Nop9 recognizes structured and single-stranded RNA elements of preribosomal RNA

JUN ZHANG,1,2 TAKAMASA TERAMOTO,1,3 CHEN QIU,1 ROBERT N. VINE,1 LAUREN E. GONZALEZ,1 SUSAN J. BASERGA,4,5,6 and TRACI M. TANAKA HALL1

1Epigenetics and Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, North Carolina 27709, USA
2Department of Chemistry, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA
3Department of Bioscience and Biotechnology, Faculty of Agriculture, Kyushu University, Fukuoka 819-0395, Japan
4Department of Genetics, Yale University School of Medicine, New Haven, Connecticut 06520, USA
5Department of Therapeutic Radiology, Yale University School of Medicine, New Haven, Connecticut 06520, USA
6Department of Molecular Biophysics and Biochemistry, Yale University School of Medicine, New Haven, Connecticut 06520, USA

ABSTRACT
Nop9 is an essential factor in the processing of preribosomal RNA. Its absence in yeast is lethal, and defects in the human ortholog are associated with breast cancer, autoimmunity, and learning/language impairment. PUF family RNA-binding proteins are best known for sequence-specific RNA recognition, and most contain eight α-helical repeats that bind to the RNA bases of single-stranded RNA. Nop9 is an unusual member of this family in that it contains eleven repeats and recognizes both RNA structure and sequence. Here we report a crystal structure of Saccharomyces cerevisiae Nop9 in complex with its target RNA within the 20S preribosomal RNA. This structure reveals that Nop9 brings together a carboxy-terminal module recognizing the 5′ single-stranded region of the RNA and a bifunctional amino-terminal module recognizing the central double-stranded stem region. We further show that the 3′ single-stranded region of the 20S target RNA adds sequence-independent binding energy to the RNA–Nop9 interaction. Both the amino- and carboxy-terminal modules retain the characteristic sequence-specific recognition of PUF proteins, but the amino-terminal module has also evolved a distinct interface, which allows Nop9 to recognize either single-stranded RNA sequences or RNAs with a combination of single-stranded and structured elements.

Keywords: Nop9; RNA; PUF protein; ribosome biogenesis

INTRODUCTION
RNA-binding proteins must select their intended targets with the appropriate fidelity to regulate specific networks of RNAs. To do so, different modules may be combined in a single protein to bring together their distinct RNA recognition capabilities and diversify the RNA sequences and structures that are bound by the whole protein versus the individual domains. Many RNA-binding proteins contain multiple copies of prominent RNA-binding domains, such as RNA recognition motifs and K-homology domains (Lunde et al. 2007). Domains contribute their individual RNA recognition properties to produce an overall specificity for the multidomain protein. The overall specificity may be additive with domains connected by flexible linkers like beads on a string and able to achieve many different arrangements. Quite often the domains are interdependent, and the specificity of the multidomain protein relies on restricted relative positions of the individual domains as they bind to their target RNAs.

Pumilio/fem-3 Binding Factor (PUF) proteins are a relatively small but unique class of modular RNA-binding proteins. PUF proteins comprise α-helical Pumilio (PUM) repeats (Barker et al. 1992; Macdonald 1992; Zamore et al. 1997; Zhang et al. 1997). Classical PUF proteins are formed from eight PUM repeats, and they bind sequence specifically to single-stranded RNA. Each PUM repeat comprises three α helices, and the second helix contains three amino acid side chains called the tripartite recognition motif (TRM) that select a specific RNA base (Edwards 2000).
et al. 2001; Wang et al. 2001, 2002; Campbell et al. 2014). The eight PUM repeats of a classical PUF protein are interdependent and arrange to form a crescent-shaped PUM homology domain (PUM-HD) with single-stranded RNA binding on the concave surface. The overall conformation of a PUM-HD correlates with the length of RNA sequence that it binds, with flatter surfaces selecting longer RNA sequences. Interaction with different length sequences is often accomplished by a two-handed mechanism: one set of PUM repeats recognizes a conserved 5′ UGUA sequence while another set of PUM repeats recognizes a 3′ element with variability interspersed (Wang et al. 2002, 2009; Miller et al. 2008; Zhu et al. 2009; Lu and Hall 2011; Qiu et al. 2012, 2019; Valley et al. 2012; Wilinski et al. 2015; Bhat et al. 2019).

The PUF protein family also includes two groups of proteins that contain PUM repeats but are structurally and functionally different from the classical PUF proteins that recognize single-stranded RNA sequences (Qiu et al. 2014; Zhang et al. 2016a). Both Nop9 and Puf-A (also known as mammalian PUM3, an ortholog of yeast Puf6) recognize structured RNA and participate in preribosomal RNA processing (Thomson et al. 2007; Li et al. 2009; Tafforeau et al. 2013; Qiu et al. 2014; Zhang et al. 2016a; Wang and Ye 2017). Ribosome biogenesis is an ordered process that requires hundreds of factors to assemble ribosomal RNAs (rRNAs) and ribosomal proteins into the large ribonucleoprotein complex for protein synthesis (Woolford and Baserga 2013; de la Cruz et al. 2015; Henras et al. 2015; Lafontaine 2015). The pre-rRNA undergoes multiple cleavage and trimming steps to remove the external and internal transcribed spacers and generate the mature rRNAs. Nop9 and Puf-A proteins are critical for small and large ribosomal subunit assembly, respectively (Thomson et al. 2007; Li et al. 2009; Tafforeau et al. 2013; Qiu et al. 2014; Zhang et al. 2016a; Wang and Ye 2017). Alterations in expression level or sequence of human NOP9 and Puf-A are associated with breast cancer, autoimmunity, and learning/language impairment, indicating the importance of these atypical PUF proteins for human health (Brickner et al. 2001; Akatsuka et al. 2003; Kuo et al. 2009; Fan et al. 2013; Turpeinen et al. 2013; Nudel et al. 2014).

Nop9 is essential for maturation of the 18S small ribosomal subunit rRNA (Thomson et al. 2007; Li et al. 2009). Depletion of Nop9 in yeast is lethal, as it abolishes synthesis of the small ribosomal subunit (Thomson et al. 2007; Li et al. 2009; Chaker-Margot et al. 2015; Zhang et al. 2016b). During ribosome biogenesis the small ribosomal subunit 20S pre-rRNA is produced in the nucleus, released to the nucleoplasm, and then exported to the cytoplasm where cleavage at site D by the endonuclease Nob1 removes internal transcribed spacer 1 (ITS1) sequences to produce the mature 18S rRNA (Fig. 1A; Fatica et al. 2003; Lamanna and Karbstein 2009; Pertschy et al. 2009; Strunk et al. 2011; Lebaron et al. 2012; Chaker-Margot et al. 2015). Although cleavage of the 20S pre-rRNA occurs in the cytoplasm, Nob1 associates with the 20S pre-rRNA in the nucleolus and therefore Nob1 nuclease activity must be prevented. Nop9 joins the assembling preribosome before Nob1. It interacts with a single-stranded RNA sequence within the 18S rRNA (nucleotides 1140–1150 in S. cerevisiae) and to a structured RNA site within the ITS1 fragment between sites D and A2 (orange, focused at nucleotides 13–19/204–212 in S. cerevisiae). Nob1 cleaves the 20S pre-rRNA at site D to produce the mature 18S rRNA. (B) Crystal packing arrangement of Nop9 and ITS1 RNA. Each asymmetric unit contains one Nop9/ITS1 complex (green Nop9 and orange ITS1) plus a second ITS1 RNA (blue) that bridges between two Nop9/ITS1 complexes. The GAAA tetraloop/tetraloop receptor interactions are indicated by red asterisks. (C) Superimposed experimental scattering curve for a Nop9:ITS1 RNA complex (black) and a calculated scattering curve for the crystal structure of Nop9/ITS1 RNA (red).

![Figure 1](image-url)
18S rRNA sequence (Thomson et al. 2007; Zhang et al. 2016a). The Nop9 binding site within ITS1 is near the Nob1 cleavage site that separates the 18S rRNA from the ITS1 region; furthermore, we have shown that Nop9 impedes Nob1 cleavage to prevent premature cleavage of the 20S pre-rRNA by Nob1 in the nucleolus (Zhang et al. 2016a).

Unlike other PUF protein families, we found that Nop9 has the ability to bind to both single-stranded RNA sequences and structured elements. We have previously determined a crystal structure of S. cerevisiae Nop9 in the absence of RNA that revealed a C-shaped protein comprising 11 PUM repeats (Zhang et al. 2016a). A small-angle X-ray scattering (SAXS) model of the Nop9–ITS1 stem–loop RNA indicated that Nop9 uses a new mode of RNA recognition for PUM repeat proteins. However, this low-resolution model could not provide the details of protein–RNA recognition. Here we present a crystal structure of Nop9 in complex with ITS1 RNA. The structure of the binary complex showed the details of how Nop9 recognizes the base of the ITS1 stem–loop using its amino-terminal PUM repeats, consistent with our previous SAXS model. In addition, the structure revealed that Nop9 uses its carboxy-terminal PUM repeats to recognize single-stranded RNA bases. This dual RNA recognition mode illustrates the evolution of PUM repeat proteins to combine recognition of sequence and structure. Moreover, we find that Nop9 adjusts the angle between the amino- and carboxy-terminal repeat modules to bind to RNAs using a two-handed mechanism like that of classical PUF proteins.

**RESULTS**

**Crystal structure of a Nop9–ITS1 protein–RNA complex**

We determined a crystal structure of *S. cerevisiae* Nop9 in complex with ITS1 RNA at 3.0 Å resolution (Table 1). Obtaining crystals of the Nop9/ITS1 complex was challenging, as demonstrated by the more than 10,000 crystallization trials with protein and RNA variants that did not yield diffracting crystals. To favor crystallization, we engineered the ITS1 RNA with guidance from existing data. Our previous small-angle X-ray scattering (SAXS) model of the Nop9/ITS1 complex showed that the distal end of the ITS1 stem–loop does not interact with Nop9 (Zhang et al. 2016a). Moreover, we had shown that truncation or mutation of the distal stem sequence did not affect binding affinity. Therefore, we substituted the distal end of the ITS1 stem–loop with a GAAA tetraloop and a tetraloop receptor (Coonrod et al. 2012). This method had been established to promote crystal packing of RNA molecules via the interaction of the tetraloop on one RNA molecule with the tetraloop receptor on another molecule. In the case of the Nop9–ITS1 stem–loop RNA complex, the engineered interaction between ITS1 RNA molecules would favor crystallization but should not interfere with protein–RNA interactions. We also engineered the protein, deleting residues 164–175, which are in a loop that was disordered in our previous structure of the Nop9 protein alone (Zhang et al. 2016a), and substituting a short linker. As expected, neither the RNA nor the protein engineering disrupted the ability of Nop9 to bind to ITS1 RNA (Table 2). On the contrary, the deletion of residues 164–175 increased the affinity of Nop9 for the RNA. These strategies allowed us to obtain diffracting crystals of a Nop9/ITS1 complex and to determine the crystal structure by molecular replacement.

The crystals of the Nop9/ITS1 complex contained two molecules of ITS1 RNA and one molecule of Nop9 protein per asymmetric unit. One molecule of ITS1 RNA was bound to Nop9 (orange in Fig. 1B), while a second bridging ITS1 molecule (blue in Fig. 1B) formed GAAA tetraloop-receptor interactions with the ITS1 RNA in two Nop9/ITS1 complexes (red asterisks in Fig. 1B). The engineered GAAA tetraloop-receptor interactions mediated lattice formation by creating a P2₁ screw axis. We confirmed that these lattice interactions did not appear to distort the Nop9/ITS1 complex relative to the structure in solution by comparing a calculated scattering curve for

| Table 1. X-ray data collection and refinement statistics |
|---------------------------------------------------------|
| **Data collection**                                     |
| Space group                                             | P2₁,2,2₁                                    |
| Cell dimensions                                         |
| a, b, c (Å)                                            | 71.9, 119.5, 142.1                         |
| α, β, γ (°)                                            | 90, 90, 90                                  |
| Resolution (Å)                                         | 50–3.02 (3.13–3.02)                        |
| R_merger                                               | 0.047 (0.75)                                |
| l/di                                                  | 22.43 (3.27)                                |
| Completeness (%)                                       | 99.1 (94.3)                                 |
| Redundancy (%)                                         | 14.4 (15.0)                                 |
| Wilson B-factor                                        | 102.8                                       |
| **Refinement**                                         |
| Resolution (Å)                                         | 35–3.02                                     |
| No. reflections                                       | 24,488                                     |
| R_work/R_free                                          | 19.9%/23.2%                                 |
| No. atoms                                              | 5913                                       |
| Macromolecules                                         | 5889                                       |
| Solvent                                                | 24                                         |
| B-factors                                              |
| Macromolecules                                         | 105.2                                      |
| Solvent                                                | 108.0                                      |
| R.m.s. deviations                                      |
| Bond lengths (Å)                                       | 0.003                                      |
| Bond angles (°)                                        | 0.583                                      |

*Values in parentheses are for the highest resolution shell, 3.13–3.02 Å.*
Nop9 interacts with the base of the ITS1 RNA duplex

The Nop9/ITS1 crystal structure revealed interactions between Nop9 and ITS1 RNA in two regions. Nop9 binds to the proximal base pairs of the ITS1 RNA stem-loop near amino-terminal PUM repeats R1 and R2 and also binds to single-stranded RNA bases using repeats R8–R11 (Fig. 2). The overall conformation of Nop9 changes with RNA binding. We compared the structure of Nop9 in complex with ITS1 with our previous structure of apo Nop9 (PDB ID: 5SVD) and found that ITS1 binding resulted in a more closed Nop9 conformation (Fig. 3A). This closure reduced the distance between the amino- and carboxy-terminal repeats from 30 to 20 Å (distance between the Cα atoms of Ala75 and Ile600). Conformational changes resulted from differences at the repeat R4–R5 and R7–R8 junctions, where we had previously noted the protein structure twisted (Zhang et al. 2016a). Closure upon RNA binding is also observed when Nop9 is bound to a single-stranded RNA sequence from the 18S rRNA (Wang and Ye 2017). Surprisingly, binding to either ITS1 or the 18S rRNA sequence induced a similar closed Nop9 conformation (Fig. 3B, RMSD 0.76 Å over 500 Cα atoms). The plant ortholog of Nop9, Arabidopsis thaliana APUM23, also binds to the single-stranded 18S RNA sequence (Bao et al. 2017). APUM23 has an overall structure that is quite similar to Nop9, but it is in a more open conformation when bound to single-stranded RNA (Fig. 3C). APUM23 uses an insertion in its third PUM repeat to interact with and position the RNA. This insertion could also help to maintain the open conformation, as it interacts with APUM23 repeats R3–R8 (Fig. 3C). Although Repeat R3 of Nop9 also contains an insertion, it was disordered or absent in all crystal structures. Nevertheless, the lack of sequence conservation between this R3 insertion in Nop9 with the R3 insertion in APUM23 suggests that this function is not retained in Nop9 (Bao et al. 2017).

Nop9 recognizes the nucleotides and structure at the base of the ITS1 duplex region (Fig. 2). Residues near the amino-terminal ends of the second α-helices of Repeats R1 and R2 interact with base pairs in the RNA minor groove (Fig. 4A). K140 is within hydrogen bonding distance of the base and the 2′-OH group of G19, and Q104 is positioned similarly to interact with the base and 2′-OH group of G19. PUM repeats in other PUF proteins contain similar conserved residues near the amino-terminal ends of the second RNA recognition helices that stack with one another and form a ridge. In human Pumilio1 and other classical PUF proteins, this ridge positions the ribose groups of single-stranded RNA rather than recognizing bases as it does in Nop9 (Fig. 4B). In our crystal structure of the atypical PUF protein, Puf-A, we observed residues at the equivalent positions interacting with the major groove of a double-stranded DNA (Fig. 4C; Qiu et al. 2014). The conservation and positioning of these residues between Nop9 and

TABLE 2. Nop9 binding affinities for ITS1 RNA

| Protein variant | RNA variant | $K_d$ (nM) | $K_{rel}$ |
|-----------------|-------------|------------|----------|
| Nop9 46–645     | ITS1 7–38, 184–206 | 11.8 ± 0.4 | 1        |
| Nop9 46–645     | ITS1 7–25 (tetraloop/receptor) 199–206 | 15.3 ± 0.5 | 1.3      |
| Nop9 46–645     | ITS1 7–38, 184–206 | 1.6 ± 0.1  | 0.14     |
| Nop9 46–645     | Synthetic ITS1 7–32, 192–206 | 14.0 ± 1.9 | 1.2      |

*In vitro transcribed RNAs contain a 5′GG preceding the U7 nucleotide, which is absent from the synthetic RNA.
Puf-A suggests that Puf-A might also recognize the nucleotides of structured RNA. Nop9 possesses a well-conserved amino-terminal peptide that interacts with the first α-helices of repeats R1 and R2, and within this peptide, the exposed aromatic ring of F52 stacks with the unpaired A20 base (Fig. 4D). This interaction appears to be non-sequence specific, which is consistent with our previous results demonstrating that substituting the AU-rich sequence in this region with a GC-rich sequence had little effect on in vitro binding affinity (Zhang et al. 2016a). Recognition of the A20 structural feature and the base-paired sequence further supports our finding that the base of the ITS1 duplex region is vital for Nop9 binding.

Nop9 recognizes the 5′ single-stranded region of ITS1 RNA

In addition to interaction by repeats R1 and R2 with the base of the ITS1 duplex, we observed electron density near the RNA-binding TRMs of repeats R8–R10 indicating that Nop9 bound to single-stranded RNA (Fig. 5A). The only single-stranded region within the crystallized RNA was the 5′ region of the ITS1 RNA, nucleotides 9–16 (Fig. 2B). The shape and size of the electron density suggested that the two 5′ nt were consecutive purines and partial electron density was present for a third nucleotide. If we placed a model of three nucleotides into this density, the shortest distance between the 3′ end of the single-stranded RNA and the 5′ end of the duplex region (U17) was 28 Å. Phosphate-to-phosphate distances within the RNA backbone of A-form RNA are ~6.0 Å. Therefore, at least 4–5 nt are needed to span the distance between the two ends. The simplest interpretation was to assign the two 5′ nucleotides as A10–A11, and we did not fit a third nucleotide into the partial density, as there were no instances of three consecutive purines in the 5′ region of the crystallized ITS1 (Fig. 2B). We considered several other possibilities. The single-stranded RNA could belong to an ITS1 molecule other than the one whose duplex region is bound to repeats R1 and R2. However, the nearest ITS1 in the lattice is 56 Å away, making this option unlikely. Alternatively, G–G–U9, including the two guanines that were added upstream of U9 to induce in vitro transcription (Fig. 2B), could fit the density. Nevertheless, upstream guanines do not appear to contribute to single-stranded RNA recognition, as a synthetic RNA beginning with the natural U7–U8 without a 5′ G–G bound to Nop9 with similar affinity as in vitro transcribed ITS1 RNA bearing the 5′ G–G (14.0 nM vs. 11.8 nM, Table 2). Finally, repeats R8–R11 could bind to an RNA sequence with three nonconsecutive purines, R–R–X–R, with a nucleotide X flipped away from the RNA-binding helices of Nop9, as has been observed for other PUF proteins (Miller et al. 2008; Wang et al. 2009; Wilinski et al. 2015). Regardless of the exact assignment, repeats R8 to R11 appear to engage several 5′ nt, which is
consistent with our previous binding assays demonstrating the importance of this region of the ITS1 RNA (Zhang et al. 2016a). Nop9 repeats R8–R10 also recognize the purine-rich 5′GGA sequence of the 18S rRNA element (Wang and Ye 2017), suggesting that these PUM repeats selectively bind single-stranded purine-rich sequences in both target RNAs.

The 5′ nucleotides in ITS1 are bound by repeats R8–R11 (Fig. 5A), similar to the "anti-parallel" recognition mode of classical PUF proteins binding to their single-stranded RNA targets. Repeats R11, R10, and R9 use the same side chains to interact with the A10 and A11 purines in ITS1 as had been shown to interact with G1140 and G1141 in the 18S rRNA (Fig. 5B; Wang and Ye 2017). The TRMs in Nop9 repeats R9 and R10 along with Gln604 and Asp608 in R11 appear capable of recognizing either purine base. Nop9 repeat R8 is divergent in amino acid sequence with a weakly conserved TRM (Zhang et al. 2016a) and a long loop preceding the RNA interaction helix. Arg485, which is in repeat R8 but not its TRM, is positioned to form a stacking interaction with the third nucleotide in the 5′ ITS1 RNA (Fig. 5A), similar to its interaction with A1142 and A1143 of the 18S rRNA (Fig. 5B). Glu492 in the repeat R8 TRM binds to 18S A1142 (Fig. 5B), and it is in position to contact the third nucleotide of the 5′ ITS1 RNA (Fig. 5A). Repeat R7 completely lacks a conserved TRM (Zhang et al. 2016a), and its inability to recognize RNA results in a break in the 1 repeat:1 base recognition pattern for both the ITS1 and 18S RNAs (Fig. 3B). We conclude that Nop9 repeats R8–R11 form a module for recognition of 5′ single-stranded purine sequences in target RNA motifs.

The 3′ end of ITS1 contributes sequence-independent Nop9 binding affinity

Due to optimization of the RNA for crystallization, the RNA in our Nop9/ITS1 complex lacked the 3′ single-stranded region of the ITS1 fragment (nucleotides 207–212). However, we had shown previously that this 3′ region does contribute to Nop9/ITS1 binding affinity (Zhang et al. 2016a). The presence of the 3′ single-stranded RNA increased affinity fivefold, from 11.8 ± 0.4 nM to 2.4 ± 0.15 nM (Tables 2, 3). In the absence of structural information, we investigated whether the sequence of the 3′ end of ITS1 influenced interaction with Nop9 by carrying out binding assays with RNA variants. We mutated each of the nucleotides from 207 to 212 and found that substitutions in this region did not dramatically perturb Nop9 binding (Table 3). Therefore, we conclude that the 3′ single-stranded region contributes to Nop9 binding affinity in a sequence-independent manner.

DISCUSSION

Our crystal structure of a Nop9/ITS1 complex provides a detailed understanding of Nop9’s ability to combine modules for sequence and structural recognition to bind to the
Nop9 recognizes pre-rRNA structure and sequence

ITS1 stem–loop RNA (Fig. 6A). This mechanism to combine amino- and carboxy-terminal modules is reminiscent of the two-handed mechanism for single-stranded RNA sequence recognition by classical PUF proteins (Wilinski et al. 2015). The Nop9 bifunctional amino-terminal PUM repeats form one module/hand that can recognize either a single-stranded RNA sequence (18S) or a structured duplex (ITS1). Repeats R1–R5 recognize the sequence of single-stranded 18S rRNA primarily using conserved TRM side chains in the second α helices of each repeat. In addition, repeats R1–R3 can recognize duplex RNA using conserved residues in an amino-terminal peptide extension and at the amino-termini of the second α helices. This amino-terminal repeat module has evolved two distinct surfaces for RNA recognition.

The Nop9 carboxy-terminal repeats form a second module/hand that recognizes purine-rich single-stranded RNA using classical PUF protein interactions. Nop9 adapts when binding to either single-stranded or stem–loop RNA by changing its overall curvature. A hinge point is located between PUM repeats R7 and R8 (Fig. 6A). This flexibility enables repositioning of the carboxy-terminal single-stranded RNA binding module relative to the amino-terminal bifunctional module, and we speculate that different relative positions of the amino- and carboxy-terminal modules allow Nop9 to bind to a range of RNA sequences/structures. For instance, Nop9 could bind to RNAs with different length linkers between a purine-rich element bound by the carboxy-terminal module and a single-stranded or stem–loop motif bound by the amino-terminal module (Fig. 6B).

Although Nop9 has not been identified in any structures of preribosomes likely due to its transient interactions (Barandun et al. 2017; Heuer et al. 2017; Sun et al. 2017; Ameismeier et al. 2018; Scaiola et al. 2018; Cheng et al. 2019; Hunziker et al. 2019), we illustrate the location of Nop9 by placing our crystal structure following the 5′ end of ITS1 in a recent cryo-electron microscopy (cryo-EM) structure of a human late pre-40S ribosome (Fig. 7; Ameismeier et al. 2018). Nop9 binds to ITS1 near the Nob1 cleavage site and inhibits premature processing in the nucleolus (Thomson et al. 2007; Zhang et al. 2016a). Nop9 binds to elements in ITS1 downstream from the binding site for Pno1, which blocks Nob1 cleavage in the cytoplasm (Turowski et al. 2014; Heuer et al. 2017; Johnson et al. 2017; Parker et al. 2019). It appears that both Nop9 and Pno1 could bind simultaneously (Fig. 7), allowing the handoff of Nob1 inhibition from Nop9 in the nucleolus to Pno1 in the cytoplasm. The other known binding site for Nop9 is at a critically important junction in pre-rRNA folding. The single-stranded binding site within the 18S rRNA (Fig. 7) includes nucleotides at the interface of two structural regions of the mature rRNA: the central domain and the 3′ major domain (Wang and Ye 2017). The 5′ nucleotides form part of the central pseudoknot and the 3′ nucleotides form one strand of helix H28. We propose that Nop9 is important for the timing of formation of these RNA structures, as its dissociation would be required to free the RNA for base-pairing. This is consistent with recent studies that

![Diagram](https://example.com/diagram)

**TABLE 3.** Binding affinity enhancement by the 3′ sequence of ITS1 is sequence independent

| Protein | RNA variant | 3′ sequence | Kd (nM) | Kass |
|---------|-------------|-------------|---------|------|
| Nop9 46–645 | ITS1 7–38, 184–212 | UACAAC | 2.4 ± 0.15 | 1.1 |
| Nop9 46–645 | U207A | AAACAAC | 2.6 ± 0.1 | 1.1 |
| Nop9 46–645 | A208U | UAACAAC | 1.8 ± 0.1 | 0.75 |
| Nop9 46–645 | C209G | UAGAAC | 4.4 ± 0.8 | 1.8 |
| Nop9 46–645 | C209A | UAACAC | 2.6 ± 0.2 | 1.1 |
| Nop9 46–645 | C209U | UUAACAC | 2.5 ± 0.3 | 1.04 |
| Nop9 46–645 | A210U, A211U | UACUUUC | 2.1 ± 0.3 | 0.9 |
| Nop9 46–645 | C212G | UACAAG | 2.9 ± 0.6 | 1.2 |

Nucleotides in the 3′ sequence that were mutated are shown in boldface.
indicate that parts of the 3′ domain assemble into the 5′-ETS structure before the central and 5′ domains are incorporated (Cheng et al. 2019). Nop9 with binding sites in both regions is positioned to play a role in ordering these processes. We can also envision that as the pre-20S rRNA is transcribed and undergoes structural changes upon ribosome maturation, it provides other potential transient binding sites for Nop9, combining 5′ purine-rich elements with downstream single-stranded or structural elements, that remain to be uncovered.

MATERIALS AND METHODS

Nop9 expression and purification

We expressed Nop9 (residues 46–645) using the pSMT3 SUMO fusion protein expression vector (kindly provided by Dr. Christopher Lima, Memorial Sloan Kettering Cancer Center). We used site-directed mutagenesis PCR with corresponding primers to create a construct, Nop9Δ164–175, encoding Nop9 with a Ser-Gly-Ser linker replacing residues 164–175. Nop9 (residues 46–645) and Nop9Δ164–175 were expressed and purified as described previously (Zhang et al. 2016a). Briefly, amino-terminal His6-SUMO-tagged Nop9 proteins were expressed at 22°C overnight in E. coli BL21-CodonPlus (DE3) in the presence of 0.4 mM IPTG, which was added when the OD600 reached 0.6. Cell pellets were resuspended in sonication buffer (25 mM HEPES pH 7.5, 1 M NaCl, 1 mM TCEP, and 25 mM imidazole) plus 1 mg/mL lysozyme and lysed by sonication, followed by centrifugation to remove cell debris. The supernatant was applied to 5 mL of HisPur Ni-NTA resin (Thermo Scientific), washed with 200 mL sonication buffer, and eluted with 25 mM HEPES pH 7.5, 500 mM NaCl, 1 mM TCEP, and 500 mM imidazole. The SUMO-tagged proteins were cleaved overnight with 2 µg/mL of Ulp1 at 4°C. The cleaved protein sample was diluted fivefold using 20 mM HEPES pH 7.5, 20 mM NaCl, and 1 mM TCEP and loaded onto a 5 mL HiTrap Heparin column (GE Healthcare). The sample was eluted with a linear gradient from 0 to 2 M NaCl in 20 mM HEPES pH 7.5 and 1 mM TCEP. Nop9 eluted from the Heparin column at ~600 mM NaCl and was further purified using a HiLoad 16/60 Superdex 200 column (GE Healthcare), equilibrated with 25 mM HEPES pH 7.5, 500 mM NaCl, and 1 mM TCEP. The identities of the proteins were confirmed by mass spectrometry, and the purities were >95% based on SDS-PAGE.

RNA preparation

The template DNA for in vitro transcription of an ITS1 RNA with a tetraloop and tetraloop receptor to facilitate crystallization was purchased from Eurofins MWG Operon with the coding strand sequence shown in Figure 2B. The T7 promoter sequence (GAAATTAATACGACTCACTATA) was used to initiate in vitro transcription using T7 RNA polymerase, and two 5′ guanine nucleotides were included to promote transcription. To increase the homogeneity of the 3′ end of the product RNA, we added two 2′-O-methylguanines at the 5′ end of the template strand (Kao et al. 1999). In vitro transcription reactions were incubated at 37°C overnight in 100 mM Tris-HCl pH 8.0, 20 mM MgCl2, 1 mM TCEP, 2 mM spermidine, 3% (w/v) PEG 8000, 0.01% (v/v) Triton X-100, 4 mM NTPs, 2 units of inorganic pyrophosphatase,
0.6 μM dsDNA template, and 0.06 mg/mL T7 RNA polymerase. In vitro transcription products were purified by 15% polyacrylamide gel electrophoresis (30 × 40 × 1.6 mm) in the presence of 8 M urea and 1× TBE. After a 1 h prerun to warm the gel to 50°C, RNA samples were resolved under constant power of 80 watts for 10 h, during which target RNAs migrated about three quarters of the gel length. RNA was eluted from the excised slices of gel containing target RNAs using a Whatman Elutrap system. Purified RNAs were diluted to 5 μM and incubated at 90°C for 2 min, followed by a snap-cooling step on ice to refold the RNA. The homogeneity of refolded RNA was confirmed by single bands on 10% polyacrylamide native TBE gels (Invitrogen). RNA secondary structures were predicted using Mfold (Zuker 2003). The graphic of the RNA secondary structure in Figure 2B was prepared using VARNa (Darty et al. 2009) and edited with Adobe Illustrator, as needed.

Protein crystallization and structure determination

We crystallized Nop9Δ164–175 (residues 46–163; 176–645) in complex with ITS1 RNA (9–25; 199–206) with a tetraloop and tetra-loop receptor (Fig. 2B). Initial crystals were obtained by hanging drop vapor diffusion at 22°C with a crystallization solution containing 18% (w/v) PEG 3350, 0.2 M ammonium citrate and 0.1 M MES pH 6.75. Crystals were improved by iterative microseeding. The crystallization solution was supplemented with 10% (v/v) ethylene glycol as a cryo-protectant.

X-ray diffraction data were collected at beamline 22-ID of the Advanced Photon Source (APS) at 100 K with a wavelength of 1.000 Å. The data were processed using HKL2000 (Otwinowski and Minor 1997). Phases were determined by molecular replacement using the coordinates of apo Nop9 (PDB ID: 5SVD). Due to the conformational change of Nop9 induced by RNA binding, two fragments of Nop9 separated at residue 331 between repeats R4 and R5 were used as search models for molecular replacement. The RNA chains were manually built in Coot (Emsley et al. 2010). Iterative refinement and model building using Phenix.Refine (Adams et al. 2010) and Coot yielded final R and free R factors of 19.9% and 23.2%, respectively. Over 95% of dihedral angles are in favored regions of the Ramachandran plot, and only 0.74% are Ramachandran plot outliers.

Electrophoretic mobility shift assays

RNAs were labeled at the 5′ end with 32P-γ-ATP (PerkinElmer Life Science) by T4 polynucleotide kinase for 1 h at 37°C. Unincorporated 32P-γ-ATP was removed using Illustra MicroSpin G-25 columns. Radiolabeled RNAs (<50 pM) were incubated with protein samples at 4°C for 40 min in 20 mM HEPES pH 7.5, 150 mM NaCl, 0.02% (v/v) Tween-20, and 1 mM TCEP. The samples were resolved on 10% polyacrylamide native TBE gels at constant voltage (150 V) with 1× TBE buffer at 4°C for 30 min. The gels were dried and exposed overnight to storage phosphor screens. The phosphor screens were scanned on a Molecular Dynamics Typhoon PhosphorImager, and band intensities were quantified with ImageQuant 5.2. The data were fit using the Hill equation with GraphPad Prism 6. EMSAs were performed with three technical replicates, and mean Kd’s and standard error of the mean are reported.

DATA DEPOSITION

Atomic coordinates and structure factors for the reported crystal structure have been deposited with the Protein Data Bank under accession number 6WPI.

ACKNOWLEDGMENTS

We thank L. Pedersen and the staff of the Southeast Regional Collaborative Access Team beamlines for assistance with X-ray data collection. We thank J. Williams of the NIEHS Mass Spectrometry Research and Support Group for mass spectrometric analysis, and the NIEHS NMR Group for advice and feedback regarding experimental design. We are grateful to our colleagues, H. Wang and R. Stanley, for critical comments on the manuscript. This work was supported in part by the Intramural Research Program of the National Institutes of Health, National Institute of Environmental Health Sciences (1ZIA50165 to T.M.T. H.) and the National Institute of General Medical Sciences (R35GM131687 to S.J.B.). The Advanced Photon Source used for this study is supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract no. W-31-109-Eng-38.

Author contributions: J.Z., C.Q., T.T., and T.M.T.H. designed research; J.Z., T.T., R.W., C.Q., and L.E.G performed research; J.Z., T.T., R.W., C.Q., L.E.G, S.J.B., and T.M.T.H. analyzed data; and all authors contributed to the writing and editing of the manuscript.

Received March 18, 2020; accepted April 29, 2020.

REFERENCES

Adams PD, Afonine PV, Bunkoczi 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. doi:10.1107/S0907444909052925

Akatsuka Y, Warren EH, Gooley TA, Bricker AG, Lin MT, Hansen JA, Martin PJ, Madtes DK, Engelhard VH, Takahashi T, et al. 2003. Dispany for a newly identified minor histocompatibility antigen, HA-8, correlates with acute graft-versus-host disease after haematopoietic stem cell transplantation from an HLA-identical sibling. Br J Haematol 123:671–675. doi:10.1046/j.1365-2458.2003.04676.x

Ameismeier M, Cheng J, Berninghausen O, Beckmann R. 2018. Visualizing late states of human 40S ribosomal subunit maturation. Nature 558:249–253. doi:10.1038/s41586-018-0193-0

Bao H, Wang N, Wang C, Jiang Y, Liu J, Xu L, Wu J, Shi Y. 2017. Structural basis for the specific recognition of 18S rRNA by APUM23. Nucleic Acids Res 45:12005–12014. doi:10.1093/nar/gkw872

Barandun J, Chaker-Margot M, Hunziker M, Molloy KR, Chait BT, Klingle S. 2017. The complete structure of the small-subunit processome. Nat Struct Mol Biol 24:944–953. doi:10.1038/nsmb.3472

Barker DD, Wang C, Moore J, Dickinson LK, Lehrmann R. 1992. Pumilio is essential for function but not for distribution of the Drosophila abdominal determinant Nanos. Genes Dev 6:2312–2326. doi:10.1101/gad.6.12a.2312

Bhat VD, McCann KL, Wang Y, Fonseca DR, Shukla T, Alexander JC, Qiu C, Wickens M, Lo TW, Hall TMT, et al. 2019. Engineering a
conserved RNA regulatory protein repurposes its biological function in vivo. eLife 8: e43788. doi:10.7554/eLife.43788

Brickner AG, Warren EH, Caldwell JA, Akatsuka Y, Golovina TN, Zalinger AL, Shabanowitz J, Eisenlohr LC, Hunt DF, Engelhard VH, et al. 2001. The immunogenicity of a new human minor histocompatibility antigen results from differential antigen processing. J Exp Med 193: 195–206. doi:10.1084/jem.193.2.195

Campbell ZT, Valley CT, Wickens M. 2014. A protein-RNA specificity code enables targeted activation of an endogenous human transcript. Nat Struct Mol Biol 21: 732–738. doi:10.1038/nsmb.2847

Chaker-Margot M, Hunziker M, Barandon 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. doi:10.1038/nsmb.3111

Cheng J, Bassler J, Fischer P, Lau B, Kellner N, Kunze R, Griesel S, Kallas M, Berninghausen O, Strauss D, et al. 2019. Thermophile 90S pre-ribosomes structures reveal the reverse order of co-transcriptional 18S rRNA subdomain integration. Mol Cell 75: 1256–1269 e1257. doi:10.1016/j.molcel.2019.06.032

Coonrod LA, Lohman JR, Berglund JA. 2012. Utilizing the GAAA tetraloop/receptor to facilitate crystal packing and determination of the structure of a CUG RNA helix. Biochemistry 51: 8330–8337. doi:10.1021/bi300829w

Darty K, Denise A, Ponty Y. 2009. VARNA: interactive drawing and editing of the RNA secondary structure. Bioinformatics 25: 1974–1975. doi:10.1093/bioinformatics/btp250

de la Cruz J, Karlstad K, Woolford JL Jr. 2015. Functions of ribosomal proteins in assembly of eukaryotic ribosomes in vivo. Annu Rev Biochem 84: 93–129. doi:10.1146/annurev-biochem-060614-033917

Edwards TA, Pyle SE, Wharton RP, Aggarwal AK. 2001. Structure of Pom111p reveals similarity between RNA and peptide binding motifs. Cell 105: 281–289. doi:10.1016/S0092-8674(01)00318-X

Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66: 486–501. doi:10.1107/S0907444910007493

Fan CC, Lee LY, Yu MY, Tzen CY, Chou C, Chang MS. 2013. Upregulated hPuf-A promotes breast cancer tumorigenesis. Tumour Biol 34: 2557–2564. doi:10.1007/s13277-013-0801-4

Fatica A, Oeffinger M, Diakic M, Tollery-Dye D. 2003. Nob1p is required for cleavage of the 3′ end of 18S rRNA. Mol Cell Biol 23: 1798–1807. doi:10.1128/MCB.23.5.1798-1807.2003

Henras AK, Plisson-Chastang C, Olszewski J, Pumilio repeat protein family for pre-rRNA processing and mRNA localization. PLoS ONE 4: e753. doi:10.1371/journal.pone.0000749

Heuer A, Thomson E, Schmidt C, Berninghausen O, Becker T, Hurt E, Beckmann R. 2017. Cryo-EM structure of a late pre-40S ribosomal subunit from Saccharomyces cerevisiae. eLife 6: e30189. doi:10.7554/eLife.30189

Hunziker M, Barandon J, Buzovetsky O, Steckler C, Molina H, Klinge S. 2019. Conformational switches control early maturation of the eukaryotic small subunit ribosome. eLife 8: e45185. doi:10.7554/eLife.45185

Johnson MC, Ghalei H, Donath KA, Karbstien K, Stroup ME. 2017. Structural heterogeneity in pre-40S ribosomes. Structure 25: 329–340. doi:10.1016/j.str.2016.12.011

Kao C, Zheng M, Rudisser S. 1999. A simple and efficient method to reduce nontranslated nucleotide addition at the 3′ terminus of RNAs transcribed by T7 RNA polymerase. RNA 5: 1268–1272. doi:10.1017/S1355838299001033

Kuo MW, Wang SH, Chang JC, Chang CH, Huang LJ, Lin HH, Yu AL, Li WH, Yu J. 2009. A novel puf-A gene predicted from evolutionary analysis is involved in the development of eyes and primordial germ-cells. PLoS ONE 4: e4980. doi:10.1371/journal.pone.0004980

Lafontaine DL. 2015. Noncoding RNAs in eukaryotic ribosome biogenesis and function. Nat Struct Mol Biol 22: 11–19. doi:10.1038/nsmb.2939

Lamanna AC, Karbstien K. 2009. Nob1 binds the single-stranded cleavage site D at the 3′-end of 18S rRNA with its PIN domain. Proc Nat Acad Sci 106: 14259–14264. doi:10.1073/pnas.0905403106

Lebaron S, Schneider C, van Nues RW, Swiatkowska A, Walsh D, Bottcher B, Granneman S, Watkins NJ, Tollervey D. 2012. Proofreading of pre-40S ribosome maturation by a translation initiation factor and 60S subunits. Nat Struct Mol Biol 19: 744–753. doi:10.1038/nsmb.2308

Li Z, Lee I, Moradi E, Hung NJ, Johnson AW, Maroette EM. 2009. Rational extension of the ribosome biogenesis pathway using network-guided genetics. PLoS Biol 7: e1000213. doi:10.1371/journal.pbio.1000213

Lu G, Hall TMT. 2011. Alternate modes of cognate RNA recognition by human PUMILIO proteins. Structure 19: 361–367. doi:10.1016/j.str.2010.12.019

Lunde BM, Moore C, Varani G. 2007. RNA-binding proteins: modular design for efficient function. Nat Rev Mol Cell Biol 8: 479–490. doi:10.1038/nrm2178

Macdonald PM. 1992. The Drosophila pumilio gene: an unusually long transcription unit and an unusual protein. Development 114: 221–232.

Miller MT, Higgin JJ, Hall TMT. 2008. Basis of altered RNA-binding specificity by PUF proteins revealed by crystal structures of yeast Puf4p. Nat Struct Mol Biol 15: 397–402. doi:10.1038/nsmb.1390

Nudel R, Simpson NH, Baird G, O’Hare A, Conti-Ramsden G, Bolton PF, Hennessy ER, Consortium SLI, Ring SM, Davey Smith G, et al. 2014. Genome-wide association analyses of child genotype effects and parent-of-origin effects in specific language impairment. Genes Brain Behav 13: 418–429. doi:10.1111/gbb.12127

Otwinowski Z, Minor W. 1997. Processing of x-ray diffraction data collected in oscillation mode. Methods Enzymol 276: 307–326. doi:10.1016/S0076-6879(97)76066-X

Parker MD, Collins JC, Korona B, Ghalei H, Karbstien K. 2019. A kinase-dependent checkpoint prevents escape of immature ribosomes into the translating pool. PLoS Biol 17: e3000329. doi:10.1371/journal.pbio.3000329

Pertsch B, Schneider C, Gnagid M, Schafer T, Tollervey D, Hurt E. 2009. RNA helicase Pha1p and its co-factor Pfa1 promote 20 to 18 S rRNA processing catalyzed by the endonuclease Nob1p. J Biol Chem 284: 35079–35091. doi:10.1074/jbc.M109.040774

Qiu C, Kershner A, Wang Y, Holley CP, Wilinski D, Keles S, Kimble J, Wickens M, Hall TMT. 2012. Divergence of Pumilio/fem-3 mRNA binding factor (PUF) protein specificity through variations in an RNA-binding pocket. J Biol Chem 287: 6949–6957. doi:10.1074/jbc.M111.326264

Qiu C, McCann KL, Wine RN, Baserga SJ, Hall TMT. 2014. A divergent Pumilio repeat protein family for pre-rRNA processing and mRNA localization. Proc Natl Acad Sci 111: 18554–18559. doi:10.1073/pnas.1407634112

Qiu C, Bhat GD, Rajeev S, Zhang C, Lasley AE, Wine RN, Campbell ZT, Hall TMT. 2019. A crystal structure of a collaborative RNA regulatory complex reveals mechanisms to refine target specificity. eLife 8: e48968. doi:10.7554/eLife.48968

Scialla A, Pena C, Weisser M, Bohringer D, Leibundgut M, Klingauf-Golle R, Nierhaus K, Gerhardy S, Panse VG, Ban N. 2018. Structure of a eukaryotic cytoplasmic pre-40S ribosomal subunit. EMBO J 37: e98499. doi:10.15252/embj.201798499
Strunk BS, Loucks CR, Su M, Vashisth H, Cheng S, Schilling J, Brooks CL III, Karbstein K, Skiniotis G. 2011. Ribosome assembly factors prevent premature translation initiation by 40S assembly intermediates. Science 333: 1449–1453. doi:10.1126/science.1208245

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. doi:10.7554/eLife.22086

Tafforeau L, Zorbas C, Langhendries JL, Mullineux ST, Stamatopoulou V, Mullier R, Wacheul L, Lafontaine DL. 2013. The complexity of human ribosome biogenesis revealed by systematic nucleolar screening of pre-rRNA processing factors. Mol Cell 51: 539–551. doi:10.1016/j.molcel.2013.08.011

Thomson E, Rappsilber J, Tollervey D. 2007. Nop9 is an RNA binding protein present in pre-40S ribosomes and required for 18S rRNA synthesis in yeast. RNA 13: 2165–2174. doi:10.1261/rna.747607

Turpeinen H, Ojala PJ, Ojala K, Miettinen M, Volin L, Partanen J. 2013. Minor histocompatibility antigens as determinants for graft-versus-host disease after allogeneic haematopoietic stem cell transplantation. Int J Immunogenet 40: 495–501. doi:10.1111/iji.12051

Valley CT, Porter DF, Qiu C, Campbell ZT, Hall TMT, Wickens M. 2012. Patterns and plasticity in RNA-protein interactions enable recruitment of multiple proteins through a single site. Proc Natl Acad Sci 109: 6054–6059. doi:10.1073/pnas.1200521109

Wang B, Ye K. 2017. Nop9 binds the central pseudoknot region of 18S rRNA. Nucleic Acids Res 45: 3559–3567. doi:10.1093/nar/gkw1323

Wang X, Zamore PD, Hall TMT. 2001. Crystal structure of a Pumilio homology domain. Mol Cell 7: 855–865. doi:10.1016/S1097-2765(01)00229-5

Wang X, McLachlan J, Zamore PD, Hall TMT. 2002. Modular recognition of RNA by a human Pumilio-homology domain. Cell 110: 501–512. doi:10.1016/S0092-8674(02)00873-5

Zhang J, McCann KL, Qiu C, Gonzalez LE, Baserga SJ, Hall TMT. 2016a. Nop9 is a PUF-like protein that prevents premature cleavage to correctly process pre-18S rRNA. Nat Commun 7: 13085. doi:10.1038/ncomms13085

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31: 3406–3415. doi:10.1093/nar/gkg595