Published in final edited form as:

Nat Struct Mol Biol. 2008 April ; 15(4): 397–402. doi:10.1038/nsmb.1390.

Basis of altered RNA-binding specificity by PUF proteins revealed by crystal structures of yeast Puf4p

Matthew T Miller, Joshua J Higgin, and Traci M Tanaka Hall

Laboratory of Structural Biology, National Institute of Environmental Health Sciences, National Institutes of Health, 111 T.W. Alexander Drive, Building 101, Room F363, Research Triangle Park, North Carolina 27709, USA

Abstract

Pumilio/FBF (PUF) family proteins are found in eukaryotic organisms and regulate gene expression post-transcriptionally by binding to sequences in the 3' untranslated region of target transcripts. PUF proteins contain an RNA binding domain that typically comprises eight α -helical repeats, each of which recognizes one RNA base. Some PUF proteins, including yeast Puf4p, have altered RNA binding specificity and use their eight repeats to bind to RNA sequences with nine or ten bases. Here we report the crystal structures of Puf4p alone and in complex with a 9-nucleotide (nt) target RNA sequence, revealing that Puf4p accommodates an 'extra' nucleotide by modest adaptations allowing one base to be turned away from the RNA binding surface. Using structural information and sequence comparisons, we created a mutant Puf4p protein that preferentially binds to an 8-nt target RNA sequence over a 9-nt sequence and restores binding of each protein repeat to one RNA base.

Gene expression can be controlled at the levels of transcription and translation. Translational regulation offers a more rapid response to environmental conditions than regulation at the transcriptional level¹. PUF family proteins, named for founding members *Drosophila melanogaster* Pumilio and *Caenorhabditis elegans* fem-3 mRNA binding factor (FBF), are translational regulators that are important for germline stem cell maintenance and embryonic development and have been identified in yeast, plants and animals (reviewed in refs. 2-3). PUF proteins are characterized by an RNA binding domain, called a Pumilio homology domain, typically formed from eight α -helical repeats⁴⁻⁸.

The crystal structures of the RNA binding domains of the *D. melanogaster* and human Pumilio proteins revealed that the eight repeats, each with three α -helices, are nestled together to form a crescent shape⁹⁻¹¹. The structures of human Pumilio homolog 1 protein (PUM1) in complex with RNA ligand showed a simple pattern of sequence-specific RNA recognition that could be used to design RNA sequence specificity^{10,12-14}. Each repeat of PUM1 recognizes a single RNA base using three side chains from conserved positions. Thus, the eight repeats of PUM1 specifically recognize an 8-nt sequence, UGUA-U or C-AUA.

PUF proteins share the recognition sequence of UGUR (where R represents a purine) at the 5' end, which is well conserved, followed by 3' sequences that are specific to each PUF protein². *Saccharomyces cerevisiae* encode six PUF proteins, Jsn1p (also known as Puf1p),

Correspondence should be addressed to T.M.T.H. (hall4@niehs.nih.gov).

AUTHOR CONTRIBUTIONS J.J.H., M.T.M. and T.M.T.H. designed the experiments; J.J.H. crystallized native Puf4p; M.T.M. determined the structures of native and mutant Puf4p and performed the RNA binding analyses; M.T.M. and T.M.T.H. prepared the manuscript.

Accession codes Protein Data Bank: Coordinates and structure factor files have been deposited with the accession codes 3BWT (apoPuf4p), 3BX2 (Puf4p-RNA), and 3BX3 (mutPuf4-RNA).

Puf2p, Puf3p, Puf4p, Mpt5p (also known as Puf5p) and Puf6p. Previous work identified the mRNAs associated with TAP-tagged PUF proteins Puf1p to Puf5p (Mpt5p) and from these data consensus sequences for Puf3p, Puf4p and Mpt5p were derived¹⁵. The consensus sequences suggest that each of these proteins binds to RNA target sequences beginning with a UGUA sequence at the 5' end and a UA sequence at the 3' end. However, the Puf3p, Puf4p and Mpt5p consensus sequences have two, three or four bases between the conserved 5' and 3' motifs, respectively. Nevertheless, each of the proteins has eight repeats, but they recognize 8-nt (Puf3p), 9-nt (Puf4p) or 10-nt (Mpt5p) sequences. This suggests that the RNA binding domains of Puf4p and Mpt5p must accommodate binding of one or two additional nucleotides, respectively.

A similar situation exists in *C. elegans*, where Puf-8 and FBF proteins have differing specificities. Opperman, *et al.*¹² showed that Puf-8 recognizes 8-nt RNA sequences whereas FBF recognizes 9-nt RNA sequences, although both proteins have eight repeats². To determine whether the extra base in the FBF recognition sequence was accommodated within the RNA binding surface or at the 3' end, they engineered mutant FBF proteins that would lead to predictable changes in RNA binding specificity depending on where the extra RNA base was accommodated. These results allowed them to conclude that the additional base was accommodated within the RNA binding surface between the fourth and fifth repeats. A chimeric protein of Puf-8 in which repeat 5 and its flanking sequences were replaced by those from FBF bound preferentially to FBF's 9-nt recognition sequence, suggesting that modifications allowing binding of a longer recognition sequence can be contained in a relatively short stretch of the protein sequence.

Puf4p seems to bind to many mRNAs encoding nuclear or nucleolar proteins¹⁵. It was recently reported that Puf4p is important for post-transcriptional regulation of HO endonuclease expression¹⁶. HO endonuclease controls mating-type switching in yeast by introducing double-stranded DNA breaks that initiate recombination¹⁷. Both Mpt5p and Puf4p bind to sequences in the 3' untranslated region (3' UTR) of *HO* mRNA, collaborating to repress its expression^{16,18}.

To determine how Puf4p accommodates an additional nucleotide in its binding surface, we have solved the crystal structure of Puf4p in complex with a target RNA sequence from *HO* mRNA. This structure reveals subtle changes in the architecture of the PUF RNA binding domain that allow binding to a 9-nt RNA sequence. Furthermore, the structure led us to identify mutations that changed the preference of Puf4p from a 9-nt target to an 8-nt target. The structure of the mutant shows that the protein binds to the 8-nt target RNA, as was seen in the PUM1 structures, with the same regular one-repeat–one RNA base pattern.

RESULTS

Crystal structures of apoPuf4p and Puf4p–RNA complex

We determined crystal structures of the RNA binding domain of Puf4p alone (apoPuf4p) and bound to an RNA with the recognition sequence found in the 3' UTR of *HO* endonuclease mRNA, UGUAUAUUA (Puf4p–RNA)¹⁹. The structures of apoPuf4p and Puf4p–RNA were determined by molecular replacement. For the apoPuf4p structure, an optimal molecular replacement search model was determined empirically by testing models derived from the crystal structure of PUM1 with different numbers of structural repeats, truncating from either the N or C terminus. A model derived from repeats 3–8' of PUM1 was most successful. The apoPuf4p structure was used as the search model for the Puf4p–RNA structure. The structure of apoPuf4p has been refined to 2.7-Å resolution, and the Puf4p–RNA structure has been refined to 2.8-Å resolution.

Similarly to the structures of PUM1 and *D. melanogaster* Pumilio, Puf4p comprises eight α -helical repeats, and the repeats stack together to form a crescent shape. The central eight canonical repeats are flanked by a short sequence at the N terminus and an incomplete pseudorepeat at the C terminus, labeled repeat 8'. RNA binds to the concave surface of the protein (Fig. 1a). The structure of Puf4p does not seem to change upon RNA binding, as the structures of Puf4p in the apoPuf4p and Puf4p–RNA crystals are virtually identical, with an r.m.s. deviation of 0.54 Å over 325 matched CA positions (Supplementary Fig. 1 online).

Comparison of the structures of Puf4p–RNA and PUM1 in complex with RNA shows that the overall curvature of the Puf4p protein, which binds to RNA bases over a 9-nt sequence, is less than for PUM1, which binds to only 8 nt (Fig. 1b). This seems to be due to changes in the angles between repeats 3 and 4, such that in Puf4p the hinge angle is larger and there is an increased writhe relative to the long axis of the protein. This modest decrease in curvature may facilitate the binding of a 9-nt sequence by the eight repeats of Puf4p by creating a more extended RNA binding surface.

Recognition of conserved 5' and 3' RNA motifs by Puf4p

The consensus target sequence for Puf4p contains two highly conserved regions, a 5' UGUA sequence and a 3' UA sequence. The 5' UGUA sequence is common to most PUF protein targets and is recognized by repeats 5–8 in PUM1 (ref. 10). As a convention, we will refer to positions in PUF-recognition sequences numbering from the first U in UGUA. In the Puf4p–RNA structure, recognition of the 5' UGUA is similar to what was observed with PUM1 (Figs. 2a,b and Supplementary Figs. 2,3 online). Identical side chains from repeats 5–8 in each protein contact the Watson-Crick edges of the RNA bases and from repeats 6 and 8 form stacking interactions with the corresponding bases. The 2' OH of G2 makes two water-mediated contacts to Puf4p (Lys796 and Asn800); one is conserved in PUM1 (Lys1076). In repeat 5 of Puf4p, Cys724 replaces Arg1008 at the equivalent position in PUM1, which forms a stacking interaction with A4. The cysteine side chain is not long enough to form a stacking interaction, as observed with Asn800 in repeat 7 of Puf4p or Asn1080 in PUM1. Repeats 4–8 of Puf4p are similar to PUM1, with an r.m.s. deviation of 1.2 Å for 171 matched CA atoms versus an overall r.m.s. deviation of 2.2 Å for Puf4p–RNA and PUM1, including 277 matched CA positions. In addition, the conformation of the RNA is identical in the 5' UGUA region in Puf4p and PUM1 structures (Supplementary Fig. 3) but divergent in the 3' sequences²⁰.

The 3' UA sequence is conserved in many PUF protein target mRNAs. Puf4p also interacts with these two bases similarly to what was observed for PUM1 (Fig. 2c and Supplementary Figs. 2,3). The interaction of repeat 1 with A9 is modestly different from what occurs with PUM1. Gln582 of Puf4p is equivalent to Gln867 in PUM1, and this side chain is positioned similarly to its equivalent in PUM1, but the distances between possibly hydrogen-bonded groups are slightly longer in the Puf4p complex versus the PUM1 complex. The 2' OH group of A9 forms a hydrogen bond with Gln575 in Puf4p, equivalent to Gln860 in PUM1, which also contacts the 2' OH of the base interacting with repeat 1.

An additional RNA base in the central UAU sequence

Although interaction with the 5' and 3' sequence motifs is conserved between Puf4p and PUM1, interaction of Puf4p with the UAU sequence between these motifs includes features both unique to Puf4p and conserved with PUM1. Because the side chain of Cys724 does not form a stacking interaction with A4, the U5 base can form a direct stacking interaction with A4, rather than having a protein side chain sandwiched between them (Figs. 2a, d). This base-base interaction results in a repositioning of the pathway of the RNA backbone relative to what was observed for PUM1, and the Watson-Crick edge of U5 does not form hydrogen bonds with repeat 4. The equivalent residue in Mpt5p is a cysteine, suggesting that Mpt5p may also

accommodate extra RNA bases in its target sequences through a base-base interaction. Following this altered RNA structure, A6 is recognized by repeat 3 of Puf4p in much the same way as PUM1 recognizes adenosines. Gln654, Thr650 and Arg651 interact specifically with A6 via hydrogen-bond, van der Waals and stacking interactions, respectively (Fig. 2a).

From the Puf4p–RNA structure, U7 represents the additional base accommodated in the Puf4p recognition sequence. Although the electron density for this base is weak, from the RNA backbone arrangement it seems to be turned away from the RNA binding surface between repeats 2 and 3 and does not contact the protein. Arg651 contacts the 2' OH group of the backbone ribose of U7 and forms stacking interactions with the flanking A6 and U8 bases, which seem to stabilize the RNA conformation (Fig. 2d).

Because U7 is not contacted by Puf4p and the consensus sequence from genomic screening suggests that the seventh position in target RNA sequences is usually U, A or C and rarely G¹⁵, we tested whether the identity of the seventh base affects binding affinity. We used electrophoretic mobility shift assays to measure the equilibrium binding constants of Puf4p for RNA containing the 9-nt Puf4p recognition sequence found in *HO* mRNA¹⁹ and RNAs with substitutions at the seventh position (Table 1). Puf4p binds to the *HO* RNA target with a K_d of 13.6 nM, consistent with the previously reported K_d of 10 nM for Puf4p binding to the *HO* target sequence in the context of a longer RNA¹⁶. Substitution of U7 with a C or an A makes little difference to the binding affinity, but substitution with a G reduces the binding affinity approximately two-fold ($K_d = 30.0$ nM). These modest differences are consistent with the frequency of appearance of each base at the seventh position in the consensus sequence¹⁵. We also tested the effect of adding an extra base between the seventh and eighth positions, placing four nucleotides between the conserved 5' UGUA and 3' UA sequences rather than three nucleotides. Binding of this RNA was approximately two-fold weaker than the *HO* target sequence ($K_d = 26.4$ nM). This difference in binding affinity may result from the accommodation of two protruding bases between Puf4p repeats 2 and 3, but it is also possible that only one base can protrude and repeats 1 and 2 bind to UU rather than UA, reducing the binding affinity. Nevertheless, this moderate difference in *in vitro* affinity seems to reflect the consensus of putative target RNAs, with G being rare at position 7 and U being rare at position 9.

Specificity of Puf4p for 9-nt versus 8-nt target RNA

The consensus target RNA sequences of Puf3p and Puf4p indicate that Puf3p prefers an 8-nt sequence, similar to PUM1, and Puf4p prefers a 9-nt sequence that contains an extra U¹⁵. This specificity difference seems to be important for target RNA regulation *in vivo*, because when *puf3* is deleted in yeast Puf4p does not compensate for the loss of Puf3p to destabilize *cox17* mRNA²¹. We used electrophoretic mobility shift assays to compare the affinity of Puf4p for RNAs containing the 9-nt Puf4p recognition sequence found in *HO* mRNA¹⁹ and the 8-nt Puf3p recognition sequence found in *cox17* mRNA^{21,22} (Table 1 and Supplementary Fig. 4a, b online). As noted above, Puf4p binds to the *HO* RNA target with a K_d of 13.6 nM but shows a preference for the 9-nt sequence, binding approximately two-fold more tightly to its cognate RNA target than to the Puf3p *cox17* target sequence. This is a modest preference similar to that of Mpt5p, which binds 3.5-fold more tightly to its cognate RNA target versus the Puf4 *HO* target but much less than the 25-fold preference of Puf4p for its target versus the Mpt5p target sequence¹⁶.

Although the overall amino acid sequence homology between Puf3p and Puf4p is only 41%, 21 of 24 predicted RNA binding residues are identical. Two of the Puf4p RNA binding residues that differ in Puf3p are in repeats 3 and 5, which interact with the central UAU sequence, suggesting that these side chains might be responsible for the preference of Puf4p for a 9-nt RNA recognition sequence. In repeat 3, Thr650 corresponds to a cysteine in Puf3p and PUM1.

This residue is near the excluded U7 base in the Puf4p–RNA structure and could therefore be important for accommodation of the extra base. In repeat 5, Cys724 corresponds to an arginine in Puf3p and PUM1. This residue forms a stacking interaction between the bases in positions 4 and 5 in PUM1, so changing Cys724 to arginine could prevent the stacking of the fourth and fifth bases that is seen in the Puf4p–RNA structure. To test this hypothesis, we mutated these residues in Puf4p to the corresponding side chains in Puf3p, creating a T650C C724R mutant (mutPuf4p). We then analyzed the binding of mutPuf4p to the *HO* and *cox17* target RNA sequences. The mutPuf4p protein bound more weakly to the two RNAs (approximately four-fold less tightly), but nevertheless showed a two-fold preference for the Puf3p *cox17* target sequence over the *HO* RNA sequence (Table 1 and Supplementary Fig. 4c, d). Thus, mutation of these two residues to the corresponding amino acids in Puf3p has changed the specificity of Puf4p to an 8-nt RNA target sequence.

We determined the crystal structure of the mutPuf4p protein in complex with an 8-nt *cox17* target RNA (mutPuf4p–RNA) to 3.0-Å resolution (Supplementary Fig. 5 online). In this structure, Arg724 stacks between the fourth and fifth bases, and along with Arg720 forms hydrogen bonds with the 2' OH of A4 (Fig. 3a). Cys650 forms a van der Waals contact with A6, equivalent to what is seen in the PUM1 structures (Fig. 3). The overall curvature of the repeats of mutPuf4p is the same as in wild-type Puf4p, rather than being increased as it is in PUM1. The path of the RNA backbone relative to the RNA binding surface of the *cox17* RNA is similar to the nanos response element (NRE) RNA in the PUM1 structure (Supplementary Fig. 6 online). Many of the distances between mutPuf4p side chains and the *cox17* RNA are longer than in the Puf4p–RNA structure. This may be a result of the extended RNA binding surface, which may not be optimal for an 8-nt RNA target, and could account for the weaker interaction of mutPuf4p and its cognate *cox17* RNA versus Puf4p and *HO* RNA.

DISCUSSION

PUF family proteins are found in a wide range of eukaryotic organisms, with the number of unique PUF proteins varying from species to species. For example, *D. melanogaster* has only 1 PUF protein, humans have 2, *S. cerevisiae* has 6 and *C. elegans* has 11 (ref. 2). In most cases, the proteins are predicted to have eight repeats in the RNA binding domain, and the sequences of known mRNA targets of PUF proteins usually include a UGUR tetranucleotide motif that is recognized by repeats 5–8 in Puf4p and PUM1. Yeast Jsn1p and Puf2p are exceptions, as they are predicted to have only six repeats, and the predicted target RNAs do not contain clear consensus sequences. Nevertheless, the atypical PUF RNA binding domain of Jsn1p appears to interact with target RNA²³. Jsn1p and Puf2p also contain an RNA recognition motif domain that may contribute to RNA binding.

The structure of PUM1 in complex with RNA revealed a regular pattern of sequence-specific RNA recognition¹⁰. The recognition pattern is so elegant and predictable that we and others have been able to design RNA specificity of this domain, often with wild-type binding affinities^{10,12–14}. The structures of Puf4p reported here begin to reveal the complexity and adaptability of this protein family for RNA recognition, combining features conserved with PUM1 and modest substitutions that alter RNA binding specificity.

In the Puf4p–RNA structure, the bases of U5 (which stacks with A4) and U7 (which protrudes from the RNA binding surface) are not contacted by the protein and thus are not recognized specifically. This is consistent with the consensus sequence derived from genomic screening experiments to identify mRNA targets, which shows poor conservation of the bases at positions 5 and 7 (ref. 15). The consensus sequences for other PUF family members have been determined^{12,15,24–28}, and the Puf4p–RNA structure suggests that RNA base positions that are poorly conserved are not bound by the cognate PUF protein or recognized specifically. We

predict that the RNA base at position 8 in the yeast Mpt5p consensus sequence is not bound specifically, because it is not conserved, and the base at position 6 may be bound in a pocket that is restricted to pyrimidine bases, because uracil and cytosine are found equally as often at that position in target RNAs. Similarly, we predict that bases at positions 5 and 6 in FBF target RNAs, which are not conserved, are not bound specifically by the RNA binding surface²⁶. *Caenorhabditis elegans* Puf-8 has been shown to bind to an 8-nt consensus sequence¹², and we expect it to bind to its target RNAs as PUM1 does. Likewise, we anticipate that yeast Puf3p will bind to its 8-nt targets¹⁵ as PUM1 does. However, we note that a cytosine at position -2 (two bases 5' of the UGUA motif) is relatively well conserved and may be recognized specifically. Assuming the UGUA motif binds to repeats 5–8 in Puf3p, this base at the 5' end of the recognition sequence would be located near the C-terminal end of the protein, suggesting that this region of Puf3p may be adapted to create an extra binding pocket to recognize this base.

Residues that do not directly contact the RNA may also affect the specificity of RNA recognition. For example, the curvature of the protein seems to be important for the number of RNA bases that are bound by the eight repeats, and Puf4p and PUM1 do not change conformation upon RNA binding. Thus, amino acid residues that determine the overall curvature of the protein may be important for specificity. The mutPuf4p protein did not change curvature, but it bound to a shorter RNA sequence. This less-than-optimal binding surface may explain the decreased binding affinity of the mutant protein (Table 1).

The flexibility of the PUF protein RNA binding surface allows recognition of more diverse RNA targets than the one repeat–one RNA base model represented by the PUM1–RNA structures. It also seems possible that protruding bases in PUF protein–RNA complexes could be available to bind to downstream effector proteins. As these bases are poorly conserved, this provides an opportunity for a further layer of regulatory specificity, if a downstream effector protein specifically recognizes the protruding base(s). Therefore, one PUF protein could pair with different effector proteins depending on the identity of the protruding base. Alternatively, the downstream effector could bind non–sequence-specifically to the protruding RNA base(s), but formation of an effector complex could be dependent on the PUF protein binding first to an appropriate RNA target. Such a regulatory mechanism could be important in organisms with multiple PUF proteins expressed in the same cell with only modestly different affinities for cognate versus noncognate targets, such as Puf4p for 8-nt versus 9-nt targets and Mpt5p for 9-nt versus 10-nt targets.

METHODS

Cloning, protein expression and purification

A cDNA encoding the RNA binding domain of Puf4p (amino acids 536–888) was amplified from *S. cerevisiae* genomic DNA and inserted into the plasmid pENTR/TEV/D-TOPO (Invitrogen). We sequenced the entry plasmid containing the insertion and recombined the insertion into pDEST565, which encodes an N-terminal hexahistidine–glutathione *S*-transferase (GST) tag (provided by the National Institute of Environmental Health Sciences (NIEHS) Protein Expression Core Facility). The GST-Puf4p fusion protein was expressed in BL21(DE3) cells with 1 mM IPTG for 3 h at 37 °C. The cell pellets were resuspended in 25 ml of lysis buffer containing 25 mM HEPES, pH 7.5, 1 M NaCl and 1 mM DTT and frozen at –80 °C. We thawed the pellets in a water bath at room temperature after addition of one Roche complete EDTA-free protease inhibitor tablet and 2-mercaptoethanol to a concentration of 5 mM. We lysed the thawed cells by sonication in a dry ice and ethanol slurry. We purified the fusion protein from the soluble fraction by adding the supernatant from a 1 l culture to 10 ml of GST-bind resin (Invitrogen) in lysis buffer. The slurry was incubated for 15 min at 4 °C with gentle mixing and then washed twice with 50 ml of 25 mM HEPES, pH 7.5, 2 M NaCl

and 1 mM DTT. A final wash was performed to transfer the resin into TEV protease cleavage buffer, 50 mM Tris-HCl, pH 8.0, 5 mM DTT and 0.5 mM EDTA. After adding 100 μ l of 1 mg ml⁻¹ recombinant TEV protease, we incubated the resin at room temperature for 16 h. Following incubation with TEV protease, the resin was poured over a disposable column, and the flow-through containing the cleaved Puf4p protein was collected. We then passed 10 ml of lysis buffer through the column to displace any remaining cleaved protein. The cleaved protein contained non-native amino acids on the N terminus with the sequence Gly-Ser-Phe-Thr. The eluant was diluted to 150 ml with 25 mM HEPES, pH 7.5, 5 mM DTT, loaded onto a 6 ml Resource Q column (GE Healthcare) and eluted with a ten-column volume gradient from 0–2 M NaCl. Fractions containing Puf4p were pooled and diluted to 150 ml with 25 mM HEPES, pH 7.5, 1 mM DTT (final NaCl concentration of ~5 mM), loaded onto a 5 ml HiTrap heparin column and eluted with a 10-column volume gradient from 0–2 M NaCl. Fractions containing Puf4p were pooled and concentrated to 2 ml in an Amicon Ultra 10k concentrator (Millipore) and then loaded onto a Superdex 200 16/60 column (GE Healthcare) run at 1 ml min⁻¹ in 25 mM HEPES, pH 7.5, 1 M NaCl and 1 mM DTT. Fractions containing Puf4p were pooled, exchanged into 30 mM HEPES pH 7.5, 100 mM sodium acetate, 3 mM magnesium acetate, 5% (v/v) glycerol and concentrated to 10 mg ml⁻¹.

We made a T650C C724R mutPuf4p cDNA using the QuikChange site-directed mutagenesis kit (Stratagene), following the manufacturer's protocol. Mutants were confirmed by sequencing the full cDNA insert. We prepared the mutant protein as described above for the wild-type protein.

Crystallization and data collection

We crystallized Puf4p without RNA bound (apoPuf4p) by vapor diffusion in hanging drops at 22 °C, combining 1 μ l of Puf4p at 5 mg ml⁻¹ with 1 μ l of well solution (1.2 M ammonium sulfate, 100 mM sodium cacodylate, pH 6.2). Crystals with dimensions of 50 $\times$ 50 $\times$ 500 μ m³ appeared after 48 h and were flash-cooled in liquid nitrogen after incubation in a cryoprotectant solution containing the well solution supplemented with 20% (v/v) glycerol. A Puf4p–RNA complex was made by incubating the RNA (UGUAUAUUA) at 70 °C for 5 min followed by 5 min on ice before combining it at an equal molar ratio with the purified protein to a final concentration of 75 μ M. We crystallized this complex by vapor diffusion in hanging drops at 22 °C by combining 1 μ l complex with 1 μ l of well solution (1.2 M ammonium sulfate, 100 mM Tris pH 8.0). Crystals with dimensions of 200 $\times$ 100 $\times$ 50 μ m³ appeared after 48 h and were then flash-cooled in liquid nitrogen after incubation in a cryoprotectant solution containing the well solution supplemented with 20% (v/v) glycerol. We grew crystals of the T650C C724R mutPuf4p in complex with RNA (CUUGUAUAUA) under the same conditions as the wild-type Puf4p–RNA crystals. X-ray diffraction data for the apoPuf4p, Puf4p–RNA and mutPuf4p–RNA crystal structures were collected using a Rigaku microMax 007 X-ray generator with a Saturn92 CCD detector and Xstream cryo system at –180 °C, and data used for initial molecular replacement were collected at 0.920 Å at the Southeast Regional Collaborative Access Team Beamline ID-22 at the Advanced Photon Source in Argonne National Laboratories (Table 2).

Structure determination and refinement

We determined the structure of apoPuf4p by molecular replacement using a polyaniline model of the structure of PUM1 (1M8Z) including residues 925–1166 (repeats 3–8') with the program Phaser²⁹. The structure was fit to experimental electron density maps automatically by the program Solve/Resolve^{30–32}, and then the remaining unmodeled regions were built manually using Coot³³. Iterative refinement was performed using Phenix³⁴ and Refmac5 (ref. 35). We calculated phases for Puf4p–RNA and mutPuf4p–RNA structures by molecular replacement using the structure of apoPuf4p and the program Phaser²⁹. Rebuilding and refinement were

done iteratively with Coot³³ and Phenix³⁴. Noncrystallographic restraints were applied for the Puf4p–RNA and mutPuf4p–RNA structures. The apoPuf4p structure comprises residues A554–886, the Puf4p–RNA structure comprises residues A560–887 and B562–885, and the mutPuf4p–RNA structure comprises residues A562–886 and B586–885. The two 5′-most nucleotide bases, positions –1 and –2, were disordered in the crystal structures of Puf4p–RNA and mutPuf4p–RNA. The Puf4p–RNA and mutPuf4p–RNA structures contain two molecules in the asymmetric unit; in the results section, we report the details of the complexes with the lower overall B factor, although most details are observed in both complexes. The structures were analyzed using MolProbity²⁰. Ramachandran analysis showed that the Puf4p–RNA structure had 94.3% residues in the favored position with 0.5% outliers, the apoPuf4p structure had 98.5% residues in the favored position with no outliers and the mutPuf4p–RNA structure had 94.7% residues in the favored position with 0.3% outliers.

Electrophoretic mobility shift assays

We determined equilibrium dissociation constants for Puf4p and mutPuf4p proteins by electrophoretic mobility shift binding assays. RNA oligonucleotides were obtained from Dharmacon, Inc. and radiolabeled at the 5′ end using ³²P-γ-ATP (PerkinElmer Life Sciences) and T4 polynucleotide kinase (New England BioLabs) following manufacturer instructions.

Binding reactions included 100 pM radiolabeled target RNA and serially diluted concentrations of protein (Supplementary Fig. 4 online) incubated for 1 h at 4 °C in buffer containing 15 mM HEPES, pH 7.5, 50 mM sodium acetate, 1.5 mM magnesium acetate, 3% (v/v) glycerol, 50 μg ml^{−1} heparin, 25 μg ml^{−1} yeast tRNA, 0.01% (v/v) IGEPAL and 0.5 mM DTT. Reactions were run on 15% 29:1 (w/w) acrylamide:bisacrylamide gels with 0.5× Tris-borate with EDTA (TBE), pH 7.5, at 4 °C under 150 V constant voltage for 1 h. The gels were dried and then exposed to storage phosphor screens (GE Healthcare) and scanned on a Molecular Dynamics Typhoon phosphorimager (GE Healthcare). The intensities of bands corresponding to bound and free radiolabeled RNA were measured using GelEval (Frogsoft), and the data were plotted and analyzed using IgorPro (WaveMetrics). The equilibrium dissociation constants were calculated by fitting the plot of the fraction of RNA bound versus protein concentration to the Hill equation, assuming a Hill coefficient of 1, and the mean of three independent experiments and standard error are reported.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank K. Adelman and L. Pedersen for critical comments on the manuscript. We are grateful to A. Sigova of the University of Massachusetts Medical School for the yeast genomic DNA, A. Clark for advice on purifying yeast genomic DNA, J. Holmes for technical assistance, L. Pedersen for crystallography support and D. Esposito of the Protein Expression Lab, NCI-Frederick, for the pDEST-565 plasmid. This work was supported by the Intramural Research Program of the National Institutes of Health, National Institute of Environmental Health Sciences. The Advanced Photon Source used for this study was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract no. W-31-109-Eng-38. We thank Z. Jin for collecting these data at SER-CAT as part of the mail-in crystallography program.

REFERENCES

1. Curtis D, Lehmann R, Zamore PD. Translational regulation in development. *Cell* 1995;81:171–178. [PubMed: 7736569]
2. Wickens M, Bernstein DS, Kimble J, Parker RA. PUF family portrait: 3′UTR regulation as a way of life. *Trends Genet* 2002;18:150–157. [PubMed: 11858839]

3. Wharton RP, Aggarwal AK. mRNA regulation by Puf domain proteins. *Sci. STKE* 2006;2006:pe37. [PubMed: 17003467]
4. Zhang B, et al. A conserved RNA-binding protein that regulates sexual fates in the *C. elegans* hermaphrodite germ line. *Nature* 1997;390:477–484. [PubMed: 9393998]
5. Barker DD, Wang C, Moore J, Dickinson LK, Lehmann R. Pumilio is essential for function but not for distribution of the *Drosophila* abdominal determinant Nanos. *Genes Dev* 1992;6:2312–2326. [PubMed: 1459455]
6. Macdonald PM. The *Drosophila pumilio* gene: an unusually long transcription unit and an unusual protein. *Development* 1992;114:221–232. [PubMed: 1576962]
7. Zamore PD, Williamson JR, Lehmann R. The Pumilio protein binds RNA through a conserved domain that defines a new class of RNA-binding proteins. *RNA* 1997;3:1421–1433. [PubMed: 9404893]
8. Wharton RP, Sonoda J, Lee T, Patterson M, Murata Y. The Pumilio RNA-binding domain is also a translational regulator. *Mol. Cell* 1998;1:863–872. [PubMed: 9660969]
9. Edwards TA, Pyle SE, Wharton RP, Aggarwal AK. Structure of Pumilio reveals similarity between RNA and peptide binding motifs. *Cell* 2001;105:281–289. [PubMed: 11336677]
10. Wang X, McLachlan J, Zamore PD, Hall TMT. Modular recognition of RNA by a human Pumilio-homology domain. *Cell* 2002;110:501–512. [PubMed: 12202039]
11. Wang X, Zamore PD, Hall TMT. Crystal structure of a Pumilio homology domain. *Mol. Cell* 2001;7:855–865. [PubMed: 11336708]
12. Opperman L, Hook B, DeFino M, Bernstein DS, Wickens M. A single spacer nucleotide determines the specificities of two mRNA regulatory proteins. *Nat. Struct. Mol. Biol* 2005;12:945–951. [PubMed: 16244662]
13. Ozawa T, Natori Y, Sato M, Umezawa Y. Imaging dynamics of endogenous mitochondrial RNA in single living cells. *Nat. Methods* 2007;4:413–419. [PubMed: 17401370]
14. Cheong C-G, Hall TMT. Engineering RNA sequence specificity of Pumilio repeats. *Proc. Natl. Acad. Sci. USA* 2006;103:13635–13639. [PubMed: 16954190]
15. Gerber AP, Herschlag D, Brown PO. Extensive association of functionally and cytologically related mRNAs with Puf family RNA-binding proteins in yeast. *PLoS Biol* 2004;2:e79. [PubMed: 15024427]
16. Hook BA, Goldstrohm AC, Seay DJ, Wickens M. Two yeast PUF proteins negatively regulate a single mRNA. *J. Biol. Chem* 2007;282:15430–15438. [PubMed: 17389596]
17. Herskowitz I. Life cycle of the budding yeast *Saccharomyces cerevisiae*. *Microbiol. Rev* 1988;52:536–553. [PubMed: 3070323]
18. Tadauchi T, Matsumoto K, Herskowitz I, Irie K. Post-transcriptional regulation through the HO 3'-UTR by Mpt5, a yeast homolog of Pumilio and FBF. *EMBO J* 2001;20:552–561. [PubMed: 11157761]
19. Goldstrohm AC, Hook BA, Seay DJ, Wickens M. PUF proteins bind Pop2p to regulate messenger RNAs. *Nat. Struct. Mol. Biol* 2006;13:533–539. [PubMed: 16715093]
20. Davis IW, et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 2007;35:W375–W383. [PubMed: 17452350]
21. Olivas W, Parker R. The Puf3 protein is a transcript-specific regulator of mRNA degradation in yeast. *EMBO J* 2000;19:6602–6611. [PubMed: 11101532]
22. Jackson JS Jr, Houshmandi SS, Lopez Leban F, Olivas WM. Recruitment of the Puf3 protein to its mRNA target for regulation of mRNA decay in yeast. *RNA* 2004;10:1625–1636. [PubMed: 15337848]
23. Ulbricht RJ, Olivas WM. Puf1p acts in combination with other yeast Puf proteins to control mRNA stability. *RNA* 2008;14:246–262. [PubMed: 18094119]
24. White EK, Moore-Jarrett T, Ruley HE. PUM2, a novel murine Puf protein, and its consensus RNA-binding site. *RNA* 2001;7:1855–1866. [PubMed: 11780640]
25. Fox M, Urano J, Reijo Pera RA. Identification and characterization of RNA sequences to which human PUMILIO-2 (PUM2) and deleted in Azoospermia-like (DAZL) bind. *Genomics* 2005;85:92–105. [PubMed: 15607425]

26. Bernstein D, Hook B, Hajarnavis A, Opperman L, Wickens M. Binding specificity and mRNA targets of a *C. elegans* PUF protein, FBF-1. *RNA* 2005;11:447–458. [PubMed: 15769874]
27. Gerber AP, Luschig S, Krasnow MA, Brown PO, Herschlag D. Genome-wide identification of mRNAs associated with the translational regulator PUMILIO in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 2006;103:4487–4492. [PubMed: 16537387]
28. Seay D, Hook B, Evans K, Wickens M. A three-hybrid screen identifies mRNAs controlled by a regulatory protein. *RNA* 2006;12:1594–1600. [PubMed: 16809817]
29. McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ. Likelihood-enhanced fast translation functions. *Acta Crystallogr. D Biol. Crystallogr* 2005;61:458–464. [PubMed: 15805601]
30. Terwilliger TC. Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr. D Biol. Crystallogr* 2003;59:38–44. [PubMed: 12499537]
31. Terwilliger TC. Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr* 2000;56:965–972. [PubMed: 10944333]
32. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr* 1999;55:849–861. [PubMed: 10089316]
33. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr* 2004;60:2126–2132. [PubMed: 15572765]
34. Adams PD, et al. Recent developments in the PHENIX software for automated crystallographic structure determination. *J. Synchrotron Radiat* 2004;11:53–55. [PubMed: 14646133]
35. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr* 1997;53:240–255. [PubMed: 15299926]
36. DeLano, WL. The PyMOL Molecular Graphics System. DeLano Scientific; San Carlos, CA, USA: 2002.

Figure 1.

Crystal structure of the Puf4p RNA binding domain. **(a)** Ribbon diagram of the crystal structure of the RNA binding domain of *S. cerevisiae* Puf4p in complex with an *HO* RNA fragment (UGUAUAUUA). Protein repeats are colored alternately blue and green. The RNA is colored according to atom type (carbon, white; nitrogen, blue; oxygen, red; sulfur, yellow; and phosphorus, orange). Protein side chains that contact the RNA are shown. This figure and Figures 2b–d and 3 were prepared with PyMol³⁶; **(b)** Superposition of CA traces of the Puf4p–RNA complex (blue) and the PUM1–RNA complex (yellow; PDB ID: 1M8Y) aligned over repeats 4–8'.

Figure 2.

Recognition of RNA by the Puf4p RNA binding domain. (a) Schematic representation of the interaction between Puf4p and *HO* RNA. Distances (Å) between interacting atoms are indicated; (b) Puf4p recognition of the 5' conserved UGUA sequence (U1–A4). The protein and RNA in **b–d** are colored as in Figure 1a. (c) Puf4p recognition of the 3' conserved UA sequence (U8–A9). (d) Puf4p recognition of the central UAU sequence (U5–U7). The carbon atoms in U7 are colored yellow.

Figure 3. Crystal structure of the mutPuf4p-*cox17* RNA complex. **(a)** Interaction of repeats 3–5 of the T650C C724R mutPuf4p with *cox17* Puf3 target RNA fragment (UGUAUAUA). Protein repeats are colored alternately blue and green. The RNA is colored as in Figure 1a, except that carbon atoms are colored tan. Cys650 and Arg724 are shown in yellow; **(b)** Interaction of repeats 3–5 of PUM1 in complex with NRE RNA. Cys935 and Arg1008 are equivalent to Cys650 and Arg724 in mutPuf4p.

Table 1

RNA binding analyses of wildtype and mutant Puf4p

Protein	RNA	RNA sequence	Kd (nM)
Puf4p	<i>HO</i>	UGUAUAUUA	13.6 +/- 0.9
Puf4p	<i>hmt1</i> ^a	UGUAUACUA	15.7 +/- 4.6
Puf4p	<i>dut1</i> ^a	UGUAUAAUA	22.3 +/- 4.6
Puf4p	<i>HO-U7G</i>	UGUAUAGUA	30.0 +/- 7.6
Puf4p	<i>HO-U7UU</i>	UGUAUAUUUA	26.4 +/- 7.1
Puf4p	<i>cox17</i>	UGUAUAUA	31.6 +/- 4.3
mutPuf4p	<i>HO</i>	UGUAUAUUA	103.4 +/- 14.0
mutPuf4p	<i>cox17</i>	UGUAUAUA	57.2 +/- 7.6

^a *Hmt1* (protein arginine N-methyl transferase) and *dut1* (dUTP pyrophosphatase) are putative mRNA targets of Puf4p identified in Gerber, *et al.*¹⁵ that contain the indicated RNA sequences.

^b The bases that are inserted relative to the *cox17* sequence are in boldface.

Table 2

Data collection and refinement statistics (molecular replacement)

	apoPuf4p	Puf4p:RNA	mutPuf4p:RNA
Data collection			
Space group	C222 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	54.2, 137.7, 137.5	65.9, 136.2, 158.9	66.1, 137.1, 160.6
α , β , γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	31.2–2.7 (2.8–2.7)*	14.9–2.8 (3.0–2.8)	24.4–3.0 (3.1–3.0)
<i>R</i> _{sym} or <i>R</i> _{merge}	12.7 (42.3)	11.2 (31.8)	18.8 (52.0)
<i>I</i> / σ <i>I</i>	8.7 (3.0)	6.3 (2.4)	14.5 (3.2)
Completeness (%)	97.7 (95.2)	94.2 (94.2)	98.0 (94.7)
Redundancy	3.6 (3.1)	2.9 (1.8)	5.2 (3.4)
Refinement			
Resolution (Å)	31.2–2.7	14.9–2.8	24.4–3.0
No. reflections	14317	32060	29415
<i>R</i> _{work} / <i>R</i> _{free}	20.0/27.2	20.6/26.1	29.4/34.9
No. atoms			
Protein	2649	5313	5117
Ligand/ion	0	282	245
Water	55	58	0
<i>B</i> -factors	(Ask for input)		
Protein	28.0	39.6	40.1
Ligand/ion	na	84.3	58.5
Water	23.9	29.5	na
R.m.s. deviations			
Bond lengths (Å)	0.008	0.010	0.009
Bond angles (°)	1.062	1.239	1.383

* Number of xtals for each structure should be noted in footnote. Values in parentheses are for highest-resolution shell.