Distinct RNA-binding modules in a single PUF protein cooperate to determine RNA specificity

Chen Qiu1, Robert C. Dutcher1, Douglas F. Porter2, Yoav Arava3, Marvin Wickens4,* and Traci M. Tanaka Hall1,*

1Epigenetics and Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709, USA, 2Program in Epithelial Biology, Stanford University School of Medicine, Stanford, CA 94305, USA, 3Department of Biology, Technion—Israel Institute of Technology, Haifa 32000, Israel and 4Department of Biochemistry, University of Wisconsin, Madison, WI 53706, USA

Received February 15, 2019; Revised June 19, 2019; Editorial Decision June 22, 2019; Accepted June 24, 2019

ABSTRACT

PUF proteins, named for Drosophila Pumilio (PUM) and Caenorhabditis elegans fem-3-binding factor (FBF), recognize specific sequences in the mRNAs they bind and control. RNA binding by classical PUF proteins is mediated by a characteristic PUM homology domain (PUM-HD). The Puf1 and Puf2 proteins possess a distinct architecture and comprise a highly conserved subfamily among fungal species. Puf1/Puf2 proteins contain two types of RNA-binding domain: a divergent PUM-HD and an RNA recognition motif (RRM). They recognize RNAs containing UAAU motifs, often in clusters. Here, we report a crystal structure of the PUM-HD of a fungal Puf1 in complex with a dual UAAU motif RNA. Each of the two UAAU tetranucleotides are bound by a Puf1 PUM-HD forming a 2:1 protein-to-RNA complex. We also determined crystal structures of the Puf1 RRM domain that identified a dimerization interface. The PUM-HD and RRM domains act in concert to determine RNA-binding specificity: the PUM-HD dictates binding to UAAU, and dimerization of the RRM domain favors binding to dual UAAU motifs rather than a single UAAU. Cooperative action of the RRM and PUM-HD identifies a new mechanism by which multiple RNA-binding modules in a single protein collaborate to create a unique RNA-binding specificity.

INTRODUCTION

RNA-binding proteins (RBPs) regulate every step in the life of an mRNA, from processing to translation and destruction. As a result, they are critical in determining the level of protein that a gene produces. Dysfunctional RBPs are linked with human neurological disorders and cancer (1–4). mRBPs bind mature mRNAs, often in their 3′-untranslated regions (3′-UTRs), to control stability, translation and localization of their mRNA targets (5). The PUF family proteins, named after founding members, Drosophila melanogaster Pumilio (PUM) and Caenorhabditis elegans fem-3 binding factor (FBF) (6,7), are exemplary mRBPs, and their roles in mRNA control have been studied extensively (8–11).

PUF proteins are evolutionarily conserved throughout eukaryotes. In metazoa, PUF proteins regulate mRNAs that encode proteins involved in embryonic development, stem cell control and neurogenesis, among other roles (12–14). In Saccharomyces cerevisiae, each PUF subfamily binds multiple mRNAs, often with functional and cytotoxic relatedness (15). Their roles are widespread, and include controls of mating type, cell wall integrity, mitochondrial functions, ion sensitivity, and other cellular events (15–21).

A single eukaryotic species can encode multiple PUF proteins that comprise as many as four subgroups: classical PUF proteins, Puf6 proteins, Nop9 proteins and the fungal-specific Puf1/Puf2 proteins. PUF proteins are characterized by RNA-binding domains composed of multiple α-helical PUM repeats, and the number and arrangement of the PUM repeats varies for each of the subgroups. Although crystal structures have revealed how the PUM repeats of classical, Puf6 and Nop9 PUF proteins assemble for specific RNA recognition (22–32), no structural information is available for the Puf1/Puf2 proteins.

Classical PUF proteins, including yeast Puf3, Puf4 and Puf5, are the most widely characterized subgroup, and are marked by a sequence-specific RNA-binding domain called the Pumilio-homology domain (PUM-HD) (6,7). Crystal structures of the PUM-HD from classical PUF proteins reveal eight PUM repeats arranged in a crescent shape (22–25,28,30). Single-stranded RNA binds to the inner con-
cave face of the protein. Typically, one PUM repeat comprising three α helices contacts one RNA base. A five-residue motif of X₁X₂X₃X₄X₅ (α, hydrophobic residue) in the second α helix of a PUM repeat defines the sequence specificity of RNA base recognition (33,34). Residues X₁ and X₃ make edge-on hydrogen bond or Van der Waals interactions with the RNA base, whereas residue X₂ stacks with and often between bases. This X₁X₂X₃X₄X₅ motif is termed the tripartite recognition motif (TRM), and the RNA recognition residues are denoted X₁X₃/X₂ (35). For example, a uracil-specific TRM, NYøøQ, is denoted NQ/Y. Yeast Puf3, Puf4 and Puf5 proteins have diverged in the length and sequence of the RNAs they bind with their eight PUM repeats (15,36). For example, yeast Puf4 recognizes nucleotides 1–6 and 8–9 of its 9-nt recognition motif, but nucleotide 7 is flipped away from the RNA-binding surface. Flipping of a specific nucleotide is a key feature of the Puf4 RNA recognition pattern (23).

Yeast Puf6 and Nop9 define two atypical PUF protein subgroups, distinct from each other and the classical PUF proteins. Both proteins are nucleolar localized and involved in pre-rRNA processing (37,38,39). In contrast to the classical PUF proteins, Puf6 and Nop9 both comprise 11 PUM repeats, arranged in an ‘L’ or ‘C’-like shape, respectively (24,29). Puf6 and human homolog Puf-A bind to single- or double-stranded RNA with no apparent sequence specificity, and most of their PUM repeats lack the characteristic RNA base recognition residues that are found in the classical PUF proteins. Nop9 utilizes a combination of classical PUF protein RNA sequence recognition and structured RNA recognition to bind to its target RNAs (29,31,39).

Puf1 and Puf2 comprise a subgroup of PUF proteins unique to fungi, yet conserved for hundreds of millions of years. They differ from other PUF proteins in containing two types of RNA-binding domain: a divergent PUM-HD and an RNA recognition motif (RRM) domain (Figure 1A). Only six PUM repeats are predicted in the PUM-HD, and some of their RNA-interacting TRMs differ from classical PUF proteins. S. cerevisiae Puf1 and Puf2 interact preferentially with mRNAs encoding cell periphery proteins (15), and Puf1/Jsn1 has been shown to be localized to the cell perimeter (40). Puf1 also co-localizes with mitochondria and associates with the Arp2/3 complex (41).

Deletion of Puf1 causes defects in mitochondrial morphology and motility, but it is not known whether Puf1 RNA regulatory activity is involved. Puf1 and Puf2 mediate responses to environmental stresses (19,21) and do so via control of specific mRNAs (19). They also appear to control some target RNAs in conjunction with the yeast classical PUF proteins: Puf3, Puf4 and Puf5 (20–21,42).

Analysis of S. cerevisiae Puf2 mRNA targets from RNA immunopurification (RIP)-microarray and CLIP-seq (UV crosslinking and immunoprecipitation with deep sequencing) identified the UAAU tetranucleotide sequence as its binding motif (43–45), in contrast to the 5'-UGUA sequence motif found within the target RNAs of classical PUF proteins. The presence of more than two UAAU sequences is predominant in the top mRNA targets of Puf2, although one Puf2 molecule binds one UAAU sequence (45). The sequence and length of the linker region between the two UAAU sequences have little effect on Puf2 binding.
based on analysis by yeast three-hybrid assay (44). CLIP analysis of Puf2 mutants lacking the RRM domain results in a lower percentage of RNA targets containing the UAAU motif, indicating that the RRM is not required to bind a UAAU motif but it enhances target selection in vivo (45). Although their sequence specificity has been examined, the Puf1/Puf2 proteins are the only remaining PUF family for which we have no structural information.

We sought to determine a crystal structure of a Puf1 protein to determine how the six-repeat Puf1/Puf2 PUM-HD recognizes UAAU motifs and how their single RRM domain influences RNA-binding activity. Here, we present a crystal structure of the PUM-HD of Schizosaccharomyces pombe Puf1 in complex with target RNA containing dual UAAU motifs. The structure reveals an extended flat PUF scaffold comprising seven PUM repeats. Unlike all other PUF-RNA complexes, two Puf1 molecules bind one RNA. PUM repeats R2 to R5 form both sequence-specific and sequence-independent protein–RNA interactions. Two crystal structures of the RRM domain of Puf1 reveal a homodimer. The RRM appears to lack RNA-binding activity in vitro, but when together with the PUM-HD, it enables Puf1 to preferentially bind to RNAs with dual UAAU sites over single sites. Dimerization via the RRM is required for high-affinity binding to the dual UAAU motif. This example of cooperative action of the RRM and PUM-HD identifies a new mechanism by which multiple RNA-binding modules within an RBP may collaborate for RNA target selection.

MATERIALS AND METHODS

Protein expression and purification

A cDNA fragment encoding the PUM-HD of Puf1 (residues 109–485) was amplified by polymerase chain reaction (PCR) from an S. pombe cDNA library pTN-RC5 (obtained from the Yeast Genetic Resource Center in Japan). Attempts to amplify fragments encoding the Puf1 RRM domain (residues 1–79) and the sequence containing both the RRM and PUM-HD domains (RP, residues 1–485) from the same cDNA library were unsuccessful. Therefore, a cDNA fragment corresponding to nucleotides 1–487 of the Puf1 gene sequence was synthesized (Thermo Fisher). The synthesized DNA was used as a PCR template to amplify a cDNA encoding the RRM domain. The RP fragment was constructed by ligating the cDNAs encoding the RRM domain and the PUM-HD through a BglII restriction site. The cDNAs were cloned into the pSM3 vector that encodes an N-terminal His6-SUMO tag (kindly provided by Dr Christopher Lima) through NotI and XhoI restriction sites (46). The usage of the NotI site introduces 12 additional amino acid residues at the N terminus generated from the vector sequence that remain after Ulp1 cleavage to remove the His6-SUMO tag. The recombinant plasmids were transformed into BL21-CodonPlus (DE3)-RIL competent cells (Agilent) for protein expression.

Puf1 proteins were overexpressed by inducing cultures in log phase with 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 16°C for ∼18 h for the PUM-HD and RP or 0.4 mM IPTG at 37°C for 3 h for the RRM. Bacteria were disrupted by sonication in lysis buffer containing 20 mM Tris (pH 8.0), 0.5 M NaCl, 20 mM imidazole, 5% (v/v) glycerol and 0.1% (v/v) β-mercaptoethanol. The proteins were purified by three sequential chromatography columns: a gravity-flow Ni-NTA chelating column (Qiagen), a Hitrap Q or a Heparin column (GE Healthcare) and a Superdex 75 column (GE Healthcare). The Ni-column elution buffer contained 20 mM Tris (pH 8.0), 50 mM NaCl, 0.2 M imidazole and 1 mM dithiothreitol (DTT). The Ni-column eluent was incubated with Ulp1 protease for 2 h (RRM or RP) or overnight (PUM-HD) to remove the His6-SUMO tag before loading onto the next column. Buffer A for the Hitrap Q or Heparin column was 20 mM Tris (pH 8.0), 1 mM DTT and buffer B contained an additional 1 M NaCl. With a linear gradient from 5–100% buffer B, the RRM protein was eluted off the Hitrap Q column with ∼15% buffer B. The PUM-HD protein was eluted off the Heparin column with ∼40% buffer B. The RP protein did not bind the Q resin but passing the solution over the column removed contaminating proteins. The protein solutions were concentrated before loading onto a Superdex 75 column equilibrated in 20 mM HEPES (pH 7.4), 0.15 M NaCl and 2 mM DTT. Peak fractions were pooled and concentrated for crystallization or RNA-binding assays.

To prepare selenomethionine (SeMet)-substituted PUM-HD protein, the plasmid was transformed into B834(DE3) Escherichia coli cells. The cells were grown in a media made from SeMet medium base (Molecular Dimensions) supplemented with 10 mg/ml SeMet. The protein was expressed and purified as the native protein. Single amino-acid substitution mutants of Puf1 PUM-HD (R315E or R318E) and Puf1 RP (Y3A/I39A/K73A) were generated using site-directed mutagenesis. The mutants were expressed and purified in the same way as the wild-type protein.

Crystallization

Concentrated Puf1 PUM-HD protein (OD280 = 3.0, ~105 μM) was mixed with synthetic RNA purchased from GE Dharmacon (5′-UUAAUAACUUAAU-3′) with a molar ratio of 1:1 and incubated on ice for two hours prior to crystallization screening. Crystals were grown in a crystallization solution of 1.4 M MgSO4, 0.1 M MES (pH 5.6) with a 1:1 ratio of sample:reservoir solution by hanging drop vapor diffusion at 20°C. SeMet derivative crystals were obtained in the same conditions as native crystals. Crystals were cryoprotected in the crystallization solution supplemented with 20% (v/v) glycerol and flash frozen in liquid nitrogen.

The Puf1 RRM protein crystallized readily, however, the needle-shaped crystals were soft and did not diffract. To improve the crystal quality, we first deleted vector sequence encoding 12 N-terminal residues that remain after Ulp1 cleavage but this produced a fusion protein where the His6-SUMO tag could not be removed by proteolytic cleavage. We therefore introduced a 3-aa linker (GGG) between the His6-SUMO tag and the RRM that restored Ulp1 cleavage. The new RRM construct was purifed in the same way as the initial construct. The protein was concentrated to 15 mg/ml. Numerous conditions from sparse-matrix crystallization screens yielded crystals. Two different conditions were subjected to optimization. One condition contained
0.3 M NaCl, 0.1 M citric acid (pH 3.6), 20% (v/v) glycerol. Crystals were directly looped out of the drop and flash frozen in liquid nitrogen. The other condition contained 15% (w/v) PEG 3350, 0.1 M MES (pH 5.6). Crystals were cryoprotected in the crystallization solution supplemented with 20% (v/v) glycerol and flash frozen in liquid nitrogen.

Data collection and structure determination
X-ray diffraction data were collected at beamline 22-ID of the Advanced Photon Source (APS) or using our in-house system, a 007HF Rigaku X-ray generator equipped with a Saturn944 CCD detector. A single-wavelength anomalous dispersion (SAD) dataset was collected on a SeMet derivative PUM-HD:RNA crystal at the wavelength of 0.97942 Å. All datasets were integrated and scaled with HKL2000 (47). For the PUM-HD:RNA structure, phases were determined from the 3.6 Å-resolution SAD dataset. Phenix AutoSol (48) found all 18 selenium sites from two protein molecules. Iterative AutoBuild and manual model building with COOT (49) were performed. The final structure was refined to in an asymmetric unit. The final P21 structure was refined to

\[ R_{free} = 0.193 \]

using a dataset from a native crystal. The crystals were directly looped out of the drop and flash frozen in liquid nitrogen. The other condition contained 10% polyacrylamide native TBE gels run at constant voltage (100 V) with 1× TBE buffer at 4°C for 35 min. The gels were dried and visualized using a Typhoon PhosphorImager (GE Healthcare). Band intensities were quantified with ImageQuant 5.1. The data were fit using the Hill equation with GraphPad Prism 7. EMSAs were performed three times, and mean \( K_d \)’s and standard error of the mean are reported (Table 2 and Supplementary Table S1).

Differential scanning fluorimetry assay
Reaction mixtures (20 µl) contained ~0.2 mg/ml PuF1 PUM-HD proteins (wild-type, R315E or R318E) and Sypro Orange dye (1:1000 dilution) (Invitrogen). Fluorescent intensity was collected from 25°C to 95°C (3°C increment/min) with a real time PCR instrument (Applied Biosystems™ QuantStudio 7 Flex System) using excitation and emission wavelengths of 470 and 586 nm, respectively. The Protein Thermal Shift software (Applied Biosystems) was used to analyze protein melting curves and calculate melting temperatures (\( T_m \)).

Size exclusion chromatography-mult-angle light scattering (SEC-MALS)
Molecular masses of PuF1 proteins were assessed by SEC-MALS using an AKTA FPLC system (GE Healthcare) coupled to miniDawn TREOS and Optilab REX detectors (Wyatt Technology). The PUM-HD and RP proteins, with or without RNA, were run on a Superdex-200 10/300 GL column (GE Healthcare) in a buffer containing 20 mM HEPES (pH 7.4), 150 mM NaCl and 2 mM DTT. A Superdex 75 10/300 column with the same buffer was used for the RRm protein. All data were analyzed using ASTRA 7.1.4 software (Wyatt Technology).

CLIP-seq analyses
CLIP-seq data for S. cerevisiae PuF2 protein variants from Porter et al. (45) were reanalyzed. The S. cerevisiae PuF2 proteins lacked the polyasparagin region prior to the C-terminus of the protein, and we refer to the PuF2 construct lacking only the polyasparagin region as the wild-type protein. Yeast strains had been UV-irradiated to covalently cross-link protein with directly bound RNA. The PuF2 protein was then affinity purified, and the attached segments of RNA had been identified by deep sequencing (45,51–53). The PuF2 CLIP-seq FASTQ files with NCBI accession number GSE73273 were downloaded. Adapter sequences (TGGAAATTCTCGGGTGCCAAGG) and duplicates were removed using fastq-mcf with the parameter -D 35 (https://github.com/ExpressionAnalysis/ea-utils). Reads were mapped to the S. cerevisiae genome (EF4 release 70) using Bowtie2 (54) with the parameters ‘-5 5 –local.’ Peaks

Electrophoretic mobility shift assays
Synthetic RNAs (GE Dharmaco) were labeled with \( ^{32}P \)-\( \gamma \)-ATP by T4 polynucleotide kinase for 1 h at 37°C. Unincorporated \( ^{32}P \)-\( \gamma \)-ATP was removed using Illustra MicroSpin G-25 columns. For the PUM-HD and RP proteins, radiolabeled RNAs (100 pM) were incubated with 2-fold serially diluted protein samples from 4000 to 0.49 nM at 4°C for 1 h in 10 mM HEPES (pH 7.4), 50 mM NaCl, 0.01% (v/v) Tween-20, 0.1 mg/ml BSA, 10 µg/ml yeast tRNA and 2 mM DTT. For the RRm protein, the binding buffer was simpler with 10 mM HEPES (pH 7.4), 50 mM NaCl, and 0.01% (v/v) Tween-20. The samples were resolved on 10% polyacrylamide native TBE gels run at constant voltage (100 V) with 1× TBE buffer at 4°C for 35 min. The gels were dried and visualized using a Typhoon PhosphorImager (GE Healthcare). Band intensities were quantified with ImageQuant 5.1. The data were fit using the Hill equation with GraphPad Prism 7. EMSAs were performed three times, and mean \( K_d \)’s and standard error of the mean are reported (Table 2 and Supplementary Table S1).

Nucleic Acids Research, 2019, Vol. 47, No. 16 8773
Table 1. Data collection and refinement statistics

<table>
<thead>
<tr>
<th></th>
<th>PUM-HD with RNA</th>
<th>RRM (PEG)</th>
<th>RRM (NaCl)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.54</td>
<td>1.0</td>
<td>1.54</td>
</tr>
<tr>
<td>Space group</td>
<td>I₄̅₁</td>
<td>P₂₁</td>
<td>P₂₁₂₁₂₁</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td>a, b, c (Å)</td>
<td>159.68 215.53</td>
<td>46.85 79.54</td>
</tr>
<tr>
<td></td>
<td>α, β, γ (°)</td>
<td>90 90 90</td>
<td>90 90 90</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50.0-3.0 (3.0-3.0)</td>
<td>50.0-1.55 (1.58-1.55)</td>
<td>50.0-2.06 (2.1-2.06)</td>
</tr>
<tr>
<td>Rmerge</td>
<td>0.062 (0.526)</td>
<td>0.061 (0.549)</td>
<td>0.08 (0.206)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.9 (100.0)</td>
<td>98.1 (90.9)</td>
<td>97.9 (95.1)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>8.1 (8.1)</td>
<td>5.5 (5.0)</td>
<td>3.6 (2.2)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>32.8-3.0</td>
<td>38.2-1.55</td>
<td>24.5-2.05</td>
</tr>
<tr>
<td>No. reflections</td>
<td>27506</td>
<td>33403</td>
<td>16888</td>
</tr>
<tr>
<td>Rwork / Rfree</td>
<td>0.211 / 0.263</td>
<td>0.193 / 0.231</td>
<td>0.172 / 0.225</td>
</tr>
<tr>
<td>No. atoms</td>
<td>6099</td>
<td>2378</td>
<td>2309</td>
</tr>
<tr>
<td>Protein</td>
<td>272</td>
<td>51</td>
<td>234</td>
</tr>
<tr>
<td>RNA</td>
<td>98.2</td>
<td>98.8</td>
<td>98.8</td>
</tr>
<tr>
<td>Solvent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wilson B</td>
<td>77.8</td>
<td>11.9</td>
<td>16.8</td>
</tr>
<tr>
<td>Protein</td>
<td>92.9</td>
<td>23.2</td>
<td>17.7</td>
</tr>
<tr>
<td>RNA</td>
<td>98.2</td>
<td>98.8</td>
<td>29.4</td>
</tr>
<tr>
<td>Solvent</td>
<td></td>
<td></td>
<td>22.9</td>
</tr>
<tr>
<td>R.m.s deviations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.003</td>
<td>0.006</td>
<td>0.002</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>0.554</td>
<td>0.822</td>
<td>0.456</td>
</tr>
</tbody>
</table>

Statistics for the highest-resolution shell are shown in parentheses.

Table 2. *In vitro* RNA-binding affinities of Puf1 determined by EMSA

<table>
<thead>
<tr>
<th>RNA variant</th>
<th>Sequenceᵃ</th>
<th>PUM-HD, Kᵣ</th>
<th>Kᵣᵇ</th>
<th>Pufl RP, Kᵣ</th>
<th>Kᵣᵇ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dual UAAU</td>
<td>UUAAUAACUUAU</td>
<td>50.1 ± 1.9</td>
<td>1</td>
<td>29.5 ± 1.3</td>
<td>1</td>
</tr>
<tr>
<td>Dual UAAU-L2</td>
<td>UUAAUA - - UUAU</td>
<td>57.3 ± 0.3</td>
<td>1.1 (0.02)</td>
<td>N.D.ᵇ</td>
<td></td>
</tr>
<tr>
<td>Dual UAAU-L0</td>
<td>UUAAU - - - - UUAU</td>
<td>90.5 ± 8.7</td>
<td>1.8 (0.01)</td>
<td>43.5 ± 1.3</td>
<td>1.5 (0.0002)</td>
</tr>
<tr>
<td>Dual UAAU-L11</td>
<td>CUUAAUAAUAUUUUGACUAUU</td>
<td>48.5 ± 0.8</td>
<td>1 (0.48)</td>
<td>25.9 ± 2.4</td>
<td>0.9 (0.23)</td>
</tr>
<tr>
<td>Mono UAAU (Site 1)</td>
<td>UUAAUAACUACAG</td>
<td>71.9 ± 0.5</td>
<td>1.4 (0.0004)</td>
<td>169 ± 48</td>
<td>5.7 (0.007)</td>
</tr>
<tr>
<td>Mono UAAU (Site 2)</td>
<td>UACAGACUAAUU</td>
<td>92.8 ± 3.5</td>
<td>1.9 (0.008)</td>
<td>125 ± 6.8</td>
<td>4.2 (0.00002)</td>
</tr>
<tr>
<td>U1G, U9G</td>
<td>UUAUAACUAUUAU</td>
<td>435 ± 43</td>
<td>9 (0.0008)</td>
<td>491 ± 8.5</td>
<td>17 (1.00001)</td>
</tr>
<tr>
<td>A2G, A10G</td>
<td>UUUAACUAUUGAG</td>
<td>484 ± 6.9</td>
<td>10 (1.00000)</td>
<td>423 ± 30</td>
<td>14 (0.0002)</td>
</tr>
<tr>
<td>A3G, A11G</td>
<td>UUUAACUAUAGU</td>
<td>133 ± 8.0</td>
<td>2.6 (0.0006)</td>
<td>98 ± 7.3</td>
<td>3.3 (0.0009)</td>
</tr>
<tr>
<td>U4G, U15G</td>
<td>UUAAUACUAAAG</td>
<td>1191 ± 167</td>
<td>24 (0.002)</td>
<td>745 ± 189</td>
<td>25 (0.003)</td>
</tr>
<tr>
<td>U1G</td>
<td>UUAUAACUAUUAU</td>
<td>93 ± 6.0</td>
<td>1.8 (0.003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A2G</td>
<td>UUAUAACUAUUAU</td>
<td>80 ± 2.0</td>
<td>1.6 (0.0004)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U4G</td>
<td>UUAUAACUAUUAU</td>
<td>112 ± 5.0</td>
<td>2.2 (0.0003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U9G</td>
<td>UUAUAACUAUUAU</td>
<td>91 ± 9.1</td>
<td>1.9 (0.012)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A10G</td>
<td>UUAUAACUAUUAU</td>
<td>106 ± 5.5</td>
<td>2.1 (0.0007)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U12G</td>
<td>UUAUAACUAUUAU</td>
<td>234 ± 23</td>
<td>4.7 (0.0013)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ᵃUAAU motifs are underlined, and altered nucleotides are in boldface.
ᵇKᵣ values are calculated relative to the Kᵣ for binding to the dual UAAU RNA (top row). P-values for comparison with the Kᵣ for binding to the dual UAAU RNA (unpaired two-tailed t-tests) are in parentheses.
ᶜN.D., not determined.

were obtained as reported in Porter et al., (45). Peak regions were defined by first merging the peak lists from yeast expressing Puf2 wild-type, PUM-HD only and ∆RRM. For overlapping peaks in these data sets, the highest peak for each region was retained. Peaks that corresponded to snoRNAs and tRNAs were discarded. Peaks with widths >40 nt were also discarded, as it was difficult to differentiate a widely spaced UAAU motif cluster from independent monomeric motifs. This retained 1614 peaks. UAAU motifs in peaks were counted, and overlapping motifs (e.g. UAAUUAU) were counted as one motif. Binding at a peak was defined as the maximum read depth in the peak region. For each dataset, we normalized binding to the median peak height for that particular Puf2 protein variant.

RESULTS

Two molecules of Puf1 recognize the dual UAAU motif in target mRNA

To provide a structural basis for understanding target RNA recognition by the yeast Puf1 and Puf2 protein family, we determined a crystal structure of *S. pombe* Puf1 in complex with RNA. We attempted to crystallize *S. cerevisiae* Puf1 and Puf2, but many different expression constructs yielded
only insoluble or aggregated protein in *E. coli*. As a result, we turned to *S. pombe* Puf1 whose PUM-HD shares ~40% sequence identity (61% similarity) with *S. cerevisiae* Puf1 and ~38% sequence identity (59% similarity) with *S. cerevisiae* Puf2. Here we present a 3.0-Å resolution crystal structure of the PUM-HD of *S. pombe* Puf1 (residues 109–485) in complex with a model RNA binding element, derived from *ARF1* mRNA and analyzed previously in three-hybrid assays (45) (5′-UAAUAACUUAAAU-3′, UAAU motifs underlined) (Figure 1A and Table 1). This RNA contains a dual UAAU motif, as is common in targets of the Puf1/Puf2 family. We define the first UAAU motif as positions 1–4 and the second motif as positions 9–12.

The crystal structure of Puf1 reveals a unique PUF protein scaffold formed by one pseudo-repeat R1, consisting of two α helices, and followed by seven PUM repeats with three α helices. Although only six PUM repeats were predicted in sequences of Puf1/Puf2 family members, the structure identified seven repeats that together form an extended flat structure whose overall shape deviates from the curved crescent shape of classical PUF proteins like Puf3 (Figure 1B and Supplementary Figure S1). Unlike the highly regular α-helical structural repeats of human PUM1, Puf1 repeats are variable, especially in the loop between the second and third α helices (Supplementary Figure S1C). Repeats R3 and R8 are most distinct; their third α helices do not align well with those of the other repeats (Supplementary Figure S1D). A sequence alignment of Puf1/Puf2 homologs and yeast Puf3 indicates that the TRMs of the Puf1 N-terminal PUM repeats align best with repeats R2–R4 of Puf3. We designated the seven Puf1 PUM repeats as R2–R8 (Figure 1B and Supplementary Figure S2). These repeat designations thus are renumbered relative to those for *S. cerevisiae* Puf2 in Porter et al. (45). Repeat R2 here corresponds to Repeat R1 in their analyses. In addition to the tandem α-helical repeats, the PUM-HD of Puf1 contains an N-terminal extension (αN, residues 109–159) (Supplementary Figure S2). Residues 109–144 form a small structural subunit made of three α helices that pack against the non-RNA binding surfaces of repeats R4–R6 in the center of the structure, and residues 145–159 form a long coil that packs along the junctions between repeats R3–R4, R2–R3 and R1′-R2 and then leads into the pseudo-repeat R1′ at one end of the extended structure (Supplementary Figure S3A).

The structure of the Puf1 PUM-HD:RNA complex reveals a novel stoichiometry of PUF protein RNA recognition with two Puf1 protein molecules bound to the one RNA molecule containing two UAAU sites (Figure 1B). The presence of one protein per UAAU was predicted from compensatory mutant analyses of *S. cerevisiae* Puf2 (45), but is seen here for the first time. The two Puf1 molecules (A and B) are arranged in a V-like shape. The RNA is bent between U8 and U9, which places the two Puf1 N-terminal pseudo repeats near each other. This V-shaped arrangement is distinctive but may not be essential, as it could vary depending on the length of the linker sequence between the two UAAU motifs (see ‘Discussion’ section). Each UAAU tetranucleotide is bound by one protein molecule via PUM repeats R5 to R2. Repeats R6 to R8 do not contact RNA. Puf1 molecule A binds to U1–A2–A3–U4, and molecule B binds to U9–A10–A11–U12. The protein–RNA contacts are nearly identical between the two proteins and the two UAAU binding sites, and below we describe in detail protein–RNA interactions based on protein molecule A and RNA nucleotides U1–A2–A3–U4. The 4 nt in the RNA linker region, A5–A6–C7–U8, stack with each other. A5 stacks with the phenol ring of Tyr203 (Figure 1B and Supplementary Figure S3B). Nucleotides A6–C7–U8 and -1U make no contact with the proteins.

Sequence conservation analysis using the ConSurf server (55) indicates that the RNA-binding surface is highly conserved across fungal Puf1/Puf2 proteins (Figure 1C). We therefore predict that RNA recognition by other fungal Puf1/Puf2 proteins is similar to that of *S. pombe* Puf1. The strongest conservation is focused at the second α helices of PUM repeats, which include the RNA base-interacting TRM motifs. Repeat R8, which is not involved in RNA recognition, is less conserved than the other repeats.

**Puf1 recognizes the UAAU motif sequence and flanking structural features**

Puf1 recognizes RNA using a combination of classical PUF protein RNA base contacts and atypical interaction between divergent PUM repeats and RNA structural features. Only repeats R2–R4 in Puf1 bear RNA interacting residues similar to the classical sequence-specific PUM repeats (Figure 2A and B; Supplementary Figure S2), whereas other repeats display divergent sequences. In classical PUF proteins the set of residues in the TRM dictates the RNA base recognized by that repeat. Typically, N/Q/Y selects U and C/Q/R or S/Q/R selects A (Figure 2C) (34). Puf1 repeats R4, R3 and R2 recognize U1, A2/A3 and U4, respectively. U1 is recognized sequence-specifically by the typical U-recognizing TRM N/Q/Y of repeat R4 (Asn279, Tyr280 and Gln283) (Figure 2B). Additionally, the atypical TRM (AR/R) in repeat R5 contributes electrostatic interactions between Arg315 and Arg318 and the U1 phosphate group (Figure 2A). U4 is recognized by the TRM NQ/T in repeat R2, as had been predicted and confirmed for *S. cerevisiae* Puf2 by Porter et al. (45). The edge-interacting residues Asn206 and Gln210 form typical interactions with U4, but Thr207 fails to stack with the base due to its small side chain. There is an additional interaction between the O2 atom of U4 and Asn240 of repeat R3 (Figure 2B).

In contrast to the more typical interactions of repeats R4 and R2 with U1 and U4, respectively, the A2 and A3 bases are stacked with one another and together are bound by the A-recognizing TRM, TQ/W, of repeat R3 (Figure 2). The edge-interacting residue Gln246 of repeat R3 forms two hydrogen bonds with A2 while the base-stacking residue Trp243 of repeat R3 stacks with A3. A2 is also stacked with Tyr280 from repeat R4. The flat RNA-binding surface formed by the Puf1 PUM repeats appears to accommodate the interaction of the single R3 repeat with the directly stacked A2 and A3 bases. The second UAAU motif (U9–A10–A11–U12) is recognized by molecule B just as molecule A contacts the first UAAU motif (Supplementary Figure S3B). In addition to recognition by molecule B, A11 of the second UAAU motif is bound by residues in the N-terminal pseudo-repeat of molecule A. Gln162 of molecule A forms...
Figure 2. Pufl repeats R2-R4 recognize the UAAU motifs. (A) Interaction between Pufl molecule A and 5’UAAU sequence. RNA-binding TRM residues from Pufl are shown as stick models and gray dashed lines indicate atoms within hydrogen bonding distance of the RNA. The structure is colored as in Figure 1. (B) Schematic drawing of the interactions between Pufl and the dual UAAU motif RNA. PUM repeats are represented by rectangles and RNA nucleotides are represented by ovals. (C) A classical PUF protein RNA interaction for comparison. RNA-binding TRM residues from human Pumilio1 are shown as stick models and gray dashed lines indicate interactions between the amino acid side chains and RNA nucleotides. The RNA is colored as in panel A.

a hydrogen bond with the N6 atom of A11 and Arg169 of molecule A forms a salt bridge with the A11 phosphate group (Figure 2B and Supplementary Figure S3B).

Pufl PUM-HD RNA-binding activity depends upon UAAU motif and RNA backbone recognition

We probed the RNA-binding activity of the Pufl PUM-HD and found that it binds with high affinity and specificity to the UAAU motif. We measured in vitro RNA-binding affinity of purified S. pombe Pufl PUM-HD protein by electrophoretic mobility shift assay (EMSA) (Table 2, Supplementary Figure 4A). We used the dual UAAU motif RNA that was crystallized, which contains two UAAU sites separated by a 4-nt linker. Pufl bound the dual UAAU motif RNA with an apparent $K_d$ of 50.1 nM. This tight binding is similar to the in vitro affinity of S. cerevisiae Pufl for a dual UAAU motif in PMP2 mRNA ($K_d = 53$ nM) (44). Our crystal structure revealed sequence-specific recognition of U1/U9, A2/A10 and U4/U12 by the TRMs of repeats R2-R4, but recognition of A3/A11 appeared to be sequence independent. We substituted A with G in the third positions of both UAAU motifs of the dual UAAU RNA (A3G, A11G). These sequence changes resulted in a modest 2.6-fold reduction in affinity. In contrast, substituting A with G in the second positions of both UAAU motifs (A2G, A10G) reduced binding affinity 10-fold relative to the base dual UAAU RNA sequence (Table 2). Therefore, as suggested by the crystal structure, an A at the second position of the UAAU motif is critical while the identity of the nucleotide at the third position is less important.

The RNA-interacting TRM motif in repeat R4 of Pufl (AR/R) is distinct from classical PUM repeats, and we found that the two arginine residues that interact with the phosphate backbone are essential for RNA complex formation. Our crystal structure revealed that Arg315 and Arg318 electrostatically interact with the phosphate groups of U1 (molecule A) and U9 (molecule B) (Figure 2 and Supplementary Figure S3B). We mutated Arg315 or Arg318 to glutamic acid and measured binding of the mutant proteins to the dual UAAU RNA (Supplementary Table S1). No protein–RNA complex was detected for the R315E mutant with the dual UAAU RNA, even at a protein concentration as high as 1.5 M. We detected binding of the R318E mutant to the dual UAAU RNA, but the mutant...
protein bound 4-fold more weakly than wild-type Puf1. To ensure that these binding defects were not due to misfolding of the mutant proteins, we checked the protein stability by differential scanning fluorimetry. The R315E mutant displayed a slightly lower melting temperature ($T_m$) than wild-type (49.3 °C versus 52.2 °C), whereas R318E displayed a slightly higher $T_m$ of 54.6 °C. The small difference in protein stability of the R315E mutant does not account for the loss of RNA-binding activity. Therefore, Arg315 is a critical residue for RNA binding. Together, the interaction of Arg315 and Arg318 with the RNA backbone may support the RNA conformation at the 5′ ends of the UAAU motifs that is observed in the crystal structure.

*S. cerevisiae* Puf2 binds to dual UAAU motifs separated by linkers of a range of lengths and sequences, and we found that for *S. pombe* Puf1 the length of the linker between UAAU sites also plays a minor role in affinity. We measured binding of the *S. pombe* Puf1 PUM-HD to RNAs with a 2-nt linker (dual UAAU-L2) or an 11-nt linker (dual UAAU-L11, corresponding to a sequence in the 3′-UTR of *S. pombe* TPO1 mRNA) and found that Puf1 bound to the dual UAAU-L2 or L11 RNA with similar affinity as to the dual UAAU motif with a 4-nt linker (dual UAAU, Table 2). When we removed the entire linker sequence (dual UAAU-L0), Puf1 binding was only 2-fold weaker than binding to two additional RNA sequences, a 35 nt sequence from *S. cerevisiae* PMP2 and a 30 nt sequence modified from *S. pombe* TPO1 (Table 2). Since the mode of RNA recognition by RRMs varies from protein to protein and some RRMs do not bind RNA, we cannot predict whether or not the Puf1 RRM domain binds to RNA based solely on the apo-RRM crystal structure. We therefore tested whether purified Puf1 RRM protein bound to RNA *in vitro*. Reasoning that the RRM could bind to sequences outside the dual UAAU motif, we used an RNA similar to the dual UAAU RNA, but with additional flanking sequence (5′-ACAUAUAUGAUCUAUAUUA-3′) and measured RNA binding by EMSA. The Puf1 RRM did not bind to this RNA, even at a protein concentration of 10 μM. We also tested binding to two additional RNA sequences, a 35 nt sequence from *S. pombe* TPO1 (5′-UUUUAUCUAUGAUUUUGA UCUAACGGAUAA-3′), the UAAU-L11 sequence with 5′ and 3′ extensions (Table 2), and a 30 nt sequence modified from *S. cerevisiae* PMP2* (5′-AAUUAUGAUCUAUUAACCUAAUAUUA-3′) (44). The Puf1 RRM domain did not bind to either of these RNAs, even at a protein concentration of 20 μM. Therefore, although the Puf1/2 RRM enhances UAAU motif recognition *in vivo*, it lacked detectable *in vitro* RNA-binding activity on its own.

We hypothesized that the RRM domain could alter the RNA-binding selectivity of the PUM-HD and found that addition of the Puf1 RRM to the PUM-HD surprisingly favored binding to a dual UAAU motif over a mono UAAU site, but had little effect on overall affinity and no effect on UAAU motif sequence specificity. We measured RNA-binding affinities of a Puf1 protein containing both the RRM and PUM-HD (Puf1 RP, residues 1-485). We found that Puf1 RP bound the dual UAAU RNA with a $K_d$ of 29.5 nM, which is similar to the value we measured for the PUM-HD alone (Table 2 and Supplementary Figure S4B). We also found that the binding affinity was not substantially affected by the length of the linker separating the two UAAU motifs: Puf1 RP binding affinity for dual UAAU-L11 RNA with an 11-nt linker was unchanged and for the dual UAAU-L0 RNA with no linker was only 1.5-fold weaker than to the RNA with a 4-nt linker (Figure 4A and Table 2). Sequence substitutions in the UAAU motifs had the same effects as for the Puf1 PUM-HD: mutations of the first (U1G/U9G), second (A2G/A10G) and fourth (U4G/U12G) nucleotides weakened Puf1 RP binding ~14-fold, but mutations of the third nucleotides (A3G/A11G) only weakened binding 3-fold (Figure 4B and Table 2). These results indicate that the PUM-HD is the main con-
The Puf1 RRM promotes dimerization and stable dual UAAU motif recognition

In addition to binding RNA, RRM domains can mediate protein–protein interactions by forming homodimers, making intramolecular contacts with flanking regions, and binding other proteins (56). We therefore examined whether the Puf1 RRM domain promotes dimerization, which might favor dual UAAU recognition. We determined molecular masses of Puf1 proteins and protein–RNA complexes by SEC-MALS and found that the Puf1 RP protein forms a stable 2:1 complex with a dual UAAU RNA in solution (Table 3). The apparent molecular mass of the PUM-HD was 43.8 kDa, similar to the calculated protein molecular weight (MW) of 44.6 kDa, indicating that the protein is monomeric in solution. In contrast, the apparent molecular masses of the Puf1 RRM and RP were ~1.6–1.7 times their respective calculated MWs; 15.1 kDa for the RRM versus 8.9 kDa (calculated) and 86.1 kDa for Puf1RP versus 54.6 kDa (calculated). These intermediate molecular masses suggest that monomer and dimer forms of the Puf1 RRM and RP proteins in the absence of RNA are in a fast equilibrium, because we observed single peaks that appear to be one species (Supplementary Figure S5). When we analyzed complexes of the Puf1 PUM-HD with dual UAAU RNAs, we found that the molecular masses were similar for RNAs with linker lengths of 0, 2 and 4 nts, about 1.5 times the monomer size (Table 3). Therefore, the PUM-HD is in a rapid equilibrium of one or two protein molecules bound to the dual UAAU motif RNA. The molecular mass of a complex of the Puf1-PUM-HD with a mono UAAU RNA bearing only one intact UAAU motif (site 1) was 44.5 kDa, corresponding...
Figure 4. The Puf1 RRM promotes binding of the PUM-HD to dual UAAU motif RNA. Bar graphs illustrate binding affinities relative to that of the corresponding protein with dual UAAU motif RNA ($K_{rel}$). (A) Relative binding affinities of the Puf1 PUM-HD (gray) or RP (black) for RNA variants with dual or mono UAAU motifs, showing the effects of motif number and linker length. 'Dual UAAU' denotes an RNA with a 4-nt linker between the two UAAU motifs, 'Dual UAAU-L0' denotes no linker nucleotides and 'Dual UAAU-L11' denotes an 11-nt linker. RNA sequences and $K_d$ values are shown in Table 2. Error bars in all graphs are the standard error of the mean of three replicate experiments. The measured $K_d$ for binding to the dual UAAU motif RNA is shown above the bar. The $K_{rel}$ is the $K_d$ for binding to a particular RNA divided by the $K_d$ for binding of the same protein to the dual UAAU motif RNA. (B) Relative binding affinities of the Puf1 PUM-HD (gray) or RP (black) for RNAs bearing UAAU sequence variants, showing the effects of disrupting the UAAU motif sequences at different positions. (C) A common dimerization interface in crystal structures of the Puf1 RRM. Backbone traces of non-crystallographic dimers in crystals of the Puf1 RRM are superimposed (space group P2_1, blue and cyan; space group P2_12_1, gray). Side chains for residues that were mutated to alanine in Puf1 RP TM are shown as stick models. (D) Relative binding affinities of the Puf1 RP (black) or RP TM (dark gray) for RNAs bearing dual or mono UAAU motifs.

to a stable 1:1 complex. In contrast, the molecular mass for Puf1 RP in complex with the dual UAAU RNA was 110.6 kDa, matching the calculated MW of 113.2 kDa for two Puf1 RP molecules bound to an RNA. Thus the RRM appears to stabilize the 2:1 Puf1 RP-dual UAAU motif RNA complex in solution.

We obtained two different types of crystals of the RRM domain of S. pombe Puf1: one was grown in a crystallization condition with PEG (P2_1 space group) and the other was grown in a condition with NaCl (P2_12_1 space group). Both models contain four molecules (labeled A, B, C, D) in an asymmetric unit, and all eight molecules are highly similar (RMSD 0.2-1.4 Å over 75–78 Cα atoms). The four molecules in both structures are related by non-crystallographic symmetry, with the A and B pair equivalent to the C and D pair. Molecules A and B are arranged similarly with respect to each other in the two structures, despite different space groups (Figure 4C). The protein molecules contact each other through hydrophobic and H-bonding interactions via residues in the β-sheet and the C-terminal loop. Although RRM homodimers have been observed previously (56), a search for similar homodimers in the Protein Data Bank did not identify a similar RRM-RRM interface. The PDBePISA server (57) calculated a buried surface area of 1360–1900 Å² with a Complexation Significance Score of 1.0. This indicates that the observed homodimer AB is likely to be biologically relevant, rather than a crystal-packing artifact.

To determine the importance of the RRM homodimer, we mutated to alanine three residues located at the inter-
In vivo studies support collaboration between the Puf1/2 RRM and PUM-HD

To examine whether the Puf1/2 RRM influences RNA-binding specificity in living cells, we re-analyzed our published CLIP-seq studies of *S. cerevisiae* Puf2 (45) and found that Puf2 lacking the RRM fails to preferentially bind to sites with multiple UAAU elements. In *S. cerevisiae* and *S. pombe*, Puf2 is closely related to Puf1, with the same overall protein architecture (15,43,58). In our CLIP-seq studies we had identified *in vivo* RNA interactions for three forms of Puf2: wild-type protein containing both the RRM and PUM-HD, a mutant protein that contained essentially the RRM-HD and lack the RRM domain, and a second mutant that lacked the RRM domain (45). In light of our *in vitro* data indicating the influence of the RRM on binding to sites with dual UAAU motifs, we assessed the preferences of the wild-type and mutant proteins for single versus multiple UAAU motifs. We first compared the number of UAAU motifs present in the binding sites for the wild-type protein, and bound to a mono UAAU RNA (site 1) only 1.5-fold weaker than binding to a dual UAAU motif rather than the 6-fold stronger binding affinity of the wild-type protein for the dual UAAU motif RNA (Figure 4D). We conclude that the RRM imposes a preference *in vitro* for dual- versus mono-UAAU elements on the PUM-HD. Past work has shown that members of the Puf1/Puf2 family prefer multiple UAAU elements (45). We predict that *in vivo* preference relies on the RRM.

The SEC-MALS molecular weight (kDa) for the Puf1/2 RRM and PUM-HD

<table>
<thead>
<tr>
<th>Table 3. SEC-MALS molecular weight (kDa)</th>
<th>Calc MW</th>
<th>RNA</th>
<th>Dual UAAU</th>
<th>Dual UAAU-L2</th>
<th>Dual UAAU-L0</th>
<th>Mono UAAU</th>
</tr>
</thead>
<tbody>
<tr>
<td>PUM-HD</td>
<td>44.6</td>
<td>43.8</td>
<td>68.5</td>
<td>68.3</td>
<td>67.0</td>
<td>44.5</td>
</tr>
<tr>
<td>RRM</td>
<td>8.9</td>
<td>15.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Puf1 RP</td>
<td>54.6</td>
<td>86.1</td>
<td>110.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Puf1 RP TM</td>
<td>54.4</td>
<td>53.1</td>
<td>106.9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*aRNA MW = ~4 kDa.

*bWhen only one UAAU motif is available for binding, the Puf1 RP appears to be in an equilibrium of various species of monomer/dimer bound to the mono UAAU RNA.

Figure 5. CLIP-seq analyses support a role for the Puf1/2 RRM in binding site selection *in vivo*. (A) *S. cerevisiae* Puf2 protein shows increasing binding to sites with 2, 3 or >3 UAAU motifs, which disappears in the absence of an RRM domain. Peak heights of CLIP-seq peaks for the indicated proteins sorted by the number of UAAU motifs. The log$_2$ peak height normalized to the median peak height for each protein are shown. The ratio of PUM-HD binding to wild-type protein that possessed only the PUM-HD (Figure 5A middle, ‘PUM-HD’). Moreover, the Puf2p ΔRRM variant (which lacks only the RRM domain) also failed to exhibit increased binding to RNA sites with two or more UAAU motifs (Figure 5A right, ‘ΔRRM’). Therefore, the Puf2 mutant proteins lacking the RRM bind equally well to sites with one or multiple UAAU motifs.

We next determined, peak-by-peak, the ratio of peak heights for the mutant proteins relative to that of wild-type protein containing both the RRM and PUM-HD, a mutant protein that contained essentially the PUM-HD alone and bound to a mono UAAU RNA (site 1) only 1.5-fold weaker than binding to a dual UAAU motif rather than the 6-fold stronger binding affinity of the wild-type protein for the dual UAAU motif RNA (Figure 4D). We conclude that the RRM imposes a preference *in vitro* for dual- versus mono-UAAU elements on the PUM-HD. Past work has shown that members of the Puf1/Puf2 family prefer multiple UAAU elements (45). We predict that *in vivo* preference relies on the RRM.
type protein. For binding regions with 1 UAAU motif, the peak heights were similar for wild-type and mutant proteins (median log2 ratio of peak heights), reflecting the favorability they bind equally well to sites with a single motif (Figure 5B). In contrast, the peak heights for binding of the PUM-HD and ΔRRM proteins to regions with multiple UAAUs were lower than for the wild-type protein (negative values for median log2 ratio of peak heights), reflecting the unfavorable binding of the wild-type protein to sites with multiple UAAUs (Figure 5B) and supporting our prediction that in vivo, the PUM-HD molecules in our crystal structure or in solution could not be made. With sequences longer than 4 nt, the linker RNA may adopt different conformations. We did not observe protein-protein interactions between the two Puf1 PUM-HD molecules in our crystal structure or in solution measurements, but intermolecular interactions might occur between the two protein molecules when bound to RNAs with longer or shorter linker lengths.

Figure 6. Model for Puf1 UAAU RNA recognition selectivity. Schematic representation of RNA recognition by the Puf1 PUM-HD alone or in combination with the RRM. The PUM-HD alone binds with similar affinity to either dual or single UAAU motif RNA (top). In contrast, the RRM + PUM-HD (RP) forms a stable 2:1 complex with dual UAAU motif RNA and binds 6-fold weaker to an RNA with a single UAAU than to a dual site (bottom).

DISCUSSION

The seven yeast PUF proteins fully represent the diversity of the PUF protein family across eukaryotes. Puf3, Puf4 and Puf5 are classical sequence-specific RBPs, and Nop9 and Puf6 represent the two atypical PUF proteins common among all eukaryotes. The Puf1/Puf2 subgroup is specific to fungi and has been conserved over hundreds of millions of years. Its defining characteristics are a divergent PUM-HD, an RRM RNA-binding domain, distinctive RNA sequence specificity, and novel 2:1 protein-to-RNA stoichiometry. Our crystal structure of S. pombe Puf1 reveals that the PUM-HD directly binds the UAAU motif using PUM repeat RNA recognition residues for both classical RNA base recognition and interaction with backbone phosphate groups. The RRM also has a key role, in that it promotes preferential Puf1 binding to dual UAAU motifs over single UAAU sequences (Figure 6). Dual UAAU motifs are present frequently in high-ranked target RNAs of S. cerevisiae Puf2, while mono UAAU motifs are common in low-ranked targets (45), strongly suggesting that this function of the RRM is biologically significant.

The specificity of Puf1/Puf2 proteins for their RNA targets in vivo reflects the interplay of multiple structural features in the proteins. The Puf1 PUM-HD recognizes UAAU motif RNA by bringing together the sequence-specific RNA-binding mode adopted by classical PUF proteins with the electrostatic interaction mode of Puf6 proteins. Of the seven PUM repeats in Puf1, only repeats R2-R4 bear RNA base-interacting TRMs similar to the classical sequence-specific PUM repeats: Repeats R4, R3 and R2 recognize U1, A2/A3 and U4, respectively. In contrast, Puf1 PUM repeats R5-R8 display divergent TRM sequences. Repeats R6 to R8 of Puf1/Puf2 protein do not contact RNA, but repeat R5 uses two arginine residues in its TRM to make electrostatic contacts with the U1 and U9 phosphate groups. This interaction by repeat R5 is reminiscent of that observed in the atypical PUF protein subgroup containing Puf6 and its human ortholog Puf-A (24). Conserved basic residues within multiple TRMs interact with the phosphate backbone in a crystal structure of Puf-A in complex with double-stranded DNA.

Dual UAAU motifs contain linkers separating the two UAAU motifs ranging from 0 to 10 nt, with 3 nt being the most prevalent (44). Our structure of Puf1 bound to a dual UAAU RNA with a 4-nt linker revealed that each UAAU sequence is recognized by one Puf1 molecule, and RNA-binding assays showed that binding by the Puf1 PUM-HD with or without the RRM is tight regardless of the linker length. In our crystal structure, the RNA makes a sharp bend between U8 and U9, which are the last nucleotide of the linker and the first nucleotide of the second UAAU motif, respectively. The intervening phosphate group is contacted by arginine side chains in the repeat R5 TRM and a Mg2+ ion supports the RNA conformation. We speculate that the relative orientation of the two RNA-bound protein molecules must change to accommodate different length linkers. For example, with no linker nucleotides, the observed relative orientation of the two Puf1 molecules is impossible, and with fewer than four intervening nucleotides, the observed interactions at both ends of the linker sequence could not be made. With sequences longer than 4 nt, the linker RNA may adopt different conformations. We did not observe protein-protein interactions between the two Puf1 PUM-HD molecules in our crystal structure or in solution measurements, but intermolecular interactions might occur between the two protein molecules when bound to RNAs with longer or shorter linker lengths.

The Puf1/Puf2 RRM domain enhances target RNA selection with dimerization favoring binding of Puf1 to the dual UAAU motif, and this finding is reflected in the mRNA targets of S. cerevisiae Puf2 protein in vivo. A single UAAU motif is relatively low in sequence complexity compared to the 8–12 nt sequences recognized by classical PUF proteins. When both the Puf1 RRM and PUM-HD are present, the binding affinity is now sensitive to loss of one of the two UAAU motifs. The highest ranked mRNA targets of S. cerevisiae Puf2 in vivo, as defined by CLIP experiments, possess two UAAU motifs, which progressively declines to a plateau of an average of ~1.5 motifs in the weakest binding RNAs (45). The RRM does not appear to affect sequence specificity of the PUM-HD for a UAAU motif in vitro, but it does affect target mRNA selection in vivo. A UAAU motif is found in 81% of target RNAs of Puf2 with both RRM and PUM-HD. In contrast, only 48% of target RNAs of the Puf2 PUM-HD alone contain a UAAU motif (45). Furthermore, as shown here via in vivo CLIP studies, the RRM imposes a systematic preference for binding sites with 2, 3 or >3 UAAU elements. Collectively, these data strongly
support the view that the Puf1/Puf2 PUM-HD directs sequence specificity and that the RRM enhances recognition of multiple UAAU motifs.

The presence of both an RRM and PUM-HD, though unique to the fungal Puf1/Puf2 PUF subgroup, emphasizes the broad principle that distinct RNA-binding modules within a single protein can alter RNA recognition even without both domains interacting with RNA. RRMs previously have been shown to collaborate with other RBPs. For example, the RRM of the 68-kDa subunit of Cleavage Factor Im (CFIm68) has been reported to alter the RNA-binding properties of the 25-kDa Nudix domain subunits (CFIm25). Similar to Puf1, the CFIm68 RRM does not affect sequence specificity of CFIm25, as CFIm25 recognizes a UGU A motif in the absence of CFIm68 (59). Instead, it appears that the RRM of CFIm68 restricts binding of the complex to RNAs with two UGU A motifs separated by >9 nt. RRMs are often combined with other RNA-binding modules in RBPs. Given our findings with Puf1, we expect that some of these RRMs might collaborate with neighboring RNA-binding domains to modulate RNA recognition. This also applies to RBPs with multiple RRMs that together determine the specificity of RNA recognition. The RBP HuR contains two N-terminal RRMs in tandem and a third C-terminal RRM3. The N-terminal RRMs comprise a module that is sufficient for AU-rich or U-rich RNA element recognition, but the C-terminal RRM3 has been shown to enhance RNA-binding affinity through both dimerization and direct RNA recognition (60,61). The dimerization interface of HuR RRM3 does not obstruct its RNA-binding surface, which is different from what we observed for the Puf1 RRM. Dimerization of the Puf1 RRM involves the β-sheet RNA-binding surface and we could not demonstrate RNA binding by the Puf1 RRM, suggesting that it may not have a dual function like that of HuR RRM3. However, the apparent fast equilibrium between monomer and dimer forms of the Puf1 RRM in the absence of RNA suggests that its β-sheet surface could become available for RNA binding. A crystal structure of a Puf1 RP–RNA complex is needed to enhance understanding of the cooperative activities of the RRM and PUM-HD, but our efforts have not yet yielded crystals.

As illustrated here, deep studies of model RNA-binding proteins produce new principles of RNA recognition, in this case how two RNA-binding domains within a protein collaborate to produce a distinct regulatory motif. Encyclopedias of RNA-binding proteins continue to expand along with glimpses into their RNA recognition specificities and target networks through powerful high-throughput, genome-wide probing (62–64). The discovery of cooperative modes of RNA recognition can drive genomic studies forward by suggesting new parameters for analyzing existing data sets and refinements for generating new data.

**DATA AVAILABILITY**

Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data Bank under accession numbers 6NY5 (Puf1 PUM-HD and RNA), 6NX5 (Puf1 RRM) and 6NWW (Puf1 RRM).

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR Online.

**ACKNOWLEDGEMENTS**

We are grateful to our colleagues Lars Pedersen and Monica Pillon for critical reading of our manuscript. We thank John Gonczy for assistance with data collection at SER-CAT beamlines 22-ID and 22-BM at the Advanced Photon Source, Argonne National Laboratory, Lars Pedersen and Juno Krahn for crystallographic and data collection support at NEIHS, and Geoffrey Mueller for preliminary NMR characterization.

**FUNDING**


**Conflict of interest statement.** None declared.

**REFERENCES**


