nature chemical biology

BRIEF COMMUNICATION

PUBLISHED ONLINE: 15 MAY 2011 | DOI: 10.1038/NCHEMBIO.577

A universal code for RNA recognition by PUF proteins

Aleksandra Filipovska, Muhammad F M Razif, Karoline K A Nygård & Oliver Rackham*

The design of proteins that can bind any RNA sequence of interest has many potential biological and medical applications. Here we have expanded the recognition of Pumilio and FBF homology protein (PUF) repeats beyond adenine, guanine and uracil and evolved them to specifically bind cytosine. These repeat sequences can be used to create PUF domains capable of selectively binding RNA targets of diverse sequence and structure.

RNA–protein complexes have essential roles in the regulation of gene expression and are also vital for cell differentiation and the complex developmental programs of multicellular eukaryotes[1,2]. Proteins bind RNAs in diverse modes that are often difficult to predict[3], limiting the ability to engineer RNA-binding proteins for biotechnological and medical applications[4]. The potential of such tools is analogous to the use of short interfering RNAs and microRNAs; however, the use of short RNA duplexes to target RNAs is limited to lowering their abundance or expression in the cytoplasm, and this depends on RNA interference pathways[5]. Engineering RNA-binding proteins is attractive because they could be fused to any desired effector domain, enabling selective binding of a specific RNA target to investigate or manipulate any aspect of its metabolism.

PUF proteins are found in most eukaryotes and are typically involved in regulating embryogenesis, development and differentiation[6]. PUF proteins contain a domain that is generally composed of eight 36-amino-acid repeats; each repeat binds a single nucleotide in its RNA target (**Fig. 1a**,**b**)[7–10]. Amino acids at positions 12 and 16 of the PUF repeat bind each RNA base via hydrogen bonding or van der Waals contacts with the Watson-Crick edge, whereas the amino acid at position 13 makes a stacking interaction (**Fig. 1c**). The recognition of RNA by naturally occurring PUF domains is base specific[7], such that cysteine and glutamine bind adenine, asparagine and glutamine bind uracil, and serine and glutamate bind guanine (**Fig. 1c**,**d**). This code has been confirmed by studies showing that the specificity of individual repeats can be switched by mutating only the amino acids that make contacts with the Watson-Crick edge of the base[11], allowing engineered PUF domains to recognize endogenous RNAs composed of adenine, guanine or uracil[7,11–14]. The use of PUF domains as tools has been hampered because naturally occurring residues that recognize cytosine have not been found. This limits the potential target sites for engineered PUFs, because even in RNAs encoded by guanine- and cytosine-poor genomes, the majority of octamer sequences contain at least one cytosine. If a small RNA or defined region of a larger

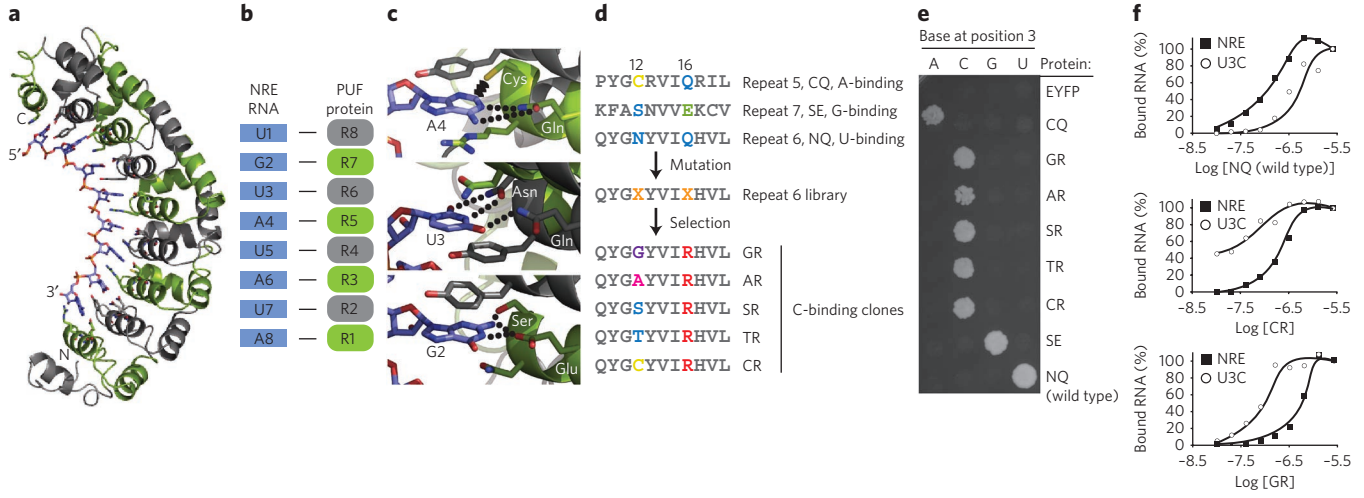


Figure 1 | Selection and characterization of PUF repeats that can specifically recognize cytosine. (**a**) Crystal structure of the human PUM1 PUF domain bound to RNA (PDB 1M8Y). (**b**) Schematic of the recognition of RNA bases in NRE RNA by the PUF repeats of PUM1. (**c**) Recognition of adenine (top), uracil (middle) and guanine (bottom) by PUF repeats 5–7 in the crystal structure of PUM1, respectively. (**d**) Sequences of RNA-binding regions of PUM1 repeats. Key hydrogen-bonding residues at positions 12 and 16 were randomized, and combinations that could recognize cytosine were selected from the library using the yeast three-hybrid system. (**e**) Specificity of the selected clones was determined by survival on medium lacking histidine and containing 0.5 mM 3-aminotriazole (a His3p competitive inhibitor). We used enhanced yellow fluorescent protein (EYFP) as a negative control. (**f**) Wild-type (NQ) and mutant (CR and GR) PUF proteins were tested against RNA probes containing uracil (NRE) or cytosine (U3C) in RNA EMSAs. Graphs depict the percentage of RNA bound by varying concentrations of each protein (gels are in **Supplementary Fig. 2**).

Western Australian Institute for Medical Research and Centre for Medical Research, The University of Western Australia, Perth, Australia.
*e-mail: rackham@waimr.uwa.edu.au

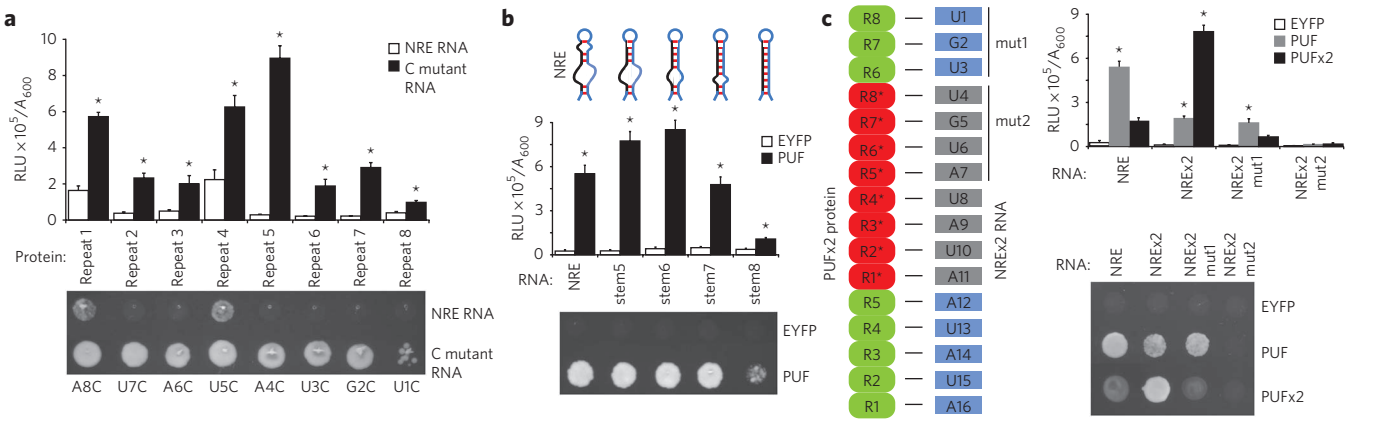


Figure 2 | General applicability of PUF domains in RNA recognition. (**a**) Engineered PUF repeats can selectively bind cytosine at all eight positions within the RNA target. Amino acids at positions 12 and 16 of individual repeats were mutated to glycine and arginine and combined with either the wild-type (NRE) RNA or a mutant NRE in which the corresponding base was changed to cytosine. (**b**) The PUF domain can bind to RNA targets located within substantially double-stranded structures. Number of bases from the PUF recognition site paired within a stem structure was increased stepwise from five (stem5) to eight (stem8). (**c**) A PUF domain comprising 16 RNA-binding repeats provides additional binding specificity and selectivity. Schematic structure of the engineered 16-repeat PUF and its cognate RNA target. Survival on medium with 0.5 mM 3-aminotriazole and lacking histidine and β-galactosidase assays were used to determine the interaction of PUF domains and their RNA targets (RLU, relative light units). The NREx2 mut1 RNA has the UGU triplet of the newly added target region mutated to CCC, and the NREx2 mut2 RNA has the UGU triplet of the native NRE mutated to CCC. Data are mean ± s.e.m. from six independent experiments. *$P < 0.01$ of cognate to noncognate RNA–protein complexes by a two-tailed Student's *t* test.

RNA were the target of interest, it could be impossible to engineer a PUF protein to bind these RNAs. To overcome this limitation, we used directed evolution to select for PUF repeat variants that can specifically recognize cytosine.

We used the yeast three-hybrid system[15] to link the interaction between PUF domains and RNA to a life-death selection in *Saccharomyces cerevisiae*. In this system, a fusion between the lexA DNA-binding domain and the MS2 coat protein effectively tethers a hybrid RNA that contains MS2 recognition sites and a PUF RNA target of interest upstream of *his3* and *lacZ* reporter genes in the yeast genome. The PUF domain is expressed as a fusion to the Gal4p transcription activator domain, so that if a PUF-RNA complex is formed, transcription of the *his3* and *lacZ* reporter genes is activated to allow survival on medium lacking histidine. In addition, the RNA-protein interaction can be quantified by measuring the activity of β-galactosidase expressed from the *lacZ* reporter gene[16]. We investigated repeat 6 of the human PUF protein, PUM1, which binds uracil in a crystal structure with a high-affinity RNA target (the nanos response element, NRE; **Fig. 1c** and **Supplementary Methods**). When the amino acid at position 12 was mutated to cysteine to resemble the adenine-binding repeat 5 of PUM1 (CQ, **Fig. 1d**) cells survived on selective medium and produced significantly higher β-galactosidase activity only when the target RNA had an adenine at the position in the RNA bound by repeat 6 (**Fig. 1e**, **Supplementary Results** and **Supplementary Fig. 1**). Furthermore, when we transplanted the guanine-recognizing amino acids from repeat 7 (SE) into repeat 6, the *his3* and *lacZ* reporters were activated only when the target RNA had a guanine in the RNA (**Fig. 1e** and **Supplementary Fig. 1**), indicating that this system provides sufficient sensitivity to enable the engineering of individual PUF repeats.

We synthesized a library based on the PUM1 PUF domain, in which positions 12 and 16 of repeat 6 were randomized to encode all possible amino acids, and combined them with an RNA target in which the corresponding base was changed to a cytosine. We selected colonies that survived on medium lacking histidine and identified five unique PUF mutants (**Fig. 1d**) that selectively interacted with RNAs containing cytosine but not those containing adenine, guanine or uracil (**Fig. 1e**). All five variants had an arginine at position 16 and an amino acid with a small or nucleophilic side chain at position 12 (glycine, alanine, serine, threonine or cysteine). In addition, we found that β-galactosidase was activated only when the PUF mutant proteins specifically interacted with target RNA containing cytosine (**Supplementary Fig. 1**). We investigated the specificity of the interactions between wild-type and mutant PUF proteins *in vitro* using RNA electrophoretic mobility shift assays (EMSAs). We overexpressed the wild-type PUF and two cytosine-recognizing PUF mutants (GR, with Gly12 and Arg16, and CR, with Cys12 and Arg16) in *Escherichia coli* and purified them to homogeneity. RNA EMSAs showed a marked specificity shift between the mutant PUF domains and their cytosine-containing target RNA and the wild-type PUF and its cognate NRE target (**Fig. 1f**). The GR mutant PUF bound its cognate cytosine-containing RNA with an affinity very similar to that of the wild-type PUF and its cognate NRE RNA, whereas the CR mutant bound its cognate cytosine-containing RNA with higher affinity. This confirms earlier observations that engineered PUF proteins do not always bind with the same affinities as wild-type proteins[11], indicating that future applications of these proteins may depend not only on binding preference but also on the affinity for their target RNAs.

To determine the general applicability of the selected sequences to the design of PUF domains with predictable binding properties, we made a set of eight *PUM1* mutants in which each repeat was sequentially modified to have a glycine at position 12 and an arginine at position 16 (GR). We found that all of these engineered PUF domains bound RNA targets with cytosine at the position in their RNA target corresponding to the mutated repeat with higher affinity than the wild-type, non–cytosine-containing RNA target (**Fig. 2a**). The magnitude of the specificity shift varied from repeat to repeat, confirming reports that not all repeat-base interactions contribute equally to the binding energy of the RNA–protein complex[10,11]; however, all mutants preferentially bound cytosine-containing RNAs.

The binding mode of PUF domains observed in different crystal structures[7,17–19] indicates that their RNA targets are exclusively single stranded in the RNA–protein complexes; however, whether their RNA targets must be single stranded before PUF domain binding is not known. We generated a series of RNA variants in which

the sequence downstream of the NRE was modified sequentially to place the NRE in increasingly base-paired structures. We found that the PUF protein could bind all of the RNA targets, including one in which every base was paired in a stem structure, albeit less efficiently (**Fig. 2b**). This indicates that PUF proteins can invade structured RNAs to bind their target sequences, presumably during the dynamic rearrangements intrinsic to RNA structures. This is relevant not only to the rational engineering of PUF domains but also to naturally occurring PUF domain proteins. For instance, it has recently been recognized that human PUM1 can bind to a 5′-UGUAUAUA-3′ sequence within a naturally occurring structured region in the 3′ untranslated region of p27 mRNA, increasing the accessibility of a miRNA target site[20]. Although this further suggests that naturally occurring PUF proteins can bind RNA sequences within complex structures, the application of these proteins is limited without the ability to bind cytosine.

Naturally occurring PUF proteins typically contain eight RNA-binding repeats. Although this is sufficient for them to selectively regulate specific developmental processes, they often do this by binding multiple different RNAs[21]. For many applications in biotechnology, synthetic biology and medicine, it would be useful to be able to target only one species of RNA within an entire transcriptome. To achieve such levels of sequence discrimination, we engineered PUFs with 16 RNA-binding repeats. We inserted sequences encoding only the RNA-binding PUF repeats, without flanking regions, from the human PUM1 cDNA between repeats 5 and 6 of a synthetic gene that encodes the same protein sequence as the PUM1 cDNA but is only 78% similar at the DNA level, to avoid potential instability of the recombinant DNA. Because the *Caenorhabditis elegans* FBF-1 and FBF-2 PUF proteins contain a short insertion close to the end of repeat 5, we reasoned that this region might tolerate the insertion of extra PUF repeats. The extended PUF bound its cognate extended RNA target in yeast and activated transcription of the β-galactosidase reporter more efficiently than the eight-repeat PUF with its cognate RNA (**Fig. 2c**). The inserted and flanking PUF repeats contributed to the binding affinity and selectivity, as separate mutation of the UGU triplets recognized by both sets of repeats significantly decreased β-galactosidase activity and growth on selective medium (**Fig. 2c**). Engineered PUF domain proteins containing 16 RNA-binding repeats provide the means to selectively bind RNAs in higher eukaryotes that have more complex transcriptomes.

Here we have expanded the base-recognition scope of PUF proteins by engineering them to specifically recognize cytosine, and we have rationally designed a PUF protein containing 16 RNA-binding repeats, thereby providing evidence that PUF proteins can be designed to bind any RNA of interest. The identification of cytosine-binding residues could facilitate computational approaches for the discovery of naturally occurring PUF proteins that bind RNA targets containing cytosine and the prediction of the target RNAs of uncharacterized PUF proteins. In addition, the availability of a code that enables the design of proteins with predictable RNA targets may provide tools to manipulate any endogenous gene at the post-transcriptional level, with many potential applications in biotechnology and medicine. Because the control of gene expression is rapid and more precise at the post-transcriptional level[22,23], engineered PUF proteins could provide unique opportunities for fine-tuning the expression of endogenous genes or transgenes. Furthermore, some aspects of gene expression can be controlled only at the RNA level, for instance the nuclear retention or cytoplasmic localization of mRNAs[2]. In synthetic biology, the availability of such designer RNA-binding proteins may provide a new tool kit for creating synthetic networks that are controlled at the RNA level and for better understanding the complex patterns of gene expression in living cells[24,25].

Received 2 December 2010; accepted 15 March 2011; published online 15 May 2011

References

1. Glisovic, T., Bachorik, J.L., Yong, J. & Dreyfuss, G. *FEBS Lett.* **582**, 1977–1986 (2008).
2. St. Johnston, D. *Nat. Rev. Mol. Cell Biol.* **6**, 363–375 (2005).
3. Auweter, S.D., Oberstrass, F.C. & Allain, F.H. *Nucleic Acids Res.* **34**, 4943–4959 (2006).
4. Lunde, B.M., Moore, C. & Varani, G. *Nat. Rev. Mol. Cell Biol.* **8**, 479–490 (2007).
5. Carthew, R.W. & Sontheimer, E.J. *Cell* **136**, 642–655 (2009).
6. Quenault, T., Lithgow, T. & Traven, A. *Trends Cell Biol.* **21**, 104–112 (2011).
7. Wang, X., McLachlan, J., Zamore, P.D. & Hall, T.M. *Cell* **110**, 501–512 (2002).
8. Wang, X., Zamore, P.D. & Hall, T.M. *Mol. Cell* **7**, 855–865 (2001).
9. Edwards, T.A., Pyle, S.E., Wharton, R.P. & Aggarwal, A.K. *Cell* **105**, 281–289 (2001).
10. Zamore, P.D., Williamson, J.R. & Lehmann, R. *RNA* **3**, 1421–1433 (1997).
11. Cheong, C.G. & Hall, T.M. *Proc. Natl. Acad. Sci. USA* **103**, 13635–13639 (2006).
12. Wang, Y., Cheong, C.G., Hall, T.M. & Wang, Z. *Nat. Methods* **6**, 825–830 (2009).
13. Tilsner, J. *et al. Plant J.* **57**, 758–770 (2009).
14. Ozawa, T., Natori, Y., Sato, M. & Umezawa, Y. *Nat. Methods* **4**, 413–419 (2007).
15. SenGupta, D.J. *et al. Proc. Natl. Acad. Sci. USA* **93**, 8496–8501 (1996).
16. Hook, B., Bernstein, D., Zhang, B. & Wickens, M. *RNA* **11**, 227–233 (2005).
17. Wang, Y., Opperman, L., Wickens, M. & Hall, T.M. *Proc. Natl. Acad. Sci. USA* **106**, 20186–20191 (2009).
18. Gupta, Y.K., Nair, D.T., Wharton, R.P. & Aggarwal, A.K. *Structure* **16**, 549–557 (2008).
19. Zhu, D., Stumpf, C.R., Krahn, J.M., Wickens, M. & Hall, T.M. *Proc. Natl. Acad. Sci. USA* **106**, 20192–20197 (2009).
20. Kedde, M. *et al. Nat. Cell Biol.* **12**, 1014–1020 (2010).
21. Gerber, A.P., Herschlag, D. & Brown, P.O. *PLoS Biol.* **2**, E79 (2004).
22. Isaacs, F.J. *et al. Nat. Biotechnol.* **22**, 841–847 (2004).
23. Zenklusen, D., Larson, D.R. & Singer, R.H. *Nat. Struct. Mol. Biol.* **15**, 1263–1271 (2008).
24. Filipovska, A. & Rackham, O. *ACS Chem. Biol.* **3**, 51–63 (2008).
25. Isaacs, F.J., Dwyer, D.J. & Collins, J.J. *Nat. Biotechnol.* **24**, 545–554 (2006).

Acknowledgments

We thank M. Wickens (University of Wisconsin–Madison) for the YBZ1 yeast strain, T. Ito (University of Tokyo) for the pGAD-RC plasmid, R. King for advice on oligonucleotide synthesis with trimer phosphoramidites and K.L. Hamilton for technical assistance. This work was supported by fellowships and grants from the Australian Research Council and the National Health and Medical Research Council.

Author contributions

A.F., M.F.M.R., K.K.A.N. and O.R. conducted experiments; A.F. and O.R. designed experiments, interpreted data and prepared the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available online at http://www.nature.com/naturechemicalbiology/. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Correspondence and requests for materials should be addressed to O.R.

SUPPLEMENTARY INFORMATION FOR

A universal code for RNA recognition by PUF proteins

Aleksandra Filipovska, Muhammad FM Razif, Karoline KA Nygård & Oliver Rackham

Western Australian Institute for Medical Research and Centre for Medical Research, University of Western Australia, Australia. Correspondence should be addressed to O.R. (rackham@waimr.uwa.edu.au)

This PDF file includes:

Supplementary Methods

Supplementary Results
(Supplementary Figure 1 and Supplementary Figure 2)

Supplementary References

Nature Chemical Biology: doi:10.1038/nchembio.577

Supplementary Methods

Plasmids. To produce a Gal4p activation domain fused to a PUF domain, a synthetic gene encoding amino acids 828 to 1176 of the human PUM1 protein (GenBank accession no. NP_001018494, GENEART) was subcloned into pJC72[1]. This plasmid was used as a template for library construction by enzymatic inverse PCR[1] using primers where the codons corresponding to amino acids 1043 and 1047 were encoded by mixtures of trimer phosphoramidites encoding all 20 amino acids (GeneWorks). This library of mutant PUF domains was subcloned into the yeast expression plasmid pGAD-RC[2]. Individual PUF domain mutants were also made by enzymatic inverse PCR. To make a 16 repeat Puf protein (PUFx2) repeats 1-8 of the human PUM1 cDNA were amplified using primers that incorporated flanking SacI sites, digested with SacI and cloned into an engineered SacI site that encodes amino acids 1030 and 1031 of the synthetic gene encoding the PUM1 PUF domain. RNA expression plasmids were made by altering the multiple cloning site of pIIIA/MS2-2[3] according to Cassiday and Maher[4] and subcloning pairs of annealed oligonucleotides corresponding to the following RNA sequences (PUF recognition sequences in bold, site specific mutations underlined):

NRE: 5'-CCGGCUAGCAAU**UGUAUAUA**UUAAUUUAAUAAAGCAUG-3';
NREU1C: 5'-CCGGCUAGCAAU**C**<u>**C**</u>**GUAUAUA**UUAAUUUAAUAAAGCAUG-3';
NREG2C: 5'-CCGGCUAGCAAU**UC**<u>**C**</u>**UAUAUA**UUAAUUUAAUAAAGCAUG-3';
NREU3A: 5'-CCGGCUAGCAAU**UG**<u>**A**</u>**AUAUA**UUAAUUUAAUAAAGCAUG-3';
NREU3C: 5'-CCGGCUAGCAAU**UG**<u>**C**</u>**AUAUA**UUAAUUUAAUAAAGCAUG-3';
NREU3G: 5'-CCGGCUAGCAAU**UG**<u>**G**</u>**AUAUA**UUAAUUUAAUAAAGCAUG-3';
NREA4C: 5'-CCGGCUAGCAAU**UGU**<u>**C**</u>**UAUA**UUAAUUUAAUAAAGCAUG-3';
NREU5C: 5'-CCGGCUAGCAAU**UGUA**<u>**C**</u>**AUA**UUAAUUUAAUAAAGCAUG-3';
NREA6C: 5'-CCGGCUAGCAAU**UGUAU**<u>**C**</u>**UA**UUAAUUUAAUAAAGCAUG-3';
NREU7C: 5'-CCGGCUAGCAAU**UGUAUA**<u>**C**</u>**A**UUAAUUUAAUAAAGCAUG-3';
NREA8C: 5'-CCGGCUAGCAAU**UGUAUAU**<u>**C**</u>UUAAUUUAAUAAAGCAUG-3';
NREstem5: 5'-CCGGCUAGCAAU**UGUAUAUA**UUAAU<u>A</u>UAAUAAAGCAUG-3';
NREstem6: 5'-CCGGCUAGCAAU**UGUAUAUA**UUAAU<u>A</u>UA<u>U</u>UAAAGCAUG-3';

NREstem7: 5'-CCGGCUAGCAAU**UGUAUAUA**UUAAUAUAUAAAGCAUG-3';
NREstem8: 5'-CCGGCUAGCAAU**UGUAUAUA**UUAAUAUAUACAAGCAUG-3';
NREx2: 5'-
CCGGCUAGCAAU**UGUUGUAUAUAAUAUA**UUAAUUUAAUAAAGCAUG-3';
NREx2mut1: 5'-
CCGGCUAGCAAU**CCCUGUAUAUAAUAUA**UUAAUUUAAUAAAGCAUG-3';
NREx2mut2: 5'-
CCGGCUAGCAAU**UGUCCCCUAUAAUAUA**UUAAUUUAAUAAAGCAUG-3'.

PUF library selections. *Saccharomyces cerevisiae* YBZ1 cells (*MAT**a**, ura3-52, leu2-3, 112, his3-200, trp1-1, ade2, LYS2 :: (LexAop)-HIS3, ura3 :: (lexA-op)-lacZ, LexA-MS2 coat (N55K)*)[5] containing the NREU3C RNA expression plasmid were transformed with the PUF domain library in pGAD-RC using the LiAc method according to Gietz and Woods[6] yielding 6 x 10^5 primary transformants. Cells were amplified by overnight growth in SC media lacking leucine and uracil, washed in TE and 1 x 10^7 CFU were plated on SC agar lacking leucine, uracil and histidine, supplemented with 0.5 mM 3-amino triazole. Colonies were picked after three days and the plasmids were isolated, transformed into *Escherichia coli* DH10B, screened by PCR to identify the PUF encoding plasmid which was sequenced and transformed into YBZ1 to analyze the specificity of the mutant PUF domains, as described below.

Yeast three-hybrid growth assays. YBZ1 transformants containing PUF domain and RNA expression plasmids were grown overnight in SC media lacking leucine and uracil, washed in SC media without amino acids, diluted to OD600 of 0.1 and replica spotted onto SC media lacking leucine and uracil (to test for cell health and plasmid maintenance) and SC agar lacking leucine, uracil and histidine, supplemented with 0.5 mM 3-amino triazole (to test for RNA-protein interactions).

ß-galactosidase assays. YBZ1 transformants containing PUF domain and RNA expression plasmids were grown overnight in SC media lacking leucine and uracil, diluted to OD600 of 0.1 and mixed with an equal volume of Beta-Glo reagent (Promega),

incubated for 1 h at room temperature and luminescence was detected using a FLUOstar OPTIMA (BMB Labtech).

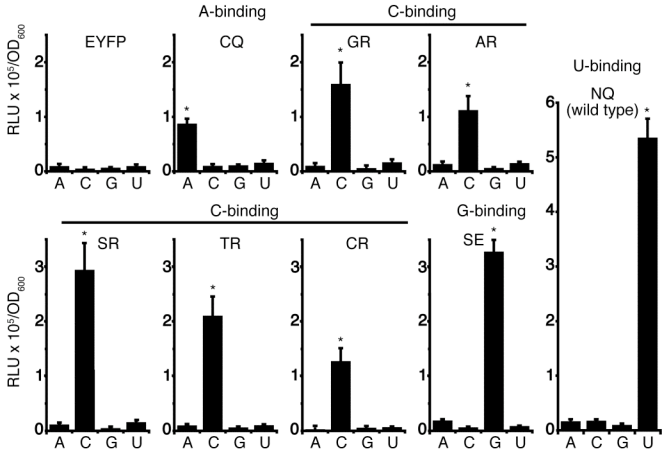
Purification of PUF proteins. PUF domains were subcloned into pTYB3 and expressed as a fusion to an intein and chitin-binding domain in *Escherichia coli* ER2566 cells (New England Biolabs). Cells were lyzed by sonication in 20 mM sodium phosphate (pH 8.0), 1 M NaCl, and 0.1 mM PMSF. Lysates were clarified by centrifugation and incubated for 40 min with chitin beads (New England Biolabs). Beads were washed twice with 20 mM sodium phosphate (pH 8.0), 1 M NaCl, and 0.1 mM PMSF, once with 20 mM sodium phosphate (pH 8.0), 0.5 M NaCl, and 0.1 mM PMSF, and once with 20 mM sodium phosphate (pH 8.0), 0.15 M NaCl, and 0.1 mM PMSF. DTT was added to the beads to 50 mM final concentration and the tube was purged with nitrogen gas before incubation at room temperature with gentle rocking for three days. Cleaved PUF domain protein, free from the intein and chitin-binding domain was collected, transferred into 10 mM Tris-HCl (pH 7.4), 150 mM NaCl, 5 mM ß-mercaptoethanol and further purified by an ÄKTA-Explorer system (GE) using a Superdex 200 10/300 column (GE) with a total bed volume of 120 ml. Pure fractions were pooled and concentrated using Microsep 10K Omega centrifugal devices (PALL). Protein concentration was determined by the bicichroninic acid (BCA) assay using bovine serum albumin (BSA) as a standard.

RNA electrophoretic mobility shift assays. Purified PUF domains were incubated at room temperature for 30 min with fluorescein labeled RNA oligonucleotides (Dharmacon) in 10 mM HEPES (pH 8.0), 1 mM EDTA, 50 mM KCl, 2 mM DTT, 0.1 mg/ml fatty acid-free BSA, and 0.02% Tween-20. The following RNA sequences were used:

NRE: 5'-(Fl)AUUGUAUAUA-3';
NREU3C: 5'-(Fl)AUUGCAUAUA-3'.

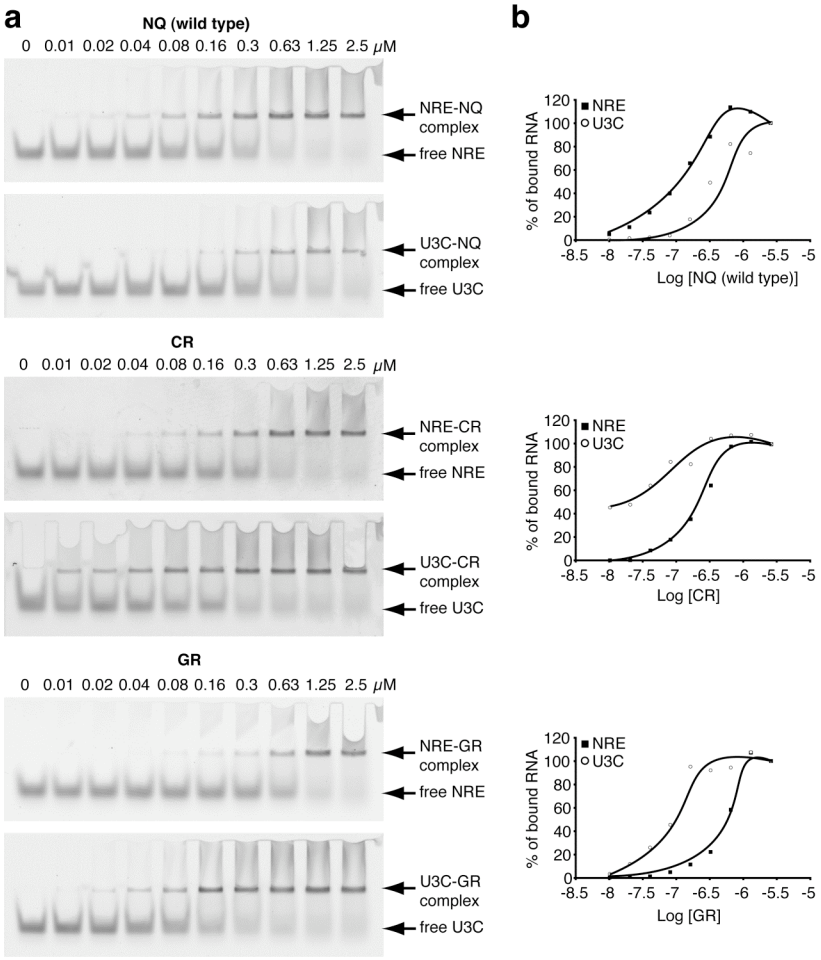
Reactions were analyzed by 10% PAGE in TAE and fluorescence was detected using a Typhoon TRIO scanner (GE).

Supplementary Results



Supplementary Figure 1 Characterization of PUF repeats that can specifically recognize cytosine. Specificity of the selected clones was quantified using ß-galactosidase assays to examine activation of a *lacZ* reporter gene. We used the enhanced yellow fluorescent protein (EYFP) as a control to show that the activation of transcription was dependent on specific RNA-protein interactions in our experiments. The fold increase in β-galactosidase activity for the selected mutants in the presence of cytosine-containing RNA was not as high as that for the wild type PUF and its cognate NRE RNA, however some of the cytosine-binding mutants in combination with their target RNAs generated activities similar to the guanine-binding mutant and higher than the adenine-binding mutant, in the presence of their cognate RNAs. Data are mean ± SEM from six independent experiments. *, $p < 0.01$ of cognate to non cognate RNA-protein complexes by a 2-tailed Student's t test.

Nature Chemical Biology: doi:10.1038/nchembio.577



Supplementary Figure 2 Specific recognition of cytosine *in vitro*. (**a**) Selected PUF domains are specific for cytosine containing RNAs, determined by RNA electrophoretic mobility shift assays. Wild type (NQ) and mutant (CR and GR) PUF proteins were tested against uracil (NRE) or cytosine (U3C) containing RNA probes. (**b**) The percentage of RNA bound by varying concentrations of each protein.

Nature Chemical Biology: doi:10.1038/nchembio.577

Supplementary References

1. Rackham, O. & Chin, J. W. A network of orthogonal ribosome•mRNA pairs. *Nat Chem Biol* **1**, 159-166 (2005).
2. Ito, T., Tashiro, K., Muta, S., Ozawa, R. et al. Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proc Natl Acad Sci U S A* **97**, 1143-7 (2000).
3. Stumpf, C. R., Opperman, L. & Wickens, M. Analysis of RNA-protein interactions using a yeast three-hybrid system. *Methods Enzymol* **449**, 295-315 (2008).
4. Cassiday, L. A. & Maher, L. J., 3rd. *In vivo* recognition of an RNA aptamer by its transcription factor target. *Biochemistry* **40**, 2433-8 (2001).
5. Hook, B., Bernstein, D., Zhang, B. & Wickens, M. RNA-protein interactions in the yeast three-hybrid system: Affinity, sensitivity, and enhanced library screening. *RNA* **11**, 227-33 (2005).
6. Gietz, R. D. & Woods, R. A. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol* **350**, 87-96 (2002).