The Plant Cell, Vol. 27: 403–416, February 2015, www.plantcell.org

Predictable Alteration of Sequence Recognition by RNA Editing Factors from Arabidopsis[OPEN]

Peter Kindgren,[a] Aaron Yap,[a] Charles S. Bond,[b] and Ian Small[a,1]

[a] Australian Research Council Centre of Excellence in Plant Energy Biology, The University of Western Australia, Crawley, WA 6009, Australia

[b] School of Chemistry and Biochemistry, The University of Western Australia, Crawley, WA 6009, Australia

ORCID IDs: 0000-0003-0947-2648 (P.K.); 0000-0002-9584-6783 (C.S.B.); 0000-0001-5300-1216 (I.S.)

RNA editing factors of the pentatricopeptide repeat (PPR) family show a very high degree of sequence specificity in the recognition of their target sites. A molecular basis for target recognition by editing factors has been proposed based on statistical correlations but has not been tested experimentally. To achieve this, we systematically mutated the pentatricopeptide motifs in the *Arabidopsis thaliana* RNA editing factor CLB19 to investigate their individual contribution to RNA recognition. We find that the motifs contributing significantly to the specificity of binding follow the previously proposed recognition rules, distinguishing primarily between purines and pyrimidines. Our results are consistent with proposals that each motif recognizes one nucleotide in the RNA target with the protein aligned parallel to the RNA and contiguous motifs aligned with contiguous nucleotides such that the final PPR motif aligns four nucleotides upstream of the edited cytidine. By altering S motifs in CLB19 and another editing factor, OTP82, and using the modified proteins to attempt to complement the respective mutants, we demonstrate that we can predictably alter the specificity of these factors in vivo.

INTRODUCTION

Members of the pentatricopeptide repeat (PPR) family of proteins can be found in all eukaryotes, from humans to algae, although they differ greatly in number between organisms. This protein family has massively expanded in terrestrial plants, which contain from ~100 (*Physcomitrella*) to over 1000 (*Selaginella*) PPR proteins (Fujii and Small, 2011). PPR proteins are targeted to organelles where they bind RNA and influence a wide range of processing events, such as formation of mRNA termini, splicing, and editing (reviewed in Barkan and Small, 2014). There are two main classes of PPR proteins, characterized by the nature of the PPR motifs they contain. P-class proteins have tandem arrays of the canonical 35-amino acid repeats (P motifs). PLS-class proteins contain characteristic triplets of P, L (35 to 36 amino acids in length), and S (31 amino acids) motifs (Lurin et al., 2004). The PLS-class proteins commonly have extra C-terminal domains that are implicated in RNA editing (Lurin et al., 2004), an alteration of specific bases in the RNA sequence that is essential for correct expression of many organellar transcripts (reviewed in Chateigner-Boutin and Small, 2010; Takenaka et al., 2013b). The process requires PPR proteins to convey the necessary RNA specificity by binding in close proximity to the cytidine that will be edited (Okuda and Shikanai, 2012). Each PPR editing factor displays a high degree of specificity, directing editing at a very limited number of RNA target sites (Hammani et al., 2009; Okuda and Shikanai, 2012).

Structurally, each PPR motif comprises two α-helices, the first of which contacts the RNA bases (Yin et al., 2013) and determines the binding specificity (Barkan et al., 2012). Both bioinformatic (Barkan et al., 2012; Takenaka et al., 2013a; Yagi et al., 2013) and structural studies (Yin et al., 2013) have indicated that the key positions for RNA base recognition (in order of importance) are position 6 within the first helix, position 1′ at the end of the loop immediately preceding this helix, and position 3 within the first helix, which appears to intercalate between adjacent bases in the RNA (numbering of positions follows that of Barkan et al., 2012). Several related RNA recognition codes have been proposed that give the correspondences between combinations of amino acids at these positions and specific ribonucleotides (Barkan et al., 2012; Takenaka et al., 2013a; Yagi et al., 2013). Position 6 appears to be important for distinguishing between purines (A, G) and pyrimidines (U, C), while position 1′ directs recognition of amino (A, C) or keto (G, U) groups (Yagi et al., 2013). Most of these correspondences are based on statistical correlations in alignments between PPR binding factors and their RNA targets, but only a few combinations have been confirmed experimentally, and these only for P motifs, and only in vitro (Barkan et al., 2012; Okuda et al., 2014). Based on statistical correlations, S motifs appear to deviate from the P motif code in some important respects and L motifs differ substantially from P and S motifs at the key position 6 (Barkan et al., 2012; Takenaka et al., 2013a). Therefore, there is good reason to expect that the experimental correspondences concerning base recognition for P-motifs will not entirely hold for S and L motifs.

In this study, we systematically mutated the P, L, and S motifs in the *Arabidopsis thaliana* RNA editing factor CLB19 to investigate their individual contributions to RNA recognition in vitro. By using variants of CLB19 and a second editing factor, OTP82, to attempt to complement the respective mutants, we also demonstrate that we can predictably alter the specificity of these factors in vivo.

[1] Address correspondence to ian.small@uwa.edu.au.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Ian Small (ian.small@uwa.edu.au).
[OPEN]Articles can be viewed online without a subscription.
www.plantcell.org/cgi/doi/10.1105/tpc.114.134189

RESULTS

Recombinant CLB19 Binds Specifically to the *clpP* and *rpoA* Editing Sites

The Arabidopsis plastid protein CLB19 consists of 10 PPR motifs and is required for editing of two plastid transcripts, *clpP* and *rpoA* (Chateigner-Boutin et al., 2008). Alignments of CLB19 to both RNA targets show a mismatch and five matches (excluding L motifs) to the previously proposed RNA recognition code for PPR proteins (Figure 1A). CLB19 lacking its transit peptide was expressed in *Escherichia coli* as a fusion protein with maltose binding protein (MBP). The observed molecular mass of the purified protein corresponds well to the predicted mass of ~98 kD (Figure 1b).

RNA-PPR complexes have generally been studied using radioactively labeled probes and RNA electrophoretic mobility shift assays (REMSAs) (Barkan et al., 2012; Okuda and Shikanai, 2012). It is difficult to differentiate multiple radioactive probes in a single assay, so we developed an assay wherein it is possible to use three probes labeled with different fluorescent dyes (Cy5, Cy3, and fluorescein) in each separate reaction. The different probes can be excited independently with lasers at different wavelengths and visualized directly on the gel. To verify that the labels do not differentially affect CLB19 binding, three identical *rpoA* probes with different labels were incubated with protein and analyzed by REMSA (Supplemental Figure 1). CLB19 showed identical affinity to all three *rpoA* probes.

Using the three fluorescent labels, binding of CLB19 to its two natural targets (*clpP* and *rpoA*) and a non-target editing site (in *rpoB*) were assessed simultaneously (Figures 1C and 1D). The *rpoA* oligonucleotide was slightly preferred over the *clpP* target. Recognition of the non-target *rpoB* sequence was negligible under these conditions, showing the specificity of the binding.

Absolute K_d values estimated from the binding curves varied considerably depending on the protein preparation and the age of the preparation. We interpret this to mean that a variable proportion of the protein in the preparations was active for binding and that this proportion decreased with even short periods of storage. More extensive purification proved counterproductive, presumably due to the lengthier procedures required. Despite the variability of the absolute K_d values, relative K_d values of CLB19 and its variants for different RNA oligonucleotides were highly reproducible (Supplemental Figure 2). All of the experiments described subsequently were therefore designed to allow

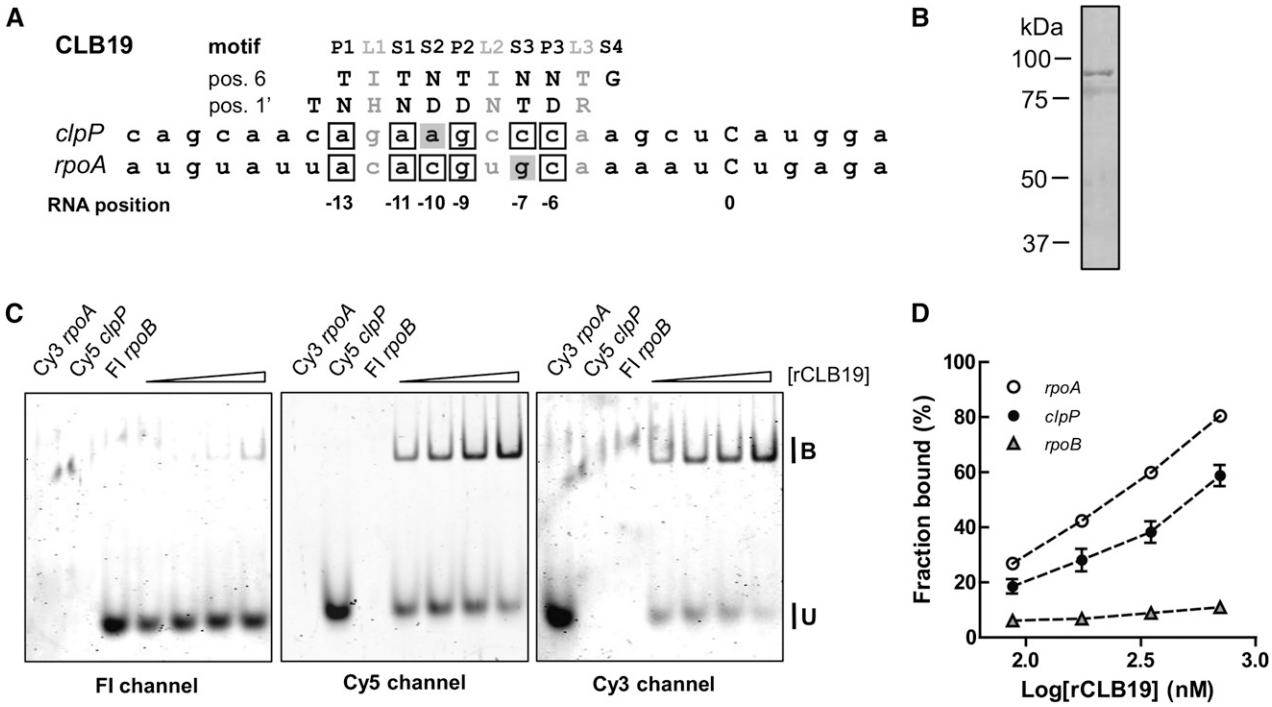

Figure 1. Alignment and Binding of CLB19 to Editing Sites in *clpP* and *rpoA*.

(A) Alignment between CLB19 and its RNA targets, *clpP* and *rpoA*. The edited C is position 0 in the RNA. The key amino acids at positions 6 and 1′ of each motif are indicated. The aligned nucleotides are boxed to indicate a match to the proposed recognition code or shaded gray to indicate a mismatch.

(B) Purified MBP-CLB19 analyzed by SDS-PAGE and stained with Coomassie blue (R 250). The molecular mass standards are indicated on the left. MBP-CLB19 shows as a single band with a molecular mass of ~98 kD, close to the predicted mass.

(C) Simultaneous binding of CLB19 to *clpP*, *rpoA*, and *rpoB* sequences. The panels to the left show the same gel visualized with different filters to reveal (from left to right) the fluorescein-labeled *rpoB* probe, the Cy5-labeled *clpP* probe, and the Cy3-labeled *rpoA* probe. Increasing concentrations of protein (87.5, 175, 350, and 700 nM) were incubated with 700 pM probe.

(D) Binding was quantified as the fraction bound (shifted) compared with unbound probe. Each reaction was done in triplicate. Error bars indicate ± SD.

measurement of K_d values for two or three oligonucleotides simultaneously, such that K_d values relative to that measured for the wild-type sequence could be calculated.

PPR Motifs Differ in Their Contributions to the CLB19-RNA Interaction

To investigate the contribution of individual PPR motifs in CLB19 to sequence-specific binding, systematic mutations at positions 6 and 1′ were made in each P and S motif individually. The first P motif (P1, containing a threonine at position 6 and an asparagine at position 1′) is expected to recognize an A in position −13 from the edited C of the RNA target (Barkan et al., 2012; Yagi et al., 2013). Two variants of CLB19 (P1[TD] and P1[NN]) containing aspartate at position 1′ or asparagine at position 6 in P1 were expressed in *E. coli* and purified. P1[TD] would be expected to recognize a G (Barkan et al., 2012), so a target oligonucleotide (*clpP*-13G) containing a G at this position was synthesized (Figure 2A). The wild type and the P1[TD] variant version of CLB19 were incubated with differentially labeled RNA targets (Figure 2B).

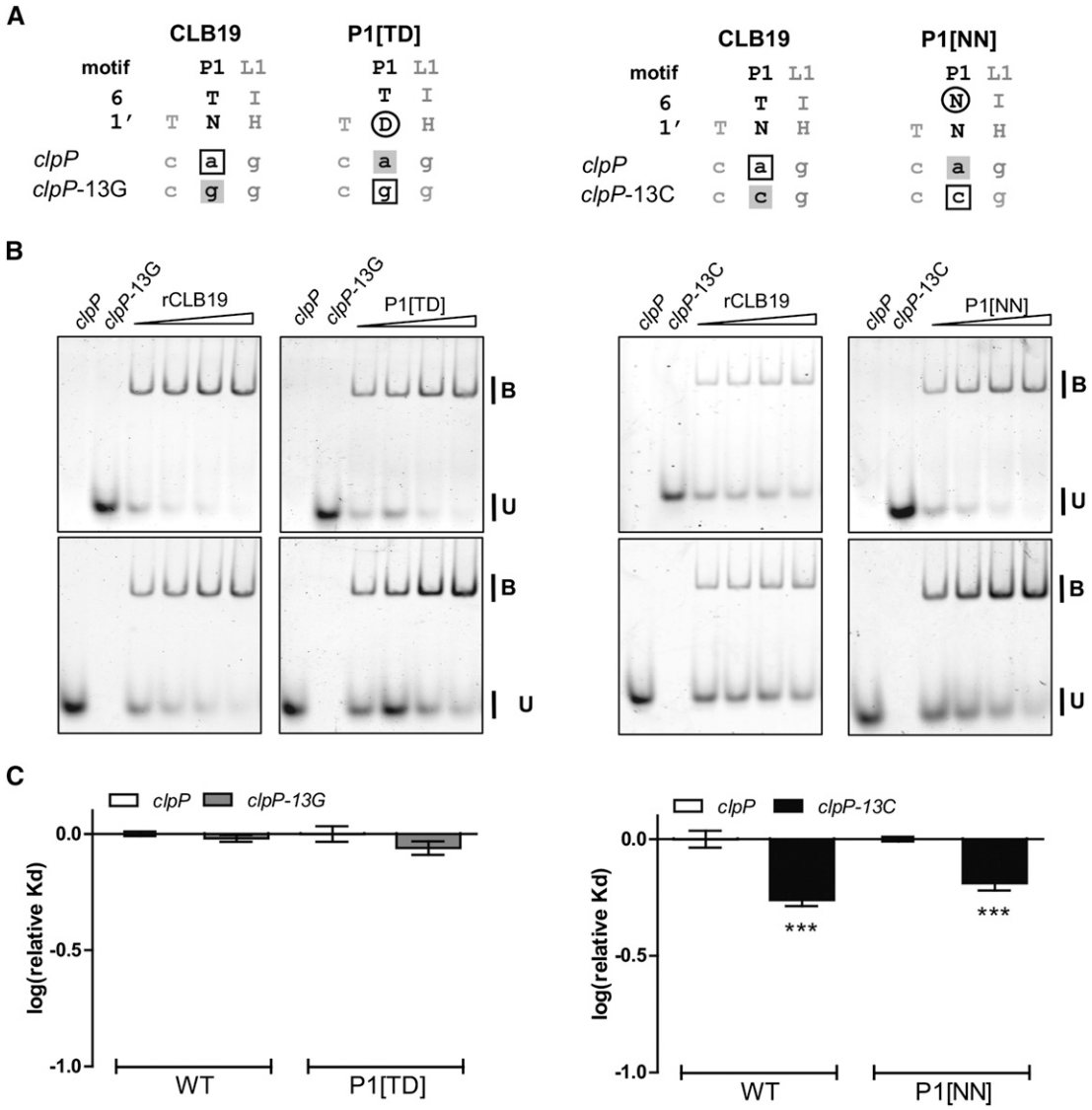

Figure 2. Analysis of the Role of Motif P1.

(A) Alignments of the P1 motif in CLB19 and the variants P1[TD] and P1[NN] to *clpP* RNA and the variants *clpP*-13G and *clpP*-13C. Altered amino acid residues are indicated by a black ring. Boxed RNA bases denote matches to the aligned motif, and gray shading indicates a mismatch.
(B) REMSA showing binding of CLB19 and the variants P1[TD] and P1[NN] to Cy5-labeled *clpP* and fluorescein-labeled *clpP*-13G or *clpP*-13C RNA oligonucleotides. Protein concentrations varied from 87.5 to 700 nM. Labeled oligonucleotides were at 750 pM. Each reaction was done in triplicate.
(C) Log(relative K_d) normalized to the K_d of *clpP* for each variant of CLB19 ± SD (n = 3). Significant preferences (one-way ANOVA, Tukey's comparison test) are indicated by asterisks (***P < 0.001).

Neither protein showed a preference between *clpP* and *clpP*-13G (Figure 2C). A similar experiment was performed with P1[NN], which was shown to prefer *clpP*-13C as expected (Figure 2C), but this preference was shared by the wild-type protein, suggesting that it was not due to base-specific recognition. Below, we demonstrate that this preference is probably due to the effect of the base change on RNA secondary structure. The lack of significant binding preferences in accordance with the recognition code suggests that P1 does not make a substantial contribution to the specificity of target recognition under these conditions.

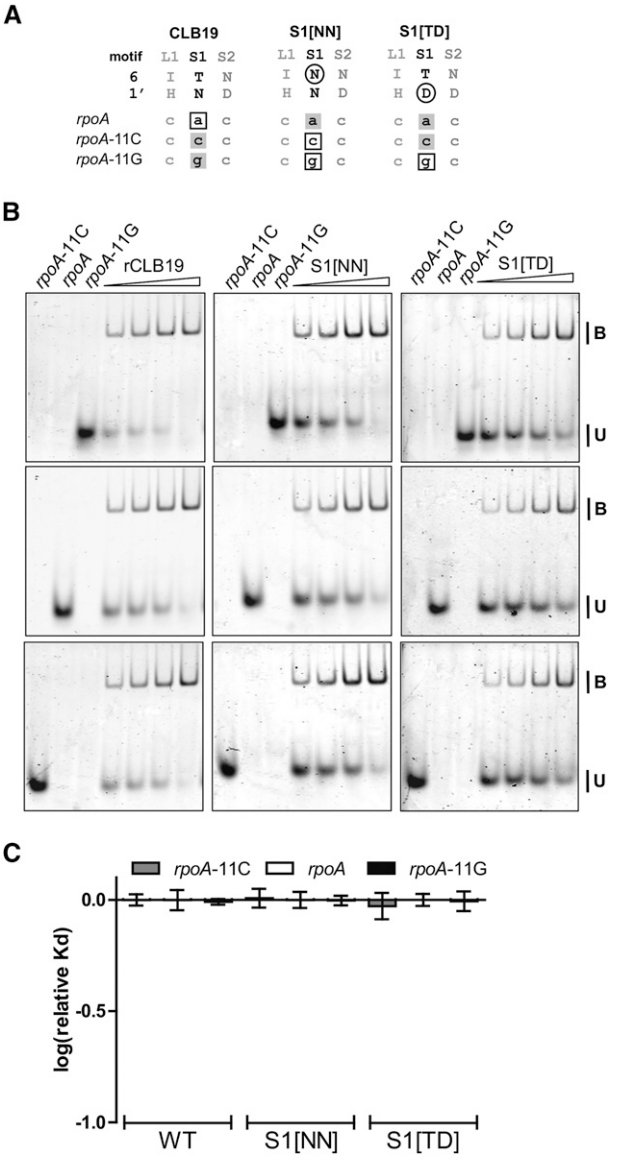

Figure 3. Analysis of the Role of Motif S1.

(A) Alignments of the S1 motif in CLB19 and the variants S1[NN] and S1[TD] to *rpoA* RNA and the variants *rpoA*-11C and *rpoA*-11G. Altered amino acid residues are indicated by a black ring. Boxed RNA bases denote matches to the aligned motif, and gray shading indicates a mismatch.
(B) REMSA showing binding of CLB19 and the variants S1[NN] and S1[TD] to Cy5-labeled *rpoA*, Cy3-labeled *rpoA*-11C, and fluorescein-labeled *rpoA*-11G RNA oligonucleotides. Protein concentrations varied from 87.5 to 700 nM. Labeled oligonucleotides were at 700 pM. Each reaction was done in triplicate.
(C) Log(relative K_d) normalized to the K_d of *rpoA* for each variant of CLB19 ± SD (n = 3). No significant preferences were observed.

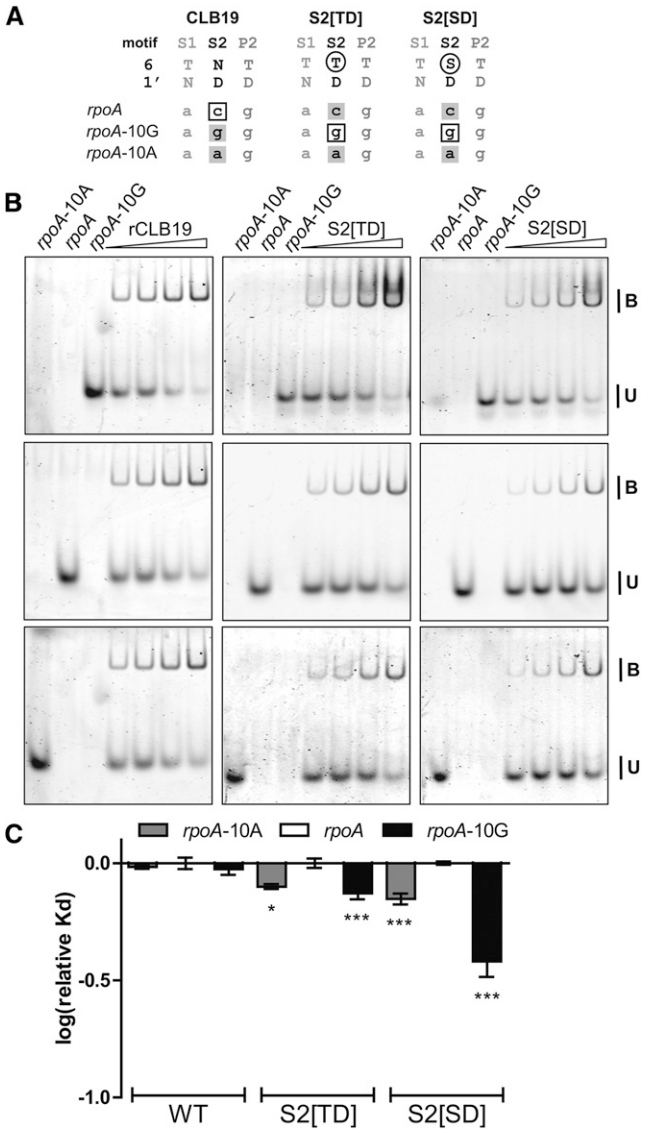

Figure 4. Analysis of the Role of Motif S2.

(A) Alignments of the S2 motif in CLB19 and the variants S2[TD] and S2 [SD] to *rpoA* RNA and the variants *rpoA*-10A and *rpoA*-10G. Altered amino acid residues are indicated by a black ring. Boxed RNA bases denote matches to the aligned motif, and gray shading indicates a mismatch.
(B) REMSA showing binding of CLB19 and the variants S2[TD] and S2[SD] to Cy5-labeled *rpoA*, Cy3-labeled *rpoA*-10A, and fluorescein-labeled *rpoA*-10G RNA oligonucleotides. Protein concentrations varied from 87.5 to 700 nM. Labeled oligonucleotides were at 700 pM. Each reaction was done in triplicate.
(C) Log(relative K_d) normalized to the K_d of *rpoA* for each variant of CLB19 ± SD (n = 3). Significant preferences (one-way ANOVA, Tukey's comparison test) are indicated asterisks: *P < 0.05 and ***P < 0.001. Both S2[TD] and S2[SD] prefer *rpoA*-10A and *rpoA*-10G more strongly than the wild type does.

Comparable results were obtained with an analogous set of experiments on motif S1 using the variant proteins S1[NN] and S1[TD] with the variant RNAs *rpoA*-11C and *rpoA*-11G (Figure 3). None of the wild-type or variant proteins showed significant oligonucleotide preferences. We conclude that motif S1 does not play a major role in target sequence recognition under these conditions.

The second S motif in CLB19, S2, is predicted to align to nucleotides that differ between the *clpP* and *rpoA* targets. The [ND] combination in wild-type CLB19 is expected to prefer pyrimidines, but whereas the *rpoA* sequence contains a C at the corresponding position, the *clpP* sequence contains an A. We attempted to create the variant S2[TN], expected to recognize A at −10, but the resulting protein did not stably accumulate in *E. coli*. We then created and expressed the variant proteins S2[TD] and S2[SD], both predicted to recognize a G at position −10 (Figure 4A). Wild-type, S2[TD], and S2[SD] proteins were incubated with *rpoA* oligos containing C, G, or A at position −10 of the RNA (Figure 4B). Wild-type CLB19 showed no binding preferences, but the variant proteins clearly preferred the sequences with a purine at position −10, particularly a G (Figure 4C). To further verify the alignment of S2 to position −10 in the RNA, we incubated S2[TD] with *clpP* and *clpP*-9C, a position that is expected to align with P2. S2[TD] showed similar affinity for both oligos (Supplemental Figure 3). Thus, altering S2 alters affinities for oligonucleotides modified at position −10, but not those similarly modified at position −9. These results imply that the predicted alignment of motif S2 to the position −10 is correct and that the S2 motif can contribute to target recognition.

The following motif, P2, has a $T_6D_{1'}$ combination expected to recognize G, the base present at the aligned position (−9G) in both *rpoA* and *clpP*. The variant protein P2[ND], expected to bind −9C better than −9G, indeed showed a significant preference for the *clpP*-9C oligonucleotide (Figure 5). Both wild-type CLB19 and the P2[TN] variant showed a clear preference for *clpP*-9A, but this

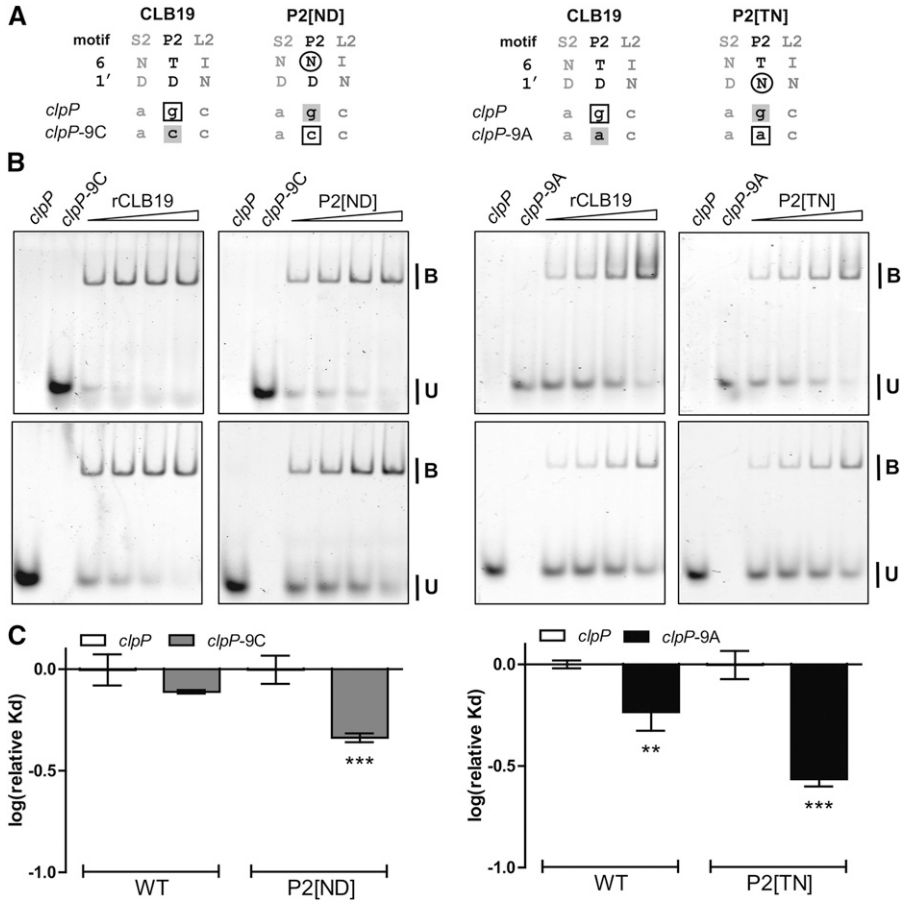

Figure 5. Analysis of the Role of Motif P2.

(A) Alignments of the P2 motif in CLB19 and the variants P2[ND] and P2[TN] to *clpP* RNA and the variants *clpP*-9C and *clpP*-9A. Altered amino acid residues are indicated by a black ring. Boxed RNA bases denote matches to the aligned motif, and gray shading indicates a mismatch.
(B) REMSA showing binding of CLB19 and the variants P2[ND] and P2[TN] to Cy5-labeled *clpP* and fluorescein-labeled *clpP*-9C or *clpP*-9A RNA oligonucleotides. Protein concentrations varied from 87.5 to 700 nM. Labeled oligonucleotides were at 750 pM. Each reaction was done in triplicate.
(C) Log(relative K_d) normalized to the K_d of *clpP* for each variant of CLB19 ± SD (n = 3). Significant preferences (one-way ANOVA, Tukey's comparison test) are indicated by asterisks: **P < 0.01 and ***P < 0.001. P2[ND] prefers *clpP*-9C (P < 0.01) more strongly than the wild type does, and P2[TN] prefers *clpP*-9A (P < 0.001) more strongly than the wild type does.

preference was stronger in the case of P2[TN], as expected from the recognition code. This confirms the alignment of position −9 in the RNA to P2 and implies that P2 can influence sequence recognition.

The third S motif (S3) has an $N_6T_{1'}$ combination expected to recognize C. −7C is present in the respective position in the

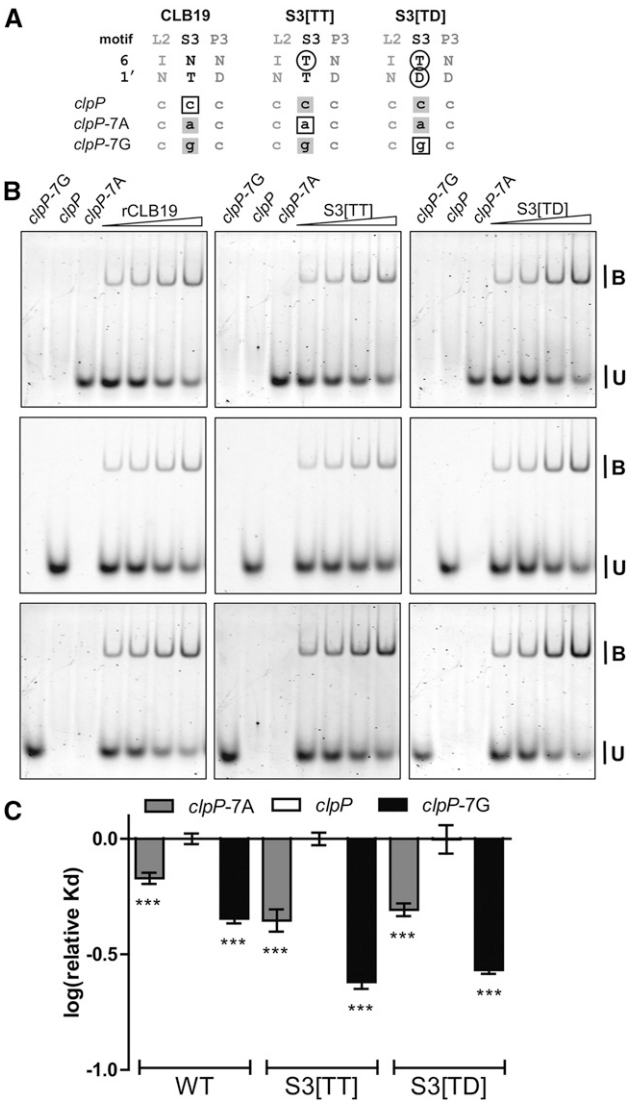

Figure 6. Analysis of the Role of Motif S3.

(A) Alignments of the S3 motif in CLB19 and the variants S3[TT] and S3 [TD] to *clpP* RNA and the variants *clpP*-7A and *clpP*-7G. Altered amino acid residues are indicated by a black ring. Boxed RNA bases denote matches to the aligned motif, and gray shading indicates a mismatch.
(B) REMSA showing binding of CLB19 and the variants S3[TT] and S3[TD] Cy5-labeled *clpP*, fluorescein-labeled *clpP*-7A, and Cy3-labeled *clpP*-7G RNA oligonucleotides. Protein concentrations varied from 87.5 to 700 nM. Labeled oligonucleotides were at 700 pM. Each reaction was done in triplicate.
(C) Log(relative K_d) normalized to the K_d of *clpP* for each variant of CLB19 ± SD (n = 3). Significant preferences (one-way ANOVA, Tukey's comparison test) are indicated by asterisks: ***P < 0.001. S3[TT] prefers *clpP*-7A (P < 0.01) and *clpP*-7G (P < 0.001) more strongly than the wild type does. S3[TD] also prefers both oligonucleotides (P < 0.001) more strongly than the wild type.

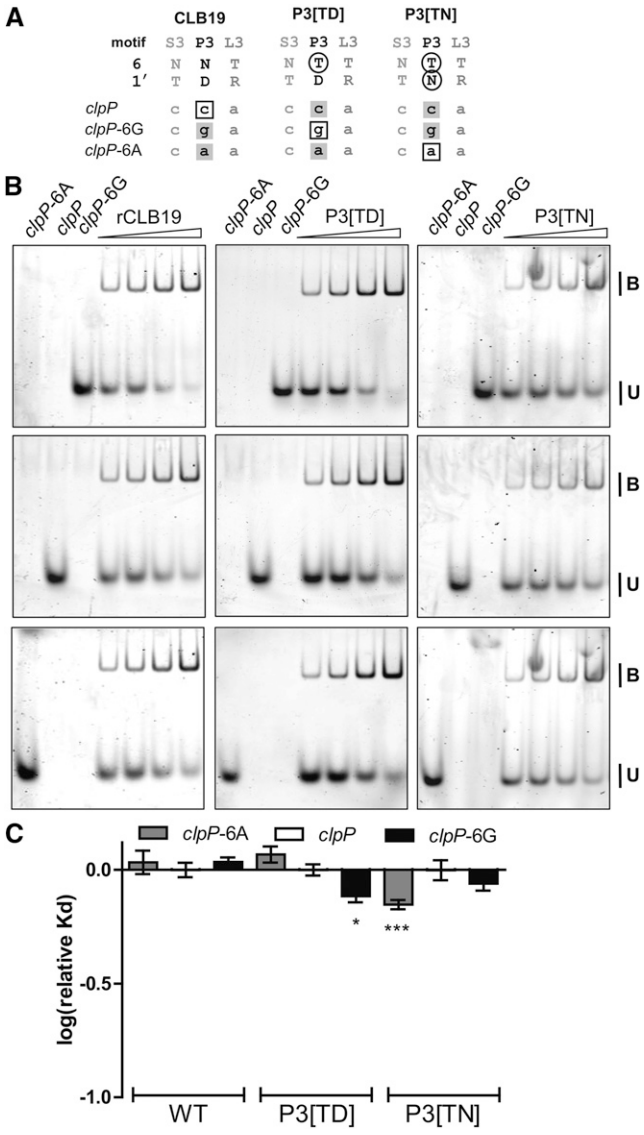

Figure 7. Analysis of the Role of Motif P3.

(A) Alignments of the P3 motif in CLB19 and the variants P3[TD] and P3[TN] to *clpP* RNA and the variants *clpP*-6G and *clpP*-6A. Altered amino acid residues are indicated by a black ring. Boxed RNA bases denote matches to the aligned motif, and gray shading indicates a mismatch.
(B) REMSA showing binding of CLB19 and the variants P3[TD] and P3 [TN] to Cy5-labeled *clpP*, fluorescein-labeled *clpP*-6G, and Cy3-labeled *clpP*-6A RNA oligonucleotides. Protein concentrations varied from 87.5 to 700 nM. Labeled oligonucleotides were at 700 pM. Each reaction was done in triplicate.
(C) Log(relative K_d) normalized to the K_d of *clpP* for each variant of CLB19 ± SD (n = 3). Significant preferences (one-way ANOVA, Tukey's comparison test) are indicated by asterisks: *P < 0.05 and ***P < 0.001. P3[TD] prefers *clpP*-6G (P < 0.001) and P3[TN] prefers *clpP*-6A (P < 0.001) more strongly than the wild type does.

clpP target but *rpoA* has −7G instead. To investigate the contribution of S3 to target recognition, we tested the variant S3[TD], expected to recognize the −7G in *rpoA*, but introducing a "mismatch" to the −7C in *clpP*. Another variant, S3[TT], would be predicted to recognize an A in position −7. To test the involvement of motif S3 in sequence recognition, wild-type CLB19 was incubated with wild-type *clpP* sequence in addition to *clpP*-7A and *clpP*-7G variants (Figure 6). In parallel, S3[TT] and S3[TD] were incubated with the same sequences. All of the tested proteins showed a preference for the variant RNAs, particularly

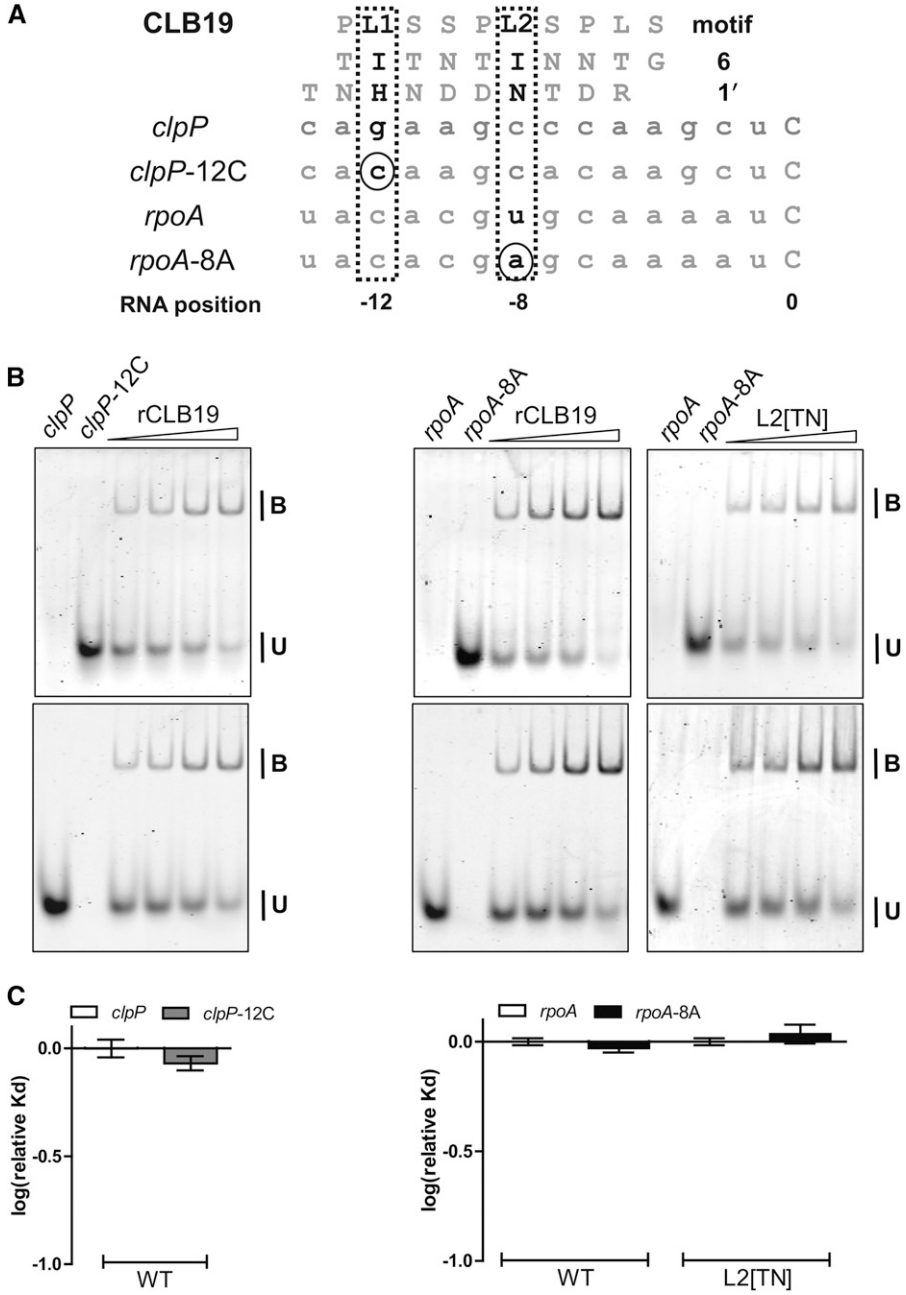

Figure 8. The L1 and L2 Motifs Do Not Contribute to the RNA Binding Specificity of CLB19.

(A) Alignments of the L1 and L2 motifs in CLB19 and the variant L2[TN] to *clpP* and *rpoA* RNA and the variants *clpP*-12C and *rpoA*-8A. Altered amino acid residues are indicated by a black ring. Boxed RNA bases denote matches to the aligned motif, and gray shading indicates a mismatch.
(B) REMSA showing binding of CLB19 and the variant L2[TN] to Cy5-labeled *clpP* or *rpoA* RNA oligonucleotides and fluorescein-labeled *clpP*-12C or *rpoA*-8A RNA oligonucleotides. Protein concentrations varied from 87.5 to 700 nM. Labeled oligonucleotides were at 750 pM. Each reaction was done in triplicate.
(C) Log(relative K_d) normalized to the K_d of *clpP* or *rpoA* for each variant of CLB19 ± SD (n = 3). No significant preferences were observed.

clpP-7G. Agreeing with expectations, the variant proteins' preferences were significantly stronger than those of the wild-type protein, although S3[TT] did not show the predicted preference for *clpP*-7A over *clpP*-7G. S3[TT] was tested with the *clpP* and *clpP*-6A variants where the nucleotide change is not expected to align with the S3 motif (Supplemental Figure 4). In this case, the nucleotide alteration did not affect S3[TT] binding. Thus, S3 is involved in base-specific recognition of *clpP* and the alignment of S3 to position −7C is confirmed.

The last P motif (P3) in CLB19 has an $N_6D_{1'}$ combination expected to recognize pyrimidines and aligning with −6C in both *rpoA* and *clpP*. Altering the asparagine at position 5 to a threonine creates a $T_6D_{1'}$ combination expected to recognize −6G instead (Figure 7A). In addition, a P3[TN] variant was expressed which should recognize −6A. Wild-type CLB19, P3[TD], and P3[TN] were expressed and incubated with probes that included C, G, or A at position −6 of the RNA target (Figure 7). The wild-type protein did not show a strong preference for any of the RNA variants, but the P3[TD] variant showed a significant preference for *clpP*-6G and the P3[TN] variant showed a significant preference *clpP*-6A (Figure 7C). This result confirms the alignment of P3 with position −6 in the RNA target and shows that P3 can participate in base-specific target recognition.

Altering Bases That Align with L Motifs Does Not Affect CLB19 Binding Affinity

L motifs in PPR proteins appear to differ from P and S motifs in their contribution to RNA recognition (Barkan et al., 2012). Instead of the hydrogen bonding asparagine, threonine, and serine residues usually found at position 6 in P and S motifs, L motifs often contain residues with very different properties. For example, both L1 and L2 in CLB19 have an isoleucine at position 6, whose side chain is incapable of forming hydrogen bonds to an aligned base. To test the interaction of L motifs with RNA, labeled *clpP* and *rpoA* oligos with variant bases aligning to L1 and L2, respectively, were

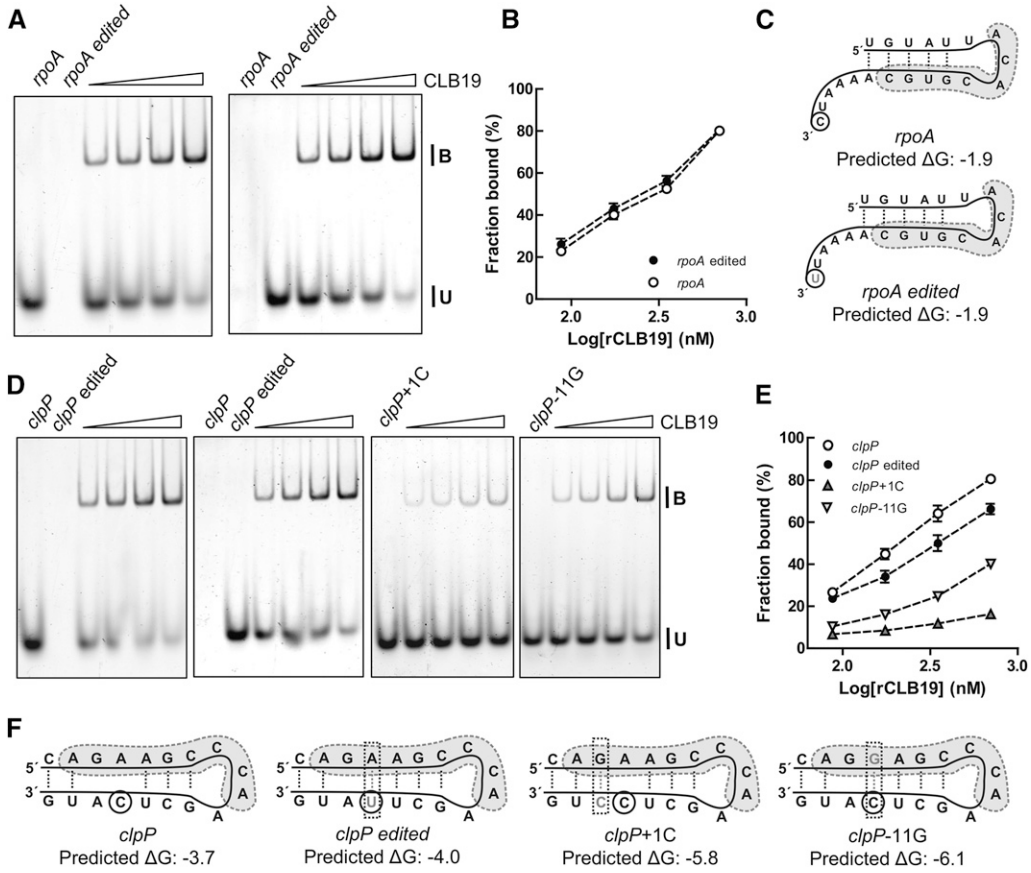

Figure 9. RNA Secondary Structure Affects Binding of CLB19.

(A) and **(D)** REMSAs of different concentrations of MBP-CLB19 (87.5, 175, 350, and 700 nM) that were incubated with 750 pM each of the fluorescein- and Cy5-labeled oligonucleotides. B indicates PPR-RNA complex and U denotes unbound probe.

(B) and **(E)** Binding curves for the different oligonucleotides indicated in **(A)** and **(D)**, respectively. The experiments were repeated in triplicate. The graphs show mean value (n = 3) ± SD.

(C) and **(F)** Predicted secondary structure of the different oligonucleotides used. The predicted binding site of CLB19 to the oligonucleotides is indicated by the gray shading. The edited C is indicated by the black ring. The mutated bases in the RNA are in gray and new hydrogen bonds are indicated by a gray box. Predicted ΔG is indicated below each structure.

incubated with wild-type CLB19 (Figure 8). No difference in binding affinity could be seen between wild-type RNA and the variant oligonucleotides. In addition, we changed the isoleucine in position 6 of L2 to a threonine, creating a $T_6N_{1'}$ combination. Although this combination is often seen in P and S motifs and is expected to recognize an A, there was no difference in affinity for the oligonucleotides with different bases in position −8 (Figure 8), further suggesting that the L motifs in CLB19 do not take part in any base-specific interactions with the target RNA.

The Binding Affinity of CLB19 Is Influenced by RNA Secondary Structure

The predicted alignment of CLB19 to its target sites does not involve any nucleotides in the immediate vicinity of the C to be edited. Thus, the expected effect of mimicking the editing event itself (C to U) would be no effect on binding. To test this, we analyzed the binding of CLB19 to *clpP*, *clpP_edited*, *rpoA*, and *rpoA_edited* (Figure 9). Whereas the *rpoA* variants were bound with equal affinity as expected, to our initial surprise there was a significant preference for the unedited form of the *clpP* sequence. As this effect was specific to the *clpP* target, we hypothesized that it might be due to an effect on the secondary structure of the RNA. The U in the edited variant of *clpP* can hydrogen bond with an A to strengthen a potential hairpin loop (Figures 9C and 9F), suggesting that CLB19 might prefer the target with the weaker RNA secondary structure. We thus tested several variants of the *clpP* target with an increasingly strong propensity to form an RNA hairpin (Figure 9). The results were clear: CLB19 showed much lower affinity for sequences predicted to form RNA hairpins. We conclude that RNA secondary structure influences the binding affinity of CLB19 and that this needs to be taken into account when interpreting binding to different RNA sequences.

Assaying the Effect of Mutating Individual PPR Motifs in Vivo

The recognition code of PPR proteins has not previously been tested in vivo where numerous complicating factors may affect the

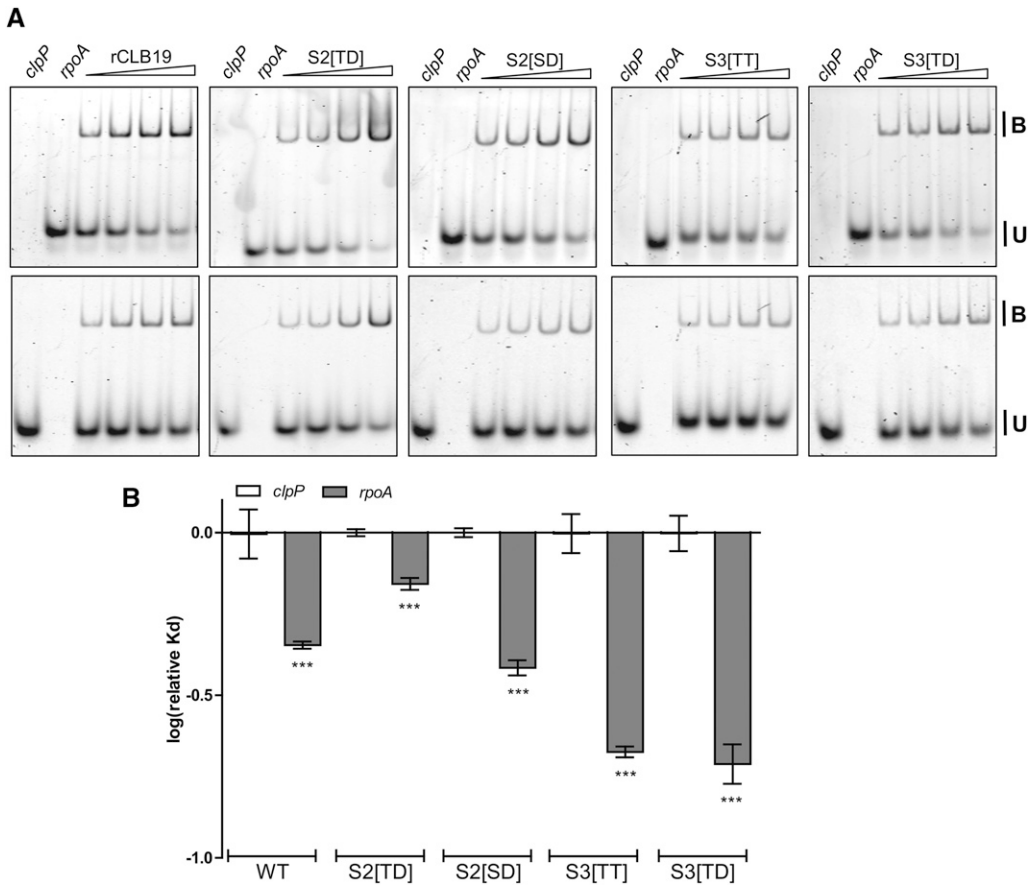

Figure 10. Variants of S2 and S3 Have Modified Preferences for *clpP* and *rpoA*.

(A) REMSA showing binding of CLB19 and the variants S2[TD], S2[SD], S3[TT], and S3[TD] to Cy5-labeled *clpP* and fluorescein-labeled *rpoA* RNA oligonucleotides. Protein concentrations varied from 87.5 to 700 nM. Labeled oligonucleotides were at 750 pM. Each reaction was done in triplicate.
(B) Log(relative K_d) normalized to the K_d of *clpP* for each variant of CLB19 ± SD (n = 3). Significant preferences (one-way ANOVA, Tukey's comparison test) are indicated by asterisks: ***P < 0.001. The wild type prefers *rpoA* more strongly than S2[TD] does (P < 0.001). Both S3[TT] and S3[TD] prefer *rpoA* (P < 0.001) more strongly than the wild type does.

results. As it is not easy to introduce variant RNA sequences into Arabidopsis plastids in vivo, we were obliged to make use of the natural sequence differences between the *clpP* and *rpoA* targets. Wild-type CLB19 shows a preference for the *rpoA* target (Figures 1 and 10). We investigated if the different S2 and S3 variants of CLB19 had altered *clpP*/*rpoA* preference in vitro (Figure 10). As expected from our earlier in vitro results (Figures 4 and 6), S2[TD] showed a significant shift in preference toward *clpP*, while S3[TN] and S3[TD] showed a significant shift in preference toward *rpoA*. The S2[SD] variant did not show a significant change in target preference. The S2[TD], S2[SD], S3[TT], and S3[TD] variants were introduced into a *clb19* null background, as was the S2[TN] variant that we could not test in vitro. All variants rescued the *clb19* phenotype (Figure 11), showing that the proteins were active in RNA editing in vivo.

Binding to RNA targets in vivo was assessed by quantifying editing of the respective RNAs using poisoned primer extension (PPE) analysis (Supplemental Figures 5 and 6). Absolute levels of editing might vary due to differences in expression levels or protein activity; to avoid this complication, we calculated the ratio of *clpP* to *rpoA* editing (Figure 11). The fact that CLB19 edits two slightly different target sequences is an advantage for these experiments as the second site provides an internal control of protein function; a change in editing of one target while the other is unaltered implies that the editing factor has altered sequence specificity with no change in its ability to induce cytidine deamination. As a control, we used wild-type CLB19 (expressed from the same promoter as the variants) to complement *clb19*, and the transformants were analyzed in the same way. Plants complemented with wild-type CLB19 showed a *clpP*/*rpoA* editing ratio not significantly different from that seen in wild-type Columbia-0 (Col-0) plants (Figure 11). Plants expressing the S2 variants showed a small, but significant, shift toward favoring *clpP* editing. The S2[SD] variant, which showed unexpectedly poor binding and discrimination in vitro (Figure 10),

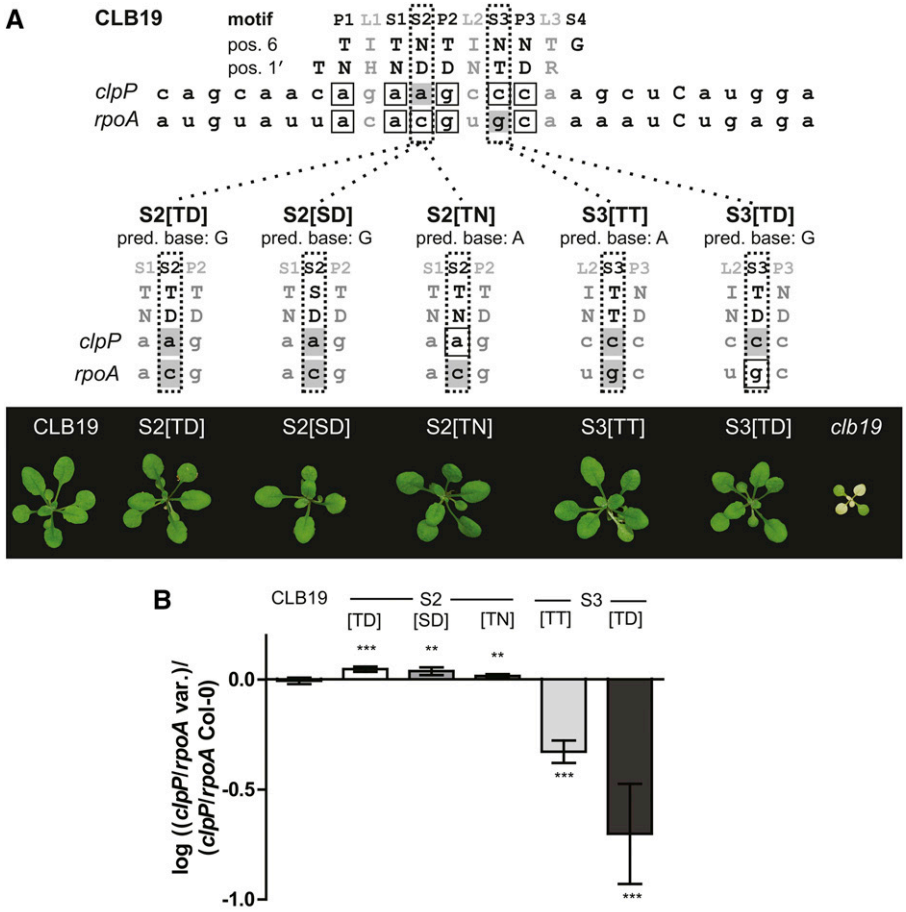

Figure 11. In Vivo Editing Induced by Variants of CLB19.

(A) Amino acids in positions 6 and 1′ in CLB19 and the variants of the protein used for in vivo analysis and how they align with *clpP* and *rpoA*. Gray shading denotes a mismatch to the aligned motif, and boxed RNA bases indicate a match. Below each alignment are images showing the phenotype of 3-week-old *clb19* plants complemented by the various constructs in comparison with wild-type CLB19 and uncomplemented *clb19*.

(B) The complemented plants were analyzed by PPE analysis. The PPE gels are shown in Supplemental Figures 5 and 6. The ratio of *clpP* editing to *rpoA* editing was calculated for each sample and then normalized to the ratio observed in wild-type Col-0 plants. The charts show mean log ratios ± SD (n = 3 to 5). Statistical significance was calculated with Student's t test: **P < 0.01 and ***P < 0.01.

showed a significant drop in both *rpoA* and *clpP* editing, suggesting that it also worked less effectively in vivo (Supplemental Figures 5 and 6). In accordance with the changes to binding affinity seen in vitro (Figures 6 and 10), the S3 variants showed a strong and significant preference for the *rpoA* site.

To see if these results could be generalized to a second editing factor, we also analyzed the plastid protein OTP82 (Okuda et al., 2010). OTP82 consists of 13 PPR motifs in tandem and is required for editing of the sites *ndhG* and *ndhB-9* (Okuda et al., 2010). Alignment of OTP82 to its targets using the proposed RNA recognition code (Figure 12A) reveals two mismatches and six matches to *ndhB-9*, while the *ndhG* alignment shows eight matches, excluding the L motifs. We altered the third S motif in OTP82, where the *ndhB-9* target has a predicted mismatch, to an amino acid combination that would be predicted to recognize the −10G nucleotide in *ndhB* but not the −10U in *ndhG*. We predicted this would alter the protein to show a preference for *ndhB* over *ndhG*. The S3[TD] was expressed in an *otp82* background and editing at the two sites assayed by PPE (Supplemental Figure 7). In accordance with the predictions, editing of *ndhB-9* was restored to wild-type levels but editing of *ndhG* was decreased by ~40% compared with the wild type, leading to a significant decrease in the *ndhG/ndhB-9* editing ratio (Figure 12).

DISCUSSION

The remarkable sequence specificity of plant RNA editing factors has long puzzled researchers. Recent proposals of how they might recognize their RNA targets are compelling (Barkan et al., 2012; Takenaka et al., 2013a; Yagi et al., 2013), but rest on statistical correlations between the protein sequences and the nucleotide sequence in their RNA ligand. Experimental data in favor of these binding models has come from in vitro binding assays with the PPR protein PPR10 (Barkan et al., 2012), which is not closely related to editing factors, and analysis of crystals of a variant of the same protein bound to one of its RNA targets (Yin et al., 2013). Given the many differences between P-class proteins such as PPR10 and PLS-class editing factors in their sequence, structure, and RNA binding activities (Barkan and Small, 2014), it was not clear that the PPR10 results could be extrapolated to them.

The most detailed analysis of RNA binding by PLS-class proteins delimited the minimum binding sites for the editing factors CRR28, OTP85, CRR21, and OTP80 (Okuda et al., 2014). This analysis showed that each of these factors binds an RNA sequence containing approximately as many nucleotides as the factor contains PPR motifs, evidence in favor of one motif:one nucleotide binding. The same work confirmed the placement of the editing factor binding site as extending upstream from approximately four nucleotides before the edited C. However, this study examined sequence recognition for only a single P motif and therefore could not make firm conclusions concerning the orientation or spacing of the motif:nucleotide interactions, nor about the role of L or S motifs in sequence recognition.

In this study, we provide experimental evidence in favor of the current hypotheses concerning target recognition by RNA editing factors and also demonstrate that editing specificity in vivo is consistent with these same hypotheses. First, our evidence is in favor of the proposed alignment of editing factors with their RNA targets such that the protein is aligned parallel to the RNA (N terminus toward the 5′ end; C terminus toward the 3′ end). Second, our evidence is in favor of the proposals that the terminal PPR motif is aligned with the nucleotide at −4 with respect to the edited C and with the other motifs aligning contiguously from this point with each motif aligned to one nucleotide. We reach these conclusions

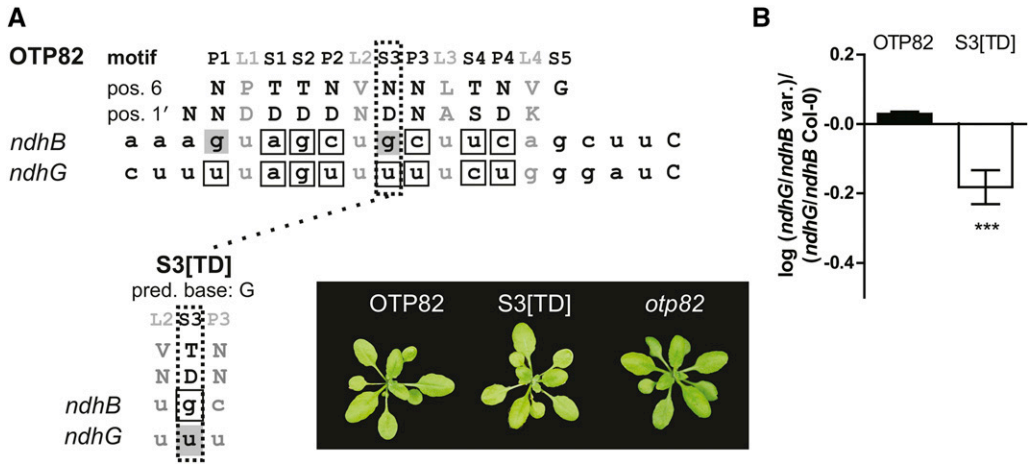

Figure 12. In Vivo Editing Induced by Variants of OTP82.

(A) Amino acids in positions 6 and 1′ in OTP82 and the S3[TD] variant of the protein used for in vivo analysis and how they align with *ndhG* and *ndhB*. Gray shading denotes a mismatch to the aligned motif, and boxed RNA bases indicate a match. Images of the phenotype of 3-week-old *otp82* and *otp82* complemented with S3[TD] and OTP82.

(B) OTP82 and the S3[TD] variant were analyzed by PPE analysis. The PPE gels are shown in Supplemental Figure 7. The ratio of *ndhG* editing to *ndhB* editing was calculated for each sample and then normalized to the ratio observed in wild-type Col-0 plants. The charts show mean log ratios ± SD ($n = 3$ to 5). Statistical significance was calculated with a Student's *t* test: ***$P < 0.001$.

based on the specific interactions noted between motifs S2, S3, P2, and P3 in CLB19 and nucleotides −10, −9, −7, and −6, respectively, in the RNA target. Third, we provide evidence in favor of some of the base recognition rules proposed for P and S motifs. Our results include both in vitro and in vivo experiments confirming that position 6 is the primary residue for discriminating between ribonucleotides, particularly between purines and pyrimidines, with position 1′ playing a less critical role. Of the combinations we tested, we have shown that $T_6D_{1'}$ in motifs S2, S3, and P3 shows a preference for G (Figures 4, 6, and 7), that $S_6D_{1'}$ in motif S2 also shows a preference for G (Figure 4), that $N_6D_{1'}$ in P2 shows a preference for a pyrimidine (Figure 5), that the combinations $T_6N_{1'}$ and $T_6T_{1'}$ in motif S3 show a preference for purines (Figure 6), and that $T_6N_{1'}$ in motif P3 shows a preference for A (Figure 7). These preferences are similar to those observed for P motifs in PPR10 (Barkan et al., 2012) and postulated for P and S motifs in editing factors (Barkan et al., 2012; Takenaka et al., 2013a; Yagi et al., 2013).

Only four motifs (S2, S3, P2, and P3) out of the eight tested were shown to affect sequence-specific binding to the RNA ligands in vitro. This appears to be consistent with some prior observations; for example, only 6 out of 19 motifs in PPR10 and 1 out of 5 in THA8 are bound to RNA in canonical fashion in the published crystal structures (Ke et al., 2013; Yin et al., 2013). Systematic mutation of all PPR motifs in the P-class protein PGR3 revealed that only some motifs were required for function, both in vitro and in vivo (Fujii et al., 2013). However, the conservation of the "correct" $T_6N_{1'}$ combination in motifs P1 and S1 in CLB19 to match the aligned −13A and −11A in both the *clpP* and *rpoA* targets suggests that these motifs do play a role in sequence recognition in vivo, at least under some conditions. Our editing assays suggest that sequence recognition is more sensitive to "mismatches" in vivo than in our in vitro binding assays. A drop in *clpP* editing of 70% was found with a single amino acid change in CLB19 that led to a relatively minor decrease in affinity for the *clpP* ligand in vitro. Nevertheless, we have shown that binding of CLB19 to its targets is relatively robust to single changes in either the RNA sequences or the protein itself, consistent with the common ability of editing factors to recognize more than one editing site, often with quite divergent sequences (Hammani et al., 2009; Okuda and Shikanai, 2012). Overall, the results suggest that individual PPR motifs contribute to different extents to RNA recognition. A future challenge will be to predict in advance of any analysis which motifs are the most important.

We have shown that potential secondary structures in the RNA ligands need to be considered when interpreting in vitro binding results. The binding affinity of CLB19 is clearly reduced by sequence changes that are predicted to promote hairpin formation in the target RNA (Figure 9). In fact, unintended differences in secondary structure are a plausible explanation for some of the minor variations in binding affinity seen in some of our experiments. For example, the *rpoA* probe has a predicted weaker secondary structure compared with the *clpP* probe and CLB19 showed a higher affinity for *rpoA* when incubated together with *clpP* (Figure 1C). In addition, CLB19 had an increased affinity for clpP-13C, *clpP*-9A, and *clpP*-7G, oligonucleotides with weaker secondary structure compared with the wild-type variants (Figures 2, 5, and 6). In vivo, access to RNA targets may be facilitated by RNA helicases. The relatively high requirement for nucleotide triphosphates for RNA editing in organelle extracts is almost certainly due to a need for helicase activity (Takenaka and Brennicke, 2003). RNA structure has been invoked previously to explain unusual dependencies between editing factors (Schallenberg-Rüdinger et al., 2013).

In a broader context, there is considerable biotechnological potential in the possibility of designing and constructing proteins that bind defined DNA and RNA target sequences. The proteins that are most amenable to such design are those with a modular structure, where each module recognizes one nucleotide via simple rules involving only two to three amino acids. Examples include the DNA binding transcription-activator-like effectors (TALEs) (Boch et al., 2009; Moscou and Bogdanove, 2009), the RNA binding Puf domain proteins (Lu et al., 2009), and PPR proteins (Yagi et al., 2014). Unlike the copious experimental data available on DNA binding by TALEs, the modular RNA binding protein families have been less studied. Here, we provide experimental evidence that plant RNA editing factors do bind RNA via contiguous modular one-motif:one nucleotide interactions and that sequence recognition does obey simple rules. We also demonstrate that their specificity for their RNA targets can be predictably manipulated in vivo. We believe this will stimulate research into using these proteins as the basis for tools to engineer RNA metabolism (Yagi et al., 2014).

METHODS

Plant Material and Growth Conditions

Arabidopsis thaliana seeds were surface-sterilized and germinated on half-strength Gamborg media + 1% sucrose for 10 d and then transferred to soil. Transformed seedlings were selected on plates containing hygromycin at 15 μg/mL. All seedlings were grown in long-day conditions (16-h-light/8-h-dark cycle) under ~120 μmol photons m^{-2} s^{-1}. The *clb19* mutant has a T-DNA insertion (SALK_104250) in At1g05750 and has been described elsewhere (Chateigner-Boutin et al., 2008). The *otp82* mutant used has a T-DNA insertion (SALK_027812) in At1g08070 (Okuda et al., 2010).

Construction of CLB19 and OTP82 Variants for in Vivo Studies

Full-length CLB19 and OTP82 was amplified from Arabidopsis genomic DNA (ecotype Col-0). For point mutations, internal forward and reverse primers were designed to include one or two mutated bases to introduce the mutation into the full-length protein. A full list of primers used for the different constructs can be found in Supplemental Table 1. PCR fragments were subcloned into pDONR207 (Invitrogen) and subsequently into pGWB2 (Nakagawa et al., 2007) via Gateway technology according to the manufacturer's instructions. The constructs were sequenced to confirm the mutations. The constructs were then transformed into *Agrobacterium tumefaciens* (C58C1) and used to transform *clb19* and *otp82* via the floral dip method (Clough and Bent, 1998).

RNA Extraction, cDNA Synthesis, and PPE Analysis

RNA from 3-week-old plants was extracted (Qiagen) and DNase-treated (Ambion) according to the manufacturers' instructions. Total RNA (1 μg) was reverse transcribed into cDNA with Superscript III (Invitrogen). PPE analysis was performed as described previously (Chateigner-Boutin et al., 2008).

Protein Expression and Purification

The region encoding CLB19, omitting the N-terminal transit peptide, was amplified from genomic DNA (ecotype Col-0) using the primers listed in

Supplemental Table 1. To create the mutated proteins, we used the same primers as described for the in vivo analysis. PCR fragments were subcloned into pDONR207 (Invitrogen) with BP clonase II (Invitrogen) and sequenced. Constructs were then subcloned with LR clonase II (Invitrogen) into the expression vector pETG-41K (EMBL). The expressed protein contained an N-terminal 6xHis tag used for purification followed by a MBP tag. For protein expression, the *Escherichia coli* strain C41(DE3) was used. Cells were grown in 500 mL cultures (1× Luria-Bertani and 50 mM Tris-HCl, pH 8.0) at 37°C, 220 rpm, until OD_{600} = 0.4. Cultures were then incubated at 16°C, 220 rpm, for 30 min. Protein expression was initiated by adding isopropyl β-D-1-thiogalactopyranoside to a final concentration of 0.1 mM, and cultures were grown at 16°C, 220 rpm. After 4 h, cells were harvested (3000*g*, 15 min) and pellets were dissolved in 35 mL lysis buffer (500 mM NaCl, 50 mM Tris-HCl, pH 8.3, 10 mM imidazole, and 7 mM β-mercaptoethanol). Cells were disrupted by homogenization (Avestin C5). Soluble proteins were obtained after centrifugation (13,000*g*, 15 min, 4°C) and proteins were purified using Ni-charged resin (Bio-Rad) in batch mode. Eluted protein was dialyzed overnight at 4°C in dialysis buffer (500 mM NaCl, 50 mM Tris-HCl, pH 8.3, 50% glycerol, 1 mM EDTA, and 7 mM β-mercaptoethanol).

RNA Electrophoresis Mobility Shift Assay

Dialyzed protein was diluted with dialysis buffer and the protein concentrations were verified with Bradford assay and a NanoDrop spectrophotometer (ND-1000; Thermo Fisher Scientific). REMSAs were run according to Schallenberg-Rüdinger et al. (2013) with a few modifications. Briefly, 10 μL binding buffer consisting of 1× THE (34 mM Tris, 66 mM HEPES, and 0.1 mM EDTA, pH 8.3) with 200 mM NaCl, 5 mM DTT, 5 mg/mL heparin, 0.1 mg/mL BSA, and 8 units RNaseOUT (Invitrogen) was mixed with 5 μL protein dilution and incubated at room temperature for 10 min. 5′-Fluorescein-labeled probes (Sigma-Aldrich) were heated for 2 min at 94°C followed by incubation on ice for at least 4 min. Denatured probes (final concentration, 1000, 750, or 700 pM each) were then added to the binding reaction (10 μL) for a total reaction volume of 25 μL. The reactions were incubated at 25°C for 15 min and 15 μL were loaded onto a prerun 5% native gel (in 1× THE) that was run at 4°C. The gels were imaged with a Typhoon Trio imager (GE Healthcare). Fluorescein-labeled oligonucleotides were excited by a 488-nm laser and detected through a 520-nm band-pass filter. For oligonucleotides labeled with Cy3 or Cy5, excitation at 532 and 633 nm was used, respectively. Emitted light was detected through a 670-nm band-pass filter for Cy5 and a 580-nm band-pass filter for Cy3. The fraction of oligonucleotide bound was determined with ImageQuant (GE Healthcare).

Accession Numbers

This work characterizes the Arabidopsis proteins encoded by the loci At1g05750 (*CLB19*) and At1g08070 (*OTP82*).

Supplemental Data

Supplemental Figure 1. CLB19 Shows No Label Preference.

Supplemental Figure 2. Relative K_d, but Not Absolute K_d, Is Reproducible between Protein Preparations.

Supplemental Figure 3. Alignment Control for S2[TD].

Supplemental Figure 4. Alignment Control for S3[TT].

Supplemental Figure 5. *clpP* Editing Induced by Variants of CLB19.

Supplemental Figure 6. *rpoA* Editing Induced by Variants of CLB19.

Supplemental Figure 7. Editing Induced by Variants of OTP82.

Supplemental Table 1. Primers and Oligonucleotides Used in This Study.

ACKNOWLEDGMENTS

This work was supported by ARC Grant DP120102870, and P.K. was supported through a postdoctoral fellowship from the Wenner-Gren Foundation.

AUTHOR CONTRIBUTIONS

P.K., A.Y., C.B., and I.S. designed the experiments. P.K. and A.Y. performed the experiments. P.K., A.Y., and I.S. analyzed the data. P.K., C.B., and I.S. wrote the article.

Received November 8, 2014; revised November 8, 2014; accepted January 16, 2015; published February 3, 2015.

REFERENCES

Barkan, A., Rojas, M., Fujii, S., Yap, A., Chong, Y.S., Bond, C.S., and Small, I. (2012). A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. PLoS Genet. **8:** e1002910.

Barkan, A. and Small, I. (2014). Pentatricopeptide repeat proteins in plants. Annu. Rev. Plant Biol. **65:** 415–442.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. Science **326:** 1509–1512.

Chateigner-Boutin, A.-L., and Small, I. (2010). Plant RNA editing. RNA Biol. **7:** 213–219.

Chateigner-Boutin, A.-L., Ramos-Vega, M., Guevara-García, A., Andrés, C., de la Luz Gutiérrez-Nava, M., Cantero, A., Delannoy, E., Jiménez, L.F., Lurin, C., Small, I., and León, P. (2008). CLB19, a pentatricopeptide repeat protein required for editing of *rpoA* and *clpP* chloroplast transcripts. Plant J. **56:** 590–602.

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J. **16:** 735–743.

Fujii, S., and Small, I. (2011). The evolution of RNA editing and pentatricopeptide repeat genes. New Phytol. **191:** 37–47.

Fujii, S., Sato, N., and Shikanai, T. (2013). Mutagenesis of individual pentatricopeptide repeat motifs affects RNA binding activity and reveals functional partitioning of Arabidopsis PROTON gradient regulation3. Plant Cell **25:** 3079–3088.

Hammani, K., Okuda, K., Tanz, S.K., Chateigner-Boutin, A.-L., Shikanai, T., and Small, I. (2009). A study of new Arabidopsis chloroplast RNA editing mutants reveals general features of editing factors and their target sites. Plant Cell **21:** 3686–3699.

Ke, J., Chen, R.-Z., Ban, T., Zhou, X.E., Gu, X., Tan, M.H.E., Chen, C., Kang, Y., Brunzelle, J.S., Zhu, J.-K., Melcher, K., and Xu, H.E. (2013). Structural basis for RNA recognition by a dimeric PPR-protein complex. Nat. Struct. Mol. Biol. **20:** 1377–1382.

Lu, G., Dolgner, S.J., and Hall, T.M.T. (2009). Understanding and engineering RNA sequence specificity of PUF proteins. Curr. Opin. Struct. Biol. **19:** 110–115.

Lurin, C., et al. (2004). Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell **16:** 2089–2103.

Moscou, M.J., and Bogdanove, A.J. (2009). A simple cipher governs DNA recognition by TAL effectors. Science **326:** 1501.

Nakagawa, T., Kurose, T., Hino, T., Tanaka, K., Kawamukai, M., Niwa, Y., Toyooka, K., Matsuoka, K., Jinbo, T., and Kimura, T. (2007). Development of series of gateway binary vectors, pGWBs,

for realizing efficient construction of fusion genes for plant transformation. J. Biosci. Bioeng. **104:** 34–41.

Okuda, K., Hammani, K., Tanz, S.K., Peng, L., Fukao, Y., Myouga, F., Motohashi, R., Shinozaki, K., Small, I., and Shikanai, T. (2010). The pentatricopeptide repeat protein OTP82 is required for RNA editing of plastid *ndhB* and *ndhG* transcripts. Plant J. **61:** 339–349.

Okuda, K., and Shikanai, T. (2012). A pentatricopeptide repeat protein acts as a site-specificity factor at multiple RNA editing sites with unrelated cis-acting elements in plastids. Nucleic Acids Res. **40:** 5052–5064.

Okuda, K., Shoki, H., Arai, M., Shikanai, T., Small, I., and Nakamura, T. (2014). Quantitative analysis of the motifs contributing to the interaction between PLS-subfamily members and their target RNA sequences in plastid RNA editing. Plant J. **80:** 870–872.

Schallenberg-Rüdinger, M., Kindgren, P., Zehrmann, A., Small, I., and Knoop, V. (2013). A DYW-protein knockout in Physcomitrella affects two closely spaced mitochondrial editing sites and causes a severe developmental phenotype. Plant J. **76:** 420–432.

Takenaka, M., and Brennicke, A. (2003). In vitro RNA editing in pea mitochondria requires NTP or dNTP, suggesting involvement of an RNA helicase. J. Biol. Chem. **278:** 47526–47533.

Takenaka, M., Zehrmann, A., Brennicke, A., and Graichen, K. (2013a). Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. PLoS ONE **8:** e65343.

Takenaka, M., Zehrmann, A., Verbitskiy, D., Härtel, B., and Brennicke, A. (2013b). RNA editing in plants and its evolution. Annu. Rev. Genet. **47:** 335–352.

Yagi, Y., Hayashi, S., Kobayashi, K., Hirayama, T., and Nakamura, T. (2013). Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. PLoS ONE **8:** e57286.

Yagi, Y., Nakamura, T., and Small I. (2014). The potential for manipulating RNA with pentatricopeptide repeat proteins. Plant J. **78:** 772–782.

Yin, P., et al. (2013). Structural basis for the modular recognition of single-stranded RNA by PPR proteins. Nature **504:** 168–171.