

Structural basis for the specific recognition of 18S rRNA by APUM23

Hongyu Bao^{1,†}, Na Wang^{1,†}, Chongyuan Wang¹, Yiyang Jiang¹, Jiuyang Liu¹, Li Xu¹, Jihui Wu^{1,*} and Yunyu Shi^{1,2,*}

¹Hefei National Laboratory for Physical Science at Microscale, School of Life Sciences, University of Science and Technology of China, Hefei, Anhui 230027, China and ²CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China

Received April 09, 2017; Revised September 14, 2017; Editorial Decision September 15, 2017; Accepted September 19, 2017

ABSTRACT

PUF (Pumilio/fem-3 mRNA binding factor) proteins, a conserved family of RNA-binding proteins, recognize specific single-strand RNA targets in a specific modular way. Although plants have a greater number of PUF protein members than do animal and fungal systems, they have been the subject of fewer structural and functional investigations. The aim of this study was to elucidate the involvement of APUM23, a nucleolar PUF protein in the plant *Arabidopsis*, in pre-rRNA processing. APUM23 is distinct from classical PUF family proteins, which are located in the cytoplasm and bind to 3'UTRs of mRNA to modulate mRNA expression and localization. We found that the complete RNA target sequence of APUM23 comprises 11 nt in 18S rRNA at positions 1141–1151. The complex structure shows that APUM23 has 10 PUF repeats; it assembles into a C-shape, with an insertion located within the inner concave surface. We found several different RNA recognition features. A notable structural feature of APUM23 is an insertion in the third PUF repeat that participates in nucleotide recognition and maintains the correct conformation of the target RNA. Our findings elucidate the mechanism for APUM23's-specific recognition of 18S rRNA.

INTRODUCTION

In eukaryotes, PUF (Pumilio/fem-3 mRNA binding factor) proteins are a conserved family of RNA-binding proteins with modular structured repeat domains. Their reported molecular functions are translational repression or activation, mRNA localization (1) and pre-ribosomal RNA (pre-rRNA) processing (2,3). Classical PUF proteins are located in the cytoplasm and regulate the stability and translation of

their target mRNAs by binding to specific sequences in the 3' untranslated regions (UTRs) (4). The PUF binding motif normally contains a conserved UGUR sequence (where R is a purine) at the 5' end (5). The typical RNA-binding domain in PUF proteins is known as a PUF repeat or Pumilio homology domain (PUM-HD) (6). Structural studies have shown that canonical PUF proteins typically have eight tandem PUF repeats with one pseudo-repeat at each terminus and that they fold into crescent shapes (6–15). In most cases, they recognize their 8-nt specific RNA targets in a modular and anti-parallel way, with each PUF repeat binding specifically to a single RNA base such that PUF repeats 1–8 sequentially bind to RNA bases 8–1 (6,10–12). However, there are exceptions. Some PUF proteins bind RNA targets longer than eight bases, with the extra RNA bases flipping away from the PUF–RNA interface (9,13,16,17); and some comprise more than eight repeats and are therefore able to bind RNA targets longer than eight bases, or even structured DNA or RNA (3,17).

PUF repeats normally contain approximately 40 residues and fold into an arrangement comprising three α -helices (6). Their RNA base recognition specificity is determined by three key residues within the five-residue motif in the second helix of each PUF repeat, known as the tripartite recognition motif (TRM) (18–22). The first and fifth residues bind the edge of the RNA base via hydrogen bonding or van der Waals contacts, and the second residue makes a stacking interaction with two adjacent bases (23). The third and fourth residues can be any hydrophobic residues. In Pumilio1 and FBF-2 (6,16), the motifs (notated as the residues in TRM positions 1, 5 and 2) bind as follows: S₁E₅-N₂ or S₁E₅-H₂ motif with guanine, N₁Q₅-H₂ or N₁Q₅-Y₂ motif with uracil and C₁Q₅-R₂ or S₁Q₅-R₂ motif with adenine. Cytosine-recognition PUF repeats are rarely seen in nature, although early studies have reported PUF repeats engineered for specific C-recognition with (G, A, S, T or C)₁R₅ motifs (18,19). Further RNA-recognition codes of PUF proteins have also been identified (21,22,24).

*To whom correspondence should be addressed. Tel: +86 551 6360 7464; Fax: +86 551 6360 1443; Email: yyshi@ustc.edu.cn

Correspondence may also be addressed to Jihui Wu. Tel: +86 551 6360 3745; Fax: +86 551 6360 1443; Email: wujihui@ustc.edu.cn

[†]These authors contributed equally to the paper as first authors.

The number of PUF proteins varies between species, with only a small number in vertebrates (4) and a greater number in plants (25). For example, the plant *Arabidopsis* alone encodes more than 25 PUF proteins, named *Arabidopsis Pumilos* (APUMs) (25–27). Despite this, PUF proteins in plants have been the subject of fewer structural and functional investigations than those in animal and fungal systems.

APUM23 is a recently characterized nucleolar PUF protein with critical roles in leaf development, plant organ polarity and pre-rRNA processing (2,28), but how it regulates these processes remains largely unknown. According to phylogenetic analysis, orthologs of APUM23 include yeast nucleolar protein 9 (Nop9) and human NOP9. It has been predicted that APUM23 contains ten PUF repeats rather than the typical eight (25). A recent study found that the target of APUM23 is a unique 10-mer sequence (5'-GAAUUGACGG-3') in the 18S rRNA (29), unlike the cognate 'UGUR' RNAs typical of cytoplasmic PUF proteins. However, how APUM23 specifically recognizes its RNA target remains unknown.

Our understanding of the molecular mechanisms that underlie the functions of APUM23 is limited by the lack of high-resolution structure of APUM23 PUF repeats. In this study, therefore, we focused on the structure of APUM23 and its RNA target. We solved several crystal structures for APUM23–RNA complexes (wild-type (WT) and mutant) and found that APUM23 has 10 PUF repeats and folds into a C-shaped structure. There is an insertion within its third PUF repeat, located at the inner concave surface. APUM23 also has an insertion within the eighth PUF repeat, located at the side face. We also found that the complete RNA target sequence of APUM23 comprises 11 nt (5'-GGAAUUGACGG-3') at positions 1141–1151 in 18S rRNA, rather than the 10 nt that had previously been suggested (29). The insertion of the third PUF repeat may function in both RNA base recognition and RNA conformation stabilization. Thus, this study elucidated the structural basis for the specific recognition of 18S rRNA by APUM23.

MATERIALS AND METHODS

Plasmid constructions

The cDNA of APUM23 was amplified from cDNA of *Arabidopsis thaliana* that was generously provided by Lab of Professor Yong Ding (University of Science and Technology of China, Hefei, China). The PUF repeats of APUM23 (residues 85–655) was cloned into a modified pET28a (Novagen) vector without a protease cleavage site after the 6 × His tag (p28a) vector. All mutants were generated using the MutanBEST kit (Takara) and verified by DNA sequencing.

Protein expression and purification

All proteins were expressed in *Escherichia coli* BL21-Gold (DE3) cells (Novagen). For native proteins, cells were cultured in LB medium at 310 K to an OD600 of 0.8–1.0, subsequently shifted to 289 K and induced with 0.4 mM isopropyl β-D-thiogalactopyranoside (IPTG) for 24 h. For Se-Met labeled proteins, cells were cultured in LeMaster and Richards minimal medium (LR medium) at 310 K to

an OD600 of 0.8–1.0, then transferred to LR medium that contained Val, Ile, Leu, Phe, Trp, Thr, Lys at the concentration of 50 mg/l and Se-Met at the concentration of 60 mg/l. After 30 min at 310 K, cells were shifted to 289 K and induced with 0.4 mM IPTG for 24 h. All proteins were purified by Ni-NTA affinity chromatography (GE Healthcare) after lysing cells by sonication in 20 mM Tris (pH 8.0) and 500 mM NaCl. After clarification at 13 000 rpm in an R22A2 Hi-tachi rotor (Hitachi, himac CR22G, Japan) for 30 min at 277 K, the supernatants were eluted in buffer containing 30–500 mM imidazole (pH 7.5) at 293 K. The eluted His-tag proteins were further purified using a Superdex 200 column (GE Healthcare) in buffer containing 20 mM Tris (pH 7.5) and 200 mM NaCl with no reducing agent at 293 K.

RNA preparation

RNA oligomers were purchased from Takara Bio, Inc., and dissolved in diethyl pyrocarbonate (DEPC)-treated water to a final concentration of 2 mM.

Crystallization and data collection

The native or Se-Met labeled APUM23^{85–655} with an N-terminal 6 × His tag was concentrated to ~20 mg/ml in buffer of 15 mM Tris (pH 7.5), 200 mM NaCl and 2 mM Dithiothreitol (DTT). The native or Se-Met labeled APUM23^{85–655}–RNA complex was prepared by mixing 10 mg/ml (final concentration) proteins with RNA in a mole ratio of 1:1.5. All crystals were formed at 293 K via sitting drop method with 1 μl each of protein–RNA complexes and reservoir solutions. The crystals of native APUM23^{85–655}–GAAUUGACGG complex were formed in mother liquor containing 0.15 M DL-Malic acid (pH 7.0) and 20% w/v Polyethylene glycol 3350 via sitting drop method. The crystals of Se-Met labeled APUM23^{85–655}–GAAUUGACGG complex were formed in mother liquor containing 0.1 M Calcium acetate hydrate, 0.1 M Sodium cacodylate (pH 5.5) and 12% w/v Polyethylene glycol 8000. The crystal of the APUM23^{85–655}–GGAAUUGACGG and APUM23^{85–655}–GGAGUUGACGG were also formed in mother liquor containing 0.15 M DL-Malic acid (pH 7.0) and 20% w/v Polyethylene glycol 3350. The crystal of the APUM23^{85–655}–GGAUUUGACGG was formed in mother liquor containing 0.2 M Sodium malonate (pH 7.0) and 20% w/v Polyethylene glycol 3350. The crystal of APUM23^{Delete-GGAUUGACGG} was formed in mother liquor containing 0.2 M Lithium sulfate monohydrate, 0.1 M Bis-Tris (pH 6.5) and 25% w/v Polyethylene glycol 3350. The cryoprotectant for crystals was 80% reservoir solution supplemented with 20% glycerol. The crystals were preserved in liquid nitrogen stream at 100 K. X-ray diffraction data of the crystals were collected at 100 K. X-ray diffraction data for the crystals were collected on beamline 17U1, 18U1 and 19U1 of the Shanghai Synchrotron Radiation Facility (SSRF). The data collection statistics are listed in Supplementary Tables S1 and 2.

Table 1. Thermodynamic parameters of the WT and mutants of the APUM23 to different RNAs by ITC experiments

Protein	RNA 5' to 3'	ΔH kcal/mol	ΔS cal/mol/K	K_D nM	K_D (mean) nM	$K_{rel}^{\#}$
WT	GGAAUUGACGG	36	-85	21	21 ± 2	1
		38	-91	20		
		36	-84	23		
WT	GAAUUGACGG	35	-82	27	32 ± 8	1.5
		33	-76	29		
		38	-95	41		
WT	GGAAUUGACGG	39	-96	12	13 ± 3	0.6
		39	-94	11		
		24	-47	35		
WT	GGAAAUGACGG	27	-58	44	43 ± 7	2.0
		25	-50	49		
		26	-61	1236		
F398A	GGAAUUGACGG	29	-71	2725	$(1.9 \pm 0.7) \times 10^3$	90
		28	-68	1783		
		28	-67	1422		
Q401E	GGAAUUGACGG	22	-48	1143	$(1.1 \pm 0.3) \times 10^3$	52
		21	-41	840		
		27	-62	205		
Delete	GGAAUUGACGG	29	-66	116	$(0.160 \pm 0.04) \times 10^3$	7.6
		28	-64	160		
		34	-86	943		
Delete	GAAUUGACGG	30	-73	1248	$(1.1 \pm 0.2) \times 10^3$	52
		32	-80	980		

$^{\#}K_{rel}$ reports the affinity of WT or mutant APUM23 for the RNA sequence, relative to the affinity of WT APUM23 for the 11-mer WT RNA 5'-GGAAUUGACGG-3'.

The ITC experiments were carried out at 293K and all proteins were in buffer conditions: 50 mM Tris (pH 7.5), 200 mM NaCl. We performed three trials for each ITC titration. WT, F398A, Q401E and Delete represent APUM23⁸⁵⁻⁶⁵⁵, APUM23⁸⁵⁻⁶⁵⁵ F398A mutant, APUM23⁸⁵⁻⁶⁵⁵ Q401E mutant and APUM23⁸⁵⁻⁶⁵⁵ without the insertion in PUF repeat R3, respectively. The average dissociation constant (K_D mean) values (\pm standard deviation) were derived from three independent experiments. The fitted curves are shown in Supplementary Figure S7.

Structure determination and refinement

Data were processed and scaled using HKL2000 and CCP4 program stackage (30). Phase was solved by the single-wavelength anomalous dispersion method using Shelx C/D/E (31) and phenix.autosol wizard (32) with the Se-Met labeled APUM23⁸⁵⁻⁶⁵⁵-GAAUUGACGG complex data. The initial model was built automatically by program Buccaneer (33). The structure of native APUM23⁸⁵⁻⁶⁵⁵-GAAUUGACGG complex was determined by program Phaser MR (34) using the model of Se-Met labeled APUM23⁸⁵⁻⁶⁵⁵-GAAUUGACGG complex as the search model. The other structures were using the refined native APUM23⁸⁵⁻⁶⁵⁵-GAAUUGACGG structure as the search model. The structures of native APUM23⁸⁵⁻⁶⁵⁵-RNA complexes were further modified and refined through a combination of the Phenix suite (35) and Coot (36). The refinement statistics are listed in Supplementary Tables S1 and 2. All structure figures were prepared with PyMOL (37).

Isothermal titration calorimetry (ITC)

Isothermal titration calorimetry (ITC) assays were carried out on a MicroCal iTC200 calorimeter (GE Healthcare) at 298 K. The buffer for proteins and RNA oligomers was 50 mM Tris (pH = 7.5), 200 mM NaCl. The titration protocol consisted of 20 injections of RNA (0.1–0.15 mM) into protein (0.01 mM). We performed three trials for each ITC titration. Curve fitting to a one-binding-site model was performed by ITC data analysis module of Origin7.5 (Origin-

Lab) provided by the manufacturer. The parameters and fitted curves are shown in Table 1 and Supplementary Figure S7.

RESULTS

Overall structure of the APUM23⁸⁵⁻⁶⁵⁵-GGAAUUGACGG complex

As reported previously (29), *Arabidopsis thaliana* APUM23 selectively interacts with a 10-mer RNA target (5'-GAAUUGACGG-3') in 18S rRNA. To gain a structural insight into how APUM23 recognizes 18S rRNA, we performed extensive crystallization trials to determine the structure of the APUM23-RNA complex. We finally determined a 2.55 Å-resolution crystal structure of the APUM23⁸⁵⁻⁶⁵⁵-GAAUUGACGG complex (Supplementary Figure S1 and Table S1) and found that the complete target of APUM23 was the 11-mer RNA 5'-GGAAUUGACGG-3' at positions 1141–1151 in 18S rRNA (referred to hereafter as '11-mer WT RNA'). This is conserved in 18S rRNA across species (Supplementary Figure S2). Thus, we determined the crystal structure of the APUM23⁸⁵⁻⁶⁵⁵-GGAAUUGACGG complex at a resolution of 2.5 Å (Figure 1A and Supplementary Table S1). APUM23 comprises ten PUF repeats and is folded into a twisted C shape, with one pseudo-repeat at each terminus. Eight of the 10 PUF repeats comprise approximately 40 amino acid residues; they share a similar structure, in the form of a typical arrangement of three α -helices. The struc-

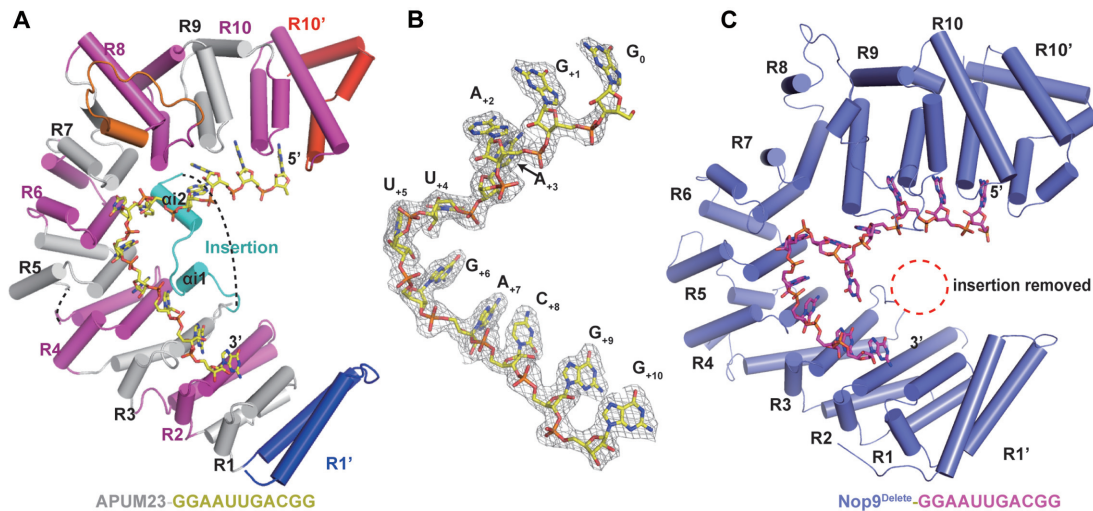


Figure 1. Overall structure of the APUM23⁸⁵⁻⁶⁵⁵-GGAAUUGACGG complex. (A) The overall structure of the complex is represented as cartoon cylindrical helices, with the pseudo-repeats at each terminus labeled R1' and R10', and the PUF repeats labeled R1 to R10. The insertion is colored in cyan and the two short α -helices are labeled $\alpha 1$ and $\alpha 2$. (B) The 2Fo-Fc electron density map around the 11-mer WT RNA 5'-GGAAUUGACGG-3', contoured at 1.0 sigma. Nucleotides G1140 to G1150 are labeled as G₀ to G₊₁₀. (C) The overall structure of the Nop9^{Delete}-GGAAUUGACGG complex is represented as cartoon cylindrical helices (PDB code: 5WTY). The insertion in the R3 PUF repeat has been removed.

tures of the other two PUF repeats (R3 and R8) differ from those of typical PUF repeats (Supplementary Figure S3). R8 contains 61 amino acid residues (residues 466–526), with an insertion between the first and second helix. This insertion folds into an α -helix at the side face of the C-shaped structure (Figure 1A). A remarkable feature of the structure of APUM23⁸⁵⁻⁶⁵⁵ is the unique characteristics of R3. This PUF repeat contains 74 amino acid residues (residues 213–286) and includes an insertion (residues 235–270; the electron density of residues 258–270 was missing) located between the second and third helix. The insertion comprises a disordered region and two short α -helices ($\alpha 1$ and $\alpha 2$), located at the inner concave surface, that participate in the recognition of RNA bases (Figure 1A and Supplementary Figure S3).

During the preparation of this manuscript, reports were published on the apo and RNA-bound structures of yeast Nop9 (17,38). The apo structure of yeast Nop9 was first reported in 2016 as having 11 PUF repeats with one N-terminal pseudo-repeat (38). However, the more recent study reported that yeast Nop9 contained 10 PUF repeats with two terminal pseudo-repeats in its RNA-bound structure (17). In the present study, we adopted this latter definition and compared APUM23 and yeast Nop9 in their recognition of the same RNA sequence (Figure 1C). The sequence position of RNA 5'-GGAAUUGACGG-3' differs between *Arabidopsis* (G1141–G1151) and *Saccharomyces cerevisiae* (G1140–G1150), so we labeled the 11 nt as G₀ to G₊₁₀ to facilitate the description of the differences in RNA recognition (Figure 1B). A search of the Dali server showed that the structure of APUM23 resembles that of yeast Nop9 (Z-score 25.4, sequence identity 18%). The C α atom root mean square deviation between the RNA-bound structure of APUM23 and yeast Nop9 is 6.673 Å (389–389 C α atoms, calculated by PyMOL). The most remarkable structural difference between yeast Nop9 and APUM23 is found in the

PUF repeat R3. In both APUM23 and yeast Nop9, repeat R3 contains an insertion between the second and the third helix; however, this insertion was invisible in the apo structure of yeast Nop9 (38) and was removed (Nop9^{Delete}) from the Nop9-RNA structure (17).

All the RNA bases of 11-mer WT RNA are positioned below APUM23's inner concave surface and the insertion in PUF repeat R3 (Supplementary Figure S4). C-shaped APUM23 interacts in a modular way with the RNA bases of G₀ to A₊₂ and U₊₅ to G₊₁₀ through PUF repeats R10 to R8 and R6 to R1, respectively, (Figure 2). R7 is not responsible for RNA interactions, and A₊₃ flips away from the inner concave surface; however, APUM23 stabilizes A₊₃ via a hydrophobic pocket that is formed mostly by the insertion (discussed in detail below), and interacts with U₊₄ via both the insertion (Arg-254) and R6 (Phe-398 and Gln-401), which also recognizes U₊₅. In addition, A₊₃ makes stacking interactions with A₊₂, and U₊₄ with U₊₅.

Yeast Nop9 and APUM23 recognize the same target RNA in different ways (Supplementary Table S3). In the Nop9^{Delete}-RNA structure, A₊₃ does not stack with A₊₂, and U₊₄ and U₊₅ flip away from the inner concave surface of Nop9, so their uracil rings do not interact with Nop9^{Delete}. APUM23 has only one PUF repeat (R7) not involved in specific RNA recognition of RNA; Nop9^{Delete} has two, R6 and R7. Furthermore, APUM23 and Nop9 use different motifs to recognize some of the same RNA bases (Supplementary Table S3).

Taken together, the ten PUF repeats and the insertion in PUF repeat R3 enable APUM23 to specifically recognize the 11-mer 18S rRNA sequence 5'-GGAAUUGACGG-3'.

Modular recognition of G₀, G₊₆, A₊₇, C₊₈, G₊₉ and G₊₁₀

G₀, G₊₆ and G₊₉ are recognized by the previously reported typical S₁E₅ motif in PUF repeats R10, R5 and R2, respectively (6). The side chain of S₁ (the residue in TRM posi-

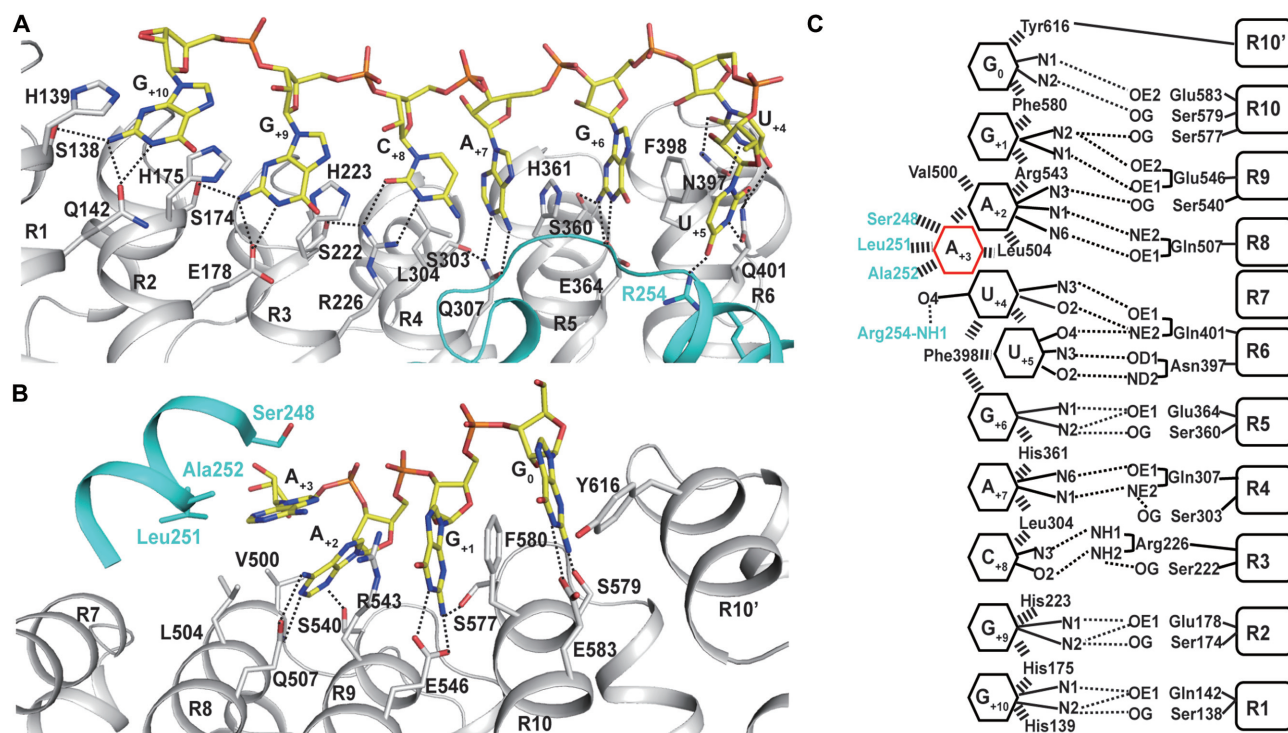


Figure 2. Recognition of target RNA by APUM23. (A and B) Interactions between APUM23 and the target RNA. The dotted lines represent hydrogen bonds. (C) Schematic representation of APUM23-RNA interactions. Hydrogen bonds and stacking interactions are shown as dotted lines and stacked bars, respectively.

tion 1) interacts with the N2 group of guanine via hydrogen bonding, and the side chain of E₅ interacts with the N1 group, or the N1 and N2 groups, of guanine (Supplementary Figure S5a, c and d). G₊₁₀ is recognized by an S₁Q₅ motif in PUF repeat R1, which normally specifies adenine (6). Nevertheless, the interaction network is identical to that for the recognition of G by the S₁E₅ motif (Supplementary Figure S5e). The S₁Q₅ motif for guanine recognition has previously been identified (8,39).

PUF repeat R4 has an S₁Q₅-L₂ motif, which recognizes A₊₇ through the side chain of Gln-307 (position 5); the side chain of Ser-303 stabilizes the conformation of Gln-307 via a hydrogen bond interaction (Supplementary Figure S5h).

Cytosine-recognition PUF repeats are rarely seen in nature. The PUF repeat R3 has an S₁R₅-H₂ motif in both APUM23 and Nop9. The NH1 and NH2 groups of R₅ form hydrogen bond interactions with the N3 and O2 groups of cytosine, and the side chain of S₁ stabilizes the conformation of R₅ via a hydrogen bond interaction (Figure 3).

Recognition of G₊₁, and A₊₂

In general, PUF proteins recognize target RNAs in a modular fashion, with each repeat binding to a single base, and a single RNA base forms hydrogen bond interactions with the TRM residues of only a single PUF repeat. However, in our complex, G₊₁ forms hydrogen bonds with residues in the two PUF repeats R9 and R10 (Supplementary Figure S5b), and A₊₂ forms hydrogen bonds with residues in both R8 and R9 (Supplementary Figure S5g).

PUF repeat R9 contains the A₁E₅-R₂ motif. However, we were unable to observe any interaction between G₊₁ and Ala-542 in R9; instead, the N2 group of G₊₁ forms a hydrogen bond with the OG group of Ser-577 in R10. At the same time, the N1 and N2 groups of G₊₁ form hydrogen bonds with the OE1 and OE2 groups of Glu-546, respectively. Thus, the G₊₁ recognition code is the S₁'E₅-R₂ motif, in which the S₁' structurally mimics the function of the residue at TRM position 1.

PUF repeat R8 contains a C₁Q₅-L₂ motif in α -helix 2. However, the N3 group of A₊₂ does not interact with Cys-503 in R8 but forms a hydrogen bond with the OG group of Ser-540 in R9. At the same time, the N1 and N6 groups of A₊₂ interact with the OE1 and NE2 group of Gln-507 in R8, respectively. Thus, the A₊₂ recognition code is S₁'Q₅-L₂. A₊₂ also makes stacking interactions with A₊₃ and Val-500 in R8.

The insertion in PUF repeat R3's involvement in the recognition of A₊₃

A₊₃ is not recognized by the APUM23 five-amino acids SGVVA motif in the second α -helix of PUF repeat R7, as would be expected. Instead, the A₊₃ adenine is stabilized by a hydrophobic pocket that comprises Ser-248, Leu-251 and Ala-252 in the α 2 helix of the insertion, Leu-504 in R8, and a stacking interaction with A₊₂ (Figure 4A). Yeast Nop9 makes a stacking interaction only with A₊₃ via the side chain of Arg-485 in PUF repeat R8 (Figure 4B). To investigate whether APUM23 shows a preference for A₊₃, we further solved the crystal structures of the APUM23⁸⁵⁻⁶⁵⁵–

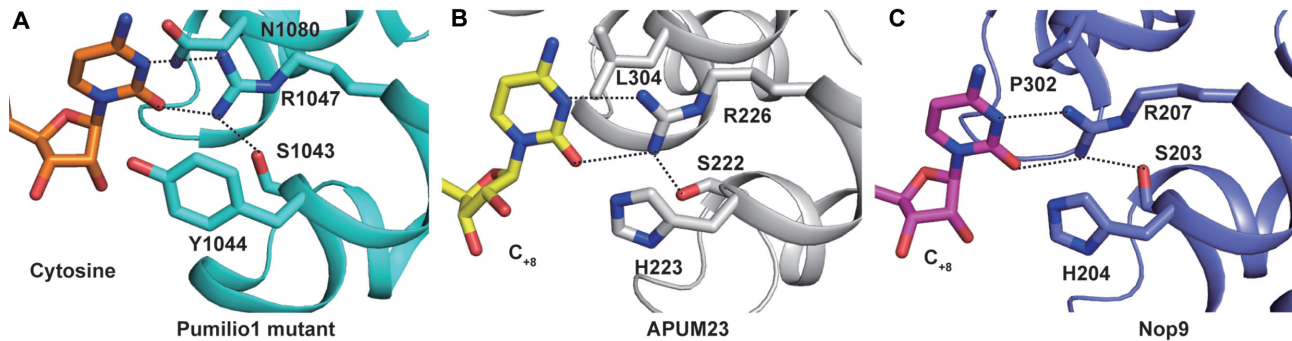


Figure 3. APUM23 and Nop9 each possess a natural cytosine-recognition PUF repeat. (A–C) Recognition of cytosine by R6 PUF repeat of a mutant of Pumilio1, R3 PUF repeat of APUM23 and R3 PUF repeat of Nop9, respectively. The dotted lines represent hydrogen bonds.

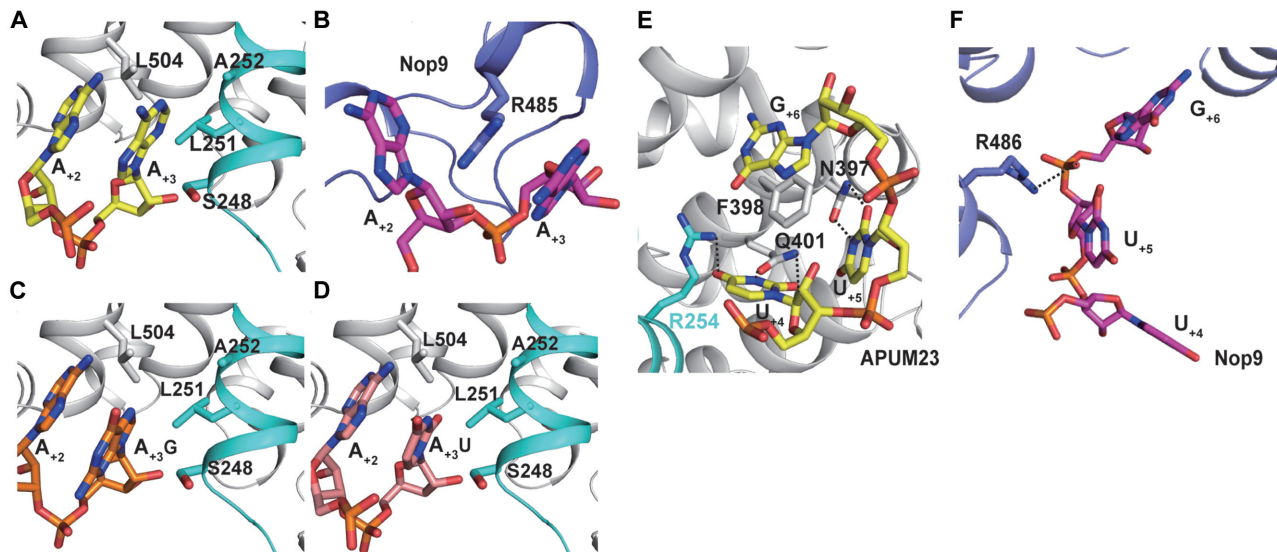


Figure 4. Recognition of A_{+3} , U_{+4} and U_{+5} by APUM23. (A and B) Recognition of A_{+3} by APUM23 and Nop9. The insertion in the R3 PUF repeat of APUM23 is colored in cyan. (C and D) Positions of $A_{+3}G$ and $A_{+3}U$ in the crystal structures of APUM23⁸⁵⁻⁶⁵⁵ in complex with 11-mer $A_{+3}G$ 5'-GGAGUUGACCG-3' and 11-mer $A_{+3}U$ 5'-GGAUUUGACCG-3', respectively. (E) APUM23 recognizes U_{+4} and U_{+5} through the R6 PUF repeat and the insertion in this repeat. Phe-398 has interactions with U_{+4} , U_{+5} and G_{+6} . (F) There are no interactions between Nop9 and U_{+4} and U_{+5} .

GGAUUGACGG (11-mer $A_{+3}U$) and APUM23⁸⁵⁻⁶⁵⁵–GGAGUGACGG (11-mer $G_{+3}U$) complexes at resolutions of 2.10 Å and 2.75 Å, respectively (Supplementary Table S2). The U_{+3} and G_{+3} positions showed no significant difference from that of A_{+3} (Figure 4C and D). ITC results showed that the binding affinity of APUM23⁸⁵⁻⁶⁵⁵ for the 11-mer $A_{+3}U$ RNA (5'-GGAUUGACCG-3') was ~1.6 times higher than that for 11-mer WT RNA (Table 1 and Supplementary Figure S7a and c). These results indicate that APUM23 has no selectivity for A_{+3} .

Recognition of U_{+4} and U_{+5} by PUF repeat R6 and the insertion in R3

U_{+5} is typically recognized by the $N_1Q_5-F_2$ motif in R6 (Figure 4E). The side chain of Asn-397 (position 1) forms hydrogen bonds with the O2 and N3 groups of U_{+5} , the side chain of Gln-401 (position 5) forms hydrogen bonds with the O4 group of Gln-401 (position 5), and the benzene ring

of Phe-398 (position 2) stacks in parallel between U_{+5} and G_{+6} .

Unexpectedly, U_{+4} is also recognized by R6 (Figure 4E). The side chain of Gln-401 forms hydrogen bonds with the O2 and N3 groups of U_{+4} and the benzene ring of Phe-398 interacts vertically with U_{+4} (Figure 4E). The side chain of Arg-254 in the $\alpha 2$ helix of the insertion in PUF repeat R3 forms a hydrogen bond with the O4 group of U_{+4} . Thus, the recognition motif for U_{+4} is $R_1'Q_5-F_2$. However, there are no interactions between Nop9 and the bases of U_{+4} and U_{+5} (Figure 4F).

The ITC results showed that the binding affinity of APUM23⁸⁵⁻⁶⁵⁵ for the 11-mer $U_{+4}A$ RNA (5'-GGAAAUGACGG-3') was approximately half that of the WT 11-mer WT RNA (Table 1 and Supplementary Figure S7a and d). This may have been because of the loss of the hydrogen bond interaction with Gln-401. Phe-398 and Gln-401 display more interactions with the RNA bases than the residues in other TRMs (Figure 4E). Phe-398 has interactions with U_{+4} , U_{+5} , and G_{+6} , and Gln-401

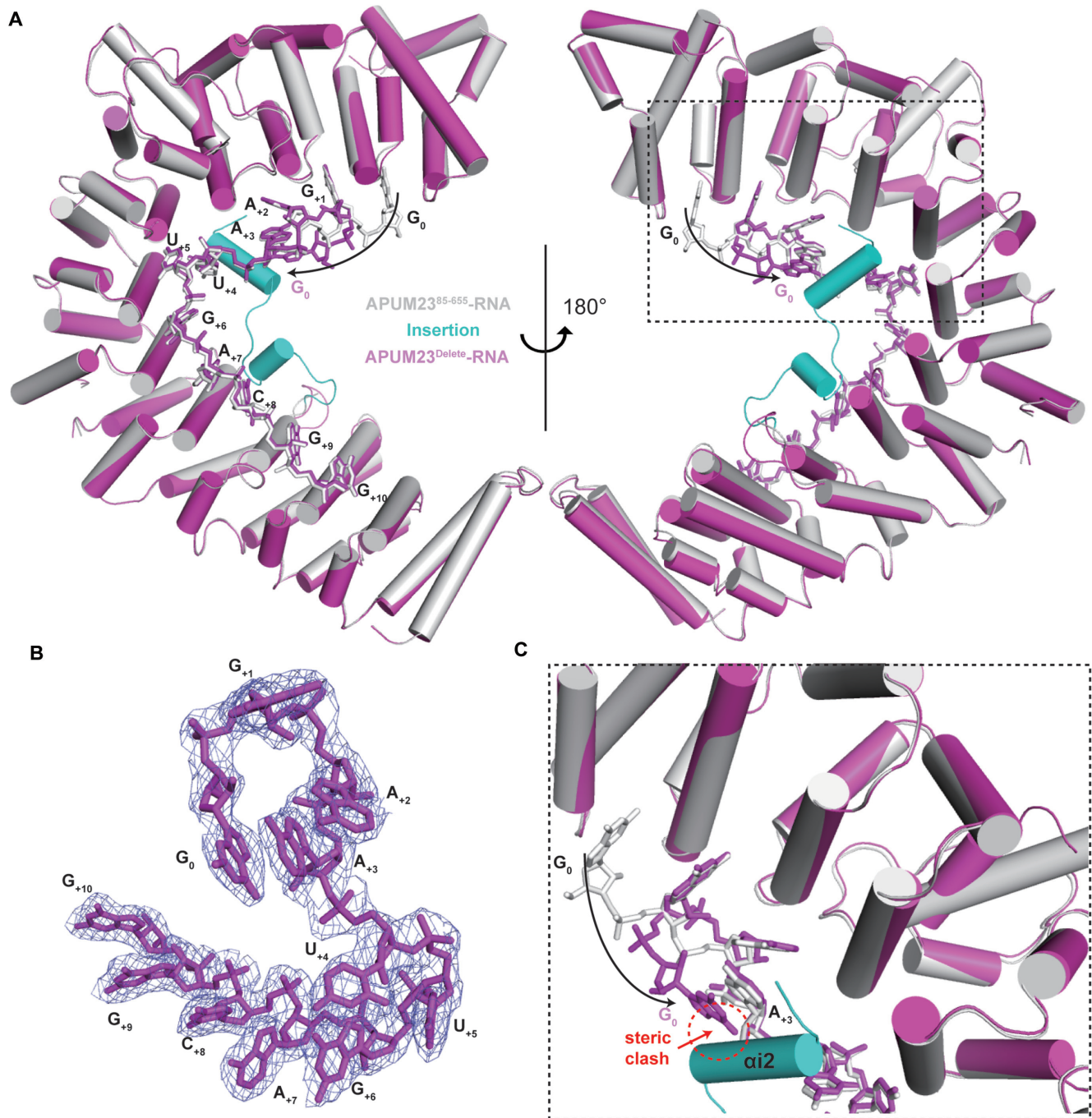


Figure 5. The insertion blocks G_0 from stacking with A_{+3} . (A) Superposition of the crystal structure of APUM23⁸⁵⁻⁶⁵⁵-GGAAUUGACGG (gray, with the insertion in cyan) and APUM23^{Delete}-GGAAUUGACGG (in magenta). (B) The 2Fo-Fc electron density map around G_0 to G_{+10} in the crystal structure of APUM23^{Delete}-GGAAUUGACGG, contoured at 1.0 sigma. (C) G_0 bends to stack against A_{+3} . The steric clash is highlighted by the red oval.

has interactions with U_{+4} and U_{+5} . These interactions indicate that Phe-398 and Gln-401 play important roles in the recognition of the 11-mer RNA target. ITC results suggested that the binding affinity of the 11-mer WT RNA for APUM23^{Q401E} and APUM23^{F398A} were ~52-fold and 90-fold weaker, respectively, than that for APUM23⁸⁵⁻⁶⁵⁵ (Table 1 and Supplementary Figure S7a, e and f).

Taken together, these findings indicate that APUM23 PUF repeat R6 is involved in the recognition of both U_{+4} and U_{+5} .

The insertion in PUF repeat R3 blocks G_0 from stacking with A_{+3}

To investigate the importance of this insertion further, we expressed an APUM23 mutant without the insertion

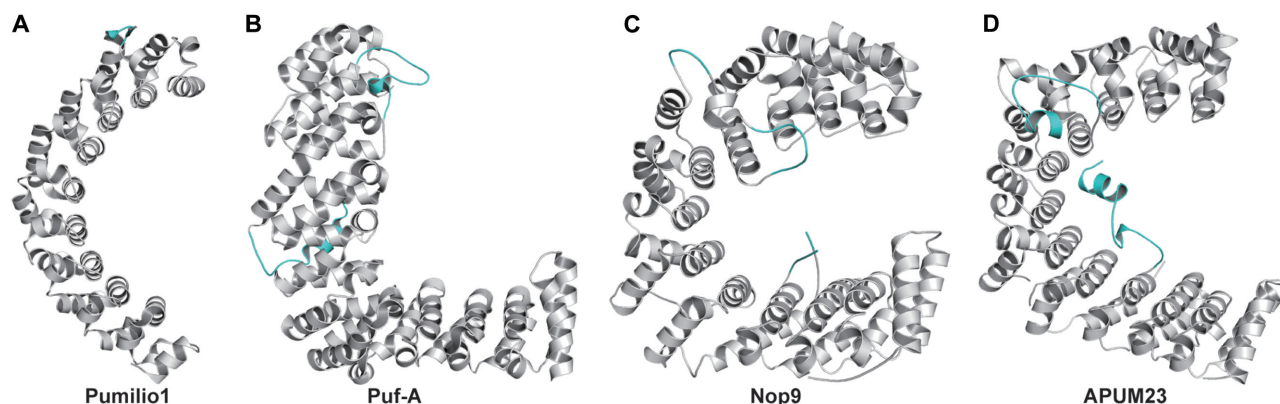


Figure 6. APUM23 varies in its structural features. (A–D) Cartoon representations of the crystal structures of Pumilio1 (PDB code: 1M8W), Puf-A (PDB code: 4WZR), Nop9 (PDB code: 5SVD) and APUM23. Insertions in each protein are colored in cyan.

(APUM23^{Delete}). Our ITC results showed that the binding affinity of the 11-mer WT RNA for APUM23^{Delete} was ~ 7.6 -fold weaker than that for APUM23⁸⁵⁻⁶⁵⁵ (Table 1 and Supplementary Figure S7a and g). We also obtained a 2.8-Å resolution crystal structure of APUM23^{Delete}-GGAAUUGACGG (Figure 5A and Supplementary Table S2). Apart from the insertion, the structure of APUM23^{Delete} is identical to that of APUM23⁸⁵⁻⁶⁵⁵. G₊₁ to G₊₁₀ are located in the same locations in the APUM23^{Delete}-RNA structure as in the APUM23⁸⁵⁻⁶⁵⁵-RNA structure (Figure 5A). Surprisingly, G₀ in the APUM23^{Delete}-RNA structure is not recognized by APUM23 R10; instead, it bends round to stack with A₊₃ (Figure 5B). By superpositioning the APUM23⁸⁵⁻⁶⁵⁵-GGAAUUGACGG and APUM23^{Delete}-GGAAUUGACGG structures, we observed a steric clash between the insertion and the bent-around G₀ (Figure 5C). This indicates that the presence of the insertion blocks G₀ from bending.

The ITC results suggested that the binding affinity of APUM23^{Delete} for 11-mer WT RNA was ~ 6.8 -fold higher than that for the 10-mer RNA 5'-GAAUUGACGG-3' (Table 1 and Supplementary Figure S7g and h). This may be due to A₊₃ instability caused by the loss of G₀ and the insertion.

These results indicated that the insertion in APUM23 PUF repeat R3 has functions that include both RNA recognition and RNA conformation maintenance.

DISCUSSION

Three main PUF protein structural features have previously been reported (Figure 6). The classic crescent-shaped PUF proteins contain eight PUF repeats and choose their specific single-stranded RNA target via modular interactions with the RNA bases (6). The L-shaped Puf-A/Puf6 contains 11 PUF repeats and binds single- or double-stranded RNA via interactions with the phosphate backbone (3). The C-shaped yeast Nop9 contains ten PUF repeats and interacts with single-stranded RNA and the duplex region of the internal transcribed spacer 1 (ITS1) pre-rRNA or the 11-mer single-stranded 18S rRNA (17,38). In the present

study, we solved the crystal structure of APUM23 in complex with 11-mer 5'-GGAAUUGACGG-3' single-stranded RNA. This crystal structure showed that the third PUF repeat of APUM23 contains an insertion located at the inner concave surface. However, the crystallization of the Nop9-RNA complex did not include this insertion. Furthermore, the conformation of the same single-stranded target RNA (5'-GGAAUUGACGG-3') differs in the structures of APUM23-RNA and Nop9-RNA, especially from A₊₃ to U₊₅. We speculate that this difference in RNA conformation is not caused by the removal of the insertion, for three reasons. First, the sequential difference between APUM23 and yeast Nop9 results in the difference in their target RNA conformations. The RNA recognition residues in the insertion in APUM23 PUF repeat R3 are not conserved in yeast Nop9 (Supplementary Figure S6), suggesting that the insertion of Nop9 may not interact with A₊₃ and U₊₄. In APUM23, the R6 PUF repeat stabilizes the conformation of both U₊₄ and U₊₅ through direct interactions. However, R6 in Nop9 has no interaction with U₊₄ or U₊₅, leaving them with greater flexibility. Second, removal of the insertion in APUM23 only affects the conformation of G₀, which bends round to stack with A₊₃. In the APUM23^{Delete}-RNA structure, PUF repeat R6 still recognizes U₊₄ and U₊₅, and G₀ replaces the insertion to stabilize the conformation of A₊₃. Third, APUM23⁸⁵⁻⁶⁵⁵ showed similar binding affinity toward both the 10-mer RNA (5'-GAAUUGACGG-3') and the 11-mer RNA (5'-GGAAUUGACGG-3'). In contrast, the binding affinity of APUM23^{Delete} for the 10-mer RNA is much weaker than that for the 11-mer RNA, possibly caused by the instability of A₊₃ resulting from the loss of both the insertion and G₀. These findings indicate that the main function of the insertion in APUM23 is to stabilize the conformation of A₊₃ and to block G₀ from bending. Taken together, the insertion in APUM23 plays roles in both stabilizing the conformation of A₊₃ and maintaining the right conformation of G₀. However, the insertion in Nop9 may not participate in the recognition of 5'-GGAAUUGACGG-3'. Further investigation is needed to elucidate the physiological function of this insertion in APUM23 and yeast Nop9.

Our complex structure included some other atypical structural features. For instance, PUF proteins usually recognize target RNAs in a modular way, with a single RNA base interacting with only one PUF repeat; however, in our complex, A₊₂ interacts with both R8 and R9, and G₊₁ with both R9 and R10. In addition, R6 is involved in the recognition of both U₊₄ and U₊₅. One advantage of PUF proteins is that they can be engineered as RNA-binding tools to monitor and control the RNA metabolism in living cells (40). Our structural data may offer an alternative template for developing such engineered RNA-binding tools.

Pre-rRNA processing is a pivotal process during eukaryotic ribosome biogenesis. In reaching their final forms, pre-rRNAs undergo multiple cleavages, modifications, and refolding, performed by endo- and exonucleases, small nucleolar RNAs, and other regulatory proteins, such as yeast Nop9 and APUM23 (41–44). Although the sequence identity between them is only 18%, APUM23 and Nop9 share some functional similarity; APUM23 expression in yeast can partially rescue Nop9 deletion (28). APUM23 and Nop9 are both involved in regulating the cleavage of 35S pre-rRNA and the mature of 18S rRNA (2,28,45). However, how they regulate pre-rRNA processing through direct interaction with 18S rRNA remains unknown. We are eager to identify other partners (protein or RNA) of APUM23 to elucidate its roles in the regulation of pre-rRNA processing. It has been proposed that Nop9 protects 20S pre-rRNA in the nucleolus from premature cleavage by maintaining the endonuclease Nob1 binding site in ITS1 (38). Whether APUM23 has other rRNA binding sites remains to be investigated.

DATA AVAILABILITY

The atomic coordinates and structure factors (codes: 5WZG, 5WZH, 5WZI, 5WZJ and 5WZK) have been deposited in the Protein Data Bank (<http://www.pdb.org/>). Primary data are available upon request- from the corresponding author.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank the staff of Beamline BL17U, BL18U and BL19U at SSRF for the assistance in data collection. We also thank Dr Fudong Li for help with the model building of Se-Met labeled APUM23.

FUNDING

Ministry of Science and Technology of China [2016YFA0500700]; Grant of the Strategic Priority Research Program of the Chinese Academy of Sciences [XDB08010101]; Chinese National Natural Science Foundation [31330018]; China Postdoctoral Science Foundation [2015M582009, 2016T90579]. Funding for open access charge: Ministry of Science and Technology of China [2016YFA0500700]; Grant of the Strategic Priority Research Program of the Chinese Academy of Sciences

[XDB08010101]; Chinese National Natural Science Foundation [31330018]; China Postdoctoral Science Foundation [2015M582009, 2016T90579].

Conflict of interest statement. None declared.

REFERENCES

1. Quenault, T., Lithgow, T. and Travençolo, A. (2011) PUF proteins: repression, activation and mRNA localization. *Trends Cell Biol.*, **21**, 104–112.
2. 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.
3. Qiu, C., McCann, K.L., Wine, R.N., Baserga, S.J. and Hall, T.M.T. (2014) A divergent Pumilio repeat protein family for pre-rRNA processing and mRNA localization. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 18554–18559.
4. 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.
5. Lu, G., Dolgner, S.J. and Hall, T.M. (2009) Understanding and engineering RNA sequence specificity of PUF proteins. *Curr. Opin. Struct. Biol.*, **19**, 110–115.
6. Wang, X., McLachlan, J., Zamore, P.D. and Hall, T.M.T. (2002) Modular recognition of RNA by a human pumilio-homology domain. *Cell*, **110**, 501–512.
7. Wang, X., Zamore, P.D. and Hall, T.M.T. (2001) Crystal structure of a Pumilio homology domain. *Mol. Cell*, **7**, 855–865.
8. Gupta, Y.K., Nair, D.T., Wharton, R.P. and Aggarwal, A.K. (2008) Structures of human Pumilio with noncognate RNAs reveal molecular mechanisms for binding promiscuity. *Structure*, **16**, 549–557.
9. 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.
10. Zhu, D., Stumpf, C.R., Krahn, J.M., Wickens, M. and Hall, T.M.T. (2009) A 5' cytosine binding pocket in Puf3p specifies regulation of mitochondrial mRNAs. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 20192–20197.
11. Lu, G. and Hall, T.M. (2011) Alternate modes of cognate RNA recognition by human PUMILIO proteins. *Structure*, **19**, 361–367.
12. Qiu, C., Kershner, A., Wang, Y., Holley, C.P., Wilinski, D., Keles, S., Kimble, J., Wickens, M. and Hall, T.M.T. (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.
13. Valley, C.T., Porter, D.F., Qiu, C., Campbell, Z.T., Hall, T.M.T. and Wickens, M. (2012) Patterns and plasticity in RNA-protein interactions enable recruitment of multiple proteins through a single site. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 6054–6059.
14. Wilinski, D., Qiu, C., Lapointe, C.P., Nevil, M., Campbell, Z.T., Tanaka, H., Hall, T.M. and Wickens, M. (2015) RNA regulatory networks diversified through curvature of the PUF protein scaffold. *Nat. Commun.*, **6**, 8213.
15. Jenkins, H.T., Baker-Wilding, R. and Edwards, T.A. (2009) Structure and RNA binding of the mouse Pumilio-2 Puf domain. *J. Struct. Biol.*, **167**, 271–276.
16. Wang, Y., Opperman, L., Wickens, M. and Hall, T.M.T. (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.
17. Wang, B. and Ye, K. (2017) Nop9 binds the central pseudoknot region of 18S rRNA. *Nucleic Acids Res.*, **45**, 3559–3567.
18. 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.
19. 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.

20. Adamala, K.P., Martin-Alarcon, D.A. and Boyden, E.S. (2016) Programmable RNA-binding protein composed of repeats of a single modular unit. *PNAS*, **113**, E2579–E2588.
21. Hall, T.M. (2016) De-coding and re-coding RNA recognition by PUF and PPR repeat proteins. *Curr. Opin. Struct. Biol.*, **36**, 116–121.
22. 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.
23. Koh, Y.Y., Wang, Y., Qiu, C., Opperman, L., Gross, L., Tanaka Hall, T.M. and Wickens, M. (2011) Stacking interactions in PUF-RNA complexes. *RNA*, **17**, 718–727.
24. Hall, T.M. (2014) Expanding the RNA-recognition code of PUF proteins. *Nat. Struct. Mol. Biol.*, **21**, 653–655.
25. Tam, P.P., Barrette-Ng, I.H., Simon, D.M., Tam, M.W., Ang, A.L. and Muench, D.G. (2010) The Puf family of RNA-binding proteins in plants: phylogeny, structural modeling, activity and subcellular localization. *BMC Plant Biol.*, **10**, 44.
26. Francischini, C.W. and Quaggio, R.B. (2009) Molecular characterization of Arabidopsis thaliana PUF proteins—binding specificity and target candidates. *FEBS J.*, **276**, 5456–5470.
27. Abbasi, N., Park, Y.-I. and Choi, S.-B. (2014) Pumilio Puf domain RNA-binding proteins in Arabidopsis. *Plant Signal. Behav.*, **6**, 364–368.
28. Huang, T., Kerstetter, R.A. and Irish, V.F. (2014) APUM23, a PUF family protein, functions in leaf development and organ polarity in Arabidopsis. *J. Exp. Bot.*, **65**, 1181–1191.
29. 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.
30. Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G. and McCoy, A. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 235–242.
31. Sheldrick, G.M. (2008) A short history of SHELX. *Acta Crystallogr. A.*, **64**, 112–122.
32. Zwart, P.H., Afonine, P.V., Grosse-Kunstleve, R.W., Hung, L.-W., Ioerger, T.R., McCoy, A.J., McKee, E., Moriarty, N.W., Read, R.J. and Sacchettini, J.C. (2008) Automated structure solution with the PHENIX suite. *Methods Mol. Biol.*, **426**, 419–435.
33. Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.*, **62**, 1002–1011.
34. 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.
35. Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J. and Grosse-Kunstleve, R.W. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.
36. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
37. DeLano, W.L. (2002) The PyMOL molecular graphics system. <http://www.pymol.org>.
38. Zhang, J., McCann, K.L., Qiu, C., Gonzalez, L.E., Baserga, S.J. and Hall, T.M. (2016) Nop9 is a PUF-like protein that prevents premature cleavage to correctly process pre-18S rRNA. *Nat. Commun.*, **7**, 13085.
39. Cheong, C.G. and Hall, T.M. (2006) Engineering RNA sequence specificity of Pumilio repeats. *PNAS*, **103**, 13635–13639.
40. Wang, Y., Wang, Z. and Tanaka Hall, T.M. (2013) Engineered proteins with Pumilio/fem-3 mRNA binding factor scaffold to manipulate RNA metabolism. *FEBS J.*, **280**, 3755–3767.
41. Woolford, J.L. Jr and Baserga, S.J. (2013) Ribosome biogenesis in the yeast *Saccharomyces cerevisiae*. *Genetics*, **195**, 643–681.
42. Weis, B.L., Kovacevic, J., Missbach, S. and Schleiff, E. (2015) Plant-Specific Features of Ribosome Biogenesis. *Trends Plant Sci.*, **20**, 729–740.
43. Henras, A.K., Plisson-Chastang, C., O'Donohue, M.F., Chakraborty, A. and Gleizes, P.E. (2015) An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley Interdiscip. Rev. RNA*, **6**, 225–242.
44. Fernandez-Pevida, A., Kressler, D. and de la Cruz, J. (2015) Processing of preribosomal RNA in *Saccharomyces cerevisiae*. *Wiley Interdiscip. Rev. RNA*, **6**, 191–209.
45. 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.