Molecular Plant
Letter to the Editor

CellPress
PARTNER JOURNAL

Specific RNA Recognition by Designer Pentatricopeptide Repeat Protein

Dear Editor,

Manipulation of gene expression through targeting specific DNA or RNA sequences is a significant challenge. In the past decade, transcription activator-like (TAL) effectors and zinc fingers (ZFs) have been successfully developed into useful tools for DNA recognition (Bogdanove and Voytas, 2011; Deng et al., 2012a, 2012b). However, little progress has been made in the realm of RNA targeting due to the lack of understanding about the modular RNA recognition mechanism. Pumilio and FBF homology (PUF) proteins and pentatricopeptide repeat (PPR) proteins are two types of sequence-specific single-strand RNA (ssRNA) binding proteins with the potential to serve as effective RNA targeting tools (Filipovska and Rackham, 2013; Campbell et al., 2014). PPR proteins, generally containing 2–30 tandem repeats, are present in terrestrial plants as a large family (Schmitz-Linneweber and Small, 2008; Barkan and Small, 2014). PPR proteins function as sequence-specific single-stranded RNA binding proteins mainly in chloroplasts and mitochondria, where they are involved in many diverse aspects of organelle RNA metabolism processes, including RNA editing, maturation, stability, and translation. Each repeat of PPR is typically composed of 35 amino acids organized into a hairpin of α helices. Previous computational and biochemical analyses suggest a model of PPR modular RNA recognition: one RNA base coordinates with one PPR motif (Barkan et al., 2012; Yagi et al., 2013). The recently reported crystal structure of PPR10 in RNA-bound state (Protein Data Bank ID 4M59) corroborates this model (Yin et al., 2013).

According to the crystal structure of PPR10 in complex with RNA, within an intact PPR repeat, amino acid residues at positions 2, 5, and 35 are responsible for sequence-specific recognition of RNA bases (Figure 1A and Supplemental Figure 1A) (Yin et al., 2013). These three amino acids were also proposed as "code" amino acids for base discrimination (Barkan et al., 2012; Yagi et al., 2013; Barkan and Small, 2014). Two hydrophobic residues at position 2 from two consecutive repeats have been highlighted as putative RNA-interacting residues sandwiching one RNA base, indicating that an amino acid at this position could influence RNA recognition of a preceding PPR motif. The major determinant is the polar amino acid at position 5. Asparagine at this position strongly correlates with pyrimidine at corresponding position of target RNA, while threonine or serine correlates with purine. Another determinant is located at position 35, appearing to stabilize the conformation of the fifth residue. All amino acids at these two positions involved in nucleotide specification have side chains that are avid hydrogen bond donors or acceptors (Yin et al., 2013). Here, based on PPR structure and related bioinformatics analysis, we developed a set of designer proteins, which possess RNA recognition specificity with their artificial PPR motifs.

To detect protein–RNA interactions, we set up an *in vitro* assay aiming to examine the ligand binding activity and specificity of designer proteins (experimental details in Supplemental Materials and Methods). We looked into the design of specific PPR motifs and took a conservative approach to construct PPR motifs. First, we analyzed repeat sequences of all P-type PPR proteins, all of which contain 35 amino acids per repeat, from *Arabidopsis thaliana* (Supplemental Figure 1B; Lurin et al., 2004). Next, we selected the most evolutionarily conserved amino acids of P-type PPR motifs as the scaffold of designer PPR motifs to build up the primary structure of the RNA base recognition units. The exceptions among these amino acids (those at position 2, 5, and 35) are RNA selection "codes" (Figure 1A).

Our *in vitro* assay requires soluble proteins and radioactive target RNA probes. To optimize the solubility and behavior of designer proteins, we fused two capping domains, one N-terminal domain (NTD), comprising amino acids 37–208 of PPR10, and one C-terminal domain (CTD), comprising amino acids 737–786 of PPR10, to the amino and carboxyl termini of multiple designer PPR motifs (Supplemental Figure 2). We synthesized a series of designer protein genes with different PPR repeat motifs and purified the proteins for further biochemical assessment (for experimental details, see Supplemental Materials and Methods). Recombinant dPPRs (designer PPR proteins) were purified to homogeneity (Supplemental Figure 3). To achieve the goal of building up motif modules with the capability of specific RNA base recognition, we constructed designer proteins containing 10 tandem identical PPR repeat motifs with a number of combinations of "code" amino acids to identify the best code for specific RNA recognition (Figure 1B and Supplemental Figure 2). Through the electrophoretic mobility shift assay (EMSA), we tested more than 10 types of the most frequent combination identified by previous bioinformatics analysis (Barkan et al., 2012). Eventually we were able to construct PPR motifs as basic RNA base recognition units that selectively recognize RNA bases A, U, and C, corresponding with amino acid codes VSN, VND, and VNS, respectively (Figure 1A). Designer proteins all contain one NTD, 10 tandem artificially designed PPR motifs, and one CTD (Figure 1B). The designer proteins depicted in the schematic diagram in Figure 1B are designated as dPPR-A10, dPPR-U10, and dPPR-C10, respectively. Designer proteins dPPR-A10, dPPR-U10, and dPPR-C10 present high RNA binding specificity according to the EMSA results. As predicted, all of the three dPPRs selectively bound their respective target RNA, whereas no significant non-specific protein–RNA binding was detected. For instance, dPPR-A10 only bound probe RNA poly A_{10}, displaying no signs of interaction

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and IPPE, SIBS, CAS.

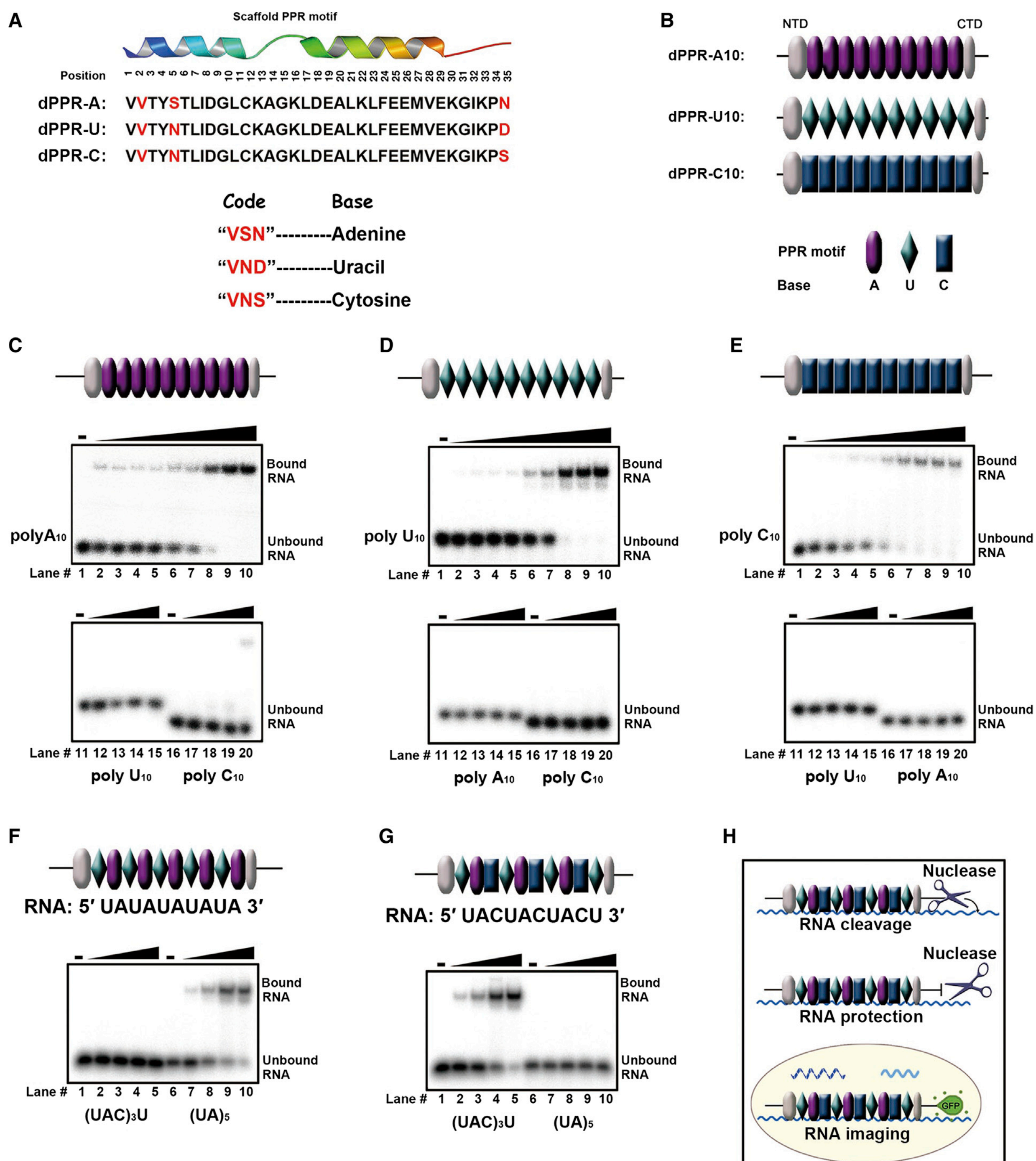

Figure 1. Specific RNA Recognition by Designer Pentatricopeptide Repeat Protein (dPPR).

(A) Sequences of scaffold PPR motif in dPPRs. A typical PPR motif, which contains 35 amino acids, assembles into a hairpin of helices. The code for specific nucleotide recognition comprising of three residues located at the 2nd, 5th, and 35th position are labeled in red (VSN, VND, and VNS recognize adenine, uracil, and cytosine, respectively).

(B) The design of engineered PPR protein. The same kind of scaffold PPR motif was tandem assembled and fused with N and C termini of PPR10. Designer PPR motifs of A, U, and C are shown in the shape of oval, rhombus, and rectangle.

(C–E) EMSA analysis of the specific RNA binding activity of dPPR-A10 **(C)**, dPPR-U10 **(D)**, and dPPR-C10 **(E)**. Designer PPR was added at a concentration of 0, 20, 30, 44, 66, 100, 150, 220, 330, and 500 nM in lanes 1–10 with target RNA, and of 0, 15, 45, 133, 400 nM in lanes 11–15 or lanes 16–20 with non-target RNA.

(F and G) Customized PPR, dPPR-UA **(F)** and dPPR-UAC **(G)** can specifically recognize their target RNA. Designer PPR protein was added at concentration of 0, 1.6, 8, 40, and 200 nM in lanes 1–5 and lanes 6–10 with ^{32}P-labeled RNA probe, with the sequence of tandem designer PPR repeat motif shown at the top. Detailed information is described in Supplemental Information.

(H) Potential applications of designer PPR protein in realms of RNA-related research.

with poly U_{10}, poly C_{10} (Figure 1C), or poly G_{10} (Supplemental Figure 4). The apparent dissociation constant of dPPR-A10 and its ligand is approximately 160 nM. Designer protein dPPR-U10 also exhibited behavior similar to that of dPPR-A10, and dPPR-C10 showed slightly higher binding affinity (Figure 1D and 1E; Supplemental Figure 4A). These results indicate that, in light of the perspective of protein engineering, 10 consecutive designer PPR motifs are sufficient to achieve specific RNA recognition. We also intend to build up specific PPR motifs recognizing RNA base G. However, we were unable to determine the appropriate combination of the three-position code, likely due to the unsuitability of the other 32 non-code amino acids as a motif scaffold or the structural instability of poly-G tracts.

We next tested the usability of designer PPR proteins with combinations of the designer PPR motifs that we determined specifically bind RNA bases A, U, and C. We designed and purified proteins containing binary and ternary blocks of designer PPR motif A, U, and C in patterns similar to those of dPPR-A/U/C10 and designated them as dPPR-UA and dPPR-UAC, respectively. dPPR-UA contains five sets of consecutive designer PPR motifs recognizing UA, while dPPR-UAC comprises three sets of designer PPR motifs UAC and a designer PPR motif U. dPPR-UA selectively bound RNA probe $(UA)_5$, not $(UAC)_3U$, and *vice versa* (Figure 1F and 1G). Neither of these two designer proteins was capable of binding RNA N_{10} (N stands for different types of RNA nucleotides) probes (Supplemental Figure 5), suggesting that for a 10-nucleotide long RNA and designer protein that contains 10 consecutive PPR repeat motifs, one corresponding motif out of every two or three is insufficient to achieve specific RNA recognition. We also attempted to discover the minimum number of PPR motif repeats required for specific RNA binding using native gel-shift assay (Supplemental Figure 6). The result indicates that the minimum numbers of PPR motifs to achieve specific binding differ for different types of RNA nucleotides, even though the apparent dissociation constants of dPPR-N10s with corresponding RNA are at similar levels (Supplemental Figure 4A). For example, 6-mers of dPPR-A motif is enough for nucleotide A recognition, whereas 8-mers of dPPR-U or dPPR-C show sufficient RNA binding activities.

In summary, we designed various types of chimeric recombinant proteins containing specific PPR motifs, which recognize RNA bases A, C, and U with a high degree of modular selectivity, and achieved specific RNA recognition by designer pentatricopeptide repeat proteins. Many obstacles still hinder the manipulation of designer proteins with target ssRNA. Several parameters of designer proteins and designer motifs remain to be optimized, such as motif numbers, amino acid sequences of the PPR motif, combinations of different motifs, and so forth. Our future research will also concentrate on designing PPR motifs that specifically bind the RNA base guanine, determining atomic structures of designer protein-target RNA complexes and quantifying these RNA recognition codes' specificities (Campbell et al., 2014). Potentially, with more thorough study into designer PPR motifs, engineered RNA editing factors could be applied to modify the amino acid sequences of organelle-encoded proteins, and domains with diversified functions (e.g. RNA-cleaving enzymes or fluorescent proteins) could be targeted to specific organellar RNAs via designer PPR tracts (Figure 1H). The development of analogous applications outside of organelles may eventually be feasible, demonstrating that PPR-based designer proteins show promise as a universal RNA targeting/processing tool in the future (Filipovska and Rackham, 2013; Barkan and Small, 2014; Yagi et al., 2014).

SUPPLEMENTAL INFORMATION

Supplemental Information is available at *Molecular Plant Online*.

FUNDING

This work was funded by the National Natural Science Foundation of China (Program No. 31200567), the Fundamental Research Funds for the Central Universities (Program No. 2014JQ001), and the Huazhong Agricultural University Scientific & Technological Self-innovation Foundation (Program No. 2013RC013).

ACKNOWLEDGMENTS

No conflict of interest declared.

Received: December 3, 2014
Revised: December 29, 2014
Accepted: January 3, 2015
Published: January 7, 2015

Cuicui Shen[1,3,6], Xiang Wang[1,2,6], Yexing Liu[5], Quanxiu Li[4,5], Zhao Yang[1,3], Nieng Yan[4,5], Tingting Zou[3] and Ping Yin[1,3,]*

[1]National Key Laboratory of Crop Genetic Improvement and National Centre of Plant Gene Research, Huazhong Agricultural University, Wuhan 430070, China
[2]College of Plant Sciences and Technology, Huazhong Agricultural University, Wuhan 430070, China
[3]College of Life Sciences and Technology, Huazhong Agricultural University, Wuhan 430070, China
[4]State Key Laboratory of Bio-membrane and Membrane Biotechnology, School of Life Sciences and School of Medicine, Tsinghua-Peking Center for Life Sciences, Beijing 100084, China
[5]Center for Structural Biology, School of Life Sciences and School of Medicine, Tsinghua-Peking Center for Life Sciences, Beijing 100084, China
[6]These authors contributed equally to this article.
***Correspondence: Ping Yin (yinping@mail.hzau.edu.cn)**
http://dx.doi.org/10.1016/j.molp.2015.01.001

REFERENCES

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

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.

Bogdanove, A.J., and Voytas, D.F. (2011). TAL effectors: customizable proteins for DNA targeting. Science **333**:1843–1846.

Campbell, Z.T., Valley, C.T., and Wickens, M. (2014). A protein-RNA specificity code enables targeted activation of an endogenous human transcript. Nat. Struct. Mol. Biol. **21**:732–738.

Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J.K., Shi, Y., and Yan, N. (2012a). Structural basis for sequence-specific recognition of DNA by TAL effectors. Science **335**:720–723.

Deng, D., Yin, P., Yan, C., Pan, X., Gong, X., Qi, S., Xie, T., Mahfouz, M., Zhu, J.K., Yan, N., et al. (2012b). Recognition of methylated DNA by TAL effectors. Cell Res. **22**:1502–1504.

Filipovska, A., and Rackham, O. (2013). Pentatricopeptide repeats: modular blocks for building RNA-binding proteins. RNA Biol. **10**:1426–1432.

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.

Schmitz-Linneweber, C., and Small, I. (2008). Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends Plant Sci. **13**:663–670.

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., Li, Q., Yan, C., Liu, Y., Liu, J., Yu, F., Wang, Z., Long, J., He, J., Wang, H.W., et al. (2013). Structural basis for the modular recognition of single-stranded RNA by PPR proteins. Nature **504**:168–171.