). The 5′ half of the NBS (1140-GGAAU-1144) pairs with the 5′ end region (10-GAUCC-14) of 18S rRNA to form the central pseudoknot (CPK) in the mature subunit. The CPK is a prominent conserved structure that organizes the global architecture of small subunits and constitutes the decoding center (33).

The CPK formation is regulated by U3 snoRNA that is a key component of 90S pre-ribosomes. U3 binds at two sites in the 5′ ETS (34–37). In addition, U3 is proposed to bind both the 5′ (nt 9–25) and 3′ sides (nt 1139–1143) of the CPK (Figure 1C) (38–40). The U3-18S interactions bring the two

sides of the CPK close and may facilitate its formation. The base pairing interaction between U3 and the 5′ side of the CPK was supported by the observation that mutations in the 5′ side of the CPK resulted in defective pre-rRNA processing at sites A1 and A2 and the phenotype can be largely rescued by the compensatory mutation in U3 (38). The mutation on the 3′ side of the CPK (3′-CPK) also inhibited cleavage at sites A1 and A2, but the inhibition cannot be rescued by the compensatory mutation in U3 (38). Our results show that mutations of the 3′-CPK would affect Nop9 binding (Figure 1D), thus providing an alternative explanation of the mutational phenotype.

Association of Nop9 would exclude U3 from binding to the 3′-CPK. Nevertheless, U3 might bind the 3′-CPK before association of Nop9 or after release of Nop9. Since Nop9 binds the 3′-CPK sequence as soon as it emerges from the pre-rRNA, the interaction of U3 with 3′-CPK is more likely to occur subsequently to Nop9 release. Nop9 might assist the duplex formation between U3 and 3′-CPK through a hand-off mechanism.

The formation of the CPK requires release of Nop9 and U3 from 18S rRNA. Nop9 is dissociated before the cleavage-competent 90S particle is assembled (7), thus freeing the 3′-CPK. U3 is released during the transition of 90S to pre-40S after the pre-rRNA is cleaved at A0, A1 and A2 sites. The RNA helicase Dhr1 has been proposed to unwind the U3-18S duplex (41).

The 3′ part of the NBS (1145-UGACGG-1150) forms helix 28 in the head domain. Obviously, the association of Nop9 would prevent the target sequence from pairing with other RNA elements and would thereby block the ribosome assembly from proceeding. In the Nop9-bound state, the CPK, the putative duplex between U3 and the 3′-CPK and helix 28 cannot be assembled. In this way, Nop9 may protect the target sequence from forming premature structures when other structural elements needed for productive assembly are not yet in place.

Nop9 and other factors are released after helix 44 is transcribed. Although it is unknown which type of structure the target sequence will form during the last assembly stage of 90S, the release of Nop9 is a prerequisite for the subsequent assembly of the target sequence. Our findings suggest that Nop9 dissociation is a key step that coordinates the concerted late reorganization events of 90S assembly.

Recently, systematic evolution of ligands by exponential enrichment (SELEX) has been applied to find the RNA target sequence of the Arabidopsis Nop9, APUM23 (42). The derived 10-nt consensus sequence closely matches the yeast target sequence identified here, further supporting our conclusion. However, the previous assignment of base-repeat pairing for the 5′ half of NBS is not consistent with our structure. APUM23 binds the target RNA much more tightly than Nop9, and the SELEX analysis of Nop9 was not successful, suggesting that APUM23 and Nop9 might not bind the target RNA in the same way (42).

The 11-nt target sequence of Nop9 is identical in the 18S rRNAs of *S. cerevisiae*, *S. pombe*, *A. thaliana*, *D. melanogaster* and *H. sapiens*. Interestingly, the TRMs of PUM repeats are generally conserved, but not identical, in the Nop9 orthologs from these organisms. For example, the Glu in the SE motif for guanine recognition is frequently replaced by Gln and Asp (Supplementary Figure S1). These natural variations demonstrate the diversity and plasticity of base recognition by PUM repeats. In addition, the function of Nop9 might not require optimal binding for all bases of the target sequence.

Because of the modular recognition of RNA sequences and the simple recognition code, PUF domains have been engineered to target designed RNA sequences for various purposes (43,44). These efforts have all been based on the classic PUF domains, which contain eight repeats and bind similar sequences that bear a 5′ UGUR (where R is a purine) motif and a 3′ UA motif. The intermediate nucleotides between the two motifs are variable in sequence and length and can be recognized in non-canonical ways (29,31,32). Nop9 displays a distinct structure, target sequence and RNA binding mode from classic PUF domains and would provide a different scaffold for engineering sequence-specific RNA-binding domains.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank the staff in the Shanghai Synchrotron Radiation Facility BL17U1 and BL18U1 for assistance in data collection and the Antibody Center of National Institute of Biological Sciences for antibody preparation.

FUNDING

National Natural Science Foundation of China [31430024, 91540201, 31325007]; Strategic Priority Research Program of the Chinese Academy of Sciences [XDB08010203]; Beijing Municipal Government. Funding for open access charge: National Natural Science Foundation of China
Conflict of interest statement. None declared.

REFERENCES

1. de la Cruz,J., Karbstein,K. and Woolford,J.L.. (2015) Functions of ribosomal proteins in assembly of eukaryotic ribosomes in vivo. *Annu. Rev. Biochem.*, **84**, 93–129.
2. Woolford,J.L. and Baserga,S.J. (2013) Ribosome biogenesis in the yeast Saccharomyces cerevisiae. *Genetics*, **195**, 643–681.
3. 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.
4. 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.
5. 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.
6. 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.

7. 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.
8. 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.
9. Thomson,E., Rappsilber,J. and 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.
10. Li,Z., Lee,I., Moradi,E., Hung,N.J., Johnson,A.W. and Marcotte,E.M. (2009) Rational extension of the ribosome biogenesis pathway using network-guided genetics. *PLoS Biol.*, **7**, e1000213.
11. Abbasi,N., Kim,H.B., Park,N.I., Kim,H.S., Kim,Y.K., Park,Y.I. and Choi,S.B. (2010) APUM23, a nucleolar Puf domain protein, is involved in pre-ribosomal RNA processing and normal growth patterning in Arabidopsis. *Plant J.*, **64**, 960–976.
12. Quenault,T., Lithgow,T. and Traven,A. (2011) PUF proteins: repression, activation and mRNA localization. *Trends Cell Biol.*, **21**, 104–112.
13. Wickens,M., Bernstein,D.S., Kimble,J. and Parker,R. (2002) A PUF family portrait: 3′UTR regulation as a way of life. *Trends Genet.*, **18**, 150–157.
14. Qiu,C., McCann,K.L., Wine,R.N., Baserga,S.J. and Hall,T.M. (2014) A divergent Pumilio repeat protein family for pre-rRNA processing and mRNA localization. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 18554–18559.
15. Wang,X., Zamore,P.D. and Hall,T.M. (2001) Crystal structure of a Pumilio homology domain. *Mol. Cell*, **7**, 855–865.
16. Edwards,T.A., Pyle,S.E., Wharton,R.P. and Aggarwal,A.K. (2001) Structure of Pumilio reveals similarity between RNA and peptide binding motifs. *Cell*, **105**, 281–289.
17. Wang,X., McLachlan,J., Zamore,P.D. and Hall,T.M. (2002) Modular recognition of RNA by a human pumilio-homology domain. *Cell*, **110**, 501–512.
18. Campbell,Z.T., Valley,C.T. and Wickens,M. (2014) A protein-RNA specificity code enables targeted activation of an endogenous human transcript. *Nat. Struct. Mol. Biol.*, **21**, 732–738.
19. Dong,S., Wang,Y., Cassidy-Amstutz,C., Lu,G., Bigler,R., Jezyk,M.R., Li,C., Hall,T.M. and Wang,Z. (2011) Specific and modular binding code for cytosine recognition in Pumilio/FBF (PUF) RNA-binding domains. *J. Biol. Chem.*, **286**, 26732–26742.
20. Filipovska,A., Razif,M.F., Nygard,K.K. and Rackham,O. (2011) A universal code for RNA recognition by PUF proteins. *Nat. Chem. Biol.*, **7**, 425–427.
21. 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.
22. Murshudov,G.N., Vagin,A.A., Lebedev,A., Wilson,K.S. and Dodson,E.J. (1999) Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr. D Biol. Crystallogr.*, **55**, 247–255.
23. 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.
24. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
25. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
26. Segerstolpe,A., Granneman,S., Bjork,P., de Lima Alves,F., Rappsilber,J., Andersson,C., Hogbom,M., Tollervey,D. and 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.
27. Lee,W.C., Zabetakis,D. and 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.
28. Kondo,K. and 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.
29. Wilinski,D., Qiu,C., Lapointe,C.P., Nevil,M., Campbell,Z.T., Tanaka Hall,T.M. and Wickens,M. (2015) RNA regulatory networks diversified through curvature of the PUF protein scaffold. *Nat. Commun.*, **6**, 8213.
30. Zhu,D., Stumpf,C.R., Krahn,J.M., Wickens,M. and Hall,T.M. (2009) A 5′ cytosine binding pocket in Puf3p specifies regulation of mitochondrial mRNAs. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 20192–20197.
31. Wang,Y., Opperman,L., Wickens,M. and Hall,T.M. (2009) Structural basis for specific recognition of multiple mRNA targets by a PUF regulatory protein. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 20186–20191.
32. Miller,M.T., Higgin,J.J. and Hall,T.M. (2008) Basis of altered RNA-binding specificity by PUF proteins revealed by crystal structures of yeast Puf4p. *Nat. Struct. Mol. Biol.*, **15**, 397–402.
33. Ben-Shem,A., Garreau de Loubresse,N., Melnikov,S., Jenner,L., Yusupova,G. and Yusupov,M. (2011) The structure of the eukaryotic ribosome at 3.0 A resolution. *Science*, **334**, 1524–1529.
34. Marmier-Gourrier,N., Clery,A., Schlotter,F., Senty-Segault,V. and 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.
35. Dutca,L.M., Gallagher,J.E. and Baserga,S.J. (2011) The initial U3 snoRNA:pre-rRNA base pairing interaction required for pre-18S rRNA folding revealed by in vivo chemical probing. *Nucleic Acids Res.*, **39**, 5164–5180.
36. Beltrame,M. and Tollervey,D. (1992) Identification and functional analysis of two U3 binding sites on yeast pre-ribosomal RNA. *EMBO J.*, **11**, 1531–1542.
37. Beltrame,M. and Tollervey,D. (1995) Base pairing between U3 and the pre-ribosomal RNA is required for 18S rRNA synthesis. *EMBO J.*, **14**, 4350–4356.
38. Sharma,K. and Tollervey,D. (1999) Base pairing between U3 small nucleolar RNA and the 5′ end of 18S rRNA is required for pre-rRNA processing. *Mol. Cell. Biol.*, **19**, 6012–6019.
39. Mereau,A., Fournier,R., Gregoire,A., Mougin,A., Fabrizio,P., Luhrmann,R. and Branlant,C. (1997) An in vivo and in vitro structure-function analysis of the Saccharomyces cerevisiae U3A snoRNP: protein-RNA contacts and base-pair interaction with the pre-ribosomal RNA. *J. Mol. Biol.*, **273**, 552–571.
40. Hughes,J.M. (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.
41. Sardana,R., Liu,X., Granneman,S., Zhu,J., Gill,M., Papoulas,O., Marcotte,E.M., Tollervey,D., Correll,C.C. and Johnson,A.W. (2015) The DEAH-box helicase Dhr1 dissociates U3 from the Pre-rRNA to promote formation of the central pseudoknot. *PLoS Biol.*, **13**, e1002083.
42. Zhang,C. and Muench,D.G. (2015) A nucleolar PUF RNA-binding protein with specificity for a unique RNA sequence. *J. Biol. Chem.*, **290**, 30108–30118.
43. Wei,H. and Wang,Z. (2015) Engineering RNA-binding proteins with diverse activities. *Wiley Interdiscip. Rev. RNA*, **6**, 597–613.
44. Cheong,C.G. and Hall,T.M. (2006) Engineering RNA sequence specificity of Pumilio repeats. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 13635–13639.