LETTER

doi:10.1038/nature12651

Structural basis for the modular recognition of single-stranded RNA by PPR proteins

Ping Yin[1,2]*, Quanxiu Li[1,2]*, Chuangye Yan[2,3], Ying Liu[1,2], Junjie Liu[2,3], Feng Yu[4], Zheng Wang[5], Jiafu Long[5], Jianhua He[4], Hong-Wei Wang[2,3], Jiawei Wang[1,2], Jian-Kang Zhu[6,7], Yigong Shi[2,3] & Nieng Yan[1,2]

Pentatricopeptide repeat (PPR) proteins represent a large family of sequence-specific RNA-binding proteins that are involved in multiple aspects of RNA metabolism. PPR proteins, which are found in exceptionally large numbers in the mitochondria and chloroplasts of terrestrial plants[1–5], recognize single-stranded RNA (ssRNA) in a modular fashion[6–8]. The maize chloroplast protein PPR10 binds to two similar RNA sequences from the *ATPI–ATPH* and *PSAJ–RPL33* intergenic regions, referred to as *ATPH* and *PSAJ*, respectively[9,10]. By protecting the target RNA elements from 5′ or 3′ exonucleases, PPR10 defines the corresponding 5′ and 3′ messenger RNA termini[9–11]. Despite rigorous functional characterizations, the structural basis of sequence-specific ssRNA recognition by PPR proteins remains to be elucidated. Here we report the crystal structures of PPR10 in RNA-free and RNA-bound states at resolutions of 2.85 and 2.45 Å, respectively. In the absence of RNA binding, the nineteen repeats of PPR10 are assembled into a right-handed superhelical spiral. PPR10 forms an antiparallel, intertwined homodimer and exhibits considerable conformational changes upon binding to its target ssRNA, an 18-nucleotide *PSAJ* element. Six nucleotides of *PSAJ* are specifically recognized by six corresponding PPR10 repeats following the predicted code. The molecular basis for the specific and modular recognition of RNA bases A, G and U is revealed. The structural elucidation of RNA recognition by PPR proteins provides an important framework for potential biotechnological applications of PPR proteins in RNA-related research areas.

PPR proteins function in multiple aspects of organelle RNA metabolism, such as RNA splicing, editing, degradation and translation[1–5]. In plants, PPR mutants may cause embryonic lethality[12–14], and a number of PPR proteins act as restorers of fertility to overcome cytoplasmic male sterility[15–19]. In humans, mutations in the mitochondrial PPR protein LRPPRC are associated with the French-Canadian-type Leigh syndrome characterized by the deficiency in Complex IV[20,21].

PPR proteins contain 2–30 tandem repeats, each typically comprising 35 amino acids that are organized into a hairpin of α-helices[1,6,22,23]. PPRs are divided into two classes: the P-class, whose members only comprise the 35-amino-acid repeats; and the PLS-class, which has repeats of 31–36 amino acids and extra domains at the carboxyl terminus[3,12]. Computational and biochemical analyses suggest that PPR proteins may recognize RNA in a modular fashion, but different from that of the RNA-binding PUF domain[6,24]. The putative RNA recognition code by PPR proteins derived from bioinformatic and biochemical analyses awaits structural corroboration[2,6–8].

To elucidate the mechanism of specific RNA recognition by PPR proteins, we sought to determine the crystal structure of well-characterized PPR proteins in complex with their target RNAs. The recombinant protein of maize chloroplast PPR10, which belongs to the P-class, specifically binds to the 17-nucleotide (nt) (*ATPH*) and 18-nt (*PSAJ*) RNA oligonucleotides (Extended Data Fig. 1a)[10]. We launched a systematic effort to determine the structures of PPR10 in both RNA-free and RNA-bound states.

The crystal structure of the RNA-free PPR10 fragment (residues 61–786) containing quadruple Cys mutations (C256S/C279S/C430S/C449S) was determined at 2.85 Å resolution. PPR10 forms a right-handed two-turn superhelical assembly, with 19 PPR motifs (residues 107–771) capped by three short α-helices at the amino-terminal domain (NTD) and a single α-helix at the C terminus (Fig. 1a). Capping motifs are known to contribute to ligand specificity for repeat proteins such as TPR (tetratricopeptide repeat)[25] and TALE (transcription activator-like effector)[26,27]. The function of the extra motifs in PPR10 remains to be determined.

The 35 amino acids in each PPR motif form a hairpin of α-helices, each containing four helical turns, followed by a five-residue loop (Fig. 1b). The two helices, designated helix a and helix b, are connected by a short turn of two amino acids. Helices a and b of each repeat constitute the inner and outer layers of the superhelical assembly, respectively (Fig. 1a). In the crystals, there is one molecule of PPR10 in each asymmetric unit, yet two symmetry-related molecules are intertwined in an antiparallel fashion. The N terminus of one molecule is in close contact with the C terminus of the other, yielding an overall appearance of an ellipsoid with a polar axis of approximately 140 Å and an equatorial diameter of 70 Å (Fig. 1c).

On the basis of the PPR10 structure, we defined the starting amino acid of helix a as the first residue in a PPR motif (Fig. 1b and Extended Data Fig. 1b). This definition results in a one-residue shift either forwards[6,12] or backwards[7,28] within each repeat compared to the previously described boundary of a PPR motif (Extended Data Fig. 1c). With the new boundary assignment of a PPR motif, the residues that were predicted to determine RNA binding specificity are all included in one structurally intact motif. We hope that this structure-based demarcation of the PPR motif will simplify future descriptions of PPR proteins.

After numerous unsuccessful crystallization trials for PPR10–*ATPH* complexes, we finally determined the structure of PPR10 (residues 69–786, C256S/C279S/C430S/C449S) in the presence of 18-nt *PSAJ* RNA (5′-GUAUUCUUUAAUUAUUUC-3′) at 2.45 Å resolution (Extended Data Table1). In the crystals, there is one antiparallel PPR10 dimer in each asymmetric unit. Analysis by sedimentation equilibrium analytical ultracentrifugation (SE-AUC) of *PSAJ*-bound PPR10 (residues 37–786, C256S/C279S/C430S/C449S) supports its dimeric existence at micromolar concentration in solution (Extended Data Fig. 2). The two PPR10 protomers can be superimposed with a root-mean-squared deviation of 1.3 Å over 629 Cα atoms (Extended Data Fig. 3). The overall appearance of the dimer has changed to a hollow cylindrical tube (Fig. 2a), and the N- and C-terminal portions of the PPR10 protomer are compressed towards the centre, resulting in a reduction of 20 Å in axial length (Fig. 2b).

[1]State Key Laboratory of Bio-membrane and Membrane Biotechnology, Tsinghua University, Beijing 100084, China. [2]Center for Structural Biology, School of Life Sciences and School of Medicine, Tsinghua-Peking Center for Life Sciences, Tsinghua University, Beijing 100084, China. [3]Ministry of Education Key Laboratory of Protein Science, Tsinghua University, Beijing 100084, China. [4]Shanghai Institute of Applied Physics, Chinese Academy of Sciences, 239 Zhangheng Road, Shanghai 201204, China. [5]State Key Laboratory of Medicinal Chemical Biology and College of Life Sciences, Nankai University, 94 Weijin Road, Tianjin 300071, China. [6]Shanghai Center for Plant Stress Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China. [7]Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette, Indiana 47907, USA.
*These authors contributed equally to this work.

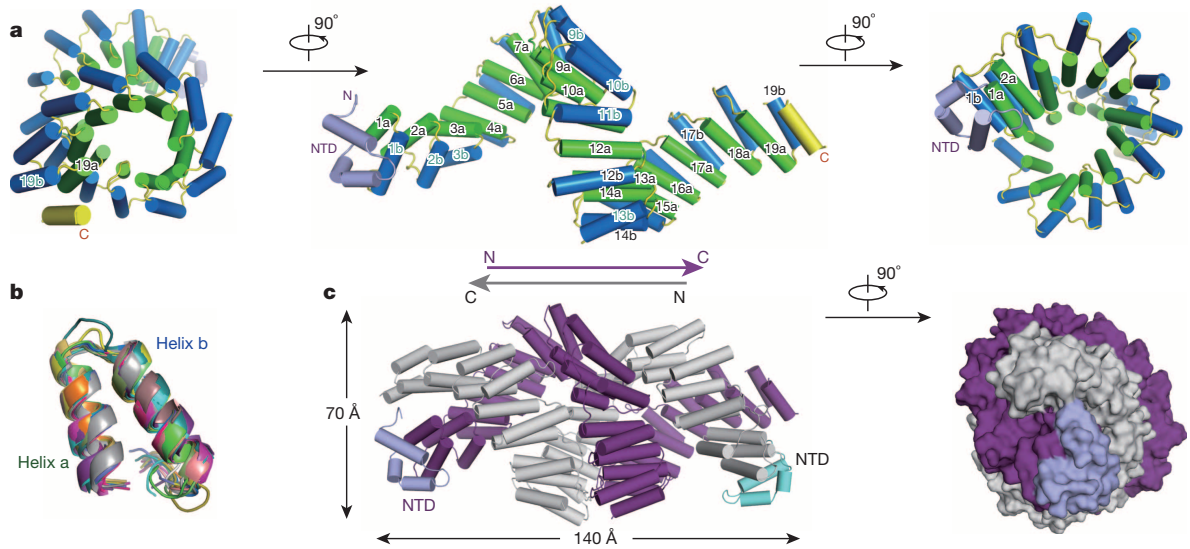


Figure 1 | Crystal structure of RNA-free PPR10. **a**, Overall structure of RNA-free PPR10. The fragment (residues 61–786, C256S/C279S/C430S/C449S) comprises 19 repeats capped by a small NTD (light purple) and a C-terminal helix (yellow). The two helices within each repeats, designated helix a and helix b, are coloured green and blue, respectively. **b**, Structural superimposition of the 19 repeats of PPR10. **c**, Overall structure of the PPR10 dimer. Two molecules from adjacent asymmetric units form an intertwined antiparallel dimer. All structure figures were prepared with PyMol[30].

Following assignment of most amino acids of PPR10 into the electron density map, strong electron densities indicative of RNA bases became clearly visible in the cavities on both ends of the cylindrical tube (Fig. 2c). Assignment of 18 and 14 nucleotides of the two bound RNA elements was validated by the anomalous signals of bromine (Br), which were collected for crystals of PPR10 bound to Br-labelled RNA oligonucleotides (Extended Data Fig. 4 and Extended Data Table 2). The 5′ and 3′ portions of the ssRNA are specifically recognized by the N-terminal repeats of one protomer and C-terminal repeats of the other. By contrast, the middle portion of the ssRNA, comprising nucleotides U5 to A10, remains largely uncoordinated by PPR10 (Fig. 2d and Extended Data Fig. 5a, b).

PPR10 has 19 repeats and the bound *PSAJ* RNA contains 18 nucleotides. Consistent with a bioinformatic prediction[6], specific recognition of the *PSAJ* RNA begins with repeat 3 (Fig. 3a). Each of the first four nucleotides on the 5′ end, 5′-GUAU-3′, is recognized by one PPR10 repeat. Such recognition exhibits a modular pattern involving residues that were predicted through biochemical and bioinformatic analyses[2,6,7]. Each RNA base is surrounded by four residues, the 2nd residues from two adjacent repeats, and the 5th and 35th residues from a corresponding repeat. In addition to base recognition, the backbone phosphate or ribose groups of the bound *PSAJ* RNA are also coordinated by charged or polar amino acids from PPR10 (Extended Data Fig. 5c).

A polar amino acid located at the 5th position in each repeat appears to be the most important determinant for RNA base specificity. Thr 178, Asn 213, Ser 249 and Asn 284 in repeats 3–6 recognize the bases G1, U2, A3 and U4, respectively, through direct hydrogen bonds (Fig. 3b). The importance of the 5th residue in RNA recognition is supported by mutational analysis. Mutating any of the 5th residues in repeats 4 (N213A), 5 (S249L) or 6 (N284A) resulted in complete abolishment of RNA binding. By contrast, substitution of the 5th residues of repeats 7, 8, 10, 11 or 13, which are not involved in RNA binding in the structure, showed little or no effect on *PSAJ* binding (Extended Data Fig. 6).

Buttressing the hydrogen bonds, five residues at the 2nd position of PPR repeats 3–7 sandwich the four bases mainly through van der Waals interactions (Fig. 3a). For example, G1 is surrounded by Arg 175/Val 210 of repeats 3 and 4. Similarly, U2, A3 and U4 are sandwiched by Val 210/Phe 246, Phe 246/Val 281 and Val 281/Val 316, respectively (Fig. 3). The 35th residue is located in the vicinity of the base. It is possible that water molecules, although invisible in the structure, may mediate hydrogen bonds between the polar residues and the bases. Importantly, Asp 244

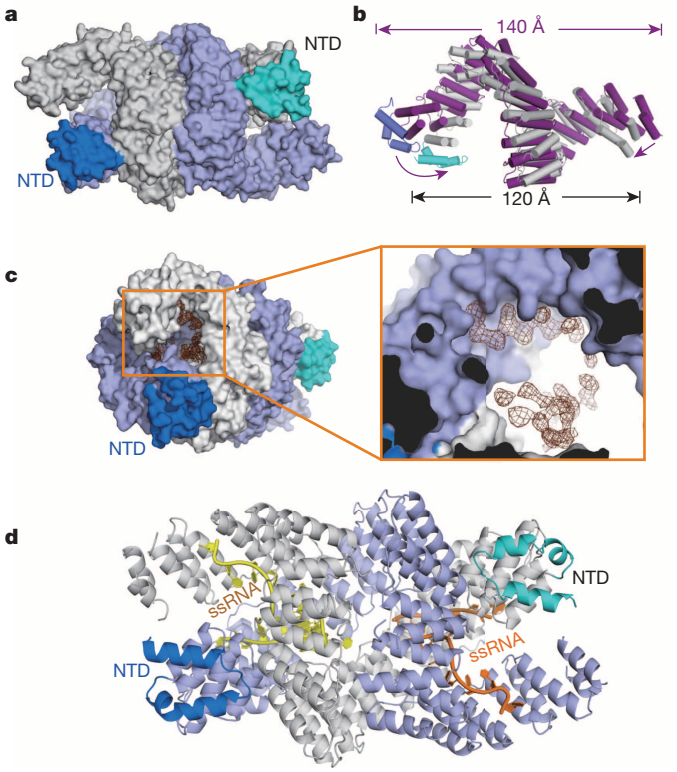


Figure 2 | Structure of PPR10 bound to an 18-nt *PSAJ* RNA element. **a**, The PPR10 dimer (residues 69–786, C256S/C279S/C430S/C449S) forms a cylindrical tube in the presence of *PSAJ*. The two protomers are coloured light purple and grey with their NTDs coloured blue and cyan. **b**, The PPR10 protomer undergoes pronounced conformational changes upon binding to *PSAJ*. The structure of RNA-free PPR10 is coloured magenta with the NTD coloured lilac. **c**, Electron densities found in the cavities on both ends of the PPR10 dimer. The 'omit' electron density, with a close-up view in the inset, is contoured at 3σ. **d**, Overall structure of the PPR10–*PSAJ* complex. The two ssRNA molecules are coloured yellow and orange.

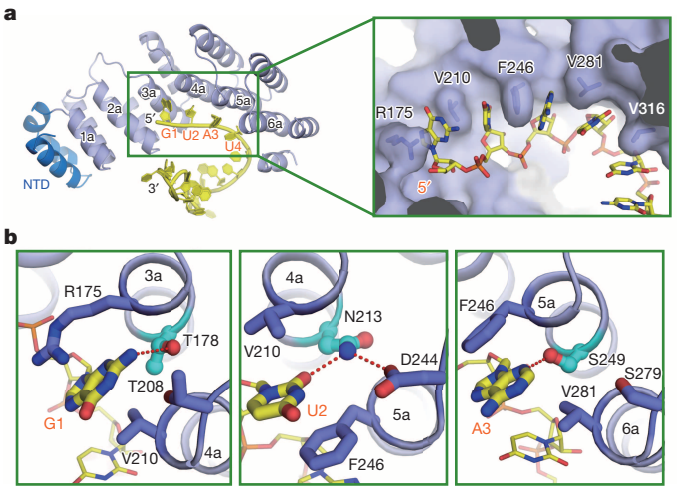


Figure 3 | Base-specific recognition of ssRNA by PPR10 repeats. **a**, The four nucleotides at the 5′ end of the *PSAJ* RNA segment are specifically recognized in a modular fashion. Inset: each of the four RNA bases at the 5′ end is sandwiched by two residues at the 2nd positions of adjacent repeats. **b**, Specific recognition of the bases G, U and A by PPR10 repeats. The side chain of the 5th residue in each repeat, which makes a direct hydrogen bond to the base, is highlighted in cyan. The hydrogen bonds are represented by red dotted lines.

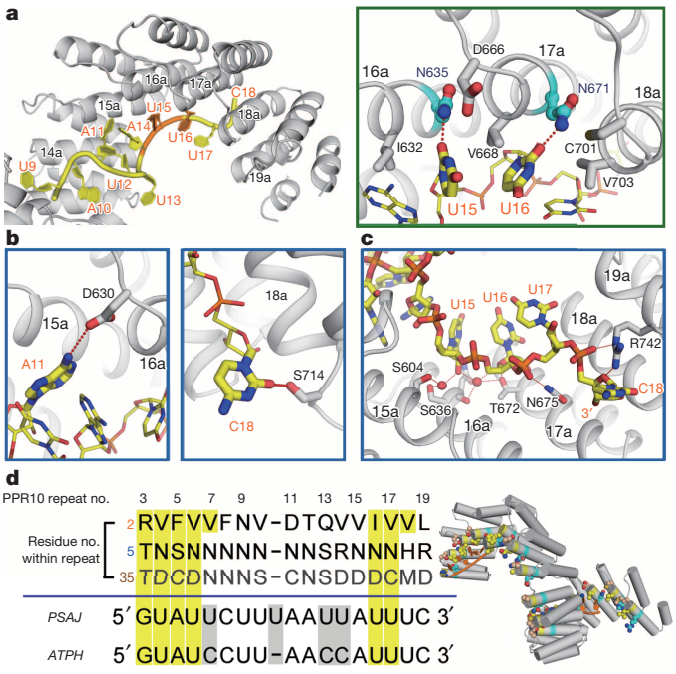


Figure 4 | Coordination of the 3′-end segment of the *PSAJ* RNA by PPR10. **a**, Recognition of the bases U15 and U16 by PPR10 follows the code discussed in Fig. 3. **b**, The bases A11 and C18 are hydrogen bonded to polar residues on PPR10 in a non-modular fashion. **c**, The backbone of the 3′-end segment of *PSAJ* is coordinated through direct or water-mediated hydrogen bonds. **d**, Summary of specific recognition of *PSAJ* RNA by PPR10. Left, the residues at the 2nd, 5th and 35th positions of PPR10 motifs and the corresponding sequences of the target RNA elements. The structurally corroborated recognition codes are shaded yellow. The difference in RNA sequences of *PSAJ* and *ATPH* is shaded grey. Right, the structure of repeats 3–19 of PPR10 with their 2nd (yellow), 5th (cyan) and 35th (wheat) residues shown in spheres, and the six recognized RNA bases shown in orange.

and Asp 314, the 35th residues in repeats 4 and 6, are respectively hydrogen bonded to Asn 213 and Asn 284, the 5th residues in the corresponding repeats, and may help to stabilize their conformation for base recognition (Fig. 3b and Extended Data Fig. 6d).

Recognition of the 3′ end of the *PSAJ* RNA by the C-terminal repeats in the other PPR10 protomer appears to be less modular except for U15 and U16, which are coordinated by repeats 16 and 17 following the described recognition pattern for U2 and U4 (Fig. 4a). The bases A11 and C18 are coordinated in a non-modular fashion. The adenine base of A11 donates a hydrogen bond to Asp 630, the 35th residue of repeat 15, whereas the cytosine base of C18 makes a hydrogen bond to Ser 714 on the last helical turn of helix 18a (Fig. 4b). A number of direct and water-mediated hydrogen bonds are found between the backbone phosphate/ribose groups of U15-U16-U17-C18 and the polar residues on repeats 15–19 of PPR10 (Fig. 4c).

In the structure of *PSAJ*-bound PPR10, only six out of 18 nucleotides in the *PSAJ* RNA element strictly follow the modular pattern (Fig. 4d). Two bases, A11 and C18, are bound by PPR10 in a non-modular fashion, and the other 10 bases are literally uncoordinated. Notably, only 17 repeats in each PPR10 protomer are available for the binding of 18 nucleotides. It remains to be seen whether the 17-nt *ATPH* binds to repeats 3–19 of PPR10 in a completely modular fashion. Binding of *ATPH* leads to dissociation of the PPR10 dimer (Extended Data Fig. 2)[6,10]. Interestingly, analysis by SE-AUC suggests that, although RNA-free PPR10 forms a stable dimer, *PSAJ* binding weakens the PPR10 dimer formation (Extended Data Fig. 2). It remains to be investigated whether *PSAJ*-bound PPR10 is a dimer under physiological conditions, in which the protein concentration can be very low. Nevertheless, the crystal structure of PPR10 bound to the 18-nt *PSAJ* RNA element reveals the molecular basis for specific recognition between a PPR protein and its target RNA sequence.

Recognition of the six RNA nucleotides 5′-G1-U2-A3-U4-3′ and 5′-U15-U16-3′ by repeats 3–6 and repeats 16 and 17, respectively, largely supports the predicted code for base discrimination in ssRNA, where the bases G, U and A are specifically recognized by the 5th residues of a PPR motif: Thr, Asn and Ser, respectively[6,7]. The 2nd and 35th residues, sitting in the vicinity of the bases, contribute to RNA binding (Fig. 4d and Extended Data Fig. 6c, d). The prediction that base C is recognized by an Asn at the 5th position can also be conveniently rationalized based on our crystal structure (Extended Data Fig. 7). One unexpected feature is that the 2nd residues from two consecutive repeats sandwich one base; therefore the identity of the 2nd residue on the next repeat must be considered for RNA binding. Base sandwiching by hydrophobic residues or Arg is also observed in the recognition of ssRNA by PUF proteins[24,29], although PPR and PUF proteins exhibit distinct RNA binding modes (Extended Data Fig. 8).

Further biochemical, computational, structural and *in vivo* characterizations are required to completely rationalize the codes for specific RNA recognition by PPRs and to engineer PPR proteins for targeted RNA manipulations. The structures reported here provide unprecedented insights into the recognition mechanism of RNA elements by PPR proteins and serve as an important foundation for understanding the function and mechanism of numerous PPR proteins in RNA metabolism, and for the potentially customized design of specific-RNA-binding PPR proteins.

METHODS SUMMARY

The codon-optimized complementary DNA of full-length *PPR10* (Gene ID: 100302579) from *Zea mays* was subcloned into pET15b vector (Novagen). Overexpression of PPR10 protein was induced in *Escherichia coli* BL21(DE3). To crystallize PPR10, we mounted a systematic protein engineering effort including a series of protein truncations and mutations of Cys residues. There are 18 Cys residues within the repeat region. We generated 18 mutants, each consisting of a single Cys to Ser mutation and tested their binding with the 17-nt *ATPH* element. For those that completely retained binding affinity, we further grouped them to double, triple and quadruple mutations. Finally, the PPR10 mutant containing C256S/C279S/C430S/C449S showed the same binding affinity as wild type and exhibited excellent protein behaviour. All the PPR10 proteins used in the manuscript contain the quadruple Cys mutations. The RNA-free PPR10 fragment (residues 61–786, C256S/C279S/C430S/C449S) was eventually crystallized in the space group $P2_12_12$. The structure

was determined by selenium-based single-wavelength anomalous diffraction and refined to 2.85 Å resolution (Extended Data Table 1). In the effort to crystallize PPR10 in complex with its target RNA, despite numerous trials, most PPR10–*ATPH* complexes defied crystallization; for those that crystallized, X-ray diffraction was consistently poor. We applied the same strategy to complexes between PPR10 and *PSAJ*. After screening more than 100,000 conditions, we were able to crystallize the complex between PPR10 (residues 69–786, C256S/C279S/C430S/C449S) and the 18-nt *PSAJ* RNA (5′-GUAUUCUUUAAUUAUUUC-3′) in the space group $P4_3$. These crystals diffract X-rays beyond 2.5 Å. The structure was determined by molecular replacement using successive segments of the RNA-free PPR10 structure, but not the entire molecule. We were able to assign all 18 nucleotides of one bound *PSAJ* RNA element, but only 14 of the other. For details of electrophoretic mobility shift assay and SE-AUC experiments, please refer to Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 May; accepted 12 September 2013.
Published online 27 October 2013.

1. Small, I. D. & Peeters, N. The PPR motif — a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem. Sci.* **25,** 45–47 (2000).
2. Nakamura, T., Yagi, Y. & Kobayashi, K. Mechanistic insight into pentatricopeptide repeat proteins as sequence-specific RNA-binding proteins for organellar RNAs in plants. *Plant Cell Physiol.* **53,** 1171–1179 (2012).
3. Schmitz-Linneweber, C. & Small, I. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends Plant Sci.* **13,** 663–670 (2008).
4. Fujii, S. & Small, I. The evolution of RNA editing and pentatricopeptide repeat genes. *New Phytol.* **191,** 37–47 (2011).
5. Kotera, E., Tasaka, M. & Shikanai, T. A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature* **433,** 326–330 (2005).
6. Barkan, A. *et al.* A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet.* **8,** e1002910 (2012).
7. Yagi, Y., Hayashi, S., Kobayashi, K., Hirayama, T. & Nakamura, T. Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS ONE* **8,** e57286 (2013).
8. Yagi, Y. *et al.* Pentatricopeptide repeat proteins involved in plant organellar RNA editing. *RNA Biol.* **10,** 1236–1242 (2013).
9. Pfalz, J., Bayraktar, O. A., Prikryl, J. & Barkan, A. Site-specific binding of a PPR protein defines and stabilizes 5′ and 3′ mRNA termini in chloroplasts. *EMBO J.* **28,** 2042–2052 (2009).
10. Prikryl, J., Rojas, M., Schuster, G. & Barkan, A. Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. *Proc. Natl Acad. Sci. USA* **108,** 415–420 (2011).
11. Zhelyazkova, P. *et al.* Protein-mediated protection as the predominant mechanism for defining processed mRNA termini in land plant chloroplasts. *Nucleic Acids Res.* **40,** 3092–3105 (2012).
12. Lurin, C. *et al.* Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* **16,** 2089–2103 (2004).
13. Cushing, D. A., Forsthoefel, N. R., Gestaut, D. R. & Vernon, D. M. *Arabidopsis emb175* and other *ppr* knockout mutants reveal essential roles for pentatricopeptide repeat (PPR) proteins in plant embryogenesis. *Planta* **221,** 424–436 (2005).
14. Khrouchtchova, A., Monde, R. A. & Barkan, A. A short PPR protein required for the splicing of specific group II introns in angiosperm chloroplasts. *RNA* **18,** 1197–1209 (2012).
15. Bentolila, S., Alfonso, A. A. & Hanson, M. R. A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. *Proc. Natl Acad. Sci. USA* **99,** 10887–10892 (2002).
16. Desloire, S. *et al.* Identification of the fertility restoration locus, *Rfo*, in radish, as a member of the pentatricopeptide-repeat protein family. *EMBO Rep.* **4,** 588–594 (2003).
17. Wang, Z. *et al.* Cytoplasmic male sterility of rice with boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. *Plant Cell* **18,** 676–687 (2006).
18. Chase, C. D. Cytoplasmic male sterility: a window to the world of plant mitochondrial–nuclear interactions. *Trends Genet.* **23,** 81–90 (2007).
19. Hu, J. *et al.* The rice pentatricopeptide repeat protein RF5 restores fertility in Hong-Lian cytoplasmic male-sterile lines via a complex with the glycine-rich protein GRP162. *Plant Cell* **24,** 109–122 (2012).
20. Mootha, V. K. *et al.* Identification of a gene causing human cytochrome c oxidase deficiency by integrative genomics. *Proc. Natl Acad. Sci. USA* **100,** 605–610 (2003).
21. Ruzzenente, B. *et al.* LRPPRC is necessary for polyadenylation and coordination of translation of mitochondrial mRNAs. *EMBO J.* **31,** 443–456 (2012).
22. Howard, M. J., Lim, W. H., Fierke, C. A. & Koutmos, M. Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5′ processing. *Proc. Natl Acad. Sci. USA* **109,** 16149–16154 (2012).
23. Ringel, R. *et al.* Structure of human mitochondrial RNA polymerase. *Nature* **478,** 269–273 (2011).
24. Wang, X., McLachlan, J., Zamore, P. D. & Hall, T. M. Modular recognition of RNA by a human pumilio-homology domain. *Cell* **110,** 501–512 (2002).
25. Grove, T. Z., Cortajarena, A. L. & Regan, L. Ligand binding by repeat proteins: natural and designed. *Curr. Opin. Struct. Biol.* **18,** 507–515 (2008).
26. Deng, D. *et al.* Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* **335,** 720–723 (2012).
27. Gao, H., Wu, X., Chai, J. & Han, Z. Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region. *Cell Res.* **22,** 1716–1720 (2012).
28. Kobayashi, K. *et al.* Identification and characterization of the RNA binding surface of the pentatricopeptide repeat protein. *Nucleic Acids Res.* **40,** 2712–2723 (2012).
29. Filipovska, A. & Rackham, O. Modular recognition of nucleic acids by PUF, TALE and PPR proteins. *Mol. Biosyst.* **8,** 699–708 (2012).
30. DeLano, W. L. The PyMOL Molecular Graphics System. http://www.pymol.org (2002).

Acknowledgements We thank X. Yu and Y. Chen at the Institute of Biophysics, Chinese Academy of Sciences, for technical support. We thank K. Hasegawa and T. Kumasaka at the SPring-8 beamline BL41XU for on-site assistance. This work was supported by funds from the Ministry of Science and Technology (grant number 2011CB910501 for N.Y.), and Projects 91017011 (N.Y.), 31070644 (N.Y.), 31021002 (Y.S., N.Y., J.W.) and 31200567 (P.Y.) of the National Natural Science Foundation of China. The research of N.Y. was supported in part by an International Early Career Scientist grant from the Howard Hughes Medical Institute.

Author Contributions P.Y., Q.L., J.-K.Z., Y.S. and N.Y. designed all experiments. P.Y., Q.L., C.Y., Y.L., J.L., F.Y., Z.W., J.L., J.H., H.-W.W., J.W. and N.Y. performed the experiments. All authors analysed the data and contributed to manuscript preparation. N.Y. wrote the manuscript.

Author Information The atomic coordinates and structure factors of RNA-free and RNA-bound PPR10 have been deposited in the Protein Data Bank (PDB) with the accession codes 4M57 and 4M59, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.Y. (nyan@tsinghua.edu.cn).

METHODS

Protein preparation. The codon-optimized complementary DNA of full-length *PPR10* (Gene ID: 100302579) from *Zea mays* was subcloned into pET15b vector (Novagen). Overexpression of PPR10 protein was induced in *E. coli* BL21(DE3) with 0.2 mM isopropyl-β-D-thiogalactoside at an $OD_{600\,nm}$ of 1.2. After growing for 16 h at 16 °C, the cells were collected, homogenized in a buffer containing 25 mM Tris-HCl, pH 8.0, and 150 mM NaCl. After sonication and centrifugation, the supernatant was applied to Ni^{2+} affinity resin (Ni-NTA, Qiagen) and further fractionated by ion-exchange chromatography (Source 15Q, GE Healthcare). The PPR10 mutants were generated using two-step PCR and subcloned, overexpressed and purified in the same way as the wild-type protein.

A systematic protein engineering effort was mounted for crystallization of RNA-free and -bound PPR10. A series of protein truncations were tested without giving rise to crystals. There are 18 Cys residues within the repeat region. It is well known that the presence of surface Cys residues, which are subject to oxidation, may lead to protein heterogeneity and impede crystallization. We therefore generated 18 mutants, each consisting of a single Cys to Ser mutation and tested their binding with the 17-nt *ATPH* element. For those that completely retained binding affinity, we further grouped them to double, triple and quadruple mutations. Finally, the PPR10 mutant containing C256S/C279S/C430S/C449S showed the same binding affinity as wild type and exhibited excellent protein behaviour. For consistency, all the PPR10 proteins used in the manuscript contain the quadruple Cys mutations.

For the crystallization trials of RNA-free PPR10 (residues 61–786, C256S/C279S/C430S/C449S), the protein was concentrated and applied to gel filtration chromatography (Superdex-200 10/30, GE Healthcare) in the buffer containing 25 mM Tris-HCl, pH 8.0, 150 mM NaCl and 10 mM dithiothreitol (DTT). Selenomethionine (Se-Met)-derived protein was purified similarly.

To obtain the crystals of protein–RNA complex, PPR10 (residues 69–786, C256S/C279S/C430S/C449S) was purified through Ni^{2+} affinity resin (Ni-NTA, Qiagen), followed by heparin affinity column (HiPrep Heparin FF 16/10, GE Healthcare). The protein was then applied to gel filtration chromatography (Superdex-200 10/30, GE Healthcare). The buffer for gel filtration contained 25 mM Tris-HCl, pH 8.0, 50 mM NaCl, 5 mM $MgCl_2$ and 10 mM DTT. The peak fractions were incubated with target RNA oligonucleotides with a molar ratio of approximately 1:1.5 at 4 °C for about 40 min before crystallization trials.

Crystallization. Both RNA-free and RNA-bound PPR10 proteins were crystallized by hanging-drop vapour-diffusion method at 18 °C. PPR10 (residues 61–786, C256S/C279S/C430S/C449S), at a concentration of approximately 6.0 mg ml^{-1}, was mixed with an equal volume of reservoir solution containing 1.8–2.1 M sodium formate, and 0.1 M Bis-Tris propane, pH 6.5. Plate-shaped crystals appeared overnight and grew to full size within 1–2 weeks. Se-Met-labelled protein was crystallized similarly.

To obtain crystals of protein–RNA complex, various combinations of protein boundaries and RNA oligonucleotides (Takara) were examined. Because the first visible residue in the structure of RNA-free PPR10 starts at position 69, we invested more effort into this construct. Finally, the protein (residues 69–786, C256S/C279S/C430S/C449S) and 18-nt RNA from the *PSAJ–RPL33* intergenic region with the sequence 5′-GUAUUCUUUAAUUAUUUC-3′ (designated *PSAJ* RNA) gave rise to crystals in the reservoir solution containing 8–10% (w/v) polyethylene glycol 3350, 8% Tacsimate, pH 6.0 (Hampton Research), and 0.1 M MES, pH 5.5.

Data collection and structural determination. All data sets were collected at SSRF beamline BL17U or SPring-8 beamline BL41XU and processed with the HKL2000 packages[31]. Further processing was carried out with programs from the CCP4 suite[32]. Data collection and structure refinement statistics are summarized in Extended Data Tables 1 and 2.

The RNA-free PPR10 structure was solved by single anomalous diffraction (SAD) of Se-Met using the program ShelxC/D/E[33]. Then a crude helical model was manually built in the program Coot[34]. Using this partial model as input, the identified Se atom positions were refined and phases were recalculated using the SAD experimental phasing module of the program Phaser[35]. With the improved map, the molecular boundary was unambiguously defined and one molecule was found in an asymmetry unit. The crude model was further rebuilt with Coot and refined with Phenix[36]. The sequence docking was aided by anomalous map of selenium.

Data sets collected from five crystals of the PPR10–RNA complex were merged for complete and better data. The structure of the PPR10–RNA complex was solved by molecular replacement with the newly solved RNA-free structure as the search model using the program Phaser[35]. To find the right solution, the structure of the RNA-free PPR10 protomer was divided into three consecutive segments. The assignment of RNA sequence was aided by the anomalous signal of bromine obtained for crystals of PPR10 in complex with Br-labelled RNA oligonucleotides, where U4/U7/U15, U5/U7/U15 or U12 were substituted by 5-bromouracil (Extended Data Table 2). The structure was manually refined with Coot and Phenix iteratively (Extended Data Table 1).

Electrophoretic mobility shift assay (EMSA). The ssRNA oligonucleotides were radiolabelled at the 5′ end with [γ-^{32}P] ATP (PerkinElmer) catalysed by T4 polynucleotide kinase (Takara). The sequences of ssRNA oligonucleotides used in EMSA are: *PSAJ*, 5′-GUAUUCUUUAAUUAUUUC-3′; and *ATPH*, 5′-GUAUCCUUA ACCAUUUC-3′.

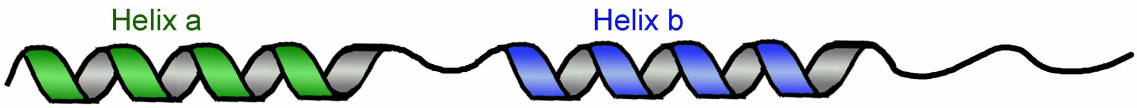
For EMSA, PPR10 (residues 37–786, C256S/C279S/C430S/C449S) and the other variants consisting of the indicated point mutations were incubated with approximately 40 pM ^{32}P-labelled probe in the final binding reactions containing 40 mM Tris-HCl, pH 7.5, 100 mM NaCl, 4 mM DTT, 0.1 mg ml^{-1} BSA, 5 μg ml^{-1} heparin and 10% glycerol at room temperature (22 °C) for 20 min. Reactions were then resolved on 6% native acrylamide gels (37.5:1 for acrylamide:bisacrylamide) in 0.5× Tris-glycine buffer under an electric field of 15 V cm^{-1} for 40 min. Vacuum-dried gels were visualized on a phosphor screen (Amersham Biosciences) with a Typhoon Trio Imager (Amersham Biosciences).

SE-AUC. The oligomeric states of PPR10 (residues 37–786, C256S/C279S/C430S/C449S) with or without target RNA oligonucleotides in solution were investigated by AUC experiments. SE-AUC experiments were performed in a Beckman Coulter XL-I analytical ultracentrifuge using six-channel centrepieces. RNA-free PPR10, *PSAJ*-bound PPR10 and *ATPH*-bound PPR10 were in solutions containing 25 mM Tris-HCl, pH 8.0, 150 mM NaCl and 2 mM DTT. The sequences of RNA oligonucleotides were identical to those used in EMSA. Data were collected by interference detection at 4 °C for all three protein concentrations (4 μM, 6 μM and 8 μM) at different rotor speeds (6,000, 8,500 and 12,000 r.p.m.). The buffer composition (density and viscosity) and protein partial specific volume (V-bar) were obtained using the SEDNTERP program (available through the Boston Biomedical Research Institute). The SE-AUC data were globally analysed using the Sedfit and Sedphat programs[37] and were fitted to a monomer–dimer equilibrium model to determine the dissociation constants (K_d) for the homodimers.

31. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
32. Collaborative Computational Project, Number 4. The *CCP*4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
33. Schneider, T. R. & Sheldrick, G. M. Substructure solution with *SHELXD*. *Acta Crystallogr. D* **58,** 1772–1779 (2002).
34. Emsley, P. & Cowtan, K. *Coot*: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
35. McCoy, A. J. *et al. Phaser* crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
36. Adams, P. D. *et al. PHENIX*: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58,** 1948–1954 (2002).
37. Schuck, P. On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal. Biochem.* **320,** 104–124 (2003).

a

b



Residue #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	PSAJ ATPH
Repeat 01			$_{107}$I	T	S	L	L	K	A	L	E	L	S	G	H	W	E	W	A	L	A	L	L	R	W	A	G	K	E	G		A	A	D_{137}	
Repeat 02	$_{138}$A	S	A	L	E	M	V	V	R	A	L	G	R	E	G	Q	H	D	A	V	C	A	L	L	D	E	T	P	L	P	P	G	S	R	L	D_{173}	
Repeat 03	$_{174}$V	R	A	Y	T	T	V	L	H	A	L	S	R	A	G	R	Y	E	R	A	L	E	L	F	A	E	L	R	R	Q	G	V	A	P	T_{208}	G G
Repeat 04	$_{209}$L	V	T	Y	N	V	V	L	D	V	Y	G	R	M	G	R	S	W	P	R	I	V	A	L	L	D	E	M	R	A	A	G	V	E	P	D_{244}	U U
Repeat 05	$_{245}$G	F	T	A	S	T	V	I	A	A	C	C	R	D	G	L	V	D	E	A	V	A	F	F	E	D	L	K	A	R	G	H	A	P	C_{279}	A A
Repeat 06	$_{280}$V	V	T	Y	N	A	L	L	Q	V	F	G	K	A	G	N	Y	T	E	A	L	R	V	L	G	E	M	E	Q	N	G	C	Q	P	D_{314}	U U
Repeat 07	$_{315}$A	V	T	Y	N	E	L	A	G	T	Y	A	R	A	G	F	F	E	E	A	A	R	C	L	D	T	M	A	S	K	G	L	L	P	N_{349}	U C
Repeat 08	$_{350}$A	F	T	Y	N	T	V	M	T	A	Y	G	N	V	G	K	V	D	E	A	L	A	L	F	D	Q	M	K	K	T	G	F	V	P	N_{384}	C C
Repeat 09	$_{385}$V	N	T	Y	N	L	V	L	G	M	L	G	K	K	S	R	F	T	V	M	L	E	M	L	G	E	M	S	R	S	G	C	T	P	N_{419}	U U
Repeat 10	$_{420}$R	V	T	W	N	T	M	L	A	V	C	G	K	R	G	M	E	D	Y	V	T	R	V	L	E	G	M	R	S	C	G	V	E	L	S_{454}	U U
Repeat 11	$_{455}$R	D	T	Y	N	T	L	I	A	A	Y	G	R	C	G	S	R	T	N	A	F	K	M	Y	N	E	M	T	S	A	G	F	T	P	C_{489}	A A
Repeat 12	$_{490}$I	T	T	Y	N	A	L	L	N	V	L	S	R	Q	G	D	W	S	T	A	Q	S	I	V	S	K	M	R	T	K	G	F	K	P	N_{524}	A A
Repeat 13	$_{525}$E	Q	S	Y	S	L	L	L	Q	C	Y	A	K	G	G	N	V	A	G	I	A	A	I	E	N	E	V	Y	G	S	G	A	V	F	P	S_{560}	U C
Repeat 14	$_{561}$W	V	I	L	R	T	L	V	I	A	N	F	K	C	R	R	L	D	G	M	E	T	A	F	Q	E	V	K	A	R	G	Y	N	P	D_{595}	U C
Repeat 15	$_{596}$L	V	I	F	N	S	M	L	S	I	Y	A	K	N	G	M	Y	S	K	A	T	E	V	F	D	S	I	K	R	S	G	L	S	P	D_{630}	A A
Repeat 16	$_{631}$L	I	T	Y	N	S	L	M	D	M	Y	A	K	C	S	E	S	W	E	A	E	K	I	L	N	Q	L	K	C	S	Q	T	M	K	P	D_{666}	U U
Repeat 17	$_{667}$V	V	S	Y	N	T	V	I	N	G	F	C	K	Q	G	L	V	K	E	A	Q	R	V	L	S	E	M	V	A	D	G	M	A	P	C_{701}	U U
Repeat 18	$_{702}$A	V	T	Y	H	T	L	V	G	G	Y	S	S	L	E	M	F	S	E	A	R	E	V	I	G	Y	M	V	Q	H	G	L	K	P	M_{736}	U U
Repeat 19	$_{737}$E	L	T	Y	R	R	V	V	E	S	Y	C	R	A	K	R	F	E	E	A	R	G	F	L	S	E	V	S	E	T	D	L	D	F	D_{771}	C C

c

1)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	1′
Repeat 03	$_{174}$V	R	A	Y	T	T	V	L	H	A	L	S	R	A	G	R	Y	E	R	A	L	E	L	F	A	E	L	R	R	Q	G	V	A	P	T_{208}	

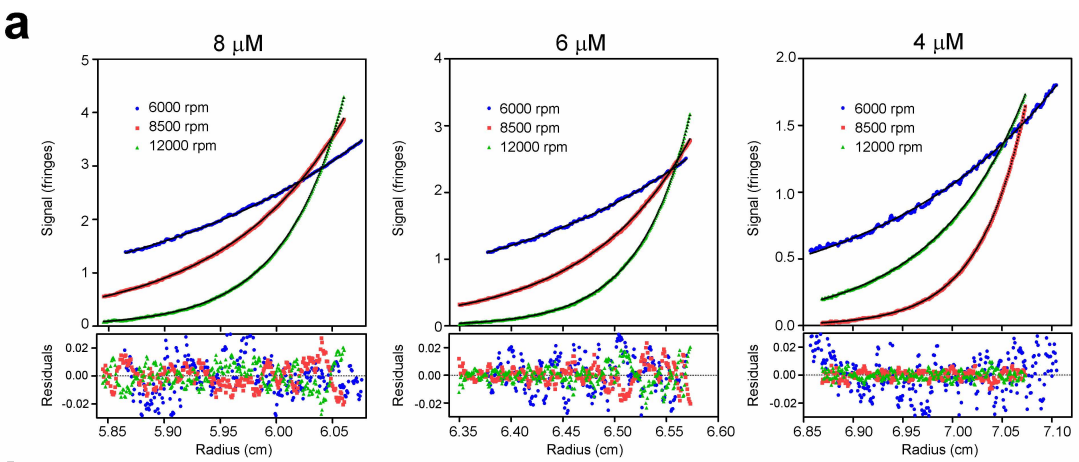
2)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	ii	35
Repeat 03	$_{174}$V	R	A	Y	T	T	V	L	H	A	L	S	R	A	G	R	Y	E	R	A	L	E	L	F	A	E	L	R	R	Q	G	V	A	P	T_{208}

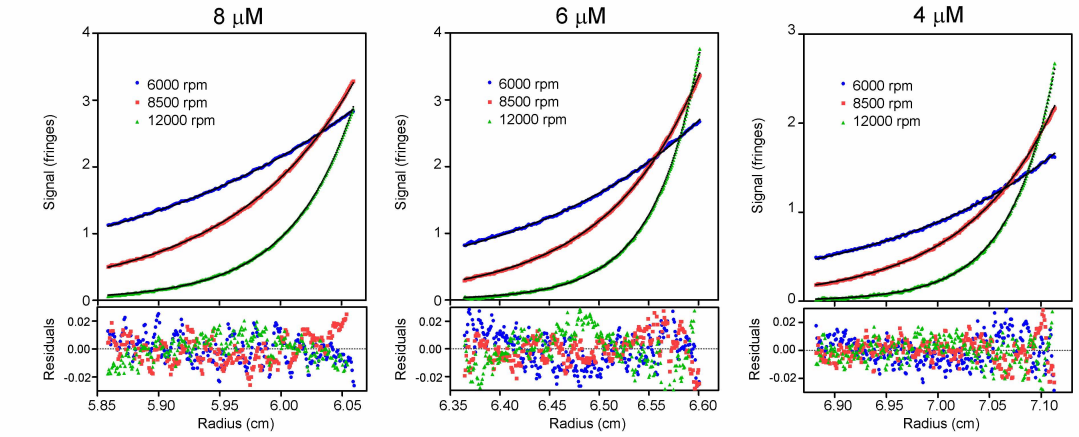
3)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
Repeat 03	$_{174}$V	R	A	Y	T	T	V	L	H	A	L	S	R	A	G	R	Y	E	R	A	L	E	L	F	A	E	L	R	R	Q	G	V	A	P	T_{208}

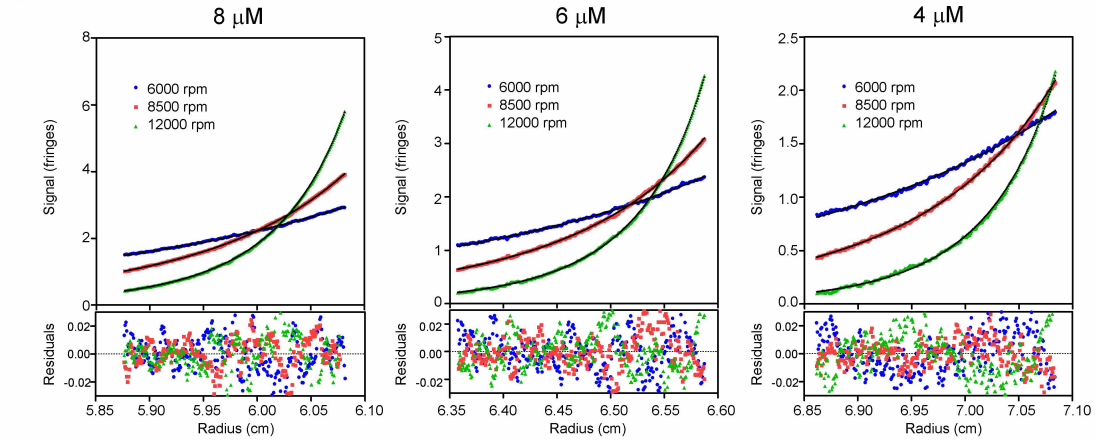
Extended Data Figure 1 | Sequence alignment of the 19 repeats of PPR10. **a**, PPR10 from maize specifically recognizes two RNA elements. The cartoon above illustrates the predicted domain organization of PPR10. 1 and 19 refer to the repeat numbers. CTP, chloroplast transit peptide. The blue brick with '?' represents a fragment of approximately 30 amino acids whose function remains to be characterized. The minimal RNA elements of *PSAJ* and *ATPH* that are targeted by PPR10 are shown below the cartoon. **b**, Sequence alignment of 19 repeats in PPR10. The secondary structural elements of a typical PPR motif are shown above. The residues at the 2nd, 5th and 35th positions which were predicted to be the molecular determinants for RNA-binding specificity are highlighted in magenta. The RNA sequences that can be recognized by PPR10 are listed on the right, 5′ to 3′ from top to bottom. The nucleotides which are recognized by PPR10 in a modular fashion in the *PSAJ*–PPR10 structure are shaded grey. **c**, The three numbering systems for a PPR motif. 1 is being used by Lurin *et al.*[12], Barkan *et al.*[6] and others; 2 is adopted by the Pfam database and being used by Kobayashi *et al.*[28], Yagi *et al.*[7] and others; and 3 is our proposed, structure-based numbering system. The residues that are predicted to specifically recognize RNA are coloured magenta.



RNA-free PPR10
Theoretical molecular weight:
82,842 Da
Fitting model:
monomer-dimer equilibrium
K_d << 100 nM

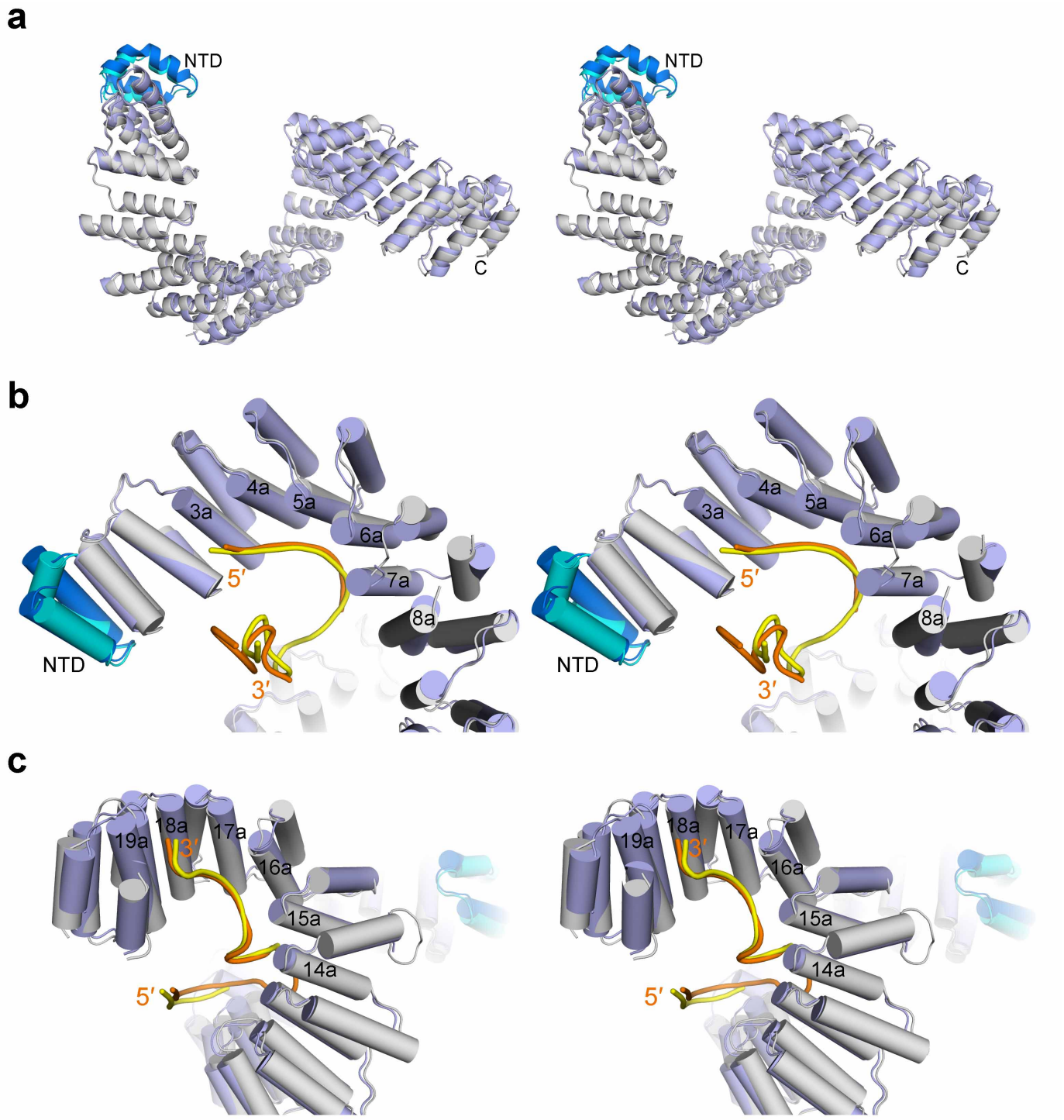


PPR10 : *PSAJ* = 1 : 1.5
Theoretical molecular weight:
88,420 Da
Fitting model:
monomer-dimer equilibrium
K_d = 102.0 ± 6.0 nM

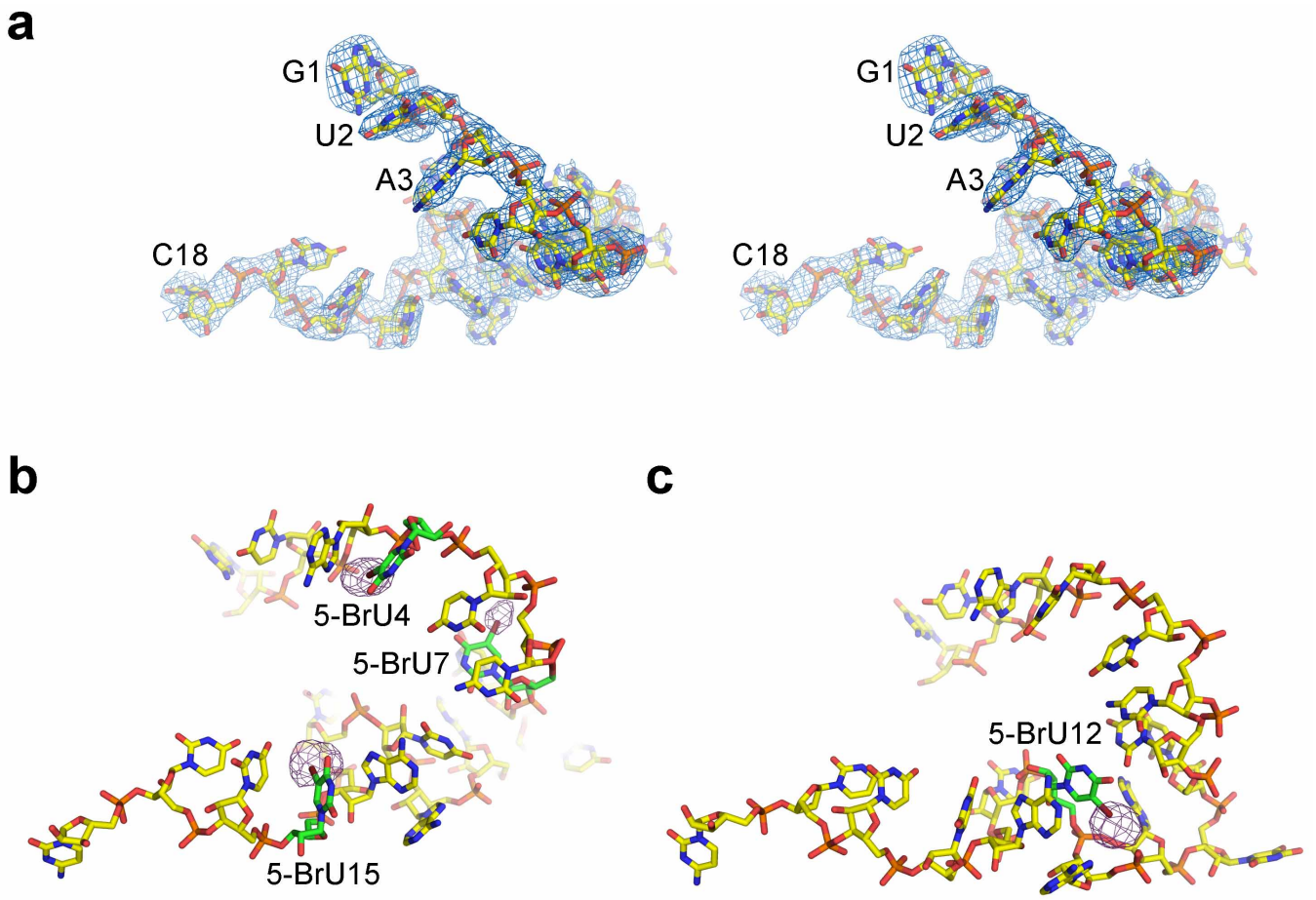


PPR10 : *ATPH* = 1 : 1.5
Theoretical molecular weight:
88,111 Da
Fitting model:
monomer-dimer equilibrium
K_d = 17.5 ± 0.2 μM

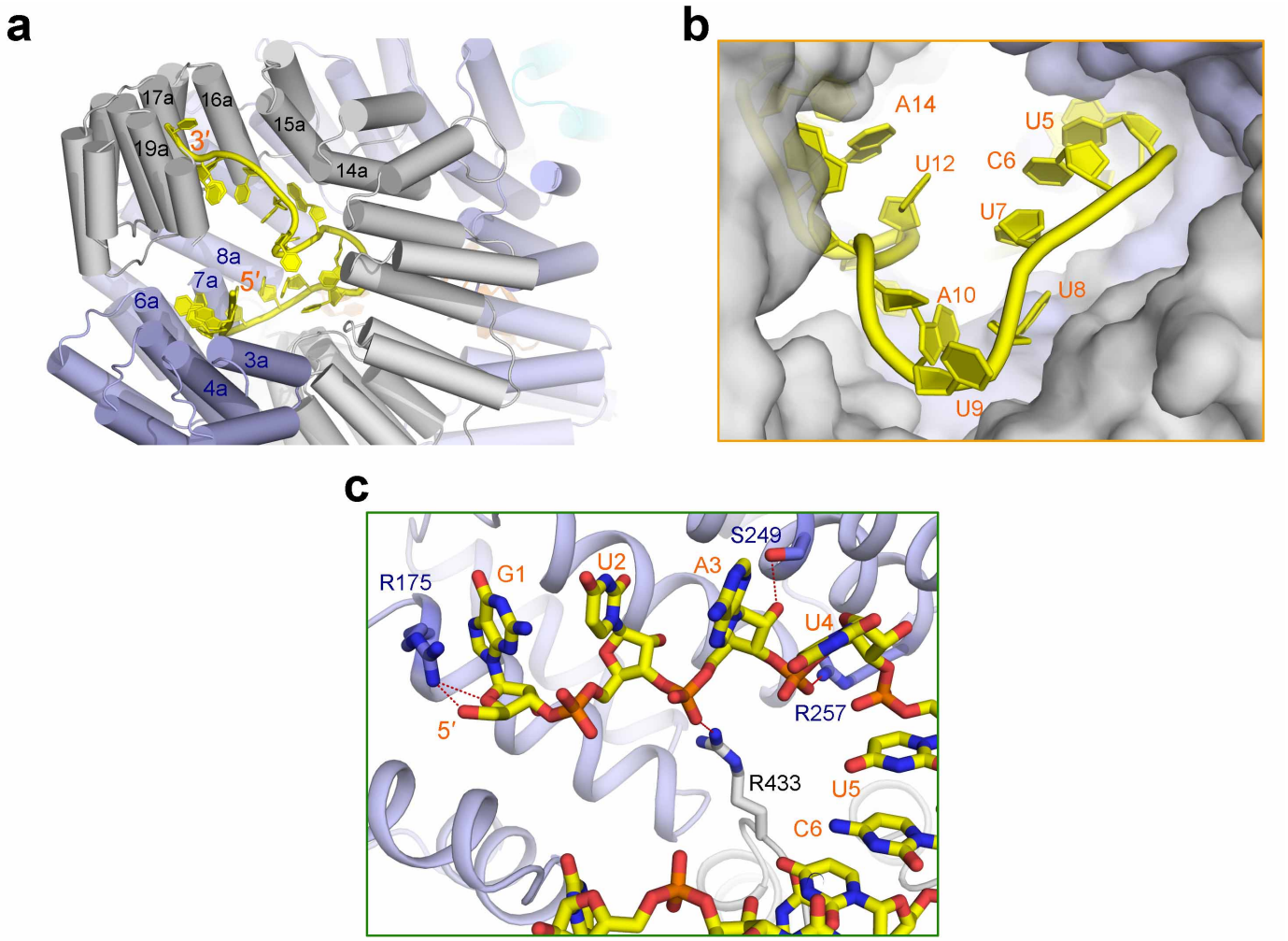
Extended Data Figure 2 | AUC-SE of PPR10 (residues 37–786, C256S/C279S/C430S/C449S) in the absence or presence of the target RNA elements. The molar concentrations of PPR10 are indicated above each panel. PPR10 and the RNA oligonucleotides were mixed at a stoichiometric ratio of approximately 1:1.5. Details of the experiments are described in Methods.



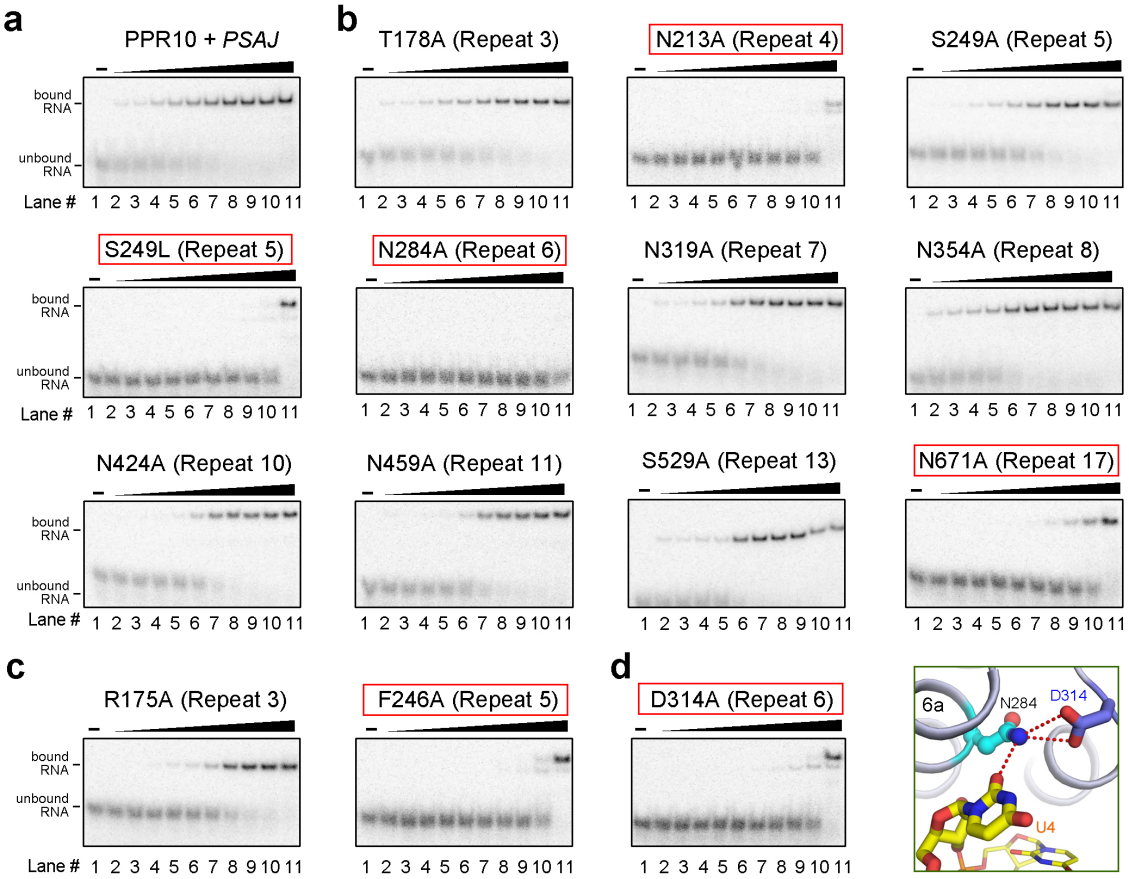
Extended Data Figure 3 | The two protomers of the RNA-bound PPR10 dimer exhibit similar conformations. **a**, The two protomers can be superimposed with a root-mean-squared deviation of 1.31 Å over 629 Cα atoms. **b**, **c**, The two ssRNA segments are coordinated by the PPR10 dimer similarly. The 5′ and 3′ segments of the bound *PSAJ* RNA are separately coordinated by the N-terminal repeats of one protomer (**b**), and the C-terminal repeats of the other protomer (**c**). Stereo-views are shown for all panels.



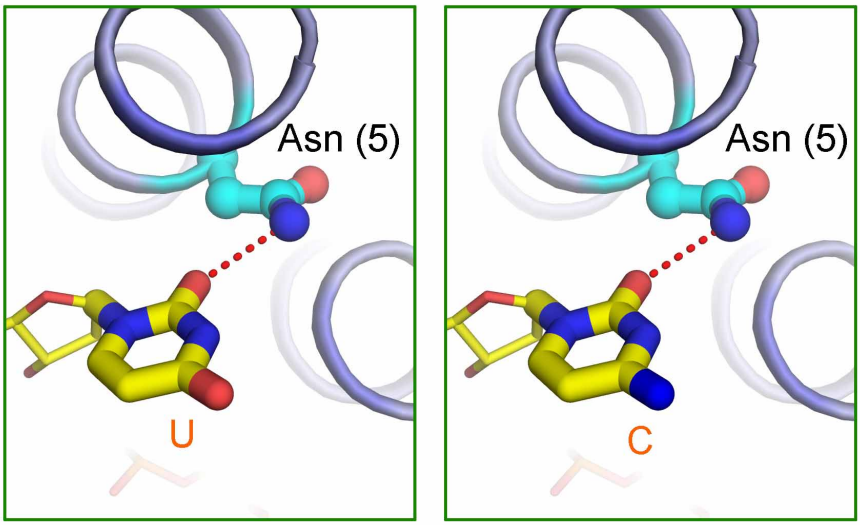
Extended Data Figure 4 | Electron density maps for a bound ssRNA segment. **a**, The $2F_o - F_c$ electron density for one segment of the bound *PSAJ* RNA. The electron density, contoured at 1σ and coloured blue, is displayed in stereo. **b**, **c**, The anomalous signals for bromine in the structures where the highlighted nucleotides were substituted with 5-bromouracil (5-BrU). The anomalous signals, shown in magenta mesh, are contoured at 5σ.



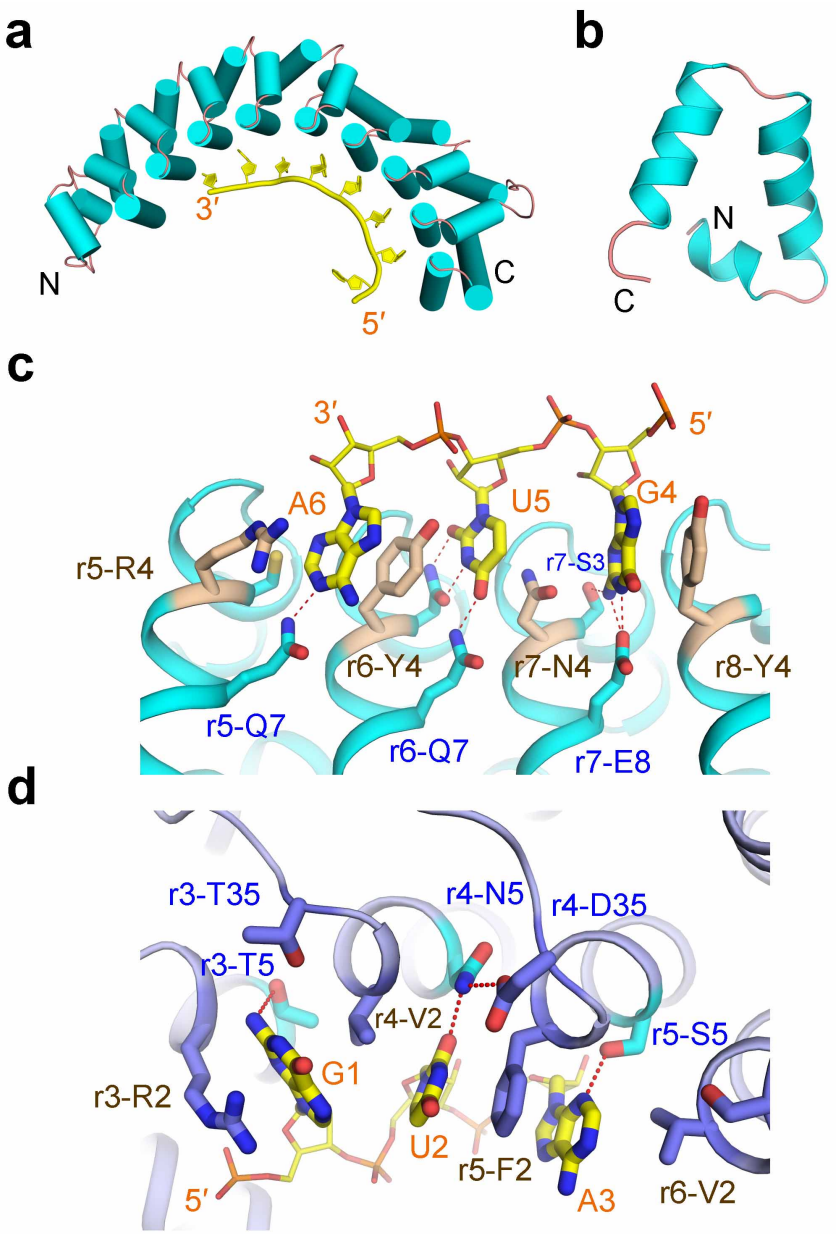
Extended Data Figure 5 | Coordination of the bound ssRNA by PPR10. **a**, The 5′ and 3′ portions of the *PSAJ* RNA element are separately bound by the N-terminal and C-terminal repeats of the two PPR10 protomers. Shown here is a close-up view of the binding of *PSAJ* by one end of the PPR10 dimer. **b**, The nucleotides U5–A10, which form a U-turn in the ssRNA, are uncoordinated in the cavity of the PPR10 dimer. The two protomers of PPR10 are shown in semi-transparent surface contour. **c**, The RNA backbone is coordinated by polar or charged residues through hydrogen bonds. The hydrogen bonds are represented by red dotted lines. The two protomers of PPR10 are coloured light purple and grey.



Extended Data Figure 6 | Mutational analysis of PPR10 residues that may be important for *PSAJ*. **a**, EMSA analysis of the interaction between PPR10 (residues 37–786, C256S/C279S/C430S/C449S) and *PSAJ* (5′-GUAUUCUUUAAUUAUUUC-3′). PPR10 was added with increasing concentrations of 0, 2, 4, 8, 16, 31, 63, 125, 250, 500, 1,000 nM in lanes 1–11 with approximately 40 pM ^{32}P-labelled *PSAJ* in each lane. **b**, Mutational analysis of the 5th residues of the indicated PPR motifs. The indicated point mutations were introduced to PPR10 (residues 37–786, C256S/C279S/C430S/C449S). **c**, Examination of the 2nd residues in repeats 3 and 5. **d**, Examination of the 35th residue of repeat 6. Note that the side group of Asp 314 is hydrogen bonded to the side chain of Asn 284, the 5th residue of repeat 6. The same structural feature is also seen in repeat 4 (Fig. 3b).



Extended Data Figure 7 | The predicted coordination of base C by an Asn at the 5th position of a PPR motif. Left, the coordination of base U by Asn observed in the structure. Right, the coordination of base C by Asn at the 5th position of a PPR motif modelled on the basis of the structure shown on the left.



Extended Data Figure 8 | Comparison of ssRNA coordination by PUF and PPR proteins. **a**, The structure of the human PUF protein PUM1 (also known as HSPUM) bound to the RNA element NRE1-19 (PDB accession code, 1M8W)[24]. The PUF repeats constitute an arc with 8-nt ssRNA bound to the concave side. Notably, the orientations of the bound RNA and the protein are antiparallel, namely the 5′ end is close to the C terminus of PUF. **b**, The structure of a PUF repeat. One canonical PUF repeat contains three helices, of which a short helix precedes a helical hairpin. **c**, Representative recognition of the RNA bases G, U, A by PUF repeats as seen in the structure of PUM1 bound to NRE1-19. The amino acids are labelled by the repeat number (r5, r6, r7, r8) followed by its one-letter code and position on the 2nd helix within a PUF repeat (S3, N4, E8, and so on). The same scheme applies to **d**. **d**, The coordination of RNA bases G, U, A by PPR10. It is noteworthy that PUF and PPR proteins share several common features for RNA binding: (1) the ssRNA elements are coordinated by the helices on the inner layer; and (2) the base is sandwiched mostly by hydrophobic residues or Arg. Yet the differences are evident between the two families of repeat proteins. As seen in **c**, the RNA base is usually coordinated by two residues that are located at the 4th and the 7th positions on helix 2 within a PUF repeat. By contrast, the base is mainly coordinated by the 5th residue of a PPR motif. The 35th residue, the last residue of a PPR motif that is located at a loop region preceding the next PPR motif, also contributes to base recognition.

Extended Data Table 1 | Statistics of data collection and refinement. Values in parentheses are for the highest resolution shell

	Se-PPR10	PPR10-*PSAJ* complex
Number of crystals	1	5
Space Group	$P2_12_12$	$P4_3$
Cell dimensions		
a, *b*, *c* (Å)	68.37, 176.54, 64.53	83.43, 83.43, 226.91
α, β, γ (°)	90, 90, 90	90, 90, 90
Wavelength (Å)	0.9793	0.9067
Resolution (Å)	40-2.85 (2.95-2.85)	40-2.45 (2.54-2.45)
R_{merge} (%)	12.7 (88.7)	8.9 (91.7)
I/sigma	28.9 (3.6)	27.7 (2.3)
Completeness (%)	100.0 (100.0)	96.3 (98.0)
Number of measured reflections	227,350	426,030
Number of unique reflections	18,683	54,087
Redundancy	12.2 (12.4)	7.9 (7.8)
Wilson B factor ($Å^2$)	73.3	82.9
R_{work} / R_{free} (%)	23.98/25.38	25.84/28.78
No. atoms		
Protein	5,326	10,384
main chain	2,808	5,480
side chain	2,518	4,904
RNA		652
water		114
B-factors		
Protein	74.8	85.45
main chain	75.2	85.13
side chain	74.3	85.81
RNA		93.84
water		67.51
R.m.s. deviations		
Bonds (Å)	0.011	0.015
Angle (°)	1.361	1.049
Ramachandran plot statistics (%)		
Most favorable	87.6	88.8
Additionally allowed	11.2	10.3
Generously allowed	1.1	0.9
Disallowed	0.0	0.0

Extended Data Table 2 | Statistics of data collection

	12-5BrU	4/7/15-5BrU	5/7/15-5BrU
Space Group	$P4_3$	$P4_3$	$P4_3$
Cell dimensions			
a, *b*, *c* (Å)	83.48, 83.48, 227.94	83.11, 83,11, 227.68	83.30, 83.30, 226.10
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Wavelength (Å)	0.9194	0.9194	0.9194
Resolution (Å)	40-3.2 (3.31-3.20)	40-3.00 (3.11-3.00)	40-3.20 (3.31-3.20)
R_{merge} (%)	12.2 (74.5)	11.7 (91.1)	14.8 (79.9)
I/sigma	21.2 (3.5)	26.2 (2.8)	20.1 (3.1)
Completeness (%)	100.0 (100.0)	99.8 (100.0)	100.0 (100.0)
Number of measured reflections	209,591	246,450	212,113
Number of unique reflections	25,708	30,609	25,347
Redundancy	8.2 (8.0)	8.1 (8.5)	8.4 (8.6)
Wilson B factor ($Å^2$)	69.0	79.6	71.4