Sequence-specific binding of a chloroplast pentatricopeptide repeat protein to its native group II intron ligand

ROSALIND WILLIAMS-CARRIER, TIFFANY KROEGER, and ALICE BARKAN

Department of Biology, Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403, USA

ABSTRACT

Pentatricopeptide repeat (PPR) proteins are defined by degenerate 35-amino acid repeats that are related to the tetratricopeptide repeat (TPR). Most characterized PPR proteins mediate specific post-transcriptional steps in gene expression in mitochondria or chloroplasts. However, little is known about the structure of PPR proteins or the biochemical mechanisms through which they act. Here we establish features of PPR protein structure and nucleic acid binding activity through *in vitro* experiments with PPR5, which binds and stabilizes a chloroplast tRNA precursor harboring a group II intron. Recombinant PPR5 was shown to be monomeric by analytical ultracentrifugation and gel filtration. Circular dichroism spectroscopy showed that PPR5 has a high content of α helices, as predicted from the similarity between PPR and TPR motifs. PPR5 and another PPR protein, CRP1, bind with high affinity to single-stranded RNA, but bind poorly to single-stranded DNA or to double-stranded RNA or DNA. A specific PPR5 binding site was identified within its group II intron ligand. The minimal site spans ~45 nucleotides, encompasses two group II intron functional motifs, and overlaps the terminus of an *in vivo* RNA decay product. These results suggest mechanisms by which PPR5 influences both RNA stability and splicing.

Keywords: pentatricopeptide repeat; RNA-binding protein; plastid; chloroplast; group II intron

INTRODUCTION

Pentatricopeptide repeat (PPR) proteins comprise a recently recognized protein class implicated in various aspects of gene expression in plants, animals, protozoa, and fungi. PPR proteins are defined by tandem arrays of a degenerate 35 amino acid motif, the pentatricopeptide repeat (Small and Peeters 2000). Most PPR proteins have an N-terminal sequence predicted to target the protein to either mitochondria or chloroplasts, followed by an average of ~12 tandem PPR motifs and a small amount of non-PPR sequence flanking the PPR tract (Small and Peeters 2000; Delannoy et al. 2007). PPR proteins are encoded by small gene families in animals and fungi (~5–10 members), by ~20 genes in trypanosomes (Mingler et al. 2006; Pusnik et al. 2007), and by over 400 genes in plants (Lurin et al. 2004; Rivals et al. 2006). The biological functions of ~20 PPR proteins have been elucidated, including representatives from animals,

plants, trypanosomes, and fungi; the vast majority of these localize to mitochondria or chloroplasts and influence specific post-transcriptional steps in gene expression (for review, see Delannoy et al. 2007). By extrapolation, it seems likely that PPR proteins play a multitude of roles in integrating nuclear and organellar functions.

Models for PPR structure have been developed from data for the related tetratricopeptide (TPR) motif, a 34 amino acid repeating motif that generally serves as a protein-interaction scaffold (D'Andrea and Regan 2003). Each repeat in a TPR tract forms a pair of antiparallel α helices, with consecutive helical hairpins stacking to form a concave substrate binding surface. PPRs are sufficiently similar to TPRs that their structures can be modeled based on established TPR structures (Delannoy et al. 2007). However, experimental data relating to the structure of a PPR tract have not been reported.

Early genetic data implicating PPR proteins in post-transcriptional steps in gene expression (Barkan et al. 1994; Manthey and McEwen 1995; Coffin et al. 1997; Fisk et al. 1999) and the recovery of PPR proteins in biochemical screens based on nucleic acid affinity (Ikeda and Gray 1999; Lahmy et al. 2000) initially suggested that PPR tracts might

Reprint requests to: Alice Barkan, Department of Biology, Institute of Molecular Biology, 297 Klamath Hall, University of Oregon, Eugene, OR 97403, USA; e-mail: abarkan@uoregon.edu; fax: (541) 346-5891.

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.1077708>.

bind RNA ligands (Small and Peeters 2000). The specificity of the RNA metabolism defects in many mutants with lesions in PPR-encoding genes (for review, see Delannoy et al. 2007), and the demonstration that several PPR proteins associate specifically with their genetically defined RNA substrates in plant extract (Schmitz-Linneweber et al. 2005, 2006; Gillman et al. 2007) suggested further that PPR proteins might bind RNA in a sequence-specific fashion. That PPR tracts can function as nucleic acid binding domains has been confirmed with a variety of in vitro assays (Ikeda and Gray 1999; Lahmy et al. 2000; Meierhoff et al. 2003; Mili and Pinol-Roma 2003; Nakamura et al. 2003; Lurin et al. 2004; Tsuchiya et al. 2004; Okuda et al. 2006; Kobayashi et al. 2007; Koussevitzky et al. 2007). However, in vitro studies to establish the substrate specificities of PPR proteins are just beginning. Preferential binding to RNA relative to DNA (Lurin et al. 2004) or to single-stranded DNA relative to double-stranded DNA (Tsuchiya et al. 2002; Nakamura et al. 2003) has been reported. Sequence-specific binding of two chloroplast PPR proteins has been inferred based on the ability of some sequences to compete more effectively than others in UV cross-linking or gel mobility shift (GMS) assays (Meierhoff et al. 2003; Nakamura et al. 2003; Okuda et al. 2006).

It is clear from these findings that many PPR tracts can bind nucleic acids and that some can do so in a sequence-specific fashion. However, studies thus far have generally not been quantitative, have involved recombinant proteins with large affinity tags, and/or have used proteins whose aggregation state was not assessed. Considering the numerous and central roles played by PPR proteins, detailed investigation of their structures and mechanisms is warranted. Among the many questions to be answered are: Do some PPR proteins bind preferentially to RNA and others to DNA, and, if so, what is the structural basis for this preference? Do PPR tracts bind RNA through a surface that is analogous to the ligand binding surface of TPR proteins? Do most PPR proteins bind nucleic acids in a sequence specific fashion, and if so, what is the structural basis of this specificity? One goal of this study was to test basic tenets of current models for PPR structure and PPR/nucleic acid interactions by using purified, untagged recombinant PPR proteins with well-defined in vivo ligands in in vitro biochemical and biophysical assays. A second goal was to pinpoint the binding site for a genetically characterized PPR protein in order to elucidate mechanisms by which PPR proteins bind specific RNA sequences and influence downstream processes.

RESULTS

Expression and purification of recombinant PPR5 and CRP1

PPR5 and CRP1 were chosen for biochemical analyses because their functions and RNA ligands are well charac-

terized. CRP1 is required for the translation of the chloroplast *petA* and *psaC* mRNAs and for the accumulation of monocistronic *petB* and *petD* mRNAs in maize (Barkan et al. 1994; Fisk et al. 1999). RNA co-immunoprecipitation assays showed that CRP1 in chloroplast extract is associated with the 5'-untranslated regions upstream of the *petA* and *psaC* coding regions (Schmitz-Linneweber et al. 2005). PPR5, in contrast, co-immunoprecipitates with unspliced *trnG*-UCC from chloroplast extract and stabilizes unspliced *trnG*-UCC transcripts in vivo (Beick et al. 2008). Mature CRP1 (i.e., lacking its cleaved transit peptide) is ~70 kDa and includes 14 tandem PPR motifs and an N-terminal hydrophobic region (Fig. 1A). Mature PPR5 is ~50 kDa and includes 10 tandem PPR motifs and a small amount of non-PPR sequence flanking the PPR tract (Fig. 1A).

When expressed in *Escherichia coli* with many standard methods, both CRP1 and PPR5 were found either entirely in inclusion bodies or in soluble microaggregates that eluted in the void volume from a Superdex 200 gel filtration column (data not shown). Exploration of numerous expression vectors, induction protocols, and purification conditions resulted in methods that yield sufficient soluble and monodisperse untagged protein for in vitro RNA binding assays (see Methods). Expression as fusions to maltose binding protein (MBP) and inclusion of a low concentration of CHAPS detergent in the lysis buffer was particularly helpful for improving protein solubility. The recombinant CRP1 and PPR5 (rCRP1 and rPPR5) used in

FIGURE 1. Purification of rCRP1 and rPPR5. (A) Domain organization of PPR5 and CRP1. PPR5 consists of 10 tandem PPR motifs preceded by ~35 amino acids at the N terminus and followed by 55 amino acids at the C terminus. CRP1 consists of 14 tandem PPR motifs preceded by a region rich in hydrophobic residues of ~80 amino acids. Both proteins are synthesized with an N-terminal chloroplast transit peptide that is not represented in the recombinant proteins. (B) Elution of rCRP1 and rPPR5 from a Superdex 200 gel filtration column. An equal proportion of contiguous column fractions spanning the elution peaks was analyzed by SDS-PAGE and staining with Coomassie brilliant blue. Elution peaks of BSA (67 kDa) and maltose binding protein (42 kDa) chromatographed under the same conditions are shown.

assays described below lacked an affinity tag, were subjected to a final purification step on a gel filtration column, were >95% pure based on visual inspection of stained gels, and had a low A_{260}/A_{280} ratio (~ 0.6), indicating negligible nucleic acid contamination. rPPR5 eluted from the gel filtration column in a well-defined peak at a position consistent with its monomeric molecular weight (Fig. 1B). Some rCRP1 was present in most column fractions eluting ahead of the monomeric species, presumably in heterogeneous microaggregates, but the only peak was at the monomeric position, and only this material was used for biochemical assays. Elution at the positions expected for their monomeric molecular weights suggests that both rCRP1 and rPPR5 do not form homomultimers at the protein concentration applied to the column (~ 5 and $1 \mu\text{M}$ for rPPR5 and rCRP1, respectively).

rPPR5 is monomeric with high α -helical content

Based on sequence similarity with the TPR motif, it has been suggested that each PPR motif folds into a pair of antiparallel α helices, and that consecutive helical hairpins in a PPR tract stack to form a superhelical structure.

Secondary structure prediction algorithms support the helical-hairpin model and predict that PPR5, for example, is $\sim 55\%$ α -helical (<http://predictprotein.org>). However, structural data for a PPR motif have not been reported. To gain insight into the physical organization of a PPR tract, we analyzed rPPR5 by circular dichroism (CD) spectroscopy and analytical ultracentrifugation. Yields of rCRP1 were insufficient for these assays.

A CD spectrum for rPPR5 is shown in Figure 2A. Deconvolution of the data using several different methods revealed that rPPR5 is between 52% and 61% α helix (see Materials and Methods). These results are consistent with the helical-hairpin model for PPR tracts, and they also provide confidence that the rPPR5 used for the RNA binding assays described below is structured.

Thermal denaturation of rPPR5 was monitored by measuring the change in the CD signal at 222 nm as the temperature was increased (Fig. 2B). A sharp transition was observed at 39°C . Denaturation was irreversible and accompanied by protein aggregation, indicating that rPPR5 is highly unstable. As rPPR5 is an RNA binding protein (see below), it was of interest to determine whether it is more stable when bound to RNA. Repetition of the melt in the

FIGURE 2. Biophysical assays of rPPR5. (A) CD spectrum of rPPR5 at 0.12 mg/mL. (B) Change in CD signal at 222 nm as rPPR5 was increased in temperature from 0°C to 70°C in the absence (black line) or presence (gray line) of an equimolar amount of the nonspecific 31-mer RNA oligonucleotide described in Materials and Methods. (C) Sedimentation velocity profile of rPPR5 during analytical ultracentrifugation. (Left) Absorbance at 229 nm plotted as a function of position from the axis of rotation at time intervals of 5 min. Black lines represent nonlinear best fits to a continuous sedimentation distribution model. Gray lines represent experimental data and are shown for every fourth scan. (Right) Sedimentation coefficient, $c(s)$, profile for rPPR5 converted into the differential molar mass distribution, $c(M)$ using SEDFIT. Uncleaved MBP-PPR5 contaminated this preparation ($\sim 5\%$ of the total) and can account for the small peak at ~ 98 kDa. The residuals were randomly distributed, with 85% of the residuals having a value $<10\%$ of the signal and a maximum spread of 20% (data not shown).

presence of an equimolar amount of an RNA oligonucleotide (which is transparent at the wavelength assayed) showed that the stability of rPPR5 did increase in the presence of RNA (Fig. 2B).

To obtain additional insight into the physical properties of rPPR5, it was subjected to sedimentation velocity analysis in an analytical ultracentrifuge. Figure 2C shows the absorbance plotted as a function of radial position from the axis of rotation at 5-min intervals (initial protein concentration 0.3 mg/mL). Every fourth scan is overlaid with the continuous size-distribution best fit. The noise in the data (RMS = 0.039) is likely to be due to the presence of dithiothreitol and to the relatively low protein concentration, which we were unable to improve due to poor protein solubility. Nonetheless, the data show that >90% of rPPR5 exists as a single species and indicate a sedimentation coefficient ($S_{20,w}$) of ~ 3.48 S for this species, which was calculated with a frictional ratio (f/f_0) of 1.38. The molecular mass of rPPR5 was calculated from these data to be 44.5 kDa, which is in good agreement with its predicted monomeric mass of 51.6 kDa. These data indicate that rPPR5 is a monomer at the concentration assayed (6 μ M). Taken together, the gel filtration, CD spectroscopy, and analytical ultracentrifugation data support the view that

PPR5 is a monomeric protein that is largely α -helical. However, it remains possible that at higher protein concentrations PPR5 could form homomultimers.

rPPR5 and rCRP1 bind nucleic acids with a strong preference for single-stranded RNA relative to other nucleic acid forms

To study their nucleic acid binding properties, rPPR5 and rCRP1 were used in GMS assays with synthetic RNA and DNA oligonucleotides (Fig. 3). These initial experiments addressed general features of their interactions with nucleic acids rather than the basis for the sequence specificity they exhibit *in vivo*. Results were similar with a variety of different nucleic acid sequences; the representative data shown here used two different sequences from the maize chloroplast *petA* 5' UTR.

Figure 3A shows that rCRP1 bound with high affinity to two single-stranded 31-mer RNA oligonucleotides (*petA*-1 and *petA*-2, ssRNA). rCRP1 had only very low affinity for single-stranded DNA 31-mers (ssDNA) of the same sequences. Double-stranded RNA and DNA substrates were generated by hybridization of radiolabeled 31-mers to a fivefold excess of the complementary unlabeled 31-mers. rCRP1

FIGURE 3. rPPR5 and rCRP1 bind preferentially to single-stranded RNA relative to single-stranded DNA or double-stranded nucleic acids. Gel mobility shift assays were performed with the indicated concentrations of rCRP1 (A) or rPPR5 (B) and 31-mer oligonucleotides in single-stranded DNA (ssDNA), double-stranded DNA (dsDNA), single-stranded RNA (ssRNA), or double-stranded RNA (dsRNA) form. The *petA*-2 and *petA*-1 nucleic acid sequences are from two different regions of the *petA* 5' untranslated region. For assays with single-stranded RNA or DNA, oligonucleotides were heated and then snap-cooled on ice immediately prior to protein addition. Double-stranded oligonucleotides were generated by solution hybridization with the complementary oligonucleotides and were used either directly in binding assays (left and middle panels), or after gel purification of the double-stranded form (right panels).

bound little, if at all, to dsRNA and dsDNA when the hybridization mixture was used directly in the binding reactions (Fig. 3A, left and middle panels) or when the double-stranded forms were gel-purified prior to use in the binding reactions (Fig. 3A, right panel). The trace binding observed in some assays using double-stranded substrates is likely due to the presence of contaminating single-stranded oligonucleotides.

rPPR5 displayed properties similar to those of rCRP1 in analogous assays (Fig. 3B). rPPR5 exhibited a strong preference for both single-stranded RNA 31-mers relative to the corresponding double-stranded RNAs, single-stranded DNAs, or double-stranded DNAs. As for rCRP1, use of gel-purified dsRNA substrates confirmed that rPPR5 binds poorly, if at all, to double-stranded RNA (Fig. 3B, right panel).

rPPR5 binds with specificity to its native RNA ligand in vitro

Co-immunoprecipitation and genetic analyses have identified the group II intron-containing precursor of chloroplast *trnG*-UCC as a ligand of PPR5: PPR5 is associated specifically with unspliced *trnG*-UCC RNA in chloroplast extract, and it is required for the accumulation of both unspliced and spliced *trnG*-UCC in vivo (Beick et al. 2008). To determine whether PPR5 has intrinsic specificity for this RNA, rPPR5 was used in GMS assays with six overlapping RNAs that together span almost all of the ~900 nucleotides (nt) of unspliced *trnG*-UCC (Fig. 4A; Supplemental Table 1). In low concentrations of heparin (2 ng/ μ L), rPPR5 bound similarly to all six RNAs (data not shown); with increased heparin (1 μ g/ μ L), however, a preference for one RNA emerged (Fig. 4A, *trnG*-4). RNA *trnG*-3, which overlaps *trnG*-4, also showed substantial binding to rPPR5, yielding a complex that migrated slightly behind unbound RNA. This interaction is consistent with the fine-mapping experiments described below, in that the overlap between *trnG*-3 and *trnG*-4 includes approximately half of the PPR5 binding site. That the *trnG*-4 RNA includes a specific binding site for rPPR5 was further supported by the results of the competition assays shown in Figure 4B. The weak interactions detected with several other RNAs likely represent nonspecific binding, as similar results were obtained with a variety of similarly sized RNAs derived from chloroplast transcripts that show

no genetic or RIP-chip evidence for interaction with PPR5 in vivo (data not shown). Nonetheless, we cannot eliminate the possibility that there are additional binding sites for PPR5 within the *trnG* intron, albeit of substantially lower affinity than that in *trnG*-4.

To pinpoint PPR5's high-affinity binding site, eight overlapping ~50-mer RNAs that spanned the region encompassed by *trnG*-3, *trnG*-4, and most of *trnG*-5 were used in GMS assays (Fig. 5A). Strong specificity for a single 51-mer RNA, *trnG*-iii, was observed. Deletion of 9 nt at the 3' end or 6 nt at the 5' end of this RNA caused a strong reduction in binding (Fig. 5B, fragments D,E). Deletion of even 4 nt at the 3' end reduced binding in comparison to the intact 51-mer, albeit modestly (Fig. 5B, cf. RNAs I and iii). These effects were not due simply to a reduction in RNA length because the loss of binding caused by deletion of 10 nt at the 5' end was not compensated by an extension of 10 nt at the 3' end (Fig. 5B, cf. RNAs N and N+10). Furthermore, A-to-C nucleotide substitutions at positions 33, 35, and 38 in the context of the complete 51-mer caused a clear decrease in rPPR5 binding (Fig. 5B, cf. RNAs iii and K). Taken together, these results show that nucleotides in *trnG*-iii near residues 5 (Fig. 5B, RNAs E,N), 35 (Fig. 5B, RNA K), and 45 (Fig. 5B, RNAs D,I) all contribute to a high-affinity interaction with PPR5. Thus, the PPR5 binding site encompasses at least 40 nt. Extension of the *trnG*-iii 51-mer sequence by 10 nt at both ends did not increase binding

FIGURE 4. Identification of a high-affinity rPPR5 binding site within the *trnG*-UCC intron. Gel mobility shift assays used body-labeled RNAs with the indicated concentrations of rPPR5. To minimize the formation of RNA structure, RNAs were heated in the absence of salts and snap-cooled immediately prior to the addition of salts and protein. (A) Preferential binding of rPPR5 to a segment of the *trnG*-UCC group II intron. The unspliced *trnG*-UCC RNA is diagrammed at the top, with exons denoted as rectangles, and the six group II intron domains indicated. The positions of the six RNAs assayed for binding are diagrammed below. The fraction of the input RNA bound to protein was quantified with a phosphorimager, and is shown below. (B) Competition assays demonstrating a sequence-specific interaction between *trnG*-4 RNA and rPPR5. Binding reactions included radioactive *trnG*-4 RNA together with a 10- or 100-fold molar excess of the indicated nonradioactive RNAs. [rPPR5] was held constant at 400 nM.

FIGURE 5. Fine-mapping the PPR5 binding site. (A) Identification of a 51-nt segment of the *trnG*-intron to which rPPR5 binds with specificity. The overlapping ~ 50 -mer RNAs diagrammed beneath the gene map were used in GMS assays. Each RNA included an additional three G residues at its 5' end, introduced by the promoter in the transcription template. (B) Mutational analyses to map the RNA sequences that contribute to sequence-specific RNA recognition by PPR5. The sequence of each RNA used in binding reactions is annotated with the position of the predicted EBS1 and α' group II intron motifs (see Fig. 7A). Each RNA except RNA E included two additional G residues at its 5' end, introduced by the promoter in the transcription template. The three point mutations in RNA K are underlined.

affinity (data not shown), suggesting that the PPR5 binding site is contained entirely within the *trnG*-iii 51-mer.

The *trnG*-iii 51-mer RNA is predicted to include a stable hairpin structure (see Fig. 7A below). The 5' and 3' deletions described above provided evidence for recognition determinants mapping near both ends of this RNA, which are predicted to be single stranded. This is consistent with rPPR5's strong preference for single-stranded RNA relative to double-stranded RNA (Fig. 3). To address whether the loop in the predicted hairpin contributes to a high-affinity interaction with rPPR5, the 4 nt in the loop were deleted (Fig. 5B, construct iii Δ) or mutated to guanines (data not shown). Neither change caused a significant reduction in binding affinity. Taken together, the binding experiments suggest that PPR5 interacts with sequences near the 5' and 3' ends of the *trnG*-iii 51-mer but not in internal sequences corresponding to a predicted loop.

The 3' end of a truncated pre-*trnG*-UCC RNA that accumulates in hypomorphic *ppr5* mutants maps to the PPR5 binding site

In wild-type maize leaves, spliced *trnG*-UCC and its unspliced precursor are the predominant transcripts from the *trnG*-UCC locus (Watkins et al. 2007). A minor transcript of intermediate size (~ 440 nt) can also be detected that contains exon 1 and approximately half of the intron (Fig. 6A; Beick et al. 2008). Hypomorphic *ppr5*

alleles cause an increase in the abundance of this truncated transcript and a corresponding decrease in the precursor form (Fig. 6A, probe *trnG*-1). However, the rate of transcription across the *trnG*-UCC locus in plants harboring these alleles appears normal in chloroplast transcription run-on experiments (Beick et al. 2008). These results suggest that the truncated transcript results from an endonucleolytic cleavage event that is inhibited by PPR5. According to this scenario, PPR5 promotes *trnG*-UCC expression by increasing the pool of intact precursor available for splicing.

To gain insight into how PPR5 might stabilize the unspliced *trnG*-UCC precursor, we mapped the 3' terminus of the truncated transcript. The size of this transcript was estimated to be ~ 440 nt based on its mobility with respect to RNA standards in agarose gels; this size placed its 3' end within or very close to the PPR5 binding site defined from the in vitro binding data (nucleotides 403–453 with respect to the *trnG*-UCC 5' end). To refine this placement, RNA gel blots were probed with several of the ~ 50 -mer oligonucleotides that were used as RNA templates in the binding assays (Fig. 6A, right panels). Probe *trnG*-i detected the truncated RNA and the intact precursor in a ratio similar to that detected with the upstream probe *trnG*-1, suggesting that this probe hybridizes across its length to the truncated RNA. The *trnG*-iii probe also detected the truncated RNA but with reduced sensitivity in comparison to the precursor, whereas the overlapping

FIGURE 6. Mapping the 3' end of a truncated *trnG-UCC* transcript whose accumulation is influenced by PPR5. (A) Total leaf RNA (5 μ g) from seedlings of the indicated genotypes was analyzed by RNA gel blot hybridization. The probes, diagrammed below a map of the *trnG-UCC* locus, were the same DNAs used to generate RNAs for the binding assays in Figure 5A. Estimated transcript sizes, indicated to the right, were based on their mobility with respect to RNA size markers. The $\sim$ 440 nt transcript was shown elsewhere to include the 5' exon and to lack the 3' exon (Beick et al. 2008). The analyzed *ppr5* mutant alleles have an incomplete loss of *ppr5* function; null mutants are uninformative because their severe plastid ribosome deficiency causes pleiotropic effects that mask defects caused directly by the absence of PPR5. (B) Total leaf RNA (2 μ g) from seedlings of the indicated genotypes was analyzed with a ribonuclease protection assay. The diagram shows the position of the 293-nt body-labeled probe with respect to the PPR5 binding site; residue 1 represents the 3' end of the probe. A reaction using an equivalent amount of yeast tRNA served as a negative control (tRNA). *hcf7* mutants exhibit a plastid ribosome deficiency of similar magnitude to that in the *ppr5* mutant alleles analyzed and were included to control for secondary effects of this ribosome deficiency. The estimated transcript sizes, indicated to the right, are based upon DNA size standards after calibration for RNA by an adjustment of 11%; this adjustment was based on the mobility of the 293-nt RNA probe with respect to the DNA standards. The protected RNAs of 293 and 272 nt both result from hybridization to intact intron RNA; the 272 fragment results from a sequence polymorphism between the genetic backgrounds used to generate the probe and the leaf RNA (data not shown).

probe trnG-iv did not detect the truncated transcript at all. These results provide strong evidence that the 3' end of the truncated transcript maps within the trnG-iii 51-mer sequence, the same sequence that harbors the PPR5 binding site. This conclusion is further supported by the results of a ribonuclease protection assay (Fig. 6B): The abundance of a protected probe fragment of $\sim$ 212 nt correlates with that of the truncated RNA detected on RNA gel blots, and the size

of this fragment is consistent with a 3' terminus mapping in the 3' half of the PPR5 binding site.

These results show that the 3' end of the truncated *trnG-UCC* transcript maps within the PPR5 binding site, which was established independently through in vitro experiments. This finding, together with the increased ratio of the truncated transcript relative to the precursor in hypomorphic *ppr5* mutants, suggests that PPR5 stabilizes unspliced trnG-UCC by binding to a nuclease-sensitive site in the intron and protecting it from endonucleolytic cleavage.

DISCUSSION

Current understanding of the structure and functions of PPR motifs is based largely on genetic data and on extrapolation from structural data for the related TPR motif. Analysis of the nucleic acid binding properties of PPR proteins has been hindered by the fact that recombinant PPR proteins are typically insoluble unless fused to solubility-enhancing affinity tags, which could alter the properties of the associated PPR tract. Thus, despite several reports that recombinant PPR proteins can bind RNA, DNA, or both, experimental data that elucidate the physical properties of PPR proteins and quantitative analyses of the substrate preferences for PPR tracts have been lacking. We present evidence here that two genetically characterized PPR proteins, PPR5 and CRP1, do not form homomultimers at physiologically relevant concentrations, that they bind single-stranded RNA with high affinity, and that they have much lower affinity for single-stranded DNA or double-stranded nucleic acids. We show further that PPR5 is largely α -helical, in agreement with structural models for PPR tracts that posit a structure akin to that adopted by TPR tracts (Small and Peeters 2000). Finally, we show that PPR5 binds with specificity to a region of $\sim$ 50 nt within its established physiological RNA ligand and that this binding site correlates with the terminus of a truncated transcript that accumulates in a PPR5-dependent manner.

Nucleic acid binding properties of PPR proteins

Preferential binding of PPR proteins to single-stranded DNA relative to double-stranded DNA (Tsuchiya et al. 2002; Nakamura et al. 2003) or to RNA homopolymers relative to DNA (Lurin et al. 2004) has been reported. However, these assays were not performed in a manner that allowed quantitative comparisons of the binding of purified PPR proteins to nucleic acids of the same sequence in these various forms, and assays that tested the binding of PPR proteins to double-stranded RNA had not been reported. Thus, one goal of this study was to better establish the preferred nucleic acid ligands of PPR tracts. rCRP1 and rPPR5 bound with high affinity to two different single-stranded RNA 31-mers, bound with much lower affinity to single-stranded DNA of the same sequences and bound

little, if at all, to the corresponding double-stranded RNAs and DNAs (Fig. 3). Strong discrimination against single-stranded DNA was also observed for the PUM-homology domain (PUM-HD), the one structurally characterized helical repeat motif that has been demonstrated to bind RNA (Wang et al. 2002). The DNA binding activity reported for several PPR proteins (Ikeda and Gray 1999; Tsuchiya et al. 2002; Nakamura et al. 2003; Koussevitzky et al. 2007) could result from a contribution by a protein region outside the PPR tract or from sufficiently high protein or nucleic acid concentrations to drive low-affinity interactions. Alternatively, PPR tracts might fall into multiple classes, with some tailored for RNA binding and others permissive for DNA binding. Direct comparisons among a variety of PPR proteins in quantitative nucleic acid binding assays will be required to distinguish among these possibilities.

Sequence-specific RNA recognition by PPR5

A significant question is whether PPR proteins generally have intrinsic specificity for the RNAs whose metabolism they influence *in vivo*. Sequence-specific binding of two PPR proteins was inferred based on the ability of some RNA sequences to compete more effectively than others in UV-crosslinking or GMS binding assays (Meierhoff et al. 2003; Nakamura et al. 2003; Okuda et al. 2006). Co-immunoprecipitation assays showed that both PPR5 and CRP1 are associated *in vivo* specifically with those chloroplast RNAs whose metabolism they influence (Schmitz-Linneweber et al. 2005; Beick et al. 2008). Like many RNA binding proteins, PPR5 binds RNA in a sequence non-specific fashion *in vitro* unless conditions are optimized to reveal sequence-specificity. Only in the presence of high concentrations of heparin and salt does rPPR5's preference for specific sequences emerge. These results show that PPR5's specificity for the *trnG-UCC* intron *in vivo* can be accounted for by properties intrinsic to PPR5.

Binding assays with variants of the 51-mer harboring the high-affinity PPR5 site showed that this site spans at least 40 nt. This is an unusually long binding site for a single-stranded RNA binding protein (for review, see Auweter et al. 2006), surpassing even the 22-nt minimal site described for the *Caenorhabditis elegans* PUM-HD protein FBF-1 (Bernstein et al. 2005). Because PPR5 has strong preference for single-stranded RNA relative to double-stranded RNA (Fig. 3), the binding assays were performed under conditions designed to minimize the formation of RNA structure: the RNA was heated in the absence of salts and snap-cooled on ice, with salt and protein then added in rapid succession. In this context, it is interesting that the PPR5 binding site is predicted to include a stable hairpin structure (Fig. 7A). The 5' deletions that impinge on the predicted hairpin cause a decrease in PPR5 binding (Fig. 5B), suggesting that some PPR5 recognition determinants can be found in this region. It seems unlikely that PPR5 binds this RNA in

FIGURE 7. Models of the PPR5 binding site. (A) Predicted secondary structure of the *trnG-iii* 51-mer RNA to which PPR5 binds with high specificity. The structure was calculated by Mfold (Zuker 2003) at <http://frontend.bioinfo.rpi.edu/applications/mfold/>, using default parameters, and has a calculated $\Delta G = -15.3$ kcal/mol. The positions of the EBS1 and α' motifs are marked; EBS1 was assigned based on the *trnG-UCC* intron model at the Comparative RNA website (<http://www.rna.cccb.utexas.edu/>); α' comes from a model generated by C. Wang and S. Zimmerly (pers. comm.). The 3' end of the truncated transcript whose accumulation is suppressed by PPR5 maps within several nucleotides of residue 30, on the 3' side of the diagrammed hairpin. (B) Models for the interaction between PPR5 and its RNA ligand. The models assume that the RNA in contact with PPR5 is single stranded, as suggested by the data in Figure 3. Each of the 10 PPR motifs in PPR5 is represented by a cylinder corresponding to predicted "helix A," which has been proposed to present the substrate binding face of PPR tracts (for review, see Delannoy et al. 2007). These models diagram a PPR5 monomer because rPPR5 is monomeric under the conditions used in the binding assays; however, cooperative binding of multiple monomers remains a possibility. The models invoke either a single contiguous binding site spanning at least 40 nt (bottom) or a discontinuous binding site that includes sequences near both ends of the 51-mer (top). A discontinuous binding site is best supported by current data.

double-stranded form, given PPR5's negligible interaction with double-stranded RNA under nonspecific binding conditions (Fig. 3). There is precedent, however, for single-stranded RNA binding proteins that can disrupt RNA structures by capturing transiently unpaired nucleotides (for review, see Cristofari and Darlix 2002; Schroeder et al. 2002). The influence of PPR5 on this hairpin may be relevant to its possible role in splicing (see below) and is under investigation.

It is interesting to consider models for RNA recognition by PPR5 that are consistent with its strong preference for single-stranded RNA and its >40 nt minimal binding site. It seems likely that PPR5 binds this RNA as a monomer because rPPR5 did not dimerize at the considerably higher protein concentrations used for gel filtration chromatography and analytical ultracentrifugation. Independent binding of several monomers to the RNA seems unlikely given

that the gel mobility shifts detected a single complex at all protein concentrations tested; however, highly cooperative binding of multiple monomers remains a possibility. PPR5 consists almost entirely of its PPR tract, with 10 tandem PPR motifs flanked by ~ 35 amino acids at its mature N terminus and 55 amino acids at its C terminus. These short flanking sequences could contribute to RNA recognition, but it seems likely that the PPR tract plays the major role. Based on current data, we favor two general models for PPR5 recognition of its specific binding site (Fig. 7B): (1) PPR5 could bind a single contiguous binding site in its RNA ligand; in this case, the entire ~ 40 -mer RNA would be spread in single-stranded form along the PPR5 surface, with each PPR repeat spanning several nucleotides along the RNA chain. The stoichiometry of protein and RNA in the complex is unknown, so speculation about the number of nucleotides spanning each repeat would be premature. (2) Alternatively, PPR5 could recognize two discrete single-stranded RNA segments within its ~ 40 nt binding site, with the intervening RNA looped out. We favor this general model because the deletion data provide strong evidence that sequences near both the 5'- and 3'-ends of the 51-mer sequence shown in Figure 7A contribute to recognition, whereas the 4 nt corresponding to the predicted internal loop (UUUA) do not. Thus, PPR5 might bind to single-stranded sequences flanking the stem-loop, possibly destabilizing the base of the hairpin such that some nucleotides that are paired in the absence of PPR5 are bound in single-stranded form to PPR5. In any case, the results showing that nucleotides near residues 5, 35, and 45 in the *trnG-iii* 51-mer all contribute to a high-affinity interaction with PPR5 (see Fig. 5, and associated text) argue against a recognition mechanism that is closely akin to that employed by a human PUM-HD, for which each of the eight helical repeats binds one of the eight contiguous nucleotides in its core binding site (Wang et al. 2002).

Role of PPR5 in metabolism of the *trnG-UCC* group II intron

The reduced level of the *trnG-UCC* precursor in *ppr5* mutants can account for their lack of spliced *trnG-UCC* (Beick et al. 2008), but it remains possible that PPR5 also plays a direct role in facilitating the splicing of the *trnG-UCC* intron. The position of the PPR5 binding site suggests how PPR5 might simultaneously stabilize the unspliced precursor and directly promote intron splicing. First, the PPR5 binding site corresponds to the 3' end of a truncated pre-*trnG-UCC* transcript (Fig. 6). The ratio of the truncated form to intact precursor is higher in *ppr5* mutants than in wild-type plants, yet transcription across the gene appears not to be PPR5-dependent (Beick et al. 2008). Together, these results support a model in which PPR5 binding masks a ribonuclease-sensitive site on the 3' side of the hairpin diagrammed in Figure 7A.

That PPR5 might also function directly in *trnG-UCC* splicing is suggested by the fact that the PPR5 binding site includes two important group II intron functional elements (Fig. 7A; for review, see Pyle and Lambowitz 2006): the EBS1 site, which pairs with the complementary IBS1 sequence in exon 1 to define the 5' splice junction, and the α' site, which pairs with the α element near the 5' end of intron domain 1. The EBS1 site in this intron is unusual: whereas in the vast majority of group II introns EBS1 is in a single-stranded loop and thus accessible for pairing with IBS1, in the *trnG-UCC* intron EBS1 is predicted to be partially sequestered in a hairpin (Fig. 7A). Thus, PPR5 binding to single-stranded RNA near EBS1 might shift the equilibrium toward an unstructured EBS1 region and thereby facilitate the EBS1/IBS1 interaction. Two evolutionary routes to the acquisition of dual roles for PPR5 seem plausible. PPR5 might have initially served to compensate for a nuclease sensitive site in the intron; this interaction might then have been permissive for the acquisition of an RNA structure that would sequester EBS1 in the absence of PPR5. Alternatively, an initial role for PPR5 in enhancing the accessibility of EBS1 by disrupting the RNA duplex could have allowed the acquisition of mutations in the PPR5 binding site that confer nuclease sensitivity in the absence of PPR5.

The results described here establish PPR5 as a promising model for the elucidation of the biochemical and structural basis of sequence-specific RNA binding by PPR proteins. Analogous studies of additional PPR proteins with well-characterized *in vivo* ligands will be necessary to deduce the general themes and variations regarding the biochemical mechanisms employed by this large and poorly understood protein family.

MATERIALS AND METHODS

Expression and purification of rCRP1 and rPPR5

Maize CRP1 is orthologous to *Arabidopsis* At5g42310 and rice Os07g36390 (<http://pogs.uoregon.edu>). A DNA fragment encoding mature CRP1 was generated by PCR from a CRP1 cDNA (GenBank accession no. AF073522) using the primers 5'-GAAGATCTCGCTACGACTTCGAGCCCCT-3' and 5'-TGCGGTCGACATATGCAACCCTCATGT-3'. This fragment was digested with BglII and SalI and cloned into the BamHI and SalI sites of pMAL-TEV to generate plasmid pmaltevCRP1(+)-N-term. pMAL-TEV is pMAL-c2X (New England Biolabs) modified to include sequences encoding a TEV protease cleavage site between the EcoRI and BamHI sites. The protein produced after TEV protease cleavage starts at CRP1 amino acid 38 (26 amino acids before the transit peptide cleavage site predicted by ChloroP [Emanuelsson and Heijne 2001]) and continues to the natural stop codon. Vector-encoded glycine and serine residues are at the N termini of both rCRP1 and rPPR5.

The MBP-CRP1 fusion protein was expressed by introducing pmaltevCRP1(+)-N-term into Rosetta 2 (DE3) cells (Novagen); cultures were grown at 37°C in LB with carbinicillin and

chloramphenicol to an OD₆₀₀ of 0.8 and transferred to ice for 20 min. Protein expression was then induced by the addition of IPTG to 1 mM and incubation for 4 h at 20°C. Harvested cells were suspended in ice cold lysis buffer (50 mM Tris-HCl, at pH 8, 750 mM NaCl, 10 mM β-mercaptoethanol, 0.13 mM phenylmethylsulfonyl fluoride, 0.5 μg/mL leupeptin, 0.06 μg/mL pepstatin, 0.01% CHAPS). Cells were lysed at 4°C by sonication, cooled on ice for 5 min, and then sonicated a second time. The lysate was cleared by centrifugation at 13,000g for 20 min. The MBP-CRP1 fusion protein was purified by incubating the cleared lysate with amylose-coupled agarose resin (New England Biolabs) at 4°C for 1 h with rotation. After one wash with lysis buffer, protein was eluted in the same buffer supplemented with 15 mM maltose. CHAPS detergent was added to the eluted protein (0.1% final concentration), and the MBP moiety was cleaved off by digestion with 15 μg of TEV protease (purified from pRK793-transformed BL21-RIL cells as described by Kapust et al. [2001]) overnight on ice. Following centrifugation at 13,000g for 10 min and filtration (Amicon Ultrafree-MC, 0.45 μm) to remove particulates, the protein was resolved on a 1 cm × 30 cm Superdex 200 column (GE Healthcare Life Sciences) in 50 mM Tris-HCl, at pH 8, 750 mM NaCl, 1% glycerol, 0.05% CHAPS, 10 mM β-mercaptoethanol. The peak of monomeric rCRP1 was dialyzed against a buffer containing 40 mM Tris-HCl, at pH 8, 750 mM NaCl, 50% glycerol, 10 mM β-mercaptoethanol, 1 mM EDTA, and stored in aliquots at -80°C. Approximately 100 μg of purified rCRP1 was recovered per liter of culture.

Maize PPR5 is orthologous to *Arabidopsis* At4g39620 and to rice Os02g51480 (<http://pogs.uoregon.edu>). To generate rPPR5, a PCR product was generated from a *ppr5* cDNA (GenBank accession no. EU037901) by PCR amplification with primers 5'-CGGGATCCGCGCGGAGGGGGTGG-3' and 5'-CGCGTCGACTTATGTAGCAGCTACATGCCA-3'. The PCR product was digested with BamHI and SalI and then cloned into Bam-HI/Sal-I cut pMAL-TEV to generate plasmid pmalteVPPR5. The protein produced after cleavage with TEV protease starts at PPR5 amino acid 46, which corresponds to the transit peptide cleavage site predicted by ChloroP, and ends at the natural stop codon.

The MBP-PPR5 fusion protein was expressed by introducing pmalteVPPR5 into Rosetta 2 (DE3) cells (Novagen). Cultures were grown and induced as for rCRP1. Harvested cells were suspended in ice-cold PPR5 lysis buffer (50 mM Na phosphate, at pH 7.5, 100 mM NaCl, 10 mM β-mercaptoethanol, 0.13 mM phenylmethylsulfonyl fluoride, 0.5 μg/mL leupeptin, 0.06 μg/mL pepstatin, 0.01% CHAPS). Cells were lysed at 4°C by sonication, cooled on ice for 5 min, and then sonicated a second time. The lysate was cleared by centrifugation at 13,000g for 20 min. The MBP-PPR5 fusion protein was purified from the soluble lysate by amylose-affinity chromatography and the MBP moiety was cleaved off as described above for rCRP1. To resolve rPPR5 from MBP, the mixture was applied to a small Superdex 200 column, from which rPPR5 eluted in the void volume due to its association with nucleic acids. To separate nucleic acids from rPPR5, NaCl was added to 300 mM, and the mixture was incubated on ice for 30 min and then applied to a Q-Sepharose fast flow ion-exchange column (GE Healthcare Life Sciences). The flowthrough was centrifuged at 13,000g for 10 min, filtered (Amicon Ultrafree-MC, 0.45 μm), and resolved on a 1 cm × 30 cm Superdex 200 column (GE Healthcare Life Sciences) in 50 mM Na phosphate, at

pH 7.5, 300 mM NaCl, 5 mM β-mercaptoethanol. rPPR5 used for nucleic acid binding assays was dialyzed against 50 mM Na phosphate, at pH 7.5, 300 mM NaCl, 5 mM β-mercaptoethanol, 50% glycerol, and stored at -20°C. Approximately 500 μg of purified rPPR5 were recovered per liter of culture. The purified proteins had an A₂₆₀/A₂₈₀ ratio=0.6, indicating little, if any, nucleic acid contamination.

CD spectroscopy

rPPR5 was dialyzed into CD buffer (10 mM Na phosphate, at pH 7.5, 100 mM NaCl, 0.5 mM DTT). rPPR5 protein concentration for CD spectroscopy was determined using a Varian 3E spectrophotometer and a computed molar absorptivity for rPPR5. CD spectra were measured using a JASCO model 710 spectropolarimeter calibrated using holmium oxide glass and magnitude-calibrated using (1S)-(+)-10-camphorsulfonic acid. Temperature was maintained at 4°C. After subtraction of baseline spectra, data were converted to [θ] and displayed against wavelength using JASCO Spectra Manager software. CD spectra were measured from 340 to 185 nm, using both 0.1- and 1-cm path lengths at a protein concentration of 0.12 mg/mL. Data from the 1- and 0.1-cm cells were internally consistent when comparing those wavelengths where both sets of data were valid. rPPR5 was found to interact strongly with the walls of the optical cell. Material immobilized on the cell walls had a diminished peptide CD spectrum. From comparison between 1- and 0.1-cm path length data, we estimate errors in the CD signal of not more than ±10% due to this effect. Secondary structure contributions to the CD spectra were deconvoluted using the K2d (Andrade et al. 1993) and CDSSTR programs at DICHROWEB (Whitmore and Wallace 2004). CDSSTR analysis using reference set 4 indicated 52% α helix, 19% β strand, 9% turn, and 20% disordered, with an NRMSD of 0.003; analysis using reference set 7 yielded 62% α helix, 15% strand, 9% turns, 15% disordered, with an NRMSD of 0.004. K2d analysis indicated 61% α helix, 6% β sheet, and 33% random coil, with an NRMSD of 0.111. Melting profiles were collected with rPPR5 at 0.12 mg/mL in a 1-cm cell at 222 nm, with temperature raised from 4°C to 75°C. The melt was repeated in the presence of an equimolar amount of a nonspecific RNA oligonucleotide (5'-UUGUCAUUCUACUGCAAUAGUUGUACUAGAA-3').

Analytical ultracentrifugation

Sedimentation velocity experiments were carried out in a Beckman XL-I analytical ultracentrifuge; 400 μL of rPPR5 (0.3 mg/mL or 6 μM) in 10 mM Na phosphate, 100 mM NaCl, 0.5 mM dithiothreitol, at pH 7.5, and a 410 μL reference sample were loaded into a conventional double sector filled Epon centerpiece (path length 1.2 cm) with quartz windows, and mounted in a Beckman An60 titanium rotor. Experiments were conducted at 20°C with a rotor speed of 55,000 rpm, and the data were collected in continuous mode for 8 h, at a single wavelength (229 nm). Multiple scans at different time points were fitted to a continuous size distribution model using the program SEDFIT (Schuck 2000). The root mean square deviation was 0.039, and the residuals were randomly distributed. Solvent densities and viscosities were computed using SEDNTERP (V1.06, developed by D. B. Haynes, T. Laue and J. Philo; <http://www.rasmb.bbri.org/>). The

partial specific volume of rPPR5 (0.736 mL/g at 20°C) was calculated from its amino acid composition.

Nucleic acid binding assays

GMS assays used gel-purified synthetic RNA and DNA 31-mers derived from two regions of the 5'-untranslated region of the maize chloroplast *petA* mRNA (*petA*-1; 5'-TTCTAGTACAAC TATTGCAGTAGAATGACAA-3') and (*petA*-2; 5'-TTAGCTACC TATCTAATTTATTGTAGAAATT-3') in both sense and antisense orientations. These RNAs are not predicted to form stable structures or to form similar weak structures by Mfold (data not shown). The sense strands were 5'-end labeled with T4 polynucleotide kinase and [γ -³²P]ATP and purified by ethanol precipitation. To generate double-stranded substrates, the radiolabeled sense oligonucleotides (125 pM) were annealed with a fivefold excess of unlabeled antisense oligonucleotide by heating the samples (10 μ L) to 95°C for 2 min, adding 10 μ L of 2 $\times$ annealing buffer (80 mM Tris-HCl, at pH 7.5, 200 mM NaCl, 10 units of RNasin [Promega], 0.05 mg/mL BSA, 10 mM DTT) and slowly cooling over 30 min to 37°C. Single-stranded substrates were heated and snap-cooled on ice prior to addition of the same buffer. rCRP1 and rPPR5 were diluted in their respective dialysis buffers, and 5 μ L of protein was mixed with 20 μ L of nucleic acid and incubated at 30°C for 30 min. The final binding reactions contained 40 mM Tris-HCl, at pH 7.5, 230 mM NaCl (rCRP1) or 140 mM NaCl (rPPR5), 4 mM DTT, 0.2 mM EDTA, 0.04 μ g/mL BSA, 10% glycerol, 10 units RNasin, 25 pM radiolabeled oligonucleotide, and the protein concentrations indicated in Figure 3. Samples were applied to a running 5% polyacrylamide (29:1 acrylamide/bis-acrylamide) gel prepared in 1 $\times$ THE (66 mM HEPES, 34 mM Tris, at pH 7.5, 0.1 mM EDTA). Gels were run at 15 W (constant power) at 4°C with buffer recirculation until the bromophenol blue dye in an adjacent lane had migrated 5 cm. Gels were fixed, dried, and imaged with a phosphorimager.

RNAs used to determine the sequence-specificity of rPPR5 were derived from the maize chloroplast *trnG-UCC* locus; their coordinates on the chloroplast genome are listed in Supplemental Table 1. These RNAs were generated by in vitro transcription from PCR products with a T7 promoter incorporated into the 5' primer sequence (Fig. 4), or from synthetic single-stranded DNA templates with an annealed T7 promoter oligonucleotide at one end (Fig. 5). The latter templates were generated by adding 1 μ L of each 70 μ M oligonucleotide stock to 33 μ L of annealing buffer (40 mM Tris-HCl, at pH 7.5, 50 mM NaCl), heating to 94°C, and slow cooling to 25°C. Transcription reactions (20 μ L) included 1 μ L of T7 RNA polymerase (Promega), 1 μ L of [α -³²P]UTP (800 Ci/mmole, 10 mCi/mL), 4 μ L of a mixture of ATP, CTP, and GTP (2.5 mM each), 1 μ L of 1 mM UTP, 1 μ L 100 mM DTT, 1 μ L RNasin (Promega), 0.2 μ M DNA template, and the transcription buffer supplied by the manufacturer. Reactions were incubated for 1 h at 37°C and then treated with 1 unit of RQ1 DNase (Promega) for 15 min at 37°C. The radiolabeled RNAs were purified on denaturing polyacrylamide gels.

Immediately prior to the binding reactions, the purified RNAs (10 μ L of a 100 pM solution) were heated to 94°C for 2 min and snap cooled on ice; 10 μ L of 2 $\times$ binding buffer (100 mM Na phosphate, at pH 7.5, 10 units RNasin, 0.1 mg/mL BSA, 10 mM DTT, 2.5 mg/mL heparin, and 225 mM; Fig. 5) or 350 mM NaCl (Fig. 4A) was then added to the RNA, followed immediately by

5 μ L of rPPR5 diluted in dialysis buffer. The mixture was transferred to a 25°C heat block and incubated for 30 min. The final binding reactions contained 50 mM Na phosphate, at pH 7.5, 150 mM (Fig. 5) or 200 mM NaCl (Fig. 4A), 4 mM DTT, 0.04 mg/mL BSA, 10% glycerol, 1 mg/mL heparin, 10 units of RNasin, 40 pM radiolabeled RNA, and protein concentrations as indicated. Samples were applied to 5% polyacrylamide gels in 1 $\times$ THE buffer as described above, and the results were imaged with a phosphorimager. For the competition assays (Fig. 4B), salt and buffer conditions were identical to those used in Figure 5, except that heparin was at 2 mg/mL. Competitor RNAs were initially quantified by their absorbance at 260 nm and were then confirmed to be intact and at the expected concentrations by agarose gel electrophoresis. Competitor RNA at the indicated molar excess was combined with the radiolabeled RNA prior to the RNA denaturation step.

Analysis of *trnG-UCC* transcripts in *ppr5* mutants

The recovery and phenotypes of maize *ppr5* insertion mutants are described in Beick et al. (2008). *ppr5-1* is a null allele with a transposon insertion in exon1 and has an albino seedling lethal phenotype. *ppr5-2* is a hypomorphic allele with a transposon insertion upstream of the open reading frame and has a pale-green seedling lethal phenotype. The noncomplementing progeny of crosses between plants heterozygous for these alleles (*ppr5-1/-2*) are intermediate in phenotype between the two parental alleles. Total leaf RNA was extracted and analyzed by RNA gel blot hybridization as described previously (Watkins et al. 2007) except that the hybridization and washing were performed at 60°C for the 50-mer oligonucleotide probes. The ribonuclease protection assay was performed as described previously (Watkins et al. 2007), with 2 μ g of total leaf RNA, a uniformly labeled probe spanning the PPR5 binding site, and ribonuclease T1.

SUPPLEMENTAL DATA

Supplemental material can be found at <http://www.rnajournal.org>.

ACKNOWLEDGMENTS

We are extremely grateful to Walt Baase, Steve Weitzel, Bryan Warf, Sandra Greive, and Kenny Watkins for their generous help and advice with CD spectroscopy and analytical ultracentrifugation. We also thank Chen Wang and Steve Zimmerly for communicating their secondary structure model of the *trnG-UCC* intron, Christian Schmitz-Linneweber for stimulating discussions and for his collaboration on the study of the in vivo function of PPR5, and Kenny Watkins for comments on the manuscript. This work was supported by USDA-NRI grant 2006-35318-17380 to A.B.

Received March 20, 2008; accepted June 2, 2008.

REFERENCES

- Andrade, M.A., Chacon, P., Merelo, J.J., and Moran, F. 1993. Evaluation of secondary structure of proteins from UV circular

- dichroism spectra using an unsupervised learning neural network. *Protein Eng.* **6**: 383–390.
- Auweter, S.D., Oberstrass, F.C., and Allain, F.H. 2006. Sequence-specific binding of single-stranded RNA: Is there a code for recognition? *Nucleic Acids Res.* **34**: 4943–4959.
- Barkan, A., Walker, M., Nolasco, M., and Johnson, D. 1994. A nuclear mutation in maize blocks the processing and translation of several chloroplast mRNAs and provides evidence for the differential translation of alternative mRNA forms. *EMBO J.* **13**: 3170–3181.
- Beick, S., Schmitz-Linneweber, C., Williams-Carrier, R., Jensen, B., and Barkan, A. 2008. The pentatricopeptide repeat protein PPR5 stabilizes a specific tRNA precursor in maize chloroplasts. *Mol. Cell. Biol.* (in press). doi: 10.1128/MCB.00563-08.
- Bernstein, D., Hook, B., Hajarnavis, A., Opperman, L., and Wickens, M. 2005. Binding specificity and mRNA targets of a *C. elegans* PUF protein, FBF-1. *RNA* **11**: 447–458.
- Coffin, J.W., Dhillon, R., Ritzel, R.G., and Nargang, F.E. 1997. The *Neurospora crassa cya-5* nuclear gene encodes a protein with a region of homology to the *Saccharomyces cerevisiae* PET309 protein and is required in a post-transcriptional step for the expression of the mitochondrially encoded COXI protein. *Curr. Genet.* **32**: 273–280.
- Cristofari, G. and Darlix, J.L. 2002. The ubiquitous nature of RNA chaperone proteins. *Prog. Nucleic Acid Res. Mol. Biol.* **72**: 223–268.
- D'Andrea, L.D. and Regan, L. 2003. TPR proteins: The versatile helix. *Trends Biochem. Sci.* **28**: 655–662.
- Delannoy, E., Stanley, W.A., Bond, C.S., and Small, I.D. 2007. Pentatricopeptide repeat (PPR) proteins as sequence-specificity factors in post-transcriptional processes in organelles. *Biochem. Soc. Trans.* **35**: 1643–1647.
- Emanuelsson, O. and Heijne, G.V. 2001. Prediction of organellar targeting signals. *Biochim. Biophys. Acta* **1541**: 114–119.
- Fisk, D.G., Walker, M.B., and Barkan, A. 1999. Molecular cloning of the maize gene *cp1* reveals similarity between regulators of mitochondrial and chloroplast gene expression. *EMBO J.* **18**: 2621–2630.
- Gillman, J.D., Bentolila, S., and Hanson, M.R. 2007. The petunia restorer of fertility protein is part of a large mitochondrial complex that interacts with transcripts of the CMS-associated locus. *Plant J.* **49**: 217–227.
- Ikedu, T. and Gray, M. 1999. Characterization of a DNA-binding protein implicated in transcription in wheat mitochondria. *Mol. Cell. Biol.* **19**: 8113–8122.
- Kapust, R., Tözsér, K., Fox, J., Anderson, D., Cherry, S., Copeland, T., and Waugh, D. 2001. Tobacco etch virus protease: Mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng.* **14**: 993–1000.
- Kobayashi, K., Suzuki, K., Tang, J., Nagata, N., Ohya, K., Seki, H., Kiuchi, R., Kaneko, Y., Nakazawa, M., Matsui, M., et al. 2007. Lovastatin insensitive 1, a novel pentatricopeptide repeat protein, is a potential regulatory factor of isoprenoid biosynthesis in *Arabidopsis*. *Plant Cell Physiol.* **48**: 322–331.
- Koussevitzky, S., Nott, A., Mockler, T.C., Hong, F., Sachetto-Martins, G., Surpin, M., Lim, J., Mittler, R., and Chory, J. 2007. Signals from chloroplasts converge to regulate nuclear gene expression. *Science* **316**: 715–719.
- Lahmy, S., Barneche, F., Derancourt, J., Filipowicz, W., Delseny, M., and Echeverria, M. 2000. A chloroplastic RNA-binding protein is a new member of the PPR family. *FEBS Lett.* **480**: 255–260.
- Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., et al. 2004. Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* **16**: 2089–2103.
- Manthey, G.M. and McEwen, J.E. 1995. The product of the nuclear gene *PET309* is required for translation of mature mRNA and stability or production of intron-containing RNAs derived from the mitochondrial *COX1* locus of *Saccharomyces cerevisiae*. *EMBO J.* **14**: 4031–4043.
- Meierhoff, K., Felder, S., Nakamura, T., Bechtold, N., and Schuster, G. 2003. HCF152, an *Arabidopsis* RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast psbB-psbT-psbH-petB-petD RNAs. *Plant Cell* **15**: 1480–1495.
- Mili, S. and Pinol-Roma, S. 2003. LRP130, a pentatricopeptide motif protein with a noncanonical RNA-binding domain, is bound in vivo to mitochondrial and nuclear RNAs. *Mol. Cell. Biol.* **23**: 4972–4982.
- Mingler, M.K., Hingst, A.M., Clement, S.L., Yu, L.E., Reifur, L., and Koslowsky, D.J. 2006. Identification of pentatricopeptide repeat proteins in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **150**: 37–45.
- Nakamura, T., Meierhoff, K., Westhoff, P., and Schuster, G. 2003. RNA-binding properties of HCF152, an *Arabidopsis* PPR protein involved in the processing of chloroplast RNA. *Eur. J. Biochem.* **270**: 4070–4081.
- Okuda, K., Nakamura, T., Sugita, M., Shimizu, T., and Shikanai, T. 2006. A pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing. *J. Biol. Chem.* **281**: 37661–37667.
- Pusnik, M., Small, I., Read, L.K., Fabbro, T., and Schneider, A. 2007. Pentatricopeptide repeat proteins in *T. brucei* function in mitochondrial ribosomes. *Mol. Cell. Biol.* **27**: 6876–6888.
- Pyle, A. and Lambowitz, A. 2006. Group II introns: Ribozymes that splice RNA and invade DNA. In *The RNA world* (eds. R. Gesteland et al.), pp. 469–506. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Rivals, E., Bruyere, C., Toffano-Nioche, C., and Lecharny, A. 2006. Formation of the *Arabidopsis* pentatricopeptide repeat family. *Plant Physiol.* **141**: 825–839.
- Schmitz-Linneweber, C., Williams-Carrier, R., and Barkan, A. 2005. RNA immunoprecipitation and microarray analysis show a chloroplast pentatricopeptide repeat protein to be associated with the 5'-region of mRNAs whose translation it activates. *Plant Cell* **17**: 2791–2804.
- Schmitz-Linneweber, C., Williams-Carrier, R.E., Williams-Voelker, P.M., Kroeger, T.S., Vichas, A., and Barkan, A. 2006. A pentatricopeptide repeat protein facilitates the trans-splicing of the maize chloroplast rps12 pre-mRNA. *Plant Cell* **18**: 2650–2663.
- Schroeder, R., Grossberger, R., Pichler, A., and Waldsich, C. 2002. RNA folding in vivo. *Curr. Opin. Struct. Biol.* **12**: 296–300.
- Schuck, P. 2000. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78**: 1606–1619.
- Small, I. and Peeters, N. 2000. The PPR motif—A TPR-related motif prevalent in plant organellar proteins. *Trends Biochem. Sci.* **25**: 46–47.
- Tsuchiya, N., Fukuda, H., Sugimura, T., Nagao, M., and Nakagama, H. 2002. LRP130, a protein containing nine pentatricopeptide repeat motifs, interacts with a single-stranded cytosine-rich sequence of mouse hypervariable minisatellite Pc-1. *Eur. J. Biochem.* **269**: 2927–2933.
- Tsuchiya, N., Fukuda, H., Nakashima, K., Nagao, M., Sugimura, T., and Nakagama, H. 2004. LRP130, a single-stranded DNA/RNA-binding protein, localizes at the outer nuclear and endoplasmic reticulum membrane, and interacts with mRNA in vivo. *Biochem. Biophys. Res. Commun.* **317**: 736–743.
- Wang, X., McLachlan, J., Zamore, P.D., and Hall, T.M. 2002. Modular recognition of RNA by a human pumilio-homology domain. *Cell* **110**: 501–512.
- Watkins, K., Kroeger, T., Cooke, A., Williams-Carrier, R., Friso, G., Belcher, S., Wijk, K.v., and Barkan, A. 2007. A ribonuclease III domain protein functions in group II intron splicing in maize chloroplasts. *Plant Cell* **19**: 2606–2623.
- Whitmore, L. and Wallace, B.A. 2004. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* **32**: W668–W673. doi: 10.1093/nar/gkh371.
- Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**: 3406–3415.

RNA

A PUBLICATION OF THE RNA SOCIETY

Sequence-specific binding of a chloroplast pentatricopeptide repeat protein to its native group II intron ligand

Rosalind Williams-Carrier, Tiffany Kroeger and Alice Barkan

RNA 2008 14: 1930-1941 originally published online July 30, 2008

Access the most recent version at doi:[10.1261/rna.1077708](https://doi.org/10.1261/rna.1077708)

Supplemental Material

<http://rnajournal.cshlp.org/content/suppl/2008/07/31/rna.1077708.DC1>

References

This article cites 37 articles, 14 of which can be accessed free at:
<http://rnajournal.cshlp.org/content/14/9/1930.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *RNA* go to:

<http://rnajournal.cshlp.org/subscriptions>