THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 288, NO. 44, pp. 31540–31548, November 1, 2013

Structure of a PLS-class Pentatricopeptide Repeat Protein Provides Insights into Mechanism of RNA Recognition

Received for publication, June 28, 2013, and in revised form, September 17, 2013 Published, JBC Papers in Press, September 18, 2013, DOI 10.1074/jbc.M113.496828

Ting Ban[‡§1], Jiyuan Ke[¶1], Runze Chen[§1], Xin Gu[¶], M. H. Eileen Tan[¶‖], X. Edward Zhou[¶], Yanyong Kang[¶], Karsten Melcher[¶], Jian-Kang Zhu[‡2], and H. Eric Xu[§¶3]**

*From the [‡]Shanghai Center for Plant Stress Biology and Shanghai Institute of Plant Physiology and Ecology, Shanghai Institutes Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China, the [§]Van Andel Research Institute/Shanghai Institute of Materia Medica (ARI/SIMM) Center, Center for Structure and Function of Drug Targets, Key Laboratory of Receptor Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China, the [¶]Laboratory of Structural Sciences, Center for Structural Biology and Drug Discovery, Van Andel Research Institute, Grand Rapids, Michigan 49503, the [‖]Department of Obstetrics and Gynecology, National University Hospital, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119228, and the **Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette, Indiana 47906*

Background: Pentatricopeptide repeat (PPR) proteins are sequence-specific RNA-binding proteins involved in organelle RNA processing.
Results: We identified RNA-binding sites of a small PPR protein (THA8L) from *Arabidopsis thaliana* and solved its crystal structure.
Conclusion: THA8L-RNA binding is dependent on a combination of specific nucleotide base interactions and nonspecific backbone interactions.
Significance: This work advances our understanding of the mechanism of PPR protein-RNA interaction.

Pentatricopeptide repeat (PPR) proteins are sequence-specific RNA-binding proteins that form a pervasive family of proteins conserved in yeast, plants, and humans. The plant PPR proteins are grouped mainly into the P and PLS classes. Here, we report the crystal structure of a PLS-class PPR protein from *Arabidopsis thaliana* called THA8L (THA8-like) at 2.0 Å. THA8L resembles THA8 (thylakoid assembly 8), a protein that is required for the splicing of specific group II introns of genes involved in biogenesis of chloroplast thylakoid membranes. The THA8L structure contains three P-type PPR motifs flanked by one L-type motif and one S-type motif. We identified several putative THA8L-binding sites, enriched with purine sequences, in the group II introns. Importantly, THA8L has strong binding preference for single-stranded RNA over single-stranded DNA or double-stranded RNA. Structural analysis revealed that THA8L contains two extensive patches of positively charged residues next to the residues that are proposed to comprise the RNA-binding codes. Mutations in these two positively charged patches greatly reduced THA8L RNA-binding activity. On the basis of these data, we constructed a model of THA8L-RNA binding that is dependent on two forces: one is the interaction between nucleotide bases and specific amino acids in the PPR motifs (codes), and the other is the interaction between the negatively charged RNA backbone and positively charged residues of PPR motifs. Together, these results further our understanding of the mechanism of PPR protein-RNA interactions.

The pentatricopeptide repeat (PPR)[4] proteins form an exceptionally large family of conserved RNA-binding proteins found in yeast, plants, and humans. In higher plants, PPR proteins are involved primarily in mitochondrial and chloroplast RNA processing, with >400 members in *Arabidopsis thaliana* (1, 2). PPR proteins are characterized by degenerate 35-amino acid repeats arranged in tandem, are localized predominantly to chloroplasts and mitochondria, and take part in virtually all processes that affect RNA metabolism in these organelles (3–5). Plant chloroplasts contain a complex genetic framework to realize the genetic information encoded in their DNA. Chloroplasts in land plants have retained ≥100 genes, most of which encode components of the basal chloroplast gene expression machinery or subunits of photosynthetic complexes. Post-transcriptional aspects of gene expression in land plant chloroplasts are particularly complex, including RNA editing, segmental mRNA stabilization, and the splicing of ≥20 group II introns (6, 7). The expression of the small chloroplast genome requires hundreds of nuclear gene products; among them are PPR proteins, which play a crucial role in recognizing and regulating RNA processing through their ability to bind target RNA in a sequence-specific manner (8–10).

PPR proteins can be separated into two main classes, denoted P and PLS, depending on the subtypes of PPR motifs. P-class PPR proteins contain tandem arrays of 35-amino acid PPR motifs, typically ended with a proline (P). Members of this class

[1] These authors contributed equally to this work.
[2] To whom correspondence may be addressed: Dept. of Horticulture and Landscape Architecture, Purdue University, West Lafayette, IN 47906. E-mail: zhu132@purdue.edu.
[3] To whom correspondence may be addressed: Lab. of Structural Sciences, Center for Structural Biology and Drug Discovery, Van Andel Research Inst., 333 Bostwick Ave. N.E., Grand Rapids, MI 49503. Tel.: 616-234-5772; Fax: 616-234-5170; E-mail: eric.xu@vai.org.

[4] The abbreviations used are: PPR, pentatricopeptide repeat; *A. thaliana* THA8, AtTHA8; SUMO, small ubiquitin-like modifier; SeMet, selenomethionine.

have been implicated in RNA stabilization, processing, splicing, and translation. PLS-class proteins contain alternating canonical P-type motifs and variant "long" (L)- and short (S)-type motifs and function mainly in RNA editing (3). PPR proteins bind RNA through contribution of individual PPR motifs, which are predicted to adopt a two-α-helix structure (11, 12). A combinatorial amino acid code for RNA recognition by PPR proteins has been recently proposed for P- and S-type motifs but not for L-type motifs, which are proposed to function as linkers that connect P- and S-type motifs in PPR proteins (1, 13). This putative code of RNA recognition is consistent with data on engineered PPR proteins and has greatly enhanced our understanding of PPR protein-RNA interactions (2, 5).

Structural studies of PPR-containing proteins have revealed the overall arrangement of PPR motifs and have shed light on the mechanism of RNA binding by PPR motifs. The crystal structure of PRORP1 (proteinaceous RNase P1) from *A. thaliana* reveals a prototypical metallonuclease domain tethered to a PPR domain by a zinc-binding domain, where the PPR domain enhances catalytic activity through interaction with pre-tRNA (14). The structure of human IFIT5 in complex with a 5′-triphosphate RNA reveals that PPR motifs interact with the triphosphate groups of nucleotides but make no base contacts (15). Thus, the structural mechanism of sequence-specific RNA binding by PPR proteins is not understood, and it is anticipated that structures of PPR proteins will contribute to our understanding of PPR protein-RNA recognition.

THA8 (*thylakoid assembly 8*) is a maize gene originally identified from a screen for nuclear mutations that cause defects in the biogenesis of chloroplast thylakoid membranes. The *THA8* gene encodes a small protein that is localized to chloroplasts, where it is required for the splicing of the *ycf3-2* and *trnA* group II introns, which contain multiple THA8-binding sites (7). Analysis of the *THA8* ortholog in *A. thaliana* showed that their molecular functions are conserved. Null mutations in *THA8* are embryo-lethal in *Arabidopsis* and seedling-lethal in maize. Whereas most PPR proteins have >10 PPR motifs, THA8 belongs to a subfamily of plant PPR proteins with a predicted four-and-a-half PPR motifs and have the potential to mediate specific RNA binding *in vivo* despite its small size. THA8L (THA8-like) protein from *A. thaliana* is a small PPR protein that resembles THA8 and is predicted to comprise four PPR motifs, but its function is unknown. In this study, we solved the THA8L crystal structure, identified its RNA-binding site, and performed mutational and biochemical analyses of THA8L-RNA interactions. These results provide important insights into the function of THA8L in specific RNA binding and processing.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification—THA8 (residues 34–263), AtTHA8 (residues 34–222), or THA8L (residues 60–257) was expressed as a His_6-small ubiquitin-like modifier (SUMO) fusion protein from the expression vector pET24a (Novagen). The modified fusion protein contains a His_6 tag (MKKGHHHHHHG) at the N terminus and a ULP1 protease site between the SUMO and PPR proteins. BL21(DE3) cells transformed with the expression plasmid were grown in LB broth at 16 °C to A_{600} ~ 1.0 and induced with 0.1 mM isopropyl 1-thio-β-D-galactopyranoside for 16 h. Cells were harvested, resuspended in 35 ml of extract buffer (20 mM Tris (pH 8.0), 200 mM NaCl, and 10% glycerol) per 2 liter of cells, and passed three times through a French press with pressure set at 1000 pacals. The lysate was centrifuged at 16,000 rpm (30,800 × *g*) in a Sorvall RC12BP rotor for 30 min, and the supernatant was loaded on a 15-ml nickel HP® column. The column was washed with 10% buffer A (20 mM Tris (pH 8.0), 200 mM NaCl, 500 mM imidazole, and 10% glycerol) for 150 ml and eluted in two steps with 50% buffer A for 100 ml and then 100% buffer A for 50 ml. The eluted His_6-SUMO-AtTHA8, His_6-SUMO-THA8L and His_6-SUMO-THA8 proteins were dialyzed against extract buffer and cleaved overnight with ULP1 at a protease/protein ratio of 1:500 in the cold room. The cleaved His_6-SUMO tag was removed by passing through a nickel HP column, and the protein was further purified by chromatography on a HiLoad 26/60 Superdex 200 gel filtration column in 25 mM Tris (pH 8.0), 200 mM ammonium acetate, 1 mM dithiothreitol, and 1 mM EDTA. Apo-AtTHA8, apo-THA8L, and apo-THA8 eluted as sharp single peaks with an estimated molecular mass of 23 kDa, suggesting that apo-AtTHA8, apo-THA8L, and apo-THA8 are monomers in solution.

To prepare selenomethionine (SeMet)-substituted THA8L protein, the pET24a-His_6-SUMO-THA8L expression plasmid was transformed into B834 methionine auxotroph cells. A single colony was inoculated into 2 liters of LB broth with antibiotics + 1% glucose and shaken overnight at 30 °C. 2 liters of cells were spun down and resuspended in 600 ml of filtered H_2O. The following solutions were added step by step and mixed well to make the 2-liter SeMet-substituted expression medium: 40 ml of solution E (50× M9 Minimal Medium Part B and 150 g of KH_2PO_4 in 1 liter of H_2O), 40 ml of solution D (50× M9 Minimal Medium Part A, 300 g of Na_2HPO_4 in 1 liter of H_2O), 20 ml of solution C (5 g of EDTA, 0.5 g of $FeCl_3$, 0.1 g of $ZnSO_4$, 0.05 g of $CuCl_2{\cdot}2H_2O$, 0.05 g of $CoCl_2$, and 0.05 g of $(NH_4)_6Mo_7O_{24}{\cdot}4H_2O$ in 1 liter of H_2O with pH adjusted to 7.0), 100 ml of solution B (600 ml of 0.5 g of 19 amino acids except methionine and 0.5 g of SeMet in 500 ml of H_2O, 0.6 g of $MgCl_2$, and 0.6 g of $CaCl_2$), 300 ml of solution A (1.8 liters (filtered) of 12 g of hydrolyzed herring sperm DNA (pH 7.0), 120 g of glucose, 250 mg of vitamin B_1, 12 g of $NH_4Cl{\cdot}6H_2O$, and appropriate antibiotics), and 1500 ml of autoclaved H_2O. 100 ml of cells were transferred into each 2-liter flask. The cells were grown at 22–25 °C until A_{600} reached 1.0–1.2. Protein expression was induced overnight with 0.1 mM isopropyl 1-thio-β-D-galactopyranoside at 16 °C. The cells were harvested the next morning and resuspended in extract buffer. The protein was purified using the same protocol as the native protein.

Crystallization—Purified THA8L protein was concentrated to ~10–15 mg/ml prior to crystallization trials. Initial screening identified that PEG 3350 (Polyethylene Glycol 3350) is favorable for crystal formation. Optimization trays using PEG were set up manually using the hanging drop method at 20 °C. Rod-shaped crystals with a length of ~200 μm for the longest dimension were obtained using 1 μl of the purified protein and 1 μl of well solution (12% (w/v) PEG 3350 and 4% (v/v) Tacsimate (pH 4.0)). These crystals diffracted x-rays to ~2.0 Å at the

TABLE 1
X-ray diffraction data collection and refinement statistics for the THA8L structure

	THA8L structure	
	Native	Selenium SAD[a] data
Data collection		
Wavelength (Å)	0.9787	0.9793 (peak)
Space group	*C*2	$P2_1$
Resolution (Å)	50-2.0	50-2.1
Cell parameters		
a, *b*, *c* (Å)	117.2, 53.3, 42.3	43.7, 53.6, 88.8
α, β, γ	90°, 99.7°, 90°	90°, 102.5°, 90°
Total/unique reflections	125,366 /17,384	174,312/23599
Completeness (%)	99.2 (95.1)	99.8 (99.5)
I/σ	15.3 (2.0)	14.0 (2.6)
Redundancy	7.2 (5.9)	7.4 (7.1)
R_{merge}	0.072 (0.839)	0.093 (0.703)
Structure determination		
Resolution (Å)	30-2.0	50-2.1
No. of reflections	16,394	22,406
No. of protein molecules	1	2
No. of residues	182	363
No. of solvent molecules	128	249
No. of non-H atoms	1634	3270
R_{cryst} (%)	20.4	0.214
R_{free} (%)	25.3	0.233
r.m.s.d. bonds (Å)	0.015	0.011
r.m.s.d. angles	1.39°	1.45°
Average *B* factor (Å^2)	44.2	36.0

[a] SAD, single-wavelength anomalous diffraction; r.m.s.d., root mean square deviation.

LS-CAT beamline of the Advanced Photon Source synchrotron. SeMet-substituted THA8L protein was crystallized using 1 μl of the purified protein and 1 μl of well solution (13% (w/v) PEG 3350, 2% (v/v) Tacsimate (pH 4.0), and 0.1 M sodium acetate (pH 4.6)). Rod-shaped crystals were obtained with a length of ~100 μm for the longest dimension.

Data Collection and Structure Determination—All crystals were transferred to well solution with 22% (v/v) ethylene glycol as cryoprotectant before flash-freezing in liquid nitrogen. Data collection was performed at beamline 21-ID-D (LS-CAT) of the Advanced Photon Source synchrotron. A native data set was collected to 2.0 Å. Initial structure determination by molecular replacement using the PPR domain from the crystal structure of PRORP1 (Protein Data Bank code 4G23; which shares ~20% sequence identity with THA8L) as a search model failed to yield any correct solution. To solve the phase problem, one data set of the SeMet-substituted THA8L crystal was collected at a wavelength of 0.9793 Å (peak wavelength) using an inversed-beam strategy to measure the selenium anomalous signal. The data were processed using XDS (16), combined using Pointless, and scaled using Scala of the CCP4 suite (17). Initial phases were established using the SHELX program (18) with the single selenium anomalous data set by single-wavelength anomalous diffraction phasing (Table 1). A total of 17 selenium sites were identified by SHELXD with a CC_{all}/CC_{weak} score of 61.0/38.0, and subsequent phasing was performed using SHELXE. Density modification for the initial electron density map was performed using DM (19). A crude model was built automatically using the CCP4 program buccaneer with R/R_{free} of 0.341/0.396. The model was further improved by several cycles of manual building using Coot (20) and refinements with the refmac program of CCP4 (21) to an *R* factor of 0.21 and an R_{free} factor of 0.23. The SeMet-substituted THA8L crystal has a space group of $P2_1$ with two molecules/asymmetric unit. One of the two models from the SeMet crystal was used to refine against the native data set, which was crystallized in the *C*2 space group with one molecule/asymmetric unit. The final structure model for THA8L was refined to an *R* factor of 0.20 and an R_{free} factor of 0.25 (Table 1).

Assays for Interactions between THA8L and Nucleic Acids—Interactions between AtTHA8, THA8L, or THA8 and biotin-Zea_{1a} RNA were assessed by luminescence-based AlphaScreen technology (PerkinElmer Life Sciences), which our group has used extensively to determine protein-protein or protein-small molecule interaction (22, 23). Biotin-RNAs were attached to streptavidin-coated donor beads, and His_6-tagged THA8 and THA8L proteins were attached to nickel-chelated acceptor beads. The donor and acceptor beads were brought into proximity by the interactions between PPR proteins and biotin-RNAs, which were measured with different concentrations of biotin-RNAs or after addition of Zea_4 ssRNA, dsRNA, RNA/DNA duplex, ssDNA, and dsDNA to the reaction system. When excited by a laser beam of 680 nm, the donor beam emits singlet oxygen that activates thioxene derivatives in the acceptor beads, which releases photons of 520–620 nm as the binding signal (see Fig. 2*B*). The experiments were conducted with 20 nM THA8L and biotin-Zea_{1a} RNA in the presence of 5 μg/ml donor and acceptor beads in 50 mM MOPS (pH 7.4), 50 mM NaF, 50 mM CHAPS, and 0.1 mg/ml bovine serum albumin. The results were based on an average of three experiments with standard errors typically <10% of the measurements. The IC_{50} values were derived from curve fitting based on a competitive inhibitor model for the binding of His_6-THA8L and biotin-RNA using GraphPad Prism.

Gel Mobility Shift Assay—A gel mobility shift assay was performed to detect RNA binding by THA8L protein. Both Zea_4 and control HB9 RNA oligoribonucleotides were 5′-end-labeled with [γ-^{32}P]ATP using T4 polynucleotide kinase according to the manufacturer's protocol (Invitrogen). The labeled RNA probe was separated from unincorporated nucleotides by centrifugation using a Sephadex G-25 quick-spin column (Roche Applied Science). The binding reaction contained 25 mM Tris (pH 8.0), 0.125 mM EDTA, 10% glycerol, 25 mM KCl, 1 ng of labeled RNA oligonucleotides, and various amounts of THA8L protein as indicated. Binding reactions were incubated for 30 min at room temperature and resolved on 6% native polyacrylamide gel running in 0.5 × Tris borate/EDTA buffer at 100 V. Results were visualized on a PhosphorImager (Fujifilm).

Analytic HPLC Gel Filtration—Size exclusion chromatography was performed using a Nanofilm SEC-500 column (Sepax Technology) in an SPD-20AV system (Shimadzu). The column was run on a Waters HPLC system at a flow rate of 0.35 ml/min with dual mode detection at 280 and 220 nm. The column was equilibrated with running buffer (20 mM Tris (pH 7.5), 50 mM NaCl, and 2% glycerol) to obtain a stable base line. After that, 50 μl of ~1.0 mg/ml wild-type or mutant THA8L protein solutions were centrifuged, and the supernatant was loaded onto the column to detect the association state and the elution time of the main peak of THA8L solutions.

Mutagenesis—Site-directed mutagenesis was carried out using the QuikChange method (Stratagene) or the GeneTailor

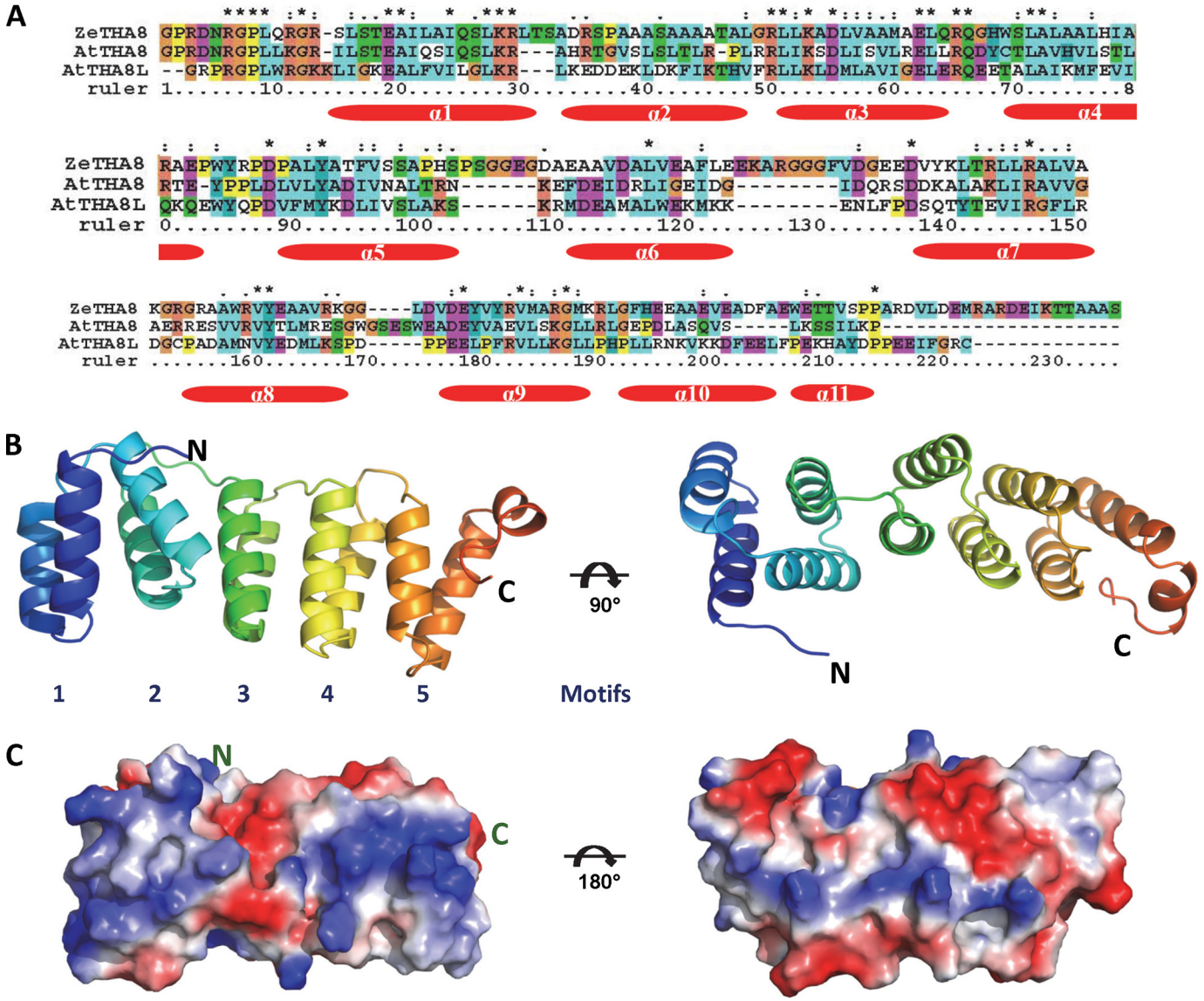

FIGURE 1. **Structure of *Arabidopsis* THA8L.** *A*, sequence alignment of *Zea mays* THA8 (*ZeTHA8*), AtTHA8, and AtTHA8L. The secondary structure elements are indicated below the sequences. Chloroplast signal peptides are excluded from the sequences. The alignment was done by ClustalW. *B*, two 90° views of the THA8L structure. Note that the THA8L structure contains 11 α-helices; the first 10 helices comprise five PPR motifs, and the final short helix helps to cap the fifth PPR motif. The first PPR motif is L-type, followed by three P-type motifs and a shorter S-type motif that has 31 amino acids. *C*, two views of THA8L surface charge potential with *blue* for positive charges, *red* for negative charges, and *white* for neutral surface.

system (Invitrogen). Mutations and all plasmid constructs were confirmed by sequencing before protein expression. Expression and purification of His-SUMO-THA8L mutant proteins were performed as described above.

RESULTS AND DISCUSSION

Structure Determination of an AtTHA8-like PPR Protein—Because THA8 is the first small PPR protein with a defined phenotype and molecular function in RNA splicing (7), we focused our studies on this subfamily of PPR proteins. Both THA8 proteins from maize and *A. thaliana* are highly conserved, with 75% sequence identify to one other (Fig. 1*A*). Sequence analysis indicated that there is a THA8L protein (gene locus At3G46870) in *A. thaliana*, with 25% sequence identity to THA8. All three PPR proteins share a similar arrangement in their secondary structural elements, indicating that they may function in a similar way in RNA recognition.

The above three proteins were expressed in *Escherichia coli* and purified for crystallization. Only THA8L (residues 60–257) produced high quality crystals that diffracted to 2.0 Å resolution. The THA8L structure was solved by the single-wavelength anomalous diffraction method using crystals grown from SeMet-substituted protein. The THA8L structure contains 11 α-helices; the first 10 helices comprise five PPR motifs, and the final short helix helps to cap the fifth PPR motif (Fig. 1*B*). The first PPR motif is L-type, followed by three P-type motifs and a shorter S-type motif that has 31 amino acids (Fig. 1*B*; see a more detailed explanation in Fig. 6*A*). The overall arrangement of the five-and-one-half PPR motif is a relatively straight rectangle box (Fig. 1*B*). The structure also reveals that one side of THA8L

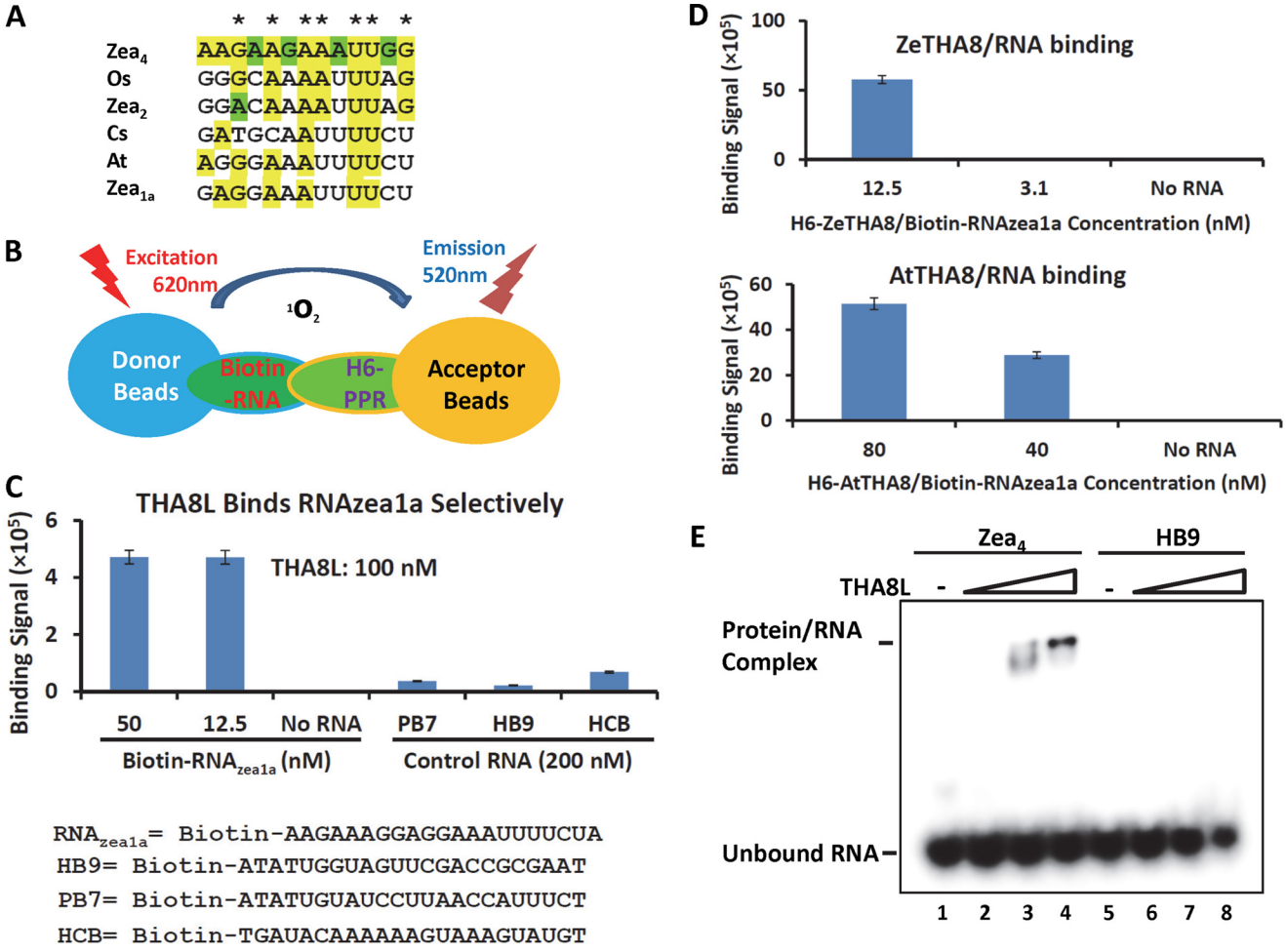

FIGURE 2. **THA8L binds to specific RNA sequences in *ycf3* intron 2.** *A*, sequence alignment of conserved RNA elements of *ycf3* intron 2 from *Z. mays* (*Zea*), *O. sativa* (*Os*), *A. thaliana* (*At*), and *G. max* (*Cs*). Zea_{1a}, Zea_2, and Zea_4 are three putative binding sites of THA8 in the *ycf3* intron 2. *B*, diagram of AlphaScreen assays for detecting THA8L-RNA binding. *C*, THA8L-RNA binding was detected by AlphaScreen assay. Significant binding signals were detected with 12.5 and 50 nM biotin-Zea_{1a} RNA and 100 nM His_6-AtTH8L. HB9, PB7, and HCB (control biotin-RNAs that bind other PPR proteins) showed weak or no binding to His_6-THA8L. All RNA sequences used for binding assays are listed below. *D*, RNA binding assay of THA8 and AtTHA8 with biotin-Zea_{1a} RNA at a ratio of 1:1 showed that THA8 has the highest affinity for biotin-Zea_{1a} RNA. AtTHA8 showed a significant RNA binding signal at a concentration of 80 nM, whereas THA8 showed significant binding to biotin-Zea_{1a} RNA at concentrations of 12.5 and 20 nM. *His_6-ZeTHA8*, His-*Z. mays* THA8. *E*, gel mobility shift assay was performed to detect RNA binding by THA8L protein. Increasing amounts of THA8L (0, 1, 4, and 8 μM) were incubated with 1 ng of labeled Zea_4 RNA or HB9 (control RNA). Protein-RNA complex formation was detected for Zea_4 RNA, but not for HB9 RNA (negative control).

has much higher surface positive charge potential than the other side of THA8L, which may be involved in RNA binding (Fig. 1*C*).

THA8L Binds to Specific RNA Sequences in ycf3 Intron 2—It has been reported that *ycf3* intron 2 contains three THA8-binding sites within fragments 1a, 2, and 4, each having a length of ~150 nucleotides (7), but the exact binding sequences were not defined. We reasoned that the corresponding THA8-binding sites in *ycf3* intron 2 should be relatively conserved in maize, rice, *Arabidopsis*, and *Glycine max*. Sequence alignment of *ycf3* intron 2 from these species indeed revealed conserved segments enriched with GAA and UU sequences (Fig. 2*A*).

To determine whether these conserved segments are the binding sites of THA8/THA8L, we developed an AlphaScreen assay to detect the PPR protein-RNA interactions, as diagramed in Fig. 2*B*. In this assay, 5′-biotinylated RNA was attached to donor beads, and His_6-tagged THA8 protein was attached to acceptor beads. When the donor and acceptors are brought into proximity by the interaction between THA8 and RNA, illuminating the sample with a 620-nm laser will cause a singlet oxygen transfer from donor beads to acceptor beads and elicit a strong emission of lights at a shorter wavelength (~520 nm). As shown in Fig. 2*C*, the Zea_{1a} RNA fragment, a 21-nucleotide fragment that encompasses the GAA conserved sequence, elicited a strong binding signal with THA8L protein at a concentration of 12.5 or 50 nM. In contrast, THA8L showed little interaction with the control RNAs of PB7, HB9, and HCB, which are the corresponding RNA-binding sites for two unrelated PPR proteins, PPR10 (PB7) and HCF152 (HB9 and HCB) (24, 25), even when these control RNAs were presented at a concentration of 200 nM, indicating specific binding of THA8L to the Zea_{1a} RNA fragment. Similarly, THA8 proteins from maize and *Arabidopsis* also showed strong interactions with the Zea_{1a} RNA fragment (Fig. 2*D*). In addition, the gel mobility shift assay confirmed the specificity of THA8L-Zea_4 binding, whereas binding between THA8L and control RNA could not be detected (Fig. 2*E*). Because our crystal structure is the THA8L protein, we focused biochemical and mutagenesis studies on this protein.

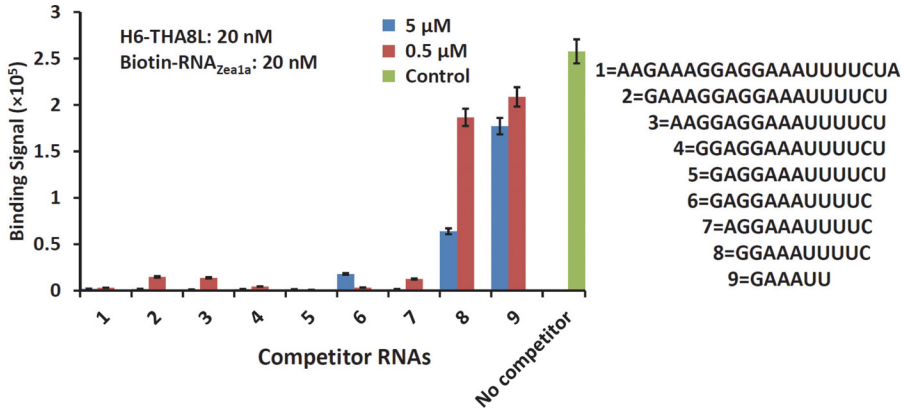

FIGURE 3. **Minimum THA8L-binding RNA sequences.** Nine RNA oligonucleotides with progressive 5′- and 3′-truncations of the original 21 nucleotides (AAGAAAGGAGGAAAUUUUCUA) without a biotin tag were used for competing the interaction between biotin-Zea$_{1a}$ RNA and His-THA8L (*H6-THA8L*). RNA-1 to RNA-7 efficiently competed for the binding between biotin-Zea$_{1a}$ RNA and His-tagged THA8L, whereas RNA-8 and RNA-9 had greatly reduced competition ability, indicating that RNA-7 (AGGAAAUUUUC) with 11 nucleotides has the minimum length for binding of THA8L. RNA sequences for the competition assay are listed.

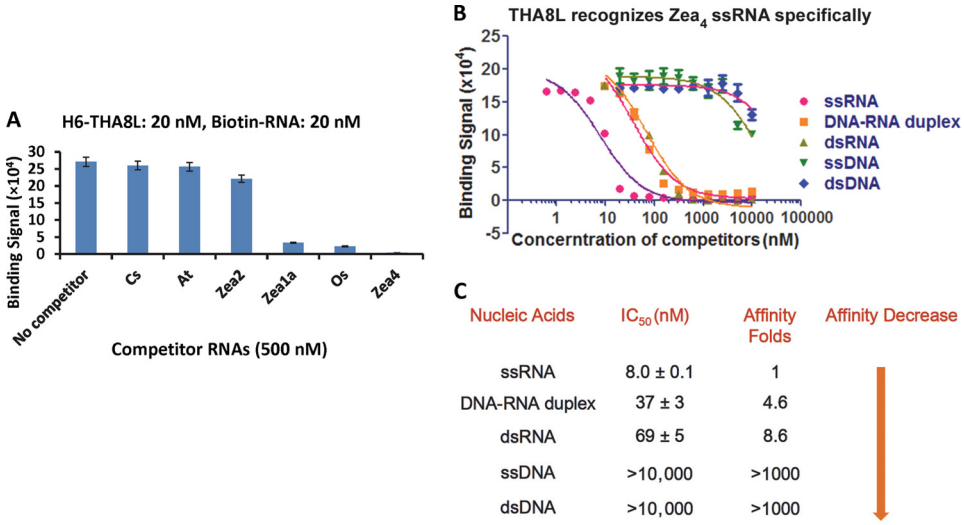

FIGURE 4. **THA8L recognizes the Zea$_4$ ssRNA fragment specifically.** *A*, relative affinities of various 13-nucleotide THA8L RNA-binding sites from different species as shown by RNA competition assays. RNA Zea$_4$ had the strongest competition ability, thus the strongest binding affinity for THA8L. *H6-THA8L*, His$_6$-THA8L. *B*, the binding signal of THA8L with different nucleic acids of Zea$_4$ demonstrated the order of competition capacity from strong to weak: ssRNA > RNA/DNA duplex > dsRNA > ssDNA > dsDNA. Different concentrations of ssRNA, dsRNA, RNA/DNA duplex, ssDNA, and dsDNA were used to compete the binding signals of His$_6$-THA8L/biotin-Zea$_{1a}$ RNA. *C*, IC$_{50}$ and -fold affinity of different nucleic acids of Zea$_4$ listed as a decreased order of competition capability. The IC$_{50}$ values were derived from curve fitting based on a competitive inhibitor model for competition of the binding of His$_6$-THA8L and biotin-Zea$_{1a}$ RNA.

Defining the Minimum THA8L-binding Site—Considering that THA8L has just five-and-a-half PPR motifs, we reasoned that only a subset of the 21 nucleotides in the Zea$_{1a}$ RNA fragment might be required for THA8L binding based on the proposed model of one PPR motif/one nucleotide (26). To determine what the minimum sequence required for THA8L binding is, we synthesized nine RNA oligonucleotides with progressive 5′- and 3′-truncations of the original 21 nucleotides (Fig. 3). These RNAs were made without a biotin tag and were used to compete for the interaction between biotin-Zea$_{1a}$ RNA and His-THA8L. As shown in Fig. 3, RNA-1 to RNA-7 efficiently competed for the binding between biotin-Zea$_{1a}$ RNA and His-THA8L, whereas RNA-8 and RNA-9 greatly lost the competition ability, indicating that 11 nucleotides (RNA-7, AGGAAAUUUUC) represent the minimum length for binding THA8L. The difference between RNA-7 and RNA-8 is a single A at the 5′-end, suggesting that this A base may play a key role in THA8L protein-RNA interaction.

RNA-5 with 13 nucleotides appeared to be a better competitor than RNA-7. We thus made the corresponding 13-nucleotide RNAs from rice, *Arabidopsis*, and *G. max* to assess which RNAs have the best binding affinity for THA8L (Fig. 2*A*). We also made the two putative 13-mer RNAs from fragments 2 and 4 in *ycf3* intron 2 (Fig. 2*A*). The relative affinities of these RNAs for THA8L were determined by their ability to compete with the binding between biotin-Zea$_{1a}$ RNA and His-THA8L using relatively low concentrations of competitor RNAs (500 nM) (Fig. 4*A*). Based on these data, the order of affinity of these RNAs for THA8L is as follows: Zea$_4$ > *Oryza sativa* > Zea$_{1a}$ ≫ Zea$_2$ > *G. max* = *A. thaliana*. It thus appears that Zea$_4$ with the sequence AAGAAGAAAUUGG has the best binding affinity for THA8L with an IC$_{50}$ value of 8 nM (Fig. 4*C*).

THA8L Specifically Recognizes ssRNA—PPR proteins are known to be RNA-binding proteins, but it is not known whether they prefer ssRNAs, hairpin RNA, or duplex RNA. To assess the binding preference of THA8L, we made Zea$_4$ ssRNA

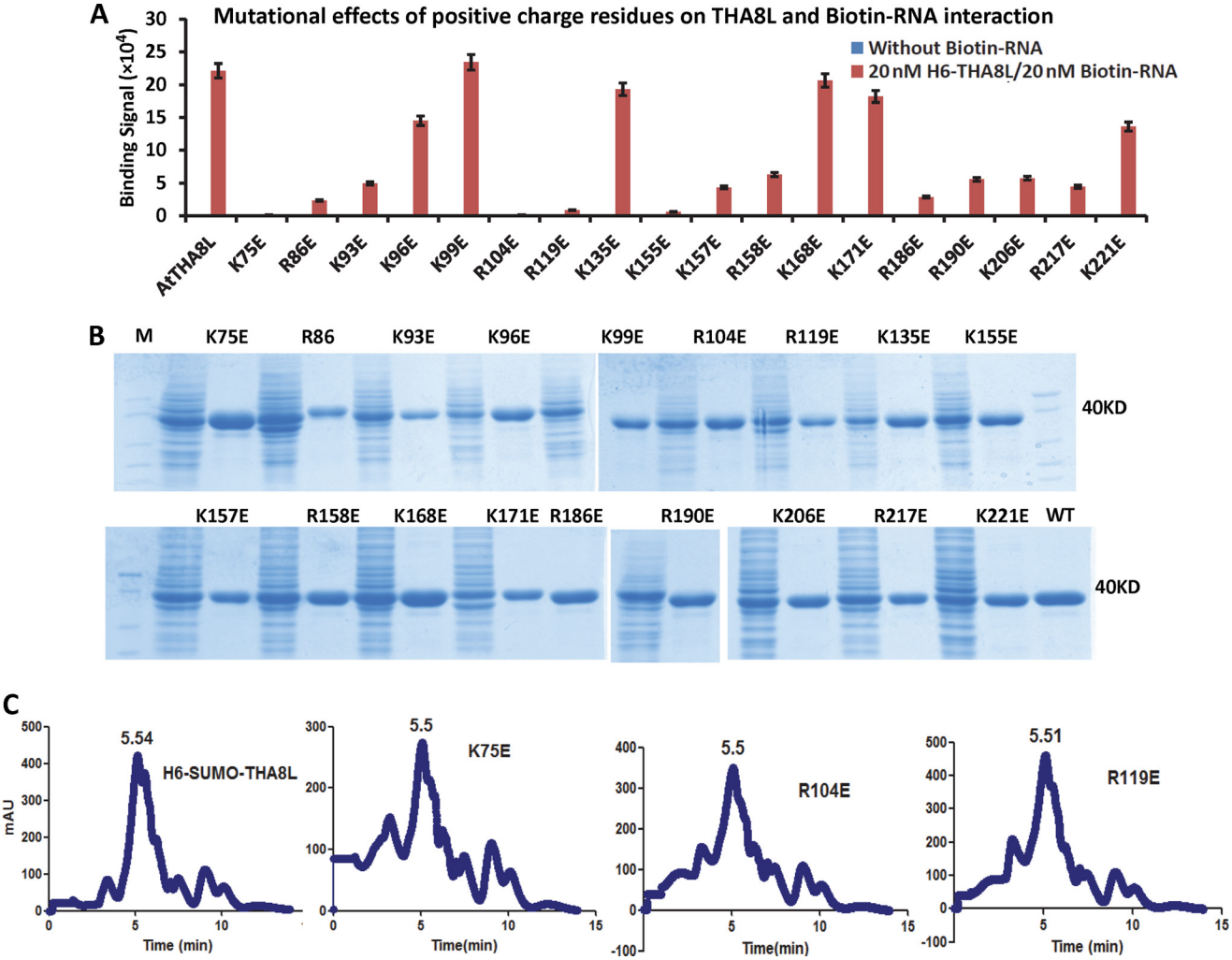

FIGURE 5. **Arg and Lys of THA8L play an important role in PPR-RNA binding.** *A*, effects of charge mutations on RNA binding measured by AlphaScreen assays. 12 of the 19 mutations reduced the binding of THA8L to RNA, and mutations at Lys-75, Arg-104, Lys-115, and Arg-119, nearly abolished RNA binding. *B*, SDS-PAGE of the purified protein samples of all 19 mutant THA8L proteins used for RNA binding assay. *C*, analytic HPLC gel filtration confirmed that all of the mutant proteins were properly folded. No aggregation/misfolding signals could be detected, and the main peaks of wild-type THA8L and mutants K75E, R104E, and R119E appeared at ~5.5 min. mAU, milli-absorbance units.

and its corresponding dsRNA as well as ssDNA and dsDNA. As shown in Fig. 4 (*B* and *C*), THA8L strongly preferred ssRNA over dsRNA or RNA/DNA duplex. THA8L did not bind to ssDNA or dsDNA, suggesting that the additional 2′-hydroxyl group in the ribose of RNA is important for the binding of THA8L. The strong preference of THA8L for ssRNA is consistent with the proposed function of the PPR motif in RNA binding and splicing.

Charge Interactions Are Important for RNA Binding by THA8L—We reasoned that the relatively high affinity binding of Zea_4 RNA by THA8L must involve extensive charge interactions between the positively charged residues of PPR motifs and the negatively charged phosphate backbone of RNA. To map these positively charged residues, we systematically mutated all arginines and lysines on the surface of the THA8L structure to the negatively charged glutamic acid. All of these mutated proteins were expressed and purified similarly as the wild-type protein for AlphaScreen RNA binding assays (Fig. 5*A*). 12 of the 19 charge reverse mutations showed reduction of RNA binding, and mutations of Lys-75, Arg-104, Lys-115, and Arg-119 nearly abolished RNA binding (Fig. 5*A*). All of these mutations are located on the same side of the surface as the predicted RNA code-determining residues of THA8L protein (Fig. 6*B*). All of mutant proteins were expressed and purified well (Fig. 5*B*), and analytic HPLC gel filtration revealed that they were eluted with nearly identical profiles as the wild-type protein (Fig. 5*C*), indicating that their loss of RNA-binding function is not due to misfolding of proteins.

THA8L RNA-binding Codes—It has been proposed that PPR tracts bind specific RNA nucleotides via the combinatorial action of two amino acids in each repeat (1, 13). The combinatorial amino acid code for nucleotide recognition by P-type PPR motifs was proposed to be as follows: $T_6D^{1'}$ = G, $T/S_6N^{1'}$ = A, $N_6D^{1'}$ = U, and $N_6N/S^{1'}$ = C, where residue 6 is the sixth residue of the first PPR motif and residue 1′ is the first residue of the next PPR motif. The S-type PPR motif may have a similar amino acid code, but there is no code proposed for L-type PPR motifs (1, 13). Sequence alignment of the five PPR motifs of THA8L indicated that motifs 2, 3, and 4 are P-type, whereas motifs 1 and 5 are L- and S-type, respectively (motif 1 does not

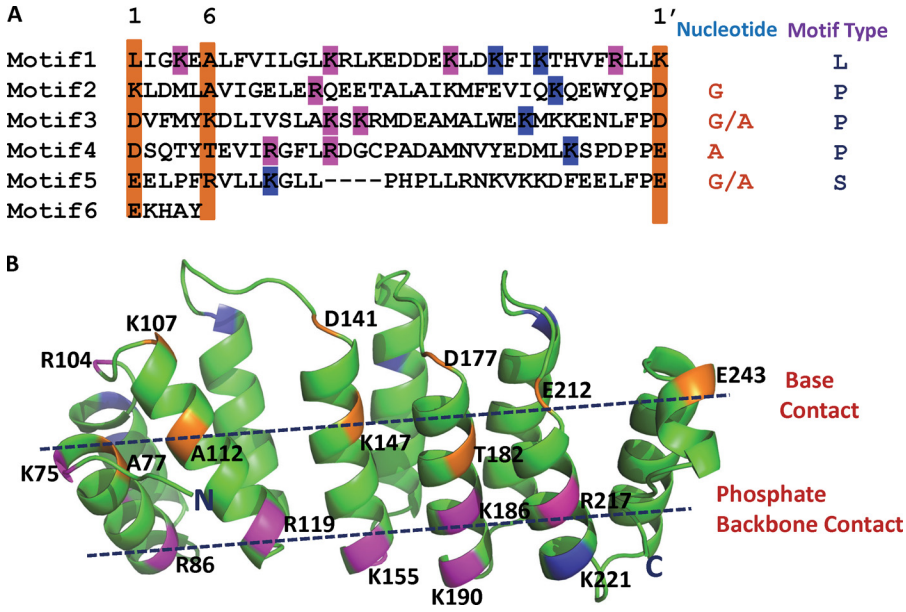

FIGURE 6. **Model of THA8L-RNA interaction.** *A*, an alignment of THA8L PPR repeats identifies the PPR motif types and the amino acid codes at positions 6 and 1′ of each PPR motif of THA8L. Residues at positions 6 and 1′ that are predicted to be the binding codes are highlighted in *orange*. Predicted binding nucleotides and PPR motif types are shown. *B*, code residues at positions 6 and 1′ are colored in *orange*. Mutations of 12 positively charged residues that reduced the binding of THA8L to RNA are colored in *magenta* and are located on the same side as the predicted code residues. Seven mutations that had little effect on RNA binding are marked in *blue* and are located on the other side of the protein. An RNA-binding model in which the specific base contacts are on top of the PPR motifs with the corresponding phosphate backbone contact near the bottom of the PPR motifs is indicated by the *two dashed lines*.

have the C-terminal conserved proline, and motif 5 contains 31 amino acids with a short three-turn helix) (Fig. 6*A*). On the basis of the proposed amino acid codes, we predicted that motif 2 with an Ala/Asp combination at position 6/1′ prefers G; motif 3 with a Lys/Asp combination prefers G/A; motif 4 with a Thr/Glu combination prefers A; and motif 5, an S-type motif with an Arg/Glu combination, prefers G/A. Our assignment of the binding code is consistent with the nucleotide preference of G/A-rich sequences by THA8L (Fig. 2*A*). As shown in Fig. 6*A*, the "coding" amino acids of THA8L are located on top of the positively charged residues whose mutations affect RNA binding. We propose a model with specific base contacts on top of the PPR motif and the corresponding phosphate backbone contact near the bottom of the PPR motifs (Fig. 6*B*).

Concluding Remarks—In this study, we solved the crystal structure of THA8L, a small protein that contains five-and-one-half PPR motifs. We also identified the small RNA-binding site for THA8L and demonstrated that THA8L has a strong preference for ssRNA over ssDNA or dsRNA. The THA8L structure reveals that PPR motifs adopt helix-turn-helix repeats that are packed into a relatively rigid rectangle structure (Fig. 1*B*). Analysis of amino acid codes for RNA recognition suggested that THA8L prefers a short G/A-rich sequence. On the basis of the location of amino acid codes and distribution of positively charged surface, we proposed a THA8L RNA-binding model with specific base contacts near the top of PPR motifs and the corresponding phosphate backbone contacts near the bottom of the PPR motifs. Validation of such a model requires the structure of a THA8L-RNA complex, which we have failed to obtain despite extensive screening of crystallization conditions with various RNA fragments. Nevertheless, the THA8L structure and the identification of its RNA-binding site as reported here provide important insights into THA8L-RNA interactions and establish THA8L as a sequence-specific RNA-binding protein that might have similar function as THA8 in RNA processing.

REFERENCES

1. 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
2. Prikryl, J., Rojas, M., Schuster, G., and Barkan, A. (2011) Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 415–420
3. Small, I. D., and Peeters, N. (2000) The PPR motif–a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem. Sci.* **25,** 46–47
4. Lurin, C., Andrés, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyère, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, M. L., Mireau, H., Peeters, N., Renou, J. P., Szurek, B., Taconnat, L., and Small, I. (2004) Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* **16,** 2089–2103
5. Fujii, S., and Small, I. (2011) The evolution of RNA editing and pentatricopeptide repeat genes. *New Phytol.* **191,** 37–47
6. Jonietz, C., Forner, J., Hildebrandt, T., and Binder, S. (2011) RNA PROCESSING FACTOR3 is crucial for the accumulation of mature *ccmC* transcripts in mitochondria of *Arabidopsis* accession Columbia. *Plant Physiol.* **157,** 1430–1439
7. Khrouchtchova, A., Monde, R. A., and Barkan, A. (2012) A short PPR protein required for the splicing of specific group II introns in angiosperm chloroplasts. *RNA* **18,** 1197–1209
8. 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
9. Schmitz-Linneweber, C., and Small, I. (2008) Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends Plant Sci.* **13,** 663–670
10. Johnson, X., Wostrikoff, K., Finazzi, G., Kuras, R., Schwarz, C., Bujaldon, S., Nickelsen, J., Stern, D. B., Wollman, F. A., and Vallon, O. (2010) MRL1, a conserved pentatricopeptide repeat protein, is required for stabilization

of *rbcL* mRNA in *Chlamydomonas* and *Arabidopsis*. *Plant Cell* **22,** 234–248
11. Loiselay, C., Gumpel, N. J., Girard-Bascou, J., Watson, A. T., Purton, S., Wollman, F. A., and Choquet, Y. (2008) Molecular identification and function of *cis*- and *trans*-acting determinants for *petA* transcript stability in *Chlamydomonas reinhardtii* chloroplasts. *Mol. Cell. Biol.* **28,** 5529–5542
12. Pfalz, J., Bayraktar, O. A., Prikryl, J., and Barkan, A. (2009) Site-specific binding of a PPR protein defines and stabilizes 5′ and 3′ mRNA termini in chloroplasts. *EMBO J.* **28,** 2042–2052
13. 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
14. Howard, M. J., Lim, W. H., Fierke, C. A., and Koutmos, M. (2012) Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5′ processing. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 16149–16154
15. Abbas, Y. M., Pichlmair, A., Górna, M. W., Superti-Furga, G., and Nagar, B. (2013) Structural basis for viral 5′-PPP-RNA recognition by human IFIT proteins. *Nature* **494,** 60–64
16. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 125–132
17. Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763
18. Sheldrick, G. M. (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 479–485
19. Cowtan, K. (1994) *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, Vol. 31, pp. 34–38, Daresbury Laboratory, Cheshire, United Kingdom
20. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126–2132
21. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53,** 240–255
22. Melcher, K., Ng, L. M., Zhou, X. E., Soon, F. F., Xu, Y., Suino-Powell, K. M., Park, S. Y., Weiner, J. J., Fujii, H., Chinnusamy, V., Kovach, A., Li, J., Wang, Y., Li, J., Peterson, F. C., Jensen, D. R., Yong, E. L., Volkman, B. F., Cutler, S. R., Zhu, J. K., and Xu, H. E. (2009) A gate-latch-lock mechanism for hormone signalling by abscisic acid receptors. *Nature* **462,** 602–608
23. Soon, F. F., Ng, L. M., Zhou, X. E., West, G. M., Kovach, A., Tan, M. H., Suino-Powell, K. M., He, Y., Xu, Y., Chalmers, M. J., Brunzelle, J. S., Zhang, H., Yang, H., Jiang, H., Li, J., Yong, E. L., Cutler, S., Zhu, J. K., Griffin, P. R., Melcher, K., and Xu, H. E. (2012) Molecular mimicry regulates ABA signaling by SnRK2 kinases and PP2C phosphatases. *Science* **335,** 85–88
24. Zhelyazkova, P., Hammani, K., Rojas, M., Voelker, R., Vargas-Suárez, M., Börner, T., and Barkan, A. (2012) Protein-mediated protection as the predominant mechanism for defining processed mRNA termini in land plant chloroplasts. *Nucleic Acids Res.* **40,** 3092–3105
25. Kobayashi, K., Kawabata, M., Hisano, K., Kazama, T., Matsuoka, K., Sugita, M., and Nakamura, T. (2012) Identification and characterization of the RNA binding surface of the pentatricopeptide repeat protein. *Nucleic Acids Res.* **40,** 2712–2723
26. Filipovska, A., and Rackham, O. (2012) Modular recognition of nucleic acids by PUF, TALE and PPR proteins. *Mol. Biosyst.* **8,** 699–708

Structure of a PLS-class Pentatricopeptide Repeat Protein Provides Insights into Mechanism of RNA Recognition

Ting Ban, Jiyuan Ke, Runze Chen, Xin Gu, M. H. Eileen Tan, X. Edward Zhou, Yanyong Kang, Karsten Melcher, Jian-Kang Zhu and H. Eric Xu

J. Biol. Chem. 2013, 288:31540-31548.
doi: 10.1074/jbc.M113.496828 originally published online September 18, 2013
