Published in final edited form as:

Curr Opin Struct Biol. 2016 February ; 36: 116–121. doi:10.1016/j.sbi.2016.01.010.

De-coding and Re-coding RNA recognition by PUF and PPR repeat proteins

Traci M. Tanaka Hall

Epigenetics and Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709, USA.

hall4@niehs.nih.gov

Abstract

PUF and PPR proteins are two families of α -helical repeat proteins that recognize single-stranded RNA sequences. Both protein families hold promise as scaffolds for designed RNA-binding domains. A modular protein RNA recognition code was apparent from the first crystal structures of a PUF protein in complex with RNA, and recent studies continue to advance our understanding of natural PUF protein recognition (decoding) and our ability to engineer specificity (re-coding). Degenerate recognition motifs make de-coding specificity of individual PPR proteins challenging. Nevertheless, recoding PPR protein specificity using a consensus recognition code has been successful.

Introduction

The molecular specificity of RNA-binding proteins is central to the fidelity of RNA regulation, and many groups are aiming to de-code the recognition properties of different families of RNA-binding proteins. Understanding RNA recognition also allows re-coding of RNA-binding proteins that may be employed to modulate RNA regulation *in vivo*. Progress toward understanding and engineering the target specificity of different RNA-binding domains has been reviewed recently [1–3]. This review focuses on recent work de-coding natural specificity and using that information to re-code specificity of two families of α -helical repeat proteins: Pumilio/*fem-3* mRNA binding factor (FBF) or PUF proteins and pentatricopeptide repeat or PPR proteins.

Natural and designed specificity of PUF and PPR proteins

Classical PUF proteins are a small family of eukaryotic proteins that bind to single-stranded RNA sequences using eight α -helical Pumilio (PUM) repeats of ~36 residues each (reviewed in [4]). PUF proteins bind to recognition sequences in their target mRNAs to control stability and translation. The PUM repeats form a crescent-shaped structure [5, 6], and RNA binds to the concave surface in an ‘antiparallel’ orientation with the 5’ end near

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

the C terminus of the protein [7]. Each PUM repeat binds to one RNA base, and the prototypical PUF protein, Pumilio1, uses each of its repeats to recognize an eight-base sequence, 5'-UGUAnAUA-3' (n=any base). Crystal structures of human Pumilio1 bound to RNA [7] illuminated the natural PUF protein recognition code for U, A, and G bases (Figure 1a). Amino acid side chains in the second of three α -helices recognize RNA bases (Figure 1b). Selection screens later identified recognition motifs for C bases (Figure 1a) [8, 9]. Site-directed mutagenesis using this recognition code predictably alters the sequence specificity of PUF proteins [7–12]. Designed RNA-recognition modules targeting endogenous RNAs have been fused to different functional modules to create engineered fluorescent tracers, splicing factors, translational regulators, and nucleases [8, 13–17].

PPR proteins are a large family of eukaryotic α -helical repeat proteins found predominantly in plants (reviewed recently in [18]). Repeats fall into three classes by sequence conservation, diverging at the repeat C-terminus: 35-residue classical P-type repeats, 31-residue S (short)-type repeats, and 35-to-36-residue L (long)-type repeats [19]. Natural PPR proteins contain from 2 to 30 repeats and are categorized as P class, if they comprise only P-type PPR motifs, or PLS class, if they bear alternating P, L, and S repeats. P-class PPR proteins commonly regulate RNA stability, processing, splicing and translation. PLS-class PPR proteins are typically involved in RNA editing. Like PUF proteins, PPR proteins bind to single-stranded RNA sequences, and their modular helical repeats suggested that they would be amenable to design for specific sequence recognition. To decode PPR protein RNA base recognition, Barkan, *et al.* [20] used bioinformatic analyses of three P-class PPR proteins (PPR10, HCF152, and CRP1) to identify conserved amino acid residues that co-vary with the RNA sequences each protein recognizes. By examining which residues correlate with the four different RNA bases and which bases correlate with particular sets of conserved side chains, an RNA recognition code for PPR proteins was predicted (Figure 1c). Residues at positions 6 and 1' (the first base of the subsequent repeat) appeared to be critical for RNA base recognition and specificity of repeats. Barkan, *et al.* mutated repeats 6 and 7 of PPR10 and validated residue combinations specific for G, A, U and C bases. Although the code can be used to design specificity, it remains difficult to predict natural PPR protein specificity, because the natural recognition code is degenerate [20].

Crystal structures of PPR proteins have revealed beautiful superhelical structures formed from the α -helical repeats [21–28]. However, structures of PPR protein:RNA complexes have been challenging to obtain, and no structure yet represents the expected biological ensemble with matching repeat specificity and RNA sequence (see below). Nevertheless, PPR protein:RNA structures provide insight into how individual repeats specifically recognize their corresponding RNA bases (Figure 1d). Each repeat comprises two α helices and a C-terminal loop region, and, as predicted, residues responsible for RNA sequence recognition are found within the first α helix (position 6) and at the end of the loop region (position 1') [26, 28]. Examples of U and G recognition from PPR10 and THA8, respectively, are shown in Figure 1d. The RNA base is situated in a pocket formed by hydrophobic interactions with side chains at position 3 in adjacent repeats. Residues at positions 6 and 1' contact the Watson-Crick edge of the RNA base. This interaction mode bears general similarity to PUF protein RNA base recognition (Figure 1b), but the specific

recognition motifs are different. Although the overall atomic-level details of PPR protein RNA recognition remain to be seen, the prospects for designing RNA-binding modules are bright [20, 23, 29–32].

An expanded PUF protein RNA recognition code

The high conservation of the RNA recognition code from Pumilio1 and other well-studied PUF proteins has emphasized the use of this code for engineered modules. However, Campbell, *et al.* recognized the diversity of RNA recognition residues in fungal PUF proteins and used this information to expand the PUF RNA recognition code [33]. They used a rigorous, systematic approach to determine the specificities of 25 different sets of RNA-binding residues culled from fungal PUF protein sequences. They substituted these combinations in repeat 7 of *C. elegans* FBF-2 and used the *in vitro* selection method, SEQRS [34], to identify how the substitutions affected RNA recognition properties. Because SEQRS determines the full nine-base recognition sequences, not just the RNA base(s) preferred at the targeted +2 position, Campbell, *et al.* also analyzed effects on specificity at the neighboring +1 and +3 positions. Recognition motifs with optimal specificity selected the expected base at the +2 position without compromising specificity for the neighboring bases (Table 1).

The study by Campbell, *et al.* [33] is rich with information, and here I highlight three findings of interest. (1) The natural PUF protein code uses the highest specificity motifs for G and U bases. The high fidelity of the natural motifs for G and U may reflect the importance of a characteristic 5'-UGU found in PUF protein recognition sequences. (2) Natural A-specific motifs show lower overall specificity than for G and U, and a new recognition motif showed the highest overall specificity. In many cases tested, the identity of the base-stacking residue has not appeared to be crucial for specificity [10, 12, 35], yet the stacking residue affects A specificity (Table 1). This finding is consistent with the idea that optimal chemical specificity is not always biologically preferable. Arginine stacking side chains, common in natural A-base recognition, may optimize protein solubility while maintaining the biologically required level of specificity. It is also possible that the stacking side chains are less important for specificity in repeats that naturally recognize A (typically repeats 1, 3 and 5) than in repeat 7 where motifs were tested. (3) Motifs that had been selected for C specificity in repeat 7 of Pumilio1 [8, 9] recognized G or U in FBF-2 repeat 7. This finding is quite surprising, since two groups independently identified the motifs. More investigations are needed to resolve this mystery.

Natural PUF protein RNA recognition modes

PUF proteins recognize RNA sequences longer than eight bases by using recognition modes different from the prototypical one repeat:one base recognition mode [11, 12, 36–41]. However, the RNA recognition side chains of the PUM repeats across PUF proteins are highly conserved, making it appear based on the recognition code that they all recognize the same eight-base sequence, despite knowledge to the contrary. As a result, it remains impossible to predict based on primary structure whether a PUF protein will bind to an eight-base sequence or a longer RNA recognition sequence. Even with knowledge of

consensus recognition sequences, binding modes cannot always be predicted. Structural and systematic biochemical studies continue to be the best ways to determine the binding mode(s) of a particular PUF protein.

Most classical PUF proteins recognize a 5'-UGUR element (R=purine) and a downstream UA element, and the characteristic recognition sequences of individual PUF proteins vary in the number of bases between the UGUR and UA elements [39]. Some of the first studies identifying target mRNAs of *S. cerevisiae* Puf4 [36] and *C. elegans* FBF [12, 42] proteins revealed that they recognize nine-base consensus sequences, one base longer than expected for their eight repeats. In-depth studies of yeast Puf4 and worm FBF demonstrated that the recognition sequences for these proteins include additional bases that are flipped away from the RNA-binding surface and not directly recognized by a PUM repeat [37, 40]. These bases appear in consensus RNA recognition elements as positions where there is no base preference. The same is true for other PUF proteins that recognize RNA sequences longer than eight bases [38, 43, 44].

An important lesson learned from de-coding the recognition preferences of PUF proteins is that manual intervention when analyzing data from genome-wide co-immunoprecipitation or *in vitro* selection methods may be needed to parse out small distinctions in RNA recognition. Straightforward analyses work well to identify consensus sequences for PUF proteins; yet in some cases, limiting the search to one consensus sequence averages out subtle features. Studies of the specificity of yeast Puf3 protein revealed a specialized binding pocket recognizing a cytosine base located two positions upstream of its 5'UGUA element (-2C) [41]. This pocket appeared to be limited to fungal PUF proteins. However, a similar pocket was found in FBF by re-examining FBF target sequences and allowing an upstream C to be present at either the -2 or -1 position [38]. In another case, yeast Puf4 was found to bind a second subset of nine-base sequences by excluding a different base from recognition [39]. This additional binding mode suggested mechanisms for co-regulation of mRNAs by multiple PUF proteins and evolution of RNA regulatory networks.

Recent studies of yeast Puf5 protein revealed that it recognizes different length RNA sequences and correspondingly regulates distinct cellular processes depending on the target sequence length [43]. The first analysis of Puf5 specificity used microarrays to identify RNAs that co-immunopurified with Puf5 [36]. Sixty-six mRNAs were identified with a 10-base consensus sequence, 5'-UGUAAYAnUA-3' (Y=pyrimidine, n=any base), two bases longer than expected for Puf5's eight PUM repeats. Now, utilizing high-throughput sequencing analyses of RNAs crosslinked to Puf5 *in vivo*, the Wickens' laboratory found that Puf5 interacts with RNA sequences ranging from 8–12 bases [43]. *In vitro* selection using the SEQRS technique also identified a range of recognition sequence lengths. The biological functions associated with mRNAs bearing different length recognition sequences suggest that Puf5 regulates diverse pathways: 8-base elements regulate mitochondrial organization, 9-base elements control ribosome biogenesis, 10-base elements regulate gene expression, and 11-base elements modulate translation. Furthermore, for 8- and 9-base elements, Puf5 may co-regulate targets together with Puf3 and Puf4, respectively. Crystal structures of Puf5 in complex with 9-, 10-, 11-, and 12-base RNA recognition elements revealed a fixed protein scaffold that recognizes conserved 5'-UGUA and downstream UA

sequence elements (Figure 2a) [43]. The RNAs of varied length adopt different conformations, while the protein conformation is unchanged. The pre-formed arrangement and curvature of PUM repeats appears to be a common feature of PUF proteins, as Pumilio1 and Puf4 proteins assume the same conformation with and without bound RNA [5, 7, 37].

PPR protein RNA target recognition

In contrast to classical PUF proteins formed from eight PUM repeats, PPR proteins contain from two to 30 repeats. Maize protein PPR10 comprises 19 PPR motifs, is localized to chloroplasts, and protects its two known mRNA targets, *PSAJ* and *ATPH*, from exonucleases. Bioinformatic analysis of PPR10 and its target RNAs guided discovery of the PPR recognition code and suggested that PPR10 uses repeats 3–19 to recognize a sequence of 17–18 bases, 5'-GUAUYCUUxAAYYAUUUC-3' (x=optional spacer RNA base) [20]. Different from PUF proteins, RNA binds to PPR proteins in a 'parallel' conformation, with the 5' end near the N-terminus of the protein [20, 26, 28]. Based on matching of PPR motifs to the RNA sequence, we expect that repeats 3–10 bind to the first eight bases and repeats 11–19 bind to the last nine bases; a spacer base may be accommodated near repeats 10 and 11.

A crystal structure of a PPR10:*PSAJ* RNA complex displayed an intertwined, antiparallel dimer of superhelical PPR10 molecules, but did not reveal the expected interaction between RNA bases and PPR motifs [28]. Two RNA molecules bound in an apparently unorthodox manner: each RNA assumes a horseshoe conformation and interacts with the N-terminal repeats of one monomer and the C-terminal repeats of the second monomer. Since many of the RNA bases are not recognized by the PPR motifs and the dimerization interferes with RNA binding to the central repeats, the overall protein:RNA conformation in this structure is likely due to the influence of crystallization contacts.

PPR10 is monomeric at lower protein concentrations and forms a 1:1 protein:RNA complex with *ATPH* or *PSAJ* RNAs [45, 46]. The optimized PPR10 protein that was crystallized with *PSAJ* RNA may have a greater propensity to dimerize due to one of the cysteine-to-serine mutations and the absence of an N-terminal extension [46]. Target RNA also induced a dimer in a crystal structure of THA8 [26], a small plant protein with five PPR motifs that is required for group II intron splicing. In the THA8 structure, RNA bases also were not aligned with the PPR motifs, making it unlikely that the structure represents the *in vivo* RNA recognition mode. Although these structures provide glimpses of PPR motif:RNA base interactions, engineering PPR10 to interfere with protein dimerization may be needed to crystallize a 1:1 biological assembly of a PPR protein with target RNA. Fully designed PPR proteins have been created [23, 30] and with optimization may also yield a structure displaying modular RNA base recognition.

The overall conformations of PPR proteins appear to be more adaptable than the fixed conformations of PUF proteins. Crystal structures of PPR10 with and without RNA are both superhelical, but the relative orientation of N- and C-terminal repeats changes in the two structures (Figure 2b) [28]. When the C-terminal repeats are aligned as in Figure 2b, the difference at the N-terminal repeats is apparent. This change in overall conformation is

likely influenced by the intertwined dimer in the PPR10:RNA crystals, but artificial PPR proteins with different sequence specificities also displayed overall conformational changes [23]. In contrast, the overall curvature of a particular PUF protein is characteristic and unchanging. Flatter RNA-binding surfaces correlate with longer RNA recognition sequences, suggesting that the repeat-to-repeat conformation influences RNA recognition specificity [37, 38, 40, 43].

Conclusions

PUF and PPR proteins show promise as engineered RNA-binding modules and their use will increase, as their RNA recognition properties are further de-coded. Despite the sense that the PUF recognition code was solved, Campbell, *et al.* [33] demonstrated that the code could be expanded. PPR proteins, natural and artificial, would benefit from determination of their *in vitro* specificities using newer selection methods that employ next generation sequencing, such as SEQRS [34], RNA Bind-N-Seq [47], and RNAcompete [48, 49]. This information would aid in de-coding specificity of the degenerate motifs and expand the PPR code for engineered proteins.

Acknowledgments

I am grateful for the input of my colleagues at NIEHS, G. Mueller and T. Teramoto. This work was supported by the Intramural Research Program of the National Institutes of Health, National Institute of Environmental Health Sciences.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest

** of outstanding interest

1. Chen Y, Varani G. Engineering RNA-binding proteins for biology. *FEBS J.* 2013; 280:3734–3754. [PubMed: 23742071]
2. Wei H, Wang Z. Engineering RNA-binding proteins with diverse activities. *Wiley Interdiscip Rev RNA.* 2015
3. Filipovska A, Rackham O. Modular recognition of nucleic acids by PUF, TALE and PPR proteins. *Mol Biosyst.* 2012; 8:699–708. [PubMed: 22234420]
4. Lu G, Dolgner SJ, Hall TMT. Understanding and engineering RNA sequence specificity of PUF proteins. *Curr Opin Struct Biol.* 2009; 19:110–115. [PubMed: 19186050]
5. Wang X, Zamore PD, Hall TMT. Crystal structure of a Pumilio homology domain. *Mol Cell.* 2001; 7:855–865. [PubMed: 11336708]
6. 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]
7. Wang X, McLachlan J, Zamore PD, Hall TM. Modular recognition of RNA by a human pumilio-homology domain. *Cell.* 2002; 110:501–512. [PubMed: 12202039]
8. Dong S, Wang Y, Cassidy-Amstutz C, Lu G, Bigler R, Jezyk MR, Li C, Hall TM, Wang Z. Specific and modular binding code for cytosine recognition in Pumilio/FBF (PUF) RNA-binding domains. *J Biol Chem.* 2011; 286:26732–26742. [PubMed: 21653694]
9. Filipovska A, Razif MF, Nygard KK, Rackham O. A universal code for RNA recognition by PUF proteins. *Nat Chem Biol.* 2011; 7:425–427. [PubMed: 21572425]

10. Cheong CG, Hall TM. Engineering RNA sequence specificity of Pumilio repeats. *Proc Natl Acad Sci U S A*. 2006; 103:13635–13639. [PubMed: 16954190]
11. Koh YY, Opperman L, Stumpf C, Mandan A, Keles S, Wickens M. A single C. elegans PUF protein binds RNA in multiple modes. *RNA*. 2009; 15:1090–1099. [PubMed: 19369425]
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. Choudhury R, Tsai YS, Dominguez D, Wang Y, Wang Z. Engineering RNA endonucleases with customized sequence specificities. *Nat Commun*. 2012; 3:1147. [PubMed: 23093184]
14. Cooke A, Prigge A, Opperman L, Wickens M. Targeted translational regulation using the PUF protein family scaffold. *Proc Natl Acad Sci U S A*. 2011; 108:15870–15875. [PubMed: 21911377]
15. 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]
16. Tilsner J, Linnik O, Christensen NM, Bell K, Roberts IM, Lacomme C, Oparka KJ. Live-cell imaging of viral RNA genomes using a Pumilio-based reporter. *Plant J*. 2009; 57:758–770. [PubMed: 18980643]
17. Wang Y, Cheong CG, Hall TM, Wang Z. Engineering splicing factors with designed specificities. *Nat Methods*. 2009; 6:825–830. [PubMed: 19801992]
18. Barkan A, Small I. Pentatricopeptide repeat proteins in plants. *Annu Rev Plant Biol*. 2014; 65:415–442. [PubMed: 24471833]
19. Lurin C, Andres C, Aubourg S, Bellaoui M, Bitton F, Bruyere C, Caboche M, Debast C, Gualberto J, Hoffmann B, et al. Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell*. 2004; 16:2089–2103. [PubMed: 15269332]
20. Barkan A, Rojas M, Fujii S, Yap A, Chong YS, Bond CS, Small I. A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet*. 2012; 8:e1002910. [PubMed: 22916040] **This manuscript describes derivation of an RNA recognition code for PPR proteins by bioinformatic analysis of three PPR proteins with known RNA targets. The authors demonstrate that this consensus code can be used to design specificity.
21. Abbas YM, Pichlmair A, Gorna MW, Superti-Furga G, Nagar B. Structural basis for viral 5'-PPP-RNA recognition by human IFIT proteins. *Nature*. 2013; 494:60–64. [PubMed: 23334420]
22. Ban T, Ke J, Chen R, Gu X, Tan MH, Zhou XE, Kang Y, Melcher K, Zhu JK, Xu HE. Structure of a PLS-class pentatricopeptide repeat protein provides insights into mechanism of RNA recognition. *J Biol Chem*. 2013; 288:31540–31548. [PubMed: 24047899]
23. Coquille S, Filipovska A, Chia T, Rajappa L, Lingford JP, Razif MF, Thore S, Rackham O. An artificial PPR scaffold for programmable RNA recognition. *Nat Commun*. 2014; 5:5729. [PubMed: 25517350] **In this manuscript, the authors create consensus repeats and use them to assemble designed PPR proteins that recognize poly-ribonucleotides.
24. Howard MJ, Lim WH, Fierke CA, Koutmos M. Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5' processing. *Proc Natl Acad Sci U S A*. 2012; 109:16149–16154. [PubMed: 22991464]
25. Katibah GE, Lee HJ, Huizar JP, Vogan JM, Alber T, Collins K. tRNA binding, structure, and localization of the human interferon-induced protein IFIT5. *Mol Cell*. 2013; 49:743–750. [PubMed: 23317505]
26. Ke J, Chen RZ, Ban T, Zhou XE, Gu X, Tan MH, Chen C, Kang Y, Brunzelle JS, Zhu JK, et al. Structural basis for RNA recognition by a dimeric PPR-protein complex. *Nat Struct Mol Biol*. 2013; 20:1377–1382. [PubMed: 24186060] *THA8 protein with 5 PPR motifs crystallizes as an RNA-induced dimer.

27. Ringel R, Sologub M, Morozov YI, Litonin D, Cramer P, Temiakov D. Structure of human mitochondrial RNA polymerase. *Nature*. 2011; 478:269–273. [PubMed: 21947009]
28. Yin P, Li Q, Yan C, Liu Y, Liu J, Yu F, Wang Z, Long J, He J, Wang HW, et al. Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature*. 2013; 504:168–171. [PubMed: 24162847] *Crystal structures of maize PPR10 protein in the presence and absence of RNA reveal an extended superhelix of 19 PPR motifs that changes conformation in the presence of RNA. Some details of PPR motif RNA recognition are revealed, but the overall structure does not appear to be the predicted biological assembly.
29. Kindgren P, Yap A, Bond CS, Small I. Predictable alteration of sequence recognition by RNA editing factors from *Arabidopsis*. *Plant Cell*. 2015; 27:403–416. [PubMed: 25649437]
30. Shen C, Wang X, Liu Y, Li Q, Yang Z, Yan N, Zou T, Yin P. Specific RNA recognition by designer pentatricopeptide repeat protein. *Mol Plant*. 2015; 8:667–670. [PubMed: 25598142] *Consensus PPR motifs are used to engineer proteins with distinct specificity. The authors show that they can design PPR proteins to selectively bind (UA)₅ vs (UAC)₃U sequences.
31. Yagi Y, Hayashi S, Kobayashi K, Hirayama T, Nakamura T. Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS One*. 2013; 8:e57286. [PubMed: 23472078]
32. Takenaka M, Zehrmann A, Brennicke A, Graichen K. Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. *PLoS One*. 2013; 8:e65343. [PubMed: 23762347]
33. Campbell ZT, Valley CT, Wickens M. A protein-RNA specificity code enables targeted activation of an endogenous human transcript. *Nat Struct Mol Biol*. 2014; 21:732–738. [PubMed: 24997599] **Diverse RNA recognition motifs in fungal PUF proteins are analyzed to establish an expanded code for RNA recognition.
34. Campbell ZT, Bhimsaria D, Valley CT, Rodriguez-Martinez JA, Menichelli E, Williamson JR, Ansari AZ, Wickens M. Cooperativity in RNA-protein interactions: global analysis of RNA binding specificity. *Cell Rep*. 2012; 1:570–581. [PubMed: 22708079]
35. Koh YY, Wang Y, Qiu C, Opperman L, Gross L, Tanaka Hall TM, Wickens M. Stacking interactions in PUF-RNA complexes. *RNA*. 2011; 17:718–727. [PubMed: 21372189]
36. 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]
37. Miller MT, Higgin JJ, Hall TM. Basis of altered RNA-binding specificity by PUF proteins revealed by crystal structures of yeast Puf4p. *Nat Struct Mol Biol*. 2008; 15:397–402. [PubMed: 18327269]
38. Qiu C, Kershner A, Wang Y, Holley CP, Wilinski D, Keles S, Kimble J, Wickens M, Hall TM. Divergence of Pumilio/fem-3 mRNA binding factor (PUF) protein specificity through variations in an RNA-binding pocket. *J Biol Chem*. 2012; 287:6949–6957. [PubMed: 22205700]
39. Valley CT, Porter DF, Qiu C, Campbell ZT, Hall TM, Wickens M. Patterns and plasticity in RNA-protein interactions enable recruitment of multiple proteins through a single site. *Proceedings of the National Academy of Sciences of the United States of America*. 2012; 109:6054–6059. [PubMed: 22467831]
40. Wang Y, Opperman L, Wickens M, Hall TM. Structural basis for specific recognition of multiple mRNA targets by a PUF regulatory protein. *Proc Natl Acad Sci U S A*. 2009; 106:20186–20191. [PubMed: 19901328]
41. Zhu D, Stumpf CR, Krahn JM, Wickens M, Hall TM. A 5' cytosine binding pocket in Puf3p specifies regulation of mitochondrial mRNAs. *Proc Natl Acad Sci U S A*. 2009; 106:20192–20197. [PubMed: 19918084]

42. 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]
43. Wilinski D, Qiu C, Lapointe CP, Nevil M, Campbell ZT, Hall TMT, Wickens M. RNA regulatory networks diversified through curvature of the PUF protein scaffold. *Nat Commun*. 2015; 6:8213. [PubMed: 26364903] **Yeast Puf5 protein is shown to recognize multiple RNA sequence lengths and regulate different biological processes depending on the target RNA length. RNA conformations adapt to a fixed and flattened Puf5 RNA-binding surface.
44. Gupta YK, Nair DT, Wharton RP, Aggarwal AK. Structures of human Pumilio with noncognate RNAs reveal molecular mechanisms for binding promiscuity. *Structure*. 2008; 16:549–557. [PubMed: 18328718]
45. Gully BS, Cowieson N, Stanley WA, Shearston K, Small ID, Barkan A, Bond CS. The solution structure of the pentatricopeptide repeat protein PPR10 upon binding atpH RNA. *Nucleic Acids Res*. 2015; 43:1918–1926. [PubMed: 25609698] *The authors demonstrate that PPR10 forms the expected monomeric 1:1 protein:RNA complex in solution.
46. Li Q, Yan C, Xu H, Wang Z, Long J, Li W, Wu J, Yin P, Yan N. Examination of the dimerization states of the single-stranded RNA recognition protein pentatricopeptide repeat 10 (PPR10). *J Biol Chem*. 2014; 289:31503–31512. [PubMed: 25231995] *The authors examine the effects of PPR10 protein constructs and mutations on dimerization state.
47. Lambert N, Robertson A, Jangi M, McGeary S, Sharp PA, Burge CB. RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. *Mol Cell*. 2014; 54:887–900. [PubMed: 24837674]
48. Ray D, Kazan H, Chan ET, Pena Castillo L, Chaudhry S, Talukder S, Blencowe BJ, Morris Q, Hughes TR. Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nat Biotechnol*. 2009; 27:667–670. [PubMed: 19561594]
49. Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, Gueroussov S, Albu M, Zheng H, Yang A, et al. A compendium of RNA-binding motifs for decoding gene regulation. *Nature*. 2013; 499:172–177. [PubMed: 23846655]

Highlights

- PUF and PPR proteins are exemplary modular α -helical repeat RNA-binding proteins.
- A highly conserved PUF protein RNA recognition code is expanded in fungal proteins.
- Yeast Puf5 protein binds 8–12 base RNA sequences using multiple recognition modes.
- PPR protein bioinformatic analysis and structures allow re-coding RNA recognition.

Figure 1.

RNA recognition by PUF and PPR proteins. **(a)** PUF RNA recognition code. Side chains at positions 1 and 5 that interact with the edges of RNA bases are indicated. The PUF recognition motif in α helix 2 is shown where ϕ is a stacking side chain. **(b)** Recognition of U (top) and G (bottom) bases by PUM repeats. **(c)** PPR RNA recognition code. Side chains at positions 6 and 1' that interact with the edges of RNA bases are indicated. The PPR recognition motif is shown. **(d)** Recognition of U (top) and G (bottom) bases by PPR

repeats. U interaction is from PPR10 repeat 4, and G interaction is from THA8 repeat 4. Panels (c) and (d) and Figure 2 were prepared with PyMOL (Schrödinger, LLC).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 2.

Overall conformational changes of PUF and PPR proteins. **(a)** PUF proteins adopt fixed conformations. Superposition of four crystal structures of Puf5 in complex with 9-, 10-, 11-, and 12-base RNAs shows unchanging protein conformation. PUM repeats are colored alternately blue and red, and helices are shown as cylinders. RNAs, which adapt conformation to the fixed Puf5 scaffold, are shown as cartoons with the conserved recognition motifs labeled. **(b)** PPR protein conformation is malleable. Superposition of crystal structures of PPR10 with (blue) and without (grey) RNA illustrates its pliability. Structures are aligned over residues 384–703 in repeats 8–16. For clarity, only the first α helices in each repeat are shown as cylinders and numbered. A fragment of RNA bound to repeats 15–18 is shown in yellow.

Table 1

Expanded PUF protein RNA recognition code. RNA recognition motifs in the Pumilio1 recognition code (Figure 1a, boldface) and those identified by Campbell, et al. with highest overall specificity for G, U, and A bases at the target base (+2) and flanking bases (+1 U and +3 U) are shown. Motifs are shown with notation defined by Campbell, *et al.*: Edge-interacting side chains at positions 1 and 5 are listed first, followed by the base-stacking side chain at position 2. Motifs that recognize C in Pumilio1 are indicated (grey shading). Overall specificity is rated +++ for specificity coefficients ≥ 2.5 , ++ for specificity coefficients ≥ 1.5 , and + for <1.5 (see Campbell, *et al.* Supplementary Figure 1). Individual base specificities are rated +++ for enrichment $\geq 80\%$, ++ for enrichment $\geq 60\%$, and + for enrichment $<60\%$ (see Campbell, *et al.* Supplementary Figure 4). Motifs with high overall specificity (+++ and ++) are highlighted yellow (G), red (U), and green (A).

Base	Motif	Overall Specificity	+1 U Specificity	+2 Specificity	+3 U Specificity
G	S₁E₅-H₂ (WT)	+++	+++	+++	+++
	S₁E₅-N₂	++	+++	+++	+++
	S ₁ R ₅ -Y ₂ (C)	+	+	+++	+++
U	N₁Q₅-T₂	++	++	+++	+++
	N₁Q₅-H₂	+++	++	+++	+++
	N₁Q₅-F₂	++	++	+++	+++
	N₁Q₅-Y₂	++	+	+++	+++
	A ₁ R ₅ -Y ₂ (C)	+	+	+++	+++
	C ₁ R ₅ -Y ₂ (C)	+	+	+++	+++
A	C₁Q₅-F₂	++	++	+++	+++
	S₁Q₅-R₂	+	+	+	+
	C₁Q₅-R₂	+	+	++	+