RNA Biology 10:9, 1417–1418; September 2013; © 2013 Landes Bioscience

Pentatricopeptide repeat proteins

A set of modular RNA-specific binders massively used for organelle gene expression

Philippe Giegé

Institute for plant molecular biology (IBMP-CNRS), University of Strasbourg

Mitochondria and chloroplasts host genomes inherited from their bacterial ancestors. Their biogenesis thus requires fully functional gene expression machineries. RNA metabolism in organelles is particular and has attracted considerable attention since it combines prokaryotic features with unique traits that evolved in their eukaryotic host cell. For a long time, most of the molecular factors governing these processes have remained elusive. Recent research has shown that proteins of the pentatricopeptide repeat (PPR) family are major players of organelle gene expression.

PPR proteins make a large group of eukaryote-specific RNA-binding proteins encoded in the nucleus and predominantly located to organelles. This protein family has been identified over a decade ago and has been the subject of a fast growing number of studies ever since.[1-4] PPR proteins have escaped identification until complete genomic sequences were available because their primary sequences are very degenerate. Their identifying features are motifs of ca. 35 amino acids occurring as tandem arrays. Contrary to the sequence, the tri-dimensional organization of these modules seems highly conserved, with each repeat arranged in a helix-turn-helix structure and the succession of PPR motifs forming a solenoid.[5]

PPR proteins are ubiquitous in eukaryotes and are totally absent from bacteria, with the exception of a small number of plant parasites such as Ralstonia that probably acquired PPR genes by horizontal gene transfer.[3] Still, the distribution of PPR genes among eukaryote clades is very uneven. They are for example present in relatively small numbers in fungi and animals, they occur in higher numbers in trypanosomes, and are present in very large numbers in higher plants. In the latter, hundreds of PPR genes are present, representing as much as 1% of nuclear genomes.[2]

The identification and functional analysis of PPR proteins has already helped to solve long-standing questions regarding organelle gene expression. It has for example enabled to understand how sequence specificity is achieved for hundreds of cytidine to uridine editing sites present in higher plant organelle transcriptomes.[6] It has also settled the long-lasting debate on the existence of protein-only enzymes holding RNase P activity.[7,8] Since their discovery, the function of PPR proteins has been associated with most aspects of gene expression in organelle, i.e., transcription, RNA splicing, RNA editing, RNA ends maturation, and translation. Research on PPR proteins is thus instrumental to understand organelle RNA metabolism.

Initial functional data on PPR proteins have mainly been derived from genetic studies. The most recent advances now unravel the mode of action of PPR proteins. Because of their modular structure, it had been proposed from the start that RNA recognition by PPR proteins would involve the recognition of specific RNA moieties by individual PPR motifs. Such a recognition code has indeed been uncovered. The occurrence of specific amino acids mainly at two to three given positions in PPR motifs accounts for the specificity of the respective motifs for individual nucleotides.[9-11]

This special issue of *RNA Biology* covers all the major aspects and recent advances on PPR proteins throughout eukaryotes. It features four original articles as well as 12 in-depth reviews. Individual reviews deal with the current knowledge on PPR proteins mode of action, on their functions in yeast, in animals, in trypanosomes, in green algae, in the moss *Physcomitrella patens*, and in higher plant chloroplasts. Other reviews describe specific functional groups of PPR proteins; i.e., P-class PPR proteins required for the end maturation of mitochondrial transcripts, PPR proteins belonging to the subgroup of "restorer of fertility" proteins, and PPR proteins involved in RNA editing. Finally, other reviews deal with PPR proteins containing specific additional domains such as the "Small MutS-Related" (SMR) domain and the "Nedd4-BP1 YacP Nuclease" (NYN) domain, characterizing PPR proteins holding RNase P activity. Original articles describe the identification of DYW domains similar to those of land plant editing factors in diverse eukaryotes, identify novel RNA editing factors in Arabidopsis, and show that in plants the dual targeting of PPR proteins to mitochondria and plastids is more widespread than previously envisaged.

Despite the progress in understanding PPR proteins biology, many important challenges remain for the future. A first aspect will be to better understand how the function of PPR proteins is integrated in a wider context, i.e., to know how PPR functions are connected with other gene expression pathways and in general with other cellular processes. Answers to this question might be brought by the determination of both protein–protein and protein–RNA interaction networks of PPR proteins and the biochemical characterization of complexes involving PPR proteins.

Correspondence to: Philippe Giegé; Email: philippe.giege@ibmp.cnrs.unistra.fr
Submitted: 08/07/2013; Accepted: 08/07/2013
http://dx.doi.org/10.4161/10.4161/rna.26081

Another critical question will be to understand at the atomic level how the PPR code functions, i.e., how base specificity is achieved by individual residues in PPR motifs. Answers to this question might come through the determination of tri-dimensional structures of PPR proteins in complex with their RNA substrates.

The full exploitation of the potential held by PPR proteins also remains a major challenge for the future. Custom designed PPR proteins obtained through the reshuffling of PPR motifs or through point mutations are expected to specifically recognize any RNA sequence and/or structure of interest. The possibility to design protein specificity toward RNA obviously opens appealing perspectives for a tremendous number of biotechnological applications.

Finally, the thorough understanding of PPR proteins mode of action might help to derive universal rules for target recognition by a much wider group of proteins. Indeed, according to their tridimensional fold, PPR proteins fall into the large family of helical-repeat proteins that bind nucleic acids, which includes PUF, TALE, HAT, mTERF, and OPR proteins families.[12] Similar to PPR proteins, recognition codes connecting individual protein motifs and nucleotides identity have been described for both PUF and TALE proteins.[13,14] It is thus possible that similar codes remain to be identified for the other protein families and that grand rules exist to define how target are recognized by all these solenoid forming proteins.

References

1. Small ID, Peeters N. The PPR motif - a TPR-related motif prevalent in plant organellar proteins. Trends Biochem Sci 2000; 25:46-7; PMID:10664580; http://dx.doi.org/10.1016/S0968-0004(99)01520-0
2. Aubourg S, Boudet N, Kreis M, Lecharny A. In Arabidopsis thaliana, 1% of the genome codes for a novel protein family unique to plants. Plant Mol Biol 2000; 42:603-13; PMID:10809006; http://dx.doi.org/10.1023/A:1006352315928
3. Lurin C, Andrés C, Aubourg S, Bellaoui M, Bitton F, Bruyère C, Caboche M, Debast C, Gualberto J, Hoffmann B, et al. Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell 2004; 16:2089-103; PMID:15269332; http://dx.doi.org/10.1105/tpc.104.022236
4. Schmitz-Linneweber C, Small I. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends Plant Sci 2008; 13:663-70; PMID:19004664; http://dx.doi.org/10.1016/j.tplants.2008.10.001
5. Filipovska A, Rackham O. Pentatricopeptide repeats: Modular blocks for building RNA-binding proteins. RNA Biol 2013; 10: In press; PMID:23635770; http://dx.doi.org/10.4161/rna.24769.
6. Castandet B, Araya A. RNA editing in plant organelles. Why make it easy? Biochemistry (Mosc) 2011; 76:924-31; PMID:22022966; http://dx.doi.org/10.1134/S0006297911080086
7. Holzmann J, Frank P, Löffler E, Bennett KL, Gerner C, Rossmanith W. RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. Cell 2008; 135:462-74; PMID:18984158; http://dx.doi.org/10.1016/j.cell.2008.09.013
8. Gobert A, Gutmann B, Taschner A, Gössringer M, Holzmann J, Hartmann RK, Rossmanith W, Giegé P. A single Arabidopsis organellar protein has RNase P activity. Nat Struct Mol Biol 2010; 17:740-4; PMID:20473316; http://dx.doi.org/10.1038/nsmb.1812
9. Barkan A, Rojas M, Fujii S, Yap A, Chong YS, Bond CS, Small I. A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. PLoS Genet 2012; 8:e1002910; PMID:22916040; http://dx.doi.org/10.1371/journal.pgen.1002910
10. 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 2013; 8:e57286; PMID:23472078; http://dx.doi.org/10.1371/journal.pone.0057286
11. Takenaka M, Zehrmann A, Brennicke A, Graichen K. Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. PLoS One 2013; 8:e65343; PMID:23762347; http://dx.doi.org/10.1371/journal.pone.0065343
12. Filipovska A, Rackham O. Modular recognition of nucleic acids by PUF, TALE and PPR proteins. Mol Biosyst 2012; 8:699-708; PMID:22234420; http://dx.doi.org/10.1039/c2mb05392f
13. Wang X, McLachlan J, Zamore PD, Hall TM. Modular recognition of RNA by a human pumilio-homology domain. Cell 2002; 110:501-12; PMID:12202039; http://dx.doi.org/10.1016/S0092-8674(02)00873-5
14. Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U. Breaking the code of DNA binding specificity of TAL-type III effectors. Science 2009; 326:1509-12; PMID:19933107; http://dx.doi.org/10.1126/science.1178811