Cooperativity in RNA-Protein Interactions: Global Analysis of RNA Binding Specificity

Zachary T. Campbell, Devesh Bhimsaria, Cary T. Valley, Jose A. Rodriguez-Martinez, Elena Menichelli, James R. Williamson, Aseem Z. Ansari, and Marvin Wickens

1Department of Biochemistry
2Graduate Program in Cellular and Molecular Biology
University of Wisconsin-Madison, 433 Babcock Drive, Madison, WI 53706-1554, USA
3Department of Molecular Biology and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA
4The Genome Center of Wisconsin, University of Wisconsin-Madison, 425G Henry Mall, Madison, WI 53706, USA

*Correspondence: ansari@biochem.wisc.edu (A.Z.A.), wickens@biochem.wisc.edu (M.W.)
DOI 10.1016/j.celrep.2012.04.003

SUMMARY

The control and function of RNA are governed by the specificity of RNA binding proteins. Here, we describe a method for global unbiased analysis of RNA-protein interactions that uses in vitro selection, high-throughput sequencing, and sequence-specificity landscapes. The method yields affinities for a vast array of RNAs in a single experiment, including both low- and high-affinity sites. It is reproducible and accurate. Using this approach, we analyzed members of the PUF (Pumilio and FBF) family of eukaryotic mRNA regulators. Our data identify effects of a specific protein partner on PUF-RNA interactions, reveal subsets of target sites not previously detected, and demonstrate that designer PUF proteins can precisely alter specificity. The approach described here is, in principle, broadly applicable for analysis of any molecule that binds RNA, including proteins, nucleic acids, and small molecules.

INTRODUCTION

RNA control pervades biology. Multiprotein complexes assemble on mRNAs to control when, where, and how much protein will be produced. These complexes are critical in a diverse range of biological contexts spanning learning, memory, development, immunity, and viral replication (Colina et al., 2008; Li and Nagy, 2011; Ule and Darnell, 2006; Wickens et al., 2000). A single regulatory protein can bind hundreds of mRNAs and coordinate their control. The specificity of proteins for particular RNA sequence elements determines which mRNAs are regulated, and is the most fundamental level of RNA control circuitry. Here, we examine the interaction between two collaborating families of mRNA regulatory proteins: PUFs (Pumilio and FBF) and CPEB (Cytoplasmic Polyadenylation Element Binding) (Richter, 2007).

PUFs are an evolutionarily widespread family of RNA binding proteins required for maintenance of diverse stem cell populations, pattern formation, learning, and memory (Ariz et al., 2009; Crittenden et al., 2002; Dubnau et al., 2003; Suh et al., 2009; Zhang et al., 1997). The PUF tertiary structure is remarkably conserved; eight repeats of three-helical bundles combine to form a crescent (Edwards et al., 2001; Wang et al., 2001, 2002; Wang et al., 2009b; Zhu et al., 2009). The concave face provides the interface with RNA, whereas the convex surface appears to be a platform for protein-protein interactions (Edwards et al., 2001, 2003; Houshmandi and Olivas, 2005). Genome-profiling experiments suggest that a single PUF protein associates with hundreds of mRNA targets, potentially regulating 7%–11% of the transcriptome (Galgano et al., 2008; Gerber et al., 2004, 2006; Hafner et al., 2010; Kershner and Kimberly, 2010; Morris et al., 2008). This association usually results in reduced mRNA stability and translation but also can affect mRNA activation and localization (Goldstrohm et al., 2007; Suh et al., 2009; Wreden et al., 1997). PUFs exert these effects on translation through collaboration with a variety of protein partners including CPEBs (Edwards et al., 2001, 2003; Goldstrohm et al., 2007). The precise effects of these protein-protein interactions on interactions with RNA generally are unclear, in part due to the difficulty of deciphering cooperative effects on binding specificity.

CPEBs are conserved among metazoans and play key roles in mRNA control (Richter and Lasko, 2011). They bind U-rich elements designated CPEs (cytoplasmic polyadenylation elements) using zinc knuckle and RRM (RNA recognition motif) domains (Besse and Ephrussi, 2008; Hake et al., 1998). CPEB proteins regulate translation, localization, and poly(A) tail length, and can either activate or repress their targets (Richter and Lasko, 2011). CPEB proteins are critical in very diverse biological contexts, from synaptic plasticity to the cell cycle, cancer progression, and cellular senescence (Burns and Richter, 2008; Ortiz-Zapater et al., 2012; Richter and Lasko, 2011; Standart and Minshall, 2008).

Our strategy to assay RNA-protein interactions (SEQRS) integrates in vitro selection, high-throughput sequencing of RNA, and SSLs (sequence specificity landscapes) (Figure 1). The three elements of our approach combine to provide a powerful level of resolution beyond existing methods. Current techniques to analyze the specificity of RNA-protein interactions are generally slow, laborious, costly, and identify only those RNAs that bind with the highest affinities; yet, lower affinity sites are often critical for regulation in vivo (Ellington and Szostak, 1990; Tuerk and...
coimmunopurification studies, lack a consensus binding site mRNAs that bind a regulatory protein in vivo, as judged by specificity of its PUF protein partner for RNA. Second, many et al., 2011). We found that a CPEB protein alters the binding interactions are presently sparse (Garvie et al., 2001; Slattery DNA binding specificity; comparable examples in RNA-protein RNA are opaque. In DNA-protein interactions, partners can alter effects on one another’s specificities and affinities for interactions. First, RNA regulatory proteins often act in complexes, yet their off-target effects. We examine this issue globally with a PUF (Cooke et al., 2011; Wang et al., 2009a). It is unclear to what extent these designer proteins bind undesired sites, eliciting off-target effects. We examine this issue globally with a PUF designer protein. The method we describe provides access to these questions through its global assessment of binding affinities.

RESULTS

The Approach

The central aim of our strategy is to determine the binding preference of a given protein for all possible sequences of a given length in a single experiment. To do so, we used a two-step strategy involving first, in vitro selections and deep sequencing, and second, analysis of the data using SSLs. We refer to this protocol as SEQRS.

We developed an iterative in vitro selection strategy, adapted from previous protocols (Figure 1A) (Ellington and Szostak, 1990; Gold, 1990). Emerging methods for analysis of DNA binding protein specificity that rely on next-generation sequencing approaches yield high-quality quantitative models of protein-nucleic acid interactions (Carlson et al., 2010; Jolma et al., 2010; Nutiu et al., 2011; Slattery et al., 2011; Stormo and Zhao, 2010; Tietjen et al., 2011). At present, visualization of the data is challenging given the number of data points per experiment. Our use of SSLs enables the detection of variant sites, unexpected new specificities, and the effects of protein partners. At the same time, it provides an intuitive and interactive graphical means of representing all data in an experiment fit to a given binding model. SEQRS is facile, rapid, reproducible, accurate, and permits identification of multiple binding modes in a single experiment.

We examine three outstanding problems in RNA-protein interactions. First, RNA regulatory proteins often act in complexes, yet their effects on one another’s specificities and affinities for RNA are opaque. In DNA-protein interactions, partners can alter DNA binding specificity; comparable examples in RNA-protein interactions are presently sparse (Garvie et al., 2001; Slattery et al., 2011). We found that a CPEB protein alters the binding specificity of its PUF protein partner for RNA. Second, many mRNAs that bind a regulatory protein in vivo, as judged by coimmunopurification studies, lack a consensus binding site (Galgano et al., 2008; Gerber et al., 2004, 2006; Hafner et al., 2010; Kershner and Kimble, 2010; Morris et al., 2008). Our approach enables us to identify previously hidden, alternative sites that not only bind the PUF but also mediate regulation in vivo. Third, designer proteins have been engineered to possess new specificities for RNA to achieve targeted regulation (Cooke et al., 2011; Wang et al., 2009a). It is unclear to what extent these designer proteins bind undesired sites, eliciting off-target effects. We examine this issue globally with a PUF designer protein. The method we describe provides access to these questions through its global assessment of binding affinities.

Figure 1. The SEQRS Approach

(A) Experimental strategy is illustrated. The DNA library contains a random region of 20 bp (orange) between two 20 bp constant regions (blue). Transcription with T7 RNA polymerase yields a pool of RNAs used as the starting material for selection. A sufficient quantity of RNA to cover all possible 20-mer sequences is incubated with recombinant protein. The protein is immobilized on resin (red circles) to enable capture of the RNA protein complex. After repeated wash steps, the remaining RNA is heat eluted from the protein. The RNA is reverse transcribed (RT) back to DNA using a primer complementary to the constant region denoted as primer B. The single-stranded DNA is amplified using a primer set that reintroduces the T7 promoter (green boxes) upstream of primer A. After the desired number of rounds, aliquots of double-stranded DNA are amplified with primers containing adapters appropriate for Illumina sequencing (pink boxes) and unique 6 bp bar code identifiers (gold boxes). The bar codes differ by at least two bases from all other bar codes to minimize misidentification due to sequencing errors. (B) Overview of computational analysis is shown. After sequencing, the 20-mer random regions are binned according to bar code. All possible k-mer sequences (ten in these experiments) from the random 20-mer are determined for each read. Enrichment over library is calculated by normalizing against the library to correct for differences in coupling efficiency for the random DNA library.

Using the n-most abundant reads (n typically = 300), sequence logos are generated. Seed motifs for specificity landscapes are generated from these logos. (C) Visualization of binding specificity. All of the data from an experiment are visualized relative to the seed motif. In this example using C. elegans FBF-2, all of the observed data are fit to the seed motif HUGURWWHD. In the linear form of the inner ring, all possible permutations are arranged in alphabetical order and then the flanking regions are considered. Each ring in the SSL represents increasing numbers of mismatches or hamming distance from the seed motif (shown in blue boxes). The height of each peak is proportional to the enrichment score of a particular sequence. A linearized rendition of the 0-mismatch (innermost) ring is shown at the top of this panel, with sequences indicated.
To identify consensus binding motifs, positional weight matrices (PWMs) were generated from the most abundant sequences in a given data set (Figure 1B). Although PWM matrixes (PWMs) were generated from the most abundant sequencing of multiple samples in parallel, and enabled deconvolution of multiple samples in parallel, and enabled deconvolution of multiplexed data.

To identify consensus binding motifs, positional weight matrices (PWMs) were generated from the most abundant sequences in a given data set (Figure 1B). Although PWM analysis identifies certain high-affinity sites, it does not capture alternate sites even when highly populated, nor does it detect context-dependent sequence features (Carlson et al., 2010; Frank et al., 1997). To enhance data analysis, we adapted the use of SSLs to enable analysis of single-stranded RNA (Carlson et al., 2010). SSLs provide a graphical representation of binding data using a series of concentric rings (Figure 1C). The innermost ring contains sequences perfectly matched to a given seed motif. Derivation of the seed motif begins with the PWM but then is optimized to yield a landscape with the greatest concentration of data in the inner ring. Subsequent rings in the SSL represent increasing numbers of mismatches from the seed motif. The z axis (height) corresponds to the number of reads of a particular sequence, normalized to the starting library. Thus, a high peak represents an RNA sequence present many times in the sequencing reads. The origin of the plot is fixed at a single position in the first, 0-mismatch ring. The sequences are arranged by motif and then the flanking regions are considered. This ordering system is maintained in subsequent rings first by the positions of the mismatches and then by the substituted nucleotide at the mismatch. As a result, the order of sequences is consistent both within and between rings.

Consider analysis of the data from a SEQRS experiment in which the consensus sequence used for seed analysis is HUGURWWWHD (Figure 1C; H = A, C, or U; W = A or U; D = A, G, or U). All sequences that correspond to that consensus are present in the first ring. An expanded view of that ring, seen in linear form, illustrates that each progressive set of sequences shift the register of the seed motif one position along the randomized sequence. The periodic peaks correspond to sequences containing C one base upstream of the UGU segment of the motif, demonstrating a preference that was not apparent in the logo. In the data shown, derived from a real experiment, the background is low: very few reads were obtained that contained more than a single mismatch indicative of high specificity. Thus, the analysis reveals a set of sequences that are represented to varying extents. We shall show later that abundance of reads is related to the affinity of the protein for the RNA.

Analysis of RNA-Protein Interactions by SEQRS

We first examined the binding specificity of a founding member of the PUF protein family, C. elegans FBF-2 (Figure 2) (Crittenden et al., 2002; Suh et al., 2009; Zhang et al., 1997). The binding specificity of FBF-2 has been analyzed extensively, providing a strong foundation for evaluating the effectiveness of the methodology (Bernstein et al., 2005; Opperman et al., 2005; Wang et al., 2009b).

The abundance of RNAs matching the consensus motif of FBF-2 was enhanced over the course of five cycles of selection (Figure 2A). The intensity of data in the inner ring of the landscape, representing precise matches to the seed motif, increased progressively, concomitant with decreases in the intensities of outer rings. Even after a single round, a motif containing the conserved UGU element had emerged (Figure 2B). However, the prevalence of FBF Binding Element (FBE) containing sequences throughout the entire data set is low. By three rounds of selection, the consensus motif strongly resembled the known optimal FBE defined previously (Figure 2B), as well as the consensus motif derived from RIP-ChIP analyses from C. elegans (Figure S1A available online) (Bernstein et al., 2005; Kershner and Kimble, 2010). The percentage of reads containing a canonical FBE was determined after each round of selection (Figure 2C) (Bernstein et al., 2005). After the first round, <0.54% had such matches; by the fifth round, nearly a quarter of the reads included an FBE.

The data obtained through SEQRS were reproducible as measured by performing three experiments using different preparations of FBF-2. Identical consensus motifs following five rounds of selection were obtained (Figure S2A). Moreover, comparisons between pairs of data sets demonstrated a high degree of correlation ($R_0 = 0.95–0.99$, Spearman’s rank). In contrast, comparison to a negative control with a different consensus motif, human PUM2, did not ($R_0 = 0.55$).

The binding profiles determined by SEQRS correlated with independent measures of binding affinity. The number of reads obtained for FBF-2 was related to binding activities measured on the C. elegans paralog (91% identical in amino acid sequence) FBF-1 using yeast three-hybrid assays (Figure 2D) (Bernstein et al., 2005; Hook et al., 2007). Similarly, the number of reads for Puf4p was related to apparent dissociation constants measured by EMSA (Figure 2E) (Hook et al., 2007; Miller et al., 2008). Comparable correlations have been reported for DNA-protein interactions (Carlson et al., 2010; Nutiu et al., 2011; Stormo and Zhao, 2010).

Ternary Complexes

CPEB proteins physically associate with PUF proteins (Richter, 2007). A minimal fragment of C. elegans CPB-1, outside of the RNA binding domain, is sufficient to bind FBF-2 in vitro and enhances FBF-2-mediated repression (Campbell et al., 2012). This 40 residue CPB-1 peptide fails to bind RNA in vitro (E.M., Z.T.C., J. Wu, J.R.W., and M.W., unpublished data). To test whether this interaction affects the binding specificity of FBF-2, CPB-1 was immobilized on glutathione resin and used to affinity select FBF-2 prior to RNA binding reactions. Unbound FBF-2 was removed by several wash steps prior to addition of RNA. In this way only FBF-2/CPB-1 complexes were detected.

After five rounds of selection with the FBF-2/CPB-1 complex, we observed a binding motif distinct from that of FBF-2 alone.
Neither CPB-1 nor FBF-2 interaction-defective mutants yielded significantly enriched motifs using MEME after five rounds of selection (Figures S3A–S3C). Similarly, their SSLs revealed little enrichment of sequences matching the FBF-2 motif. These controls demonstrate that specificities seen in Figure 3 are due to the FBF component of the FBF-2/CPB-1 complex.

Two differences are apparent in the comparison of the FBF-2/CPB-1 complex to FBF-2 alone: the ternary complex exhibits differences in preferences upstream of the UGU, and appears to be more permissive or diverse downstream. Upstream of the UGU, the most conspicuous difference is the decreased presence of a cytosine in the ternary complex as compared with FBF-2 alone. FBF-2 requires a cytosine preceding the

(Figures 3A and 3B). Neither CPB-1 nor FBF-2 interaction-defective mutants yielded significantly enriched motifs using MEME after five rounds of selection (Figures S3A–S3C). Similarly, their SSLs revealed little enrichment of sequences matching the FBF-2 motif. These controls demonstrate that specificities seen in Figure 3 are due to the FBF component of the FBF-2/CPB-1 complex.

Two differences are apparent in the comparison of the FBF-2/CPB-1 complex to FBF-2 alone: the ternary complex exhibits differences in preferences upstream of the UGU, and appears to be more permissive or diverse downstream. Upstream of the UGU, the most conspicuous difference is the decreased presence of a cytosine in the ternary complex as compared with FBF-2 alone. FBF-2 requires a cytosine preceding the
Figure 3. Analysis of the FBF-2/CPB-1 Protein Complex Reveals Changes in Specificity

(A) Analysis of FBF-2 after five rounds of selection is illustrated. The analysis reveals a motif enriched for an upstream C. The two highest intensity peaks on the SSL both represent sequences with C at the −1 position.

(B) Analysis of the CPB-1/FBF-2 complex is presented. The complex yields a distinct motif, as noted in the text.
UGU for high-affinity binding, which enhances binding ~20-fold by interacting with a specific pocket in the protein (Qiu et al., 2012; Zhu et al., 2009). Using SEQRS, FBF-2 enriched RNAs containing a cytosine at position −1 by the second round of selection (Figure 2B). Although the enrichment in the MEME-derived logo appears modest, 58% of the reads containing FBEs possessed a C at the −1 position after five rounds of selection, whereas only 14% did in the presence of CPB-1 (Figure 3C). This difference in specificity is highlighted by examining sequences in the first ring of the SSL, arranged in linear fashion (Figure 3D). Overall, the profiles are very similar, with the conspicuous exception of certain sequences that contain cytosine at the −1 position.

To test whether CPB-1 altered the binding specificity of FBF-2, we analyzed three sequences that were overrepresented (by 50-fold or more) in the protein complex relative to FBF-2 alone using a modified yeast three-hybrid assay (Figures 3E–3G).

Specificities of the PUF Family

To evaluate the utility of SEQRS in greater depth, we analyzed four additional PUF proteins: Human PUM2, C. elegans PUF-8, PUF-11, and S. cerevisiae Puf4p (Figure 4). The PUM2 binding site deduced by our approach was nearly identical to that obtained by PAR-CLIP and RIP-ChIP (Figures 4A and S1B) (Galgano et al., 2008; Hafner et al., 2010). The core motif identified (UUAUWAUA) was strikingly similar to the consensus motifs of D. melanogaster Pumilio and S. cerevisiae Puf3p, as expected (Gerber et al., 2004, 2006). The sequence logo and SSL obtained with PUF-8 were consistent with a Pumilio-like mode of RNA recognition (Figure 4B). However, PUF-8 had a more stringent requirement for a G at position 2. The motif we observed for Puf4p contained a UGUUA motif, three A/U-rich spacer nucleotides, and a terminal UA dinucleotide consistent with RIP-ChIP data (Figures 4 and S1C) (Gerber et al., 2004). C. elegans PUF-11 is unusual in its ability to accept RNA substrates with varying spacing between the UGU and AU elements (Koh et al., 2009). PUF-11 can accommodate RNAs with either two or three spacer nucleotides. Following seven rounds of selection, we identified one major motif consisting of three spacer nucleotides and an upstream C (Figure 4D).

To determine whether CPB-1 alters repression by FBF, we utilized in vitro translation assays in reticulocyte lysate (Figure 3H). The firefly luciferase reporter contained the sequence for RNA-1 in its 3′ UTR. A control reaction containing CPB-1 alone was used to normalize each sample to 1. Significant repression by FBF-2 was only observed in the presence of CPB-1 (Figure 3I). Mutants of FBF-2 that disrupted binding of FBF-2 to RNA (RNA<sup>def</sup>) or to CPB-1 (CPB<sup>def</sup>) failed to repress translation (Campbell et al., 2012). Similarly, point mutants in CPB-1 that disrupt its binding to FBF-2 abrogated repression. These data indicate that CPB-1 enhances the activity of FBF-2 on a specific mRNA in vitro.

Repression of SEQRS RNA-1 is dependent upon CPB-1. All of the samples were normalized to a mock assay containing only CPB-1. Repression by FBF-2 is insignificant in the absence of CPB-1 or the presence of an interaction-defective version of CPB-1 (FBF-2<sup>Y479A</sup>). Mutant versions of FBF-2 (Y479A, CPB-1 binding defective, CPB<sup>def</sup>, and H326A RNA binding defective, RNA<sup>def</sup>) fail to promote repression in the presence of wild-type CPB-1. Error bars indicate SD. SSLs are presented for three additional controls (Figures S2A–S2C).

(C) Analysis of the −1 position is shown. The enrichment of −1C is diminished across the entire data set for the CPB-1/FBF-2 complex.
(D) A linear representation of the 0-mismatch SSL ring is illustrated. The y axis represents the prevalence of all permutations of the HUGURHHWD motif. Note the lack of enrichment for the upstream C element for the CPB-1/FBF-2 complex.
(E) Design of the modified yeast three-hybrid assay is presented. Candidate RNAs were expressed in yeast expressing an FBF-2/AD fusion and the interacting peptide derived from CPB-1. CPB-1 was fused to an SV40 nuclear localization signal, but not to any other domain. Levels of activity of β-galactosidase, produced from the LacZ reporter gene, were used to assay FBF-2 binding to the RNA.
(F) CPB-1 enhances binding by FBF-2 to a specific RNA measured by a modified yeast three-hybrid assay. This experiment was done in presence and absence of CPB-1, as indicated below the bars. The gld-1a RNA serves as a positive control for binding. Error bars indicate SD.
(G) Additional RNAs assayed using the modified yeast three-hybrid assay are shown. The sequences of additional RNAs analyzed are provided. Data represent the ratio of β-galactosidase levels with and without CPB-1.
(H) Design of in vitro translation assays is presented. Repression by FBF-2 was assayed in the presence and absence of CPB-1 in rabbit reticulocyte lysate (RRL).
(I) Repression of SEQRS RNA-1 is dependent upon CPB-1. All of the samples were normalized to a mock assay containing only CPB-1. Repression by FBF-2 is insignificant in the absence of CPB-1 or the presence of an interaction-defective version of CPB-1 (FBF-2<sup>Y479A</sup>). Mutant versions of FBF-2 (Y479A, CPB-1 binding defective, CPB<sup>def</sup>, and H326A RNA binding defective, RNA<sup>def</sup>) fail to promote repression in the presence of wild-type CPB-1. Error bars indicate SD. SSLs are presented for three additional controls (Figures S2A–S2C).
To characterize the specificity of Puf5p in depth, we examined specificity after seven rounds of selection (Figures 5A and 5B). The sequence logo obtained was a composite of two alternate binding modes (Figure 5A). After separating the top 300 sequences based on the spacing between the UGU and UA motifs, we detected 2 distinct consensus motifs: motif A had a 4-nucleotide spacer between the UGU and UA sequences; motif B had a 5-nucleotide spacer followed by a UA. In the complete data set, reads that matched motif B were three times more abundant than those that matched motif A. We found well-populated peaks in the 1-mismatch ring of the motif B SSL containing sequences belonging to motif A (Figure 5B). These findings suggested that a cryptic alternate motif might exist in some of the mRNAs controlled by Puf5p. We reasoned that if this previously unknown motif was biologically relevant, it should be present in a large number of mRNAs bound by Puf5p in vivo.

To test whether motif A was present in mRNAs found to physically associate with Puf5p in RIP-ChIP experiments, we computationally removed transcripts containing the canonical motif (motif B) from the Puf5p RIP-ChIP data set and then searched for a common motif in the remaining transcripts using MEME (Figure S1D) (Bailey et al., 2006; Gerber et al., 2004). A single enriched motif matching motif A was found in 28% of the remaining transcripts.

The two sites we identified account for 48% of the mRNAs associated with Puf5p in RIP-ChIP studies; 52% have neither element in their 3' UTR. To identify additional motifs, we computationally removed transcripts containing either motif A or B from the data set and examined the remainder for enriched sequence elements using MEME. None was detected. As a result, we suggest that the two modes we have defined are the most common in 3' UTRs; others may exist but be poorly populated. The high “background” in SEQRS with Puf5p could be due to such sites, or to binding elements elsewhere in the mRNA: human PUM2 appears to bind sites situated in the ORFs (Hafner et al., 2010). Associations in vivo may also be indirect, or due to protein partners as observed with the CPB-1 peptide’s recruitment by FBF-2.

To determine whether Puf5p regulated mRNAs that contained motif A, we used in vitro translation assays (Figure 5C). Reporters with motif A sites were repressed (Figure 5D). Repression was specific, in that it was abrogated by mutation of UGU to ACA, which disrupts binding and repression via canonical PUF sites, such as the Puf5p site in the control mRNA, CIN8 (Chritton and Wickens, 2010; Hook et al., 2007). We conclude that the noncanonical site mediates repression by Puf5p in vitro.

Assessing the Specificities of Designer Proteins

The modular architecture of PUF proteins enables the design of proteins with new specificities (Figure 6A). This affords an opportunity to engineer custom PUF proteins to control stability, translation, or splicing of targeted RNAs (Cooke et al., 2011; Wang et al., 2009a). Previous studies examined only the designed protein bound to the new sequence; they could not assess specificity broadly for other sequences, which could cause off-target
effects. SEORS provides a means to deduce global effects of the mutations on specificity.

We characterized the binding specificity of a variant of FBF-2 containing mutations in two residues directly involved in RNA recognition (Figure 6A). In yeast three-hybrid assays the double mutant (N475S, Q479E) failed to bind the wild-type sequence, UUGUCCAUAA, and instead bound the sequences UGAACCAUA and UGGACCAUA (Opperman et al., 2005). Following five rounds of selection, we observed a strong A/G bias at position +3 (Figure 6B). Moreover, sequences containing an A at position +3 bound approximately 5-fold better than sequences containing a G in yeast three-hybrid data (Opperman et al., 2005). We observe a similar bias where an A is strongly favored over G at each position in the 0-mismatch ring when fit to the seed motif HUGDRHHWD (Figure 6C). We do not detect any other significant differences in specificity. In SSLs of the wild-type consensus (Figure 6D), the landscape indicates a poor fit at a single position.

**DISCUSSION**

Our studies reveal that one protein can affect another’s RNA binding specificity. mRNA control involves interactions among proteins assembled on the 3’ UTR, of which CPEB and PUF proteins are exemplary. The regulation of mRNA expression frequently involves coordination and competition between multiple RNA-protein complexes that can be assembled in the 3’ UTR. We found specific RNA sequences whose binding to FBF-2 was enhanced in the presence of CPB-1, as judged by the number of reads obtained. Our results reveal that CPB-1 enhances binding of FBF-2 to specific RNA sequences. The linear representation of the FBF-2 alone and FBF-2/CPB-1 complex illustrates the loss of a requirement for an upstream cytosine in the complex (Figure 3D). Using these RNA elements,
Figure 6. Specificity of Designer Proteins

(A) Schematic of PUF/RNA complexes and the structure of an FBF-2/RNA complex is presented (Wang et al., 2009b). PUF proteins all possess a similar architecture, in which eight three-helical bundles (green) are stacked into an arc. Along one face of this arc, eight α helices interact with RNA (gray), with a single helix recognizing a single base. Different PUF proteins achieve specificity in part through variations on this basic scaffold. Inset illustrates residues that were altered (N475S and Q479E) and are shown (pink) as sticks opposite the RNA base they coordinate (blue).

(B) The sequence logo of the designer protein (N475S Q479E) is shown. The +3 position opposite the site of the N475S and Q479E mutations is indicated with an arrow (pink). Data were obtained after five rounds of selection.

(C) Changes in specificity are presented. The relative abundance of sequences fit to the seed motif HUGDRHHWD is shown as a function of position in the 0-mismatch SSL ring in linear form. Sequence families are presented above the x axis.

(D and E) The SSL of N475S Q479E after five rounds of selection fit to the wild-type consensus (D) or a motif identified through MEME (E) is illustrated.
...we demonstrated that the interaction between CPB-1 and FBF-2 enhanced RNA binding affinity (Figures 3H and 3I).

Because native CPEB proteins possess their own RNA binding domains, it has been assumed that they act only when bound to the mRNA. However, we show here that a small segment of CPB-1, lacking RNA binding activity, alters the specificity of a PUF protein to which it binds. We suggest that in vivo, CPEB proteins need not bind to RNA to affect regulation; rather, they can exert their effects by altering the specificity of their PUF protein partners.

The massive amounts of data generated in SEQRS are visualized using SSLs, initially developed for duplex DNA and adapted here for single-stranded RNA. These landscapes organize binding events across a logically ordered terrain. They enable the user to readily evaluate the validity of a binding model because suboptimal motifs yield peaks in the outer rings. In analogous fashion, SSLs readily identify alternate motifs (such as Puf5p, Figure 5), reveal effects of flanking sequences that neighbor a core motif (such as with FBF-2, Figure 3), identify differences between members of the same protein family (here, PUF proteins), and enable comparison of the effects of protein partners (as with the CPB-1/FBF complex, Figure 3).

Emerging techniques such as PAR-CLIP and HITS-CLIP rely on covalent crosslinking to capture RNAs bound to a specific protein in vivo. (Hafner et al., 2010; Licatalosi et al., 2008). These methods capture RNA-protein complexes in a cellular context, enabling assignment of in vivo RNA targets to specific proteins. SEQRS is complementary: it interrogates specificity in vitro, and yields a comprehensive measure of the affinity of a single protein or complex for a wide array of sequences. Both approaches are needed to understand the biochemical basis of RNA control networks.

SEQRS provides a substantial methodological advance over existing methods. An alternative approach for analysis of RNA-protein interactions has been described in which RNAs selected in a single binding reaction are assayed by hybridization of protein bound RNAs to DNA microarrays (Ray et al., 2009). This method, termed “RNAcompete,” yielded 7-mer sequences, which are insufficient to accurately describe the binding consensus for many RNA binding proteins, including those examined in the present study. The library of 200,000 sequences used in the array experiments is considerably smaller than that used here (4^15 members). In SEQRS, successive rounds of selection substantially improve the quality of the SSLs and PVMs. This method yields binding measurements for all possible RNA sequences and the means to identify highly populated alternate binding modes. The use of multiplexed next-generation sequencing also provides practical advantages over microarrays: many more samples (up to 50) can be assayed in a single experiment, and the need for custom array synthesis and quality control is obviated.

Our experiments demonstrate the existence of multiple modes of binding for a single protein, yeast Puf5p, which correlates with association in vivo and repression activity in vitro. Puf5p and PUF-11, both of which have alternate binding modes, also have a higher background. They may bind more weakly to their targets compared to other members of the family, as previously reported for PUF-11 (approximately 50 nM). Directed evolution experiments targeted at design of alternate specificity suggest that broadened specificity is correlated to a reduction in binding affinity (Koh et al., 2009).

The design of PUFs with altered specificity, or neo-PUFs, is an attractive route to tailored control of mRNA expression, stability, or splicing (Cooke et al., 2011; Wang et al., 2009a). Such experiments confront a priori issues of off-target effects, a major issue in the analysis of designer zinc finger proteins to bind DNA and control transcription (Beerli and Barbas, 2002). Our work on a mutant form of FBF-2 reveals a very specific change in specificity, and provides a precedent for global analysis of the specificity of altered proteins on both the desired and undesired sequences.

The SEQRS method provides a rapid means to comprehensively assess RNA binding specificity of a single protein or protein complex in a single experiment. The approach should be useful in assessing the binding of any molecule to RNA, for purposes in basic biological research, drug development, and biotechnology.

**EXPERIMENTAL PROCEDURES**

**Protein Purification**

The GST fusion constructs used in the present study include C. elegans FBF-2 residues 121–632 (Bernstein et al., 2005), C. elegans CPB-1 residues 40–80, C. elegans PUF-8 residues 127–519 (Opperman et al., 2005), C. elegans PUF-11 residues 50–505 (Koh et al., 2009), Human PUM2 residues 456–1,064 (Goldstrohm et al., 2007), S. cerevisiae Puf4p residues 537–889 (Hook et al., 2007), and S. cerevisiae Puf5p residues 127–627 (Hook et al., 2007). A single MBP fusion construct containing FBF-2 residues 121–614 in pMal-c2x (New England Biolabs) was generated. MBP-FBF-2 Y479A and CPB-1 L40A mutants were created using site-directed mutagenesis. The repeat 6 specificity mutant (N475S Q479E) was previously described by Opperman et al. (2005). GST and MBP fusion proteins were purified as described using high-capacity magnetic GST-agarose beads (Sigma-Aldrich) (Bernstein et al., 2005). GST and MBP-FBF-2 fusion proteins were purified as described using high-capacity magnetic GST-agarose beads (Sigma-Aldrich) (Bernstein et al., 2005). Protein purity was >90% judged by visualization on SDS-PAGE. Aliquots of protein were stored in SEQRS buffer containing 20% glycerol prior to flash freezing and storage at −80°C. Fresh batches of protein were purified prior to each of the sequencing runs.

**In Vitro Selection**

The library contains a random region flanked by two constant regions lacking the conserved UGU trinucleotide common to the consensus sequence of virtually all members of the PUF family (Table S1). The initial library was transcribed from 1 μg of input dsDNA using the AmpliScribe T7-Flash Transcription Kit (Epicenter). The reaction was treated with RNase-free DNase to remove residual DNA and purified using the GeneJET RNA Purification Kit (Fermentas). A total of 150 ng of the purified RNA library was added to RNA binding proteins containing 50–100 nM of fusion protein. The total volume in the binding reactions was 100 μl in SEQRS buffer (50 mM HEPES, 5 mM EDTA, 150 mM NaCl, 5 mM DTT, 0.01% Tween 20, and 1% glycerol) containing 200 ng yeast RNA competitor and 0.1 U of RNase inhibitor (Promega) in eight-sample strip tubes. The samples were allowed to incubate for 30 min at ambient temperature prior to capture of the protein-RNA complex on a 96-well magnetic block. The binding reaction was aspirated, and the beads were washed four times with 200 μl of ice-cold SEQRS buffer. After the final wash step, the resin was resuspended in elution buffer (1 mM Tris [pH 8.0]) containing 10 pmol of the reverse-transcription primer (Table S1). Samples were heated to 65°C for 10 min and then cooled on ice. A 5 μL aliquot of the sample was added to a 10 μL iPrM-ol reverse-transcription reaction (Promega). The product ssDNA was used as a template for PCR.
High-Throughput Sequencing

The purity of each sample was determined by electrophoresis prior to sequencing. Individual samples were purified using Wizard SV Gel and PCR Clean-Up columns (Promega). Approximately equal amounts of bar-coded DNA were combined based on individual concentrations determined by Quant-IT PicoGreen fluorescence assays (Invitrogen). After pooling samples, 3 pmol of DNA was sequenced on an Illumina HiSeq 2000 instrument using a custom primer.

Bioinformatics

Sequences containing a bar code were identified using a custom MATLAB script (MathWorks). Exact matches to each bar code at the correct position (50–56) in the read were required for binning. The 20-mer sequences were extracted and then subdivided into all possible 10-mer. These were then counted and used for subsequent analysis. The background was corrected by division against all 10-mer represented in the library. Typically, the 300 most-abundant sequences were used to generate sequence logos (Bailey et al., 2006). SSLs were generated as described with minor modifications to enable analysis of RNA (Carlson et al., 2010).

Translation Assays

Puf5p extract assays were carried out as described by Chritton and Wickens (2010). FBF-2 and CPB-1 proteins were generated using in vitro translation of 50 ng of each mRNA. One microliter of each was added to the second reaction containing reporter RNAs. Reporter RNAs were transcribed from the pYCV2 plasmid using primers that generated the candidate regulatory elements directly after the stop codon. Renilla luciferase was transcribed from pSP65-Ren as described by Chritton and Wickens (2010). Individual reactions were assembled as previously described, assayed by the dual-luciferase assay system (Promega), and measured with a 96-well synergy-2 plate reader (Chritton and Wickens, 2010).

Yeast Three-Hybrid Assays

These experiments were conducted as described with minor adjustments (Bernstein et al., 2005; Opperman et al., 2005). CPB-1 (residues 40–80) was overexpressed from p414TEF and fused to an SV40 nuclear localization signal. These experiments were conducted as described with minor adjustments (Bernstein et al., 2005; Opperman et al., 2005). CPB-1 (residues 40–80) was overexpressed from p414TEF and fused to an SV40 nuclear localization signal. Luminescence data were collected using the i-Glo reagent (Promega) and measured with a 96-well synergy-2 plate reader.

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and one table and can be found with this article online at doi:10.1016/j.celrep.2012.04.003.

ACKNOWLEDGMENTS

We thank Marie Adams and the Biotechnology Next-Generation DNA Sequencing Facility at the University of Wisconsin-Madison. We are grateful to Laura Vanderploeg in the Biochemistry Media Lab for help with figures. M.W. is supported by NIH GM53320 and GM031892. J.R.W. is supported by NIH Grant GM53320. A.A. and D.B. are supported by NIH GM069420 and GM050942. Z.T.C. is supported to Laura Vanderploeg in the Biochemistry Media Lab for help with figures.

REFERENCES


