Biochimie 113 (2015) 93–99

Contents lists available at ScienceDirect

Biochimie

journal homepage: www.elsevier.com/locate/biochi

Review

An overview of pentatricopeptide repeat proteins and their applications

Sam Manna

Department of Microbiology, La Trobe University, Melbourne, Victoria, Australia

ARTICLE INFO

Article history:
Received 19 February 2015
Accepted 3 April 2015
Available online 14 April 2015

Keywords:
Pentatricopeptide repeat protein
Mitochondria
Chloroplast
Horizontal gene transfer
tRNA guanine methyltransferase

ABSTRACT

Pentatricopeptide repeat (PPR) proteins are a large family of modular RNA-binding proteins which mediate several aspects of gene expression primarily in organelles but also in the nucleus. These proteins facilitate processing, splicing, editing, stability and translation of RNAs. While major advances in PPR research have been achieved with plant PPR proteins, the significance of non-plant PPR proteins is becoming of increasing importance. PPR proteins are classified into different subclasses based on their domain architecture, which is often a reflection of their function. This review provides an overview of the significant findings regarding the functions, evolution and applications of PPR proteins. Horizontal gene transfer appears to have played a major role in the sporadic phylogenetic distribution of different PPR subclasses in both eukaryotes and prokaryotes. Additionally, the use of synthetic biology and protein engineering to create designer PPR proteins to control gene expression *in vivo* is discussed. This review also highlights some of the aspects of PPR research that require more attention particularly in non-plant organisms. This includes the lack of research into the recently discovered PPR-TGM subclass, which is not only the first PPR subclass absent from plants but present in economically and clinically-relevant pathogens. Investigation into the structure and function of PPR-TGM proteins in these pathogens presents a novel opportunity for the exploitation of PPR proteins as drug targets to prevent disease.

1. Pentatricopeptide repeat proteins

Due to the relatively low number of promoters in organelle genomes and the long half-lives of their RNAs, the use of transcriptional regulators to control gene expression is not sufficient. Thus, organelle transcriptomes have a great dependence on RNA-binding proteins to regulate gene expression at the post-transcriptional level [1]. One of the major mediators of organelle post-transcriptional control is the pentatricopeptide repeat (PPR) protein family. The PPR family was simultaneously discovered by two independent research groups during the sequencing of the *Arabidopsis thaliana* genome [2,3]. All PPR proteins contain tandemly repeated sequence motifs (the PPR motifs) which can vary in number [2,3]. These proteins are found in all eukaryotic lineages but appear to have undergone an expansion in terrestrial plants [4]. A small number of PPR-encoding genes have also been reported in prokaryotes (including pathogenic and symbiotic members of the genera *Rhodobacter*, *Ralstonia*, *Simkania*, *Erwinia*, and *Legionella*), but these genes are proposed to have been acquired via eukaryote-to-prokaryote horizontal gene transfer events [4–9].

Proteins containing PPR motifs are known to have roles in transcription, RNA processing, splicing, stability, editing, and translation (Table 1) [4,10]. As a result, PPR proteins are important for expression of organelle genomes and organelle biogenesis. PPR proteins can be non-catalytic where they act as adaptors by mediating interactions between cognate transcripts and their effectors. Alternatively, increasing evidence is emerging of some PPR proteins that catalyse functions such RNA processing and editing themselves [4,10].

2. Structure of PPR proteins

Amino acid alignments between the consensus PPR and the previously characterised tetratricopeptide (TPR) motif revealed that the PPR motif is a degenerate 35 amino acid motif repeated in tandem [3]. The number of PPR motifs within a protein range from 2 to over 26 [11]. The sequence similarity of these helical repeat

Abbreviations: PPR, pentatricopeptide repeat; PRORP, proteinaceous RNase P; SMR, small MutS-related; TGM, tRNA guanine-N7 methyltransferase; TPR, tetratricopeptide repeat.

E-mail address: sam_manna@y7mail.com.

http://dx.doi.org/10.1016/j.biochi.2015.04.004

Table 1
Selected PPR proteins and their respective functions in organelle gene expression.

Organism	Protein	PPR subclass	Localisation	Transcript	Function	Reference
Homo sapiens	POLRMT	mtRNAP	Mitochondria	N/A	Transcription	[18]
Homo sapiens	MRPP3	PRORP	Mitochondria	tRNAs	5′ tRNA processing	[53]
Homo sapiens	PTCD1	P	Mitochondria	tRNAs	3′ tRNA processing	[76]
Homo sapiens	PTCD2	P	Mitochondria	*ND5-Cyt b*	Non-tRNA processing	[77]
Homo sapiens	LRPPRC	P	Mitochondria	mRNAs	Polyadenylation, stability	[41,42]
Saccharomyces cerevisiae	Rpo41	mtRNAP	Mitochondria	N/A	Transcription	[61,78]
Saccharomyces cerevisiae	Ccm1p	P	Mitochondria	*cob*, *cox1*	Splicing	[79]
Saccharomyces cerevisiae	Pet309	P	Mitochondria	*cox1*	Stability, translation	[80]
Neurospora crassa	Cya-5	P	Mitochondria	*cox1*	Translation	[81]
Dictyostelium discoideum	PtcE	PPR-TGM	Mitochondria	tRNAs	tRNA processing	[28]
Trypanosoma brucei	KPAF1/KPAF2	P	Mitochondria	mRNAs	Polyadenylation, polyuridylation, translation	[82]
Trypanosoma brucei	PRORP1	PRORP	Nucleus	tRNAs	5′ tRNA processing	[59]
Trypanosoma brucei	PRORP2	PRORP	Mitochondria	tRNAs	5′ tRNA processing	[59]
Chlamydomonas reinhardtii	MCA1	P	Chloroplast	*petA*	Stability	[83]
Arabidopsis thaliana	CRR4	PLS (E)	Chloroplast	*ndhD*	Editing	[84]
Arabidopsis thaliana	PTAC2	P	Chloroplast	N/A	Transcription	[85]
Arabidopsis thaliana	MTSF1	P	Mitochondria	*nad4*	Stability	[86]
Arabidopsis thaliana	OTP51	LAGLIDADG	Chloroplast	*ycf3*	Splicing	[67]
Arabidopsis thaliana	PRORP1	PRORP	Mitochondria chloroplast	tRNAs	5′ tRNA processing	[55]
Arabidopsis thaliana	PRORP2/PRORP3	PRORP	Nucleus	tRNAs	5′ tRNA processing	[58]
Arabidopsis thaliana	SVR7	PPR-SMR	Chloroplast	*atpB*, *atpE*, *rbcL*	Translation	[70]
Zea mays	Crp1	P	Chloroplast	*petA*, *psaC*, *petD*	Translation processing	[87]
Zea mays	PPR2263	PLS (DYW)	Mitochondria	*nad5*, *cob*	Editing	[88]
Physcomitrella patens	PpPPR79	PLS (DYW)	Mitochondria	*nad5*	Editing	[89]
Physcomitrella patens	PpPPR38	P	Chloroplast	*clpP*	Splicing	[90]

families, taken together with the greater prevalence of TPR proteins in prokaryotes, led to the hypothesis that the PPR motif emerged from the TPR motif during the early stages of eukaryotic evolution [5]. Similarly to the TPR motif, the PPR motif forms two anti-parallel α-helices, which interact to produce a helix-turn-helix motif (Fig. 1) [4]. The series of helix-turn-helix motifs formed by PPR motifs throughout the protein produces a superhelix with a central groove that allows the protein to bind RNA [3,4].

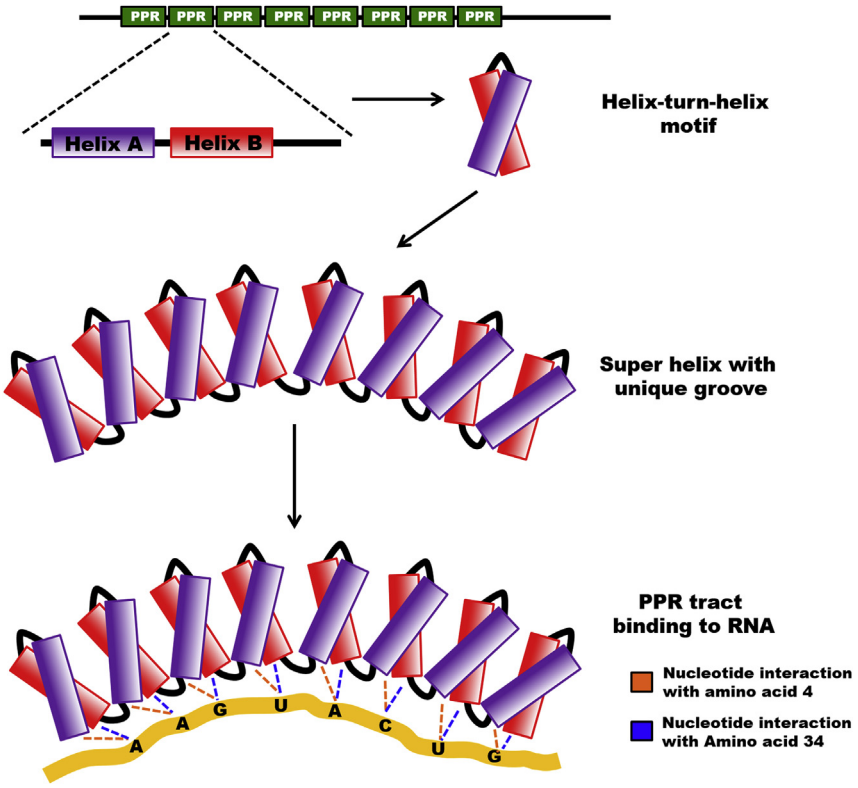

Fig. 1. PPR protein structure and the mechanism of transcript recognition. Each PPR motif forms two α-helices, which interact to form a helix turn helix motif. The series of helix turn helix motifs throughout the protein are stacked together to form a superhelix with an RNA binding groove. Modular recognition of transcripts is mediated by nucleotide interactions with the amino acids at positions 4 and 34 of each PPR motif.

The RNA-binding capabilities of PPR proteins were subsequently confirmed by several studies and supported by RNase protection experiments of the *A. thaliana* chloroplast transcriptome, in which several short (16–28 nucleotides) RNA molecules remained intact following nuclease treatment [12–16]. These RNase footprints correspond to protein target sites, where the binding of the PPR proteins provides the RNAs with protection from nucleases [16].

The insoluble nature of heterologously expressed PPR proteins has impeded studies that attempted to confirm the superhelical structure of these proteins and their mechanism of RNA recognition [17]. Over a decade after the discovery of PPR proteins, the superhelical structure was confirmed. Ringel et al. [18] reported the crystal structure of the mammalian mitochondrial RNA polymerase POLRMT, which contains two PPR motifs towards the N-terminus. The crystal structure of POLRMT confirmed predictions that each PPR motif forms pairs of interacting anti-parallel α-helices. Following the resolution of the POLRMT structure, subsequent studies have investigated the crystal structure of other PPR proteins. The protein-only mitochondrial RNase P crystal structure from *A. thaliana* (PRORP1) confirmed the role of its PPR motifs in pre-tRNA binding and suggest it has evolved independently from other RNase P proteins that rely on catalytic RNA [19]. Crystal structures of PPR proteins from *Zea mays* (PPR10) and *Brachypodium distachyon* (THA8) in their RNA-free and RNA-bound forms revealed that these proteins undergo conformational changes and form homodimers upon RNA binding [20,21].

3. Sequence-specific recognition of RNA by PPR proteins and the use of synthetic biology for *in vivo* manipulation of gene expression

PPR proteins bind their cognate RNA substrates in a sequence-specific manner. The modular recognition and combinatorial amino acid code used by PPR proteins for RNA recognition was confirmed by Barkan et al. [22] who demonstrated positions 6 and 1′ in the PROSITE model of the PPR motif (where 1′ denotes the first amino acid of the downstream motif) were responsible for RNA binding and nucleotide recognition [22]. It is important to note that some studies use the Pfam model of the PPR motif which designates the beginning and end of the PPR motif two amino acids earlier than the PROSITE model. Therefore, positions 6 and 1′ in the PROSITE model are referred to as positions 4 and 34 in studies using the Pfam model [23]. Due to its wide use in NCBI databases, this review will use the Pfam numbering system in reference to the key residues involved in nucleotide recognition from this point forward.

Using positions 4 and 34 within each motif, PPR proteins recognise their cognate transcript in a modular fashion. That is, one PPR motif recognises one nucleotide (Fig. 1). The residues at positions 4 and 34 will determine the nucleotide the motif can recognise. For example a PPR motif with a threonine at position 4, in combination with asparagine at position 34, will recognise adenine, whereas an asparagine and aspartic acid at these respective positions would recognise uracil [22,23]. Crystal structures of PPR proteins have also provided extensive insight into the interaction between PPR motifs and nucleotides, which occurs via van der Waals interactions [21]. Since PPR proteins mediate several different functions in organelle RNA metabolism, elucidation of the recognition code has begun to allow the *in silico* prediction of RNA targets and provide a better indication of the function of specific PPR proteins [22].

Additionally, resolution of the PPR recognition code makes it possible to direct a PPR protein to bind to a specific RNA sequence of interest by manipulating the residues at the aforementioned positions within the PPR motifs [22]. Compared to other helical repeat RNA-binding proteins, PPR motifs exhibit higher diversity in amino acid composition, suggesting greater versatility for engineering. The creation of designer PPR proteins that recognise an RNA sequence of interest has been achieved using two different approaches. The first strategy involved introducing mutations into the coding sequence of PPR10 from *Z. mays* to change the amino acids encoded at positions 4 and 34 within the sixth and seventh PPR motifs. Using this method, Barkan et al. [22] were able to direct the binding of the protein to an RNA sequence of interest. Rather than mutating existing PPR motifs to influence the RNA-binding preference of the protein, the second strategy involves the creation of an artificial PPR protein scaffold [24]. This was achieved using a consensus design strategy to create synthetic PPR proteins that are stable and programmable to efficiently bind RNA in a sequence-specific manner [24]. These designer PPR proteins constructed via synthetic biology and protein engineering could be used to promote or prevent endogenous gene expression, regulate endogenous RNA metabolism, edit RNA sequences in a site-specific manner or even prevent infection by RNA viruses or viroids [25,26].

4. Phenotypes associated with loss of PPR proteins

Most PPR proteins localise either to mitochondria or chloroplasts, where they modulate gene expression at the RNA level. Some of these proteins are even dual targeted to both organelles, suggesting regulated expression between mitochondria and chloroplasts [27]. Since PPR proteins have major roles in organelle gene expression, it is not surprising that defects in their function can yield phenotypes associated with organelle dysfunction. Knockdown of PPR-encoding genes in protists such as *Dictyostelium discoideum* and *Trypanosoma brucei* result in growth defects [28,29]. In plants, mutants of mitochondrial PPR proteins often exhibit defective embryo development and cytoplasmic male sterility [30–35]. Plant mitochondrial genomes often encode factors that induce cytoplasmic male sterility, the inability to produce viable pollen [36]. These factors are normally suppressed by nuclear-encoded restorer of fertility genes, such as the PPR genes [30,33,35–37]. The restorer of fertility PPR proteins appears to function by preventing the translation or accumulation of mitochondrial transcripts, whose gene products induce cytoplasmic male sterility [37]. Plant PPR mutants often have no mechanism to supersede these cytoplasmic male sterility-inducing mitochondrial factors, and therefore cannot produce functional pollen. Mutants of chloroplast PPR proteins in plants also exhibit various phenotypes including defective embryo development, aberrant photosynthetic ability and changes in seed pigmentation [4,13,38]. Rapid methods for the screening and identification of mutations in PPR genes that affect the viability of crops therefore have significant implications in agriculture.

In humans, a point mutation in the gene encoding the leucine-rich pentatricopeptide repeat cassette (LRPPRC) protein is associated with the French-Canadian variant of Leigh syndrome, a neurodegenerative disorder [39]. The mutation results in a cytochrome c oxidase deficiency caused by reduced levels of cytochrome c oxidase mRNAs, suggesting a role for LRPPRC in the stability of these transcripts [39]. The stability of other mitochondrial transcripts is also negatively affected upon knockdown of *LRPPRC* expression [40]. Ruzzenente et al. [41] demonstrated a role for LRPPRC in polyadenylation, which supports this notion [41]. The role of LRPPRC in facilitating mitochondrial gene expression has also been associated with tumourigenesis, in which the overexpression of LRPPRC in tumours can prevent apoptosis and enhance mitochondrial function [42]. LRPPRC has also been proposed to function in mitochondrial transcription, however this is currently under dispute [43,44]. Thus, further investigation into

LRPPRC is required to elucidate its precise function in mitochondrial RNA metabolism and the molecular pathologies underlying the French-Canadian variant of Leigh syndrome.

5. Types of PPR proteins

PPR proteins are classified based on their domain architecture (Fig. 2). P-class PPR proteins possess the canonical 35 amino acid motif and normally lack additional domains [4]. Members of this class have functions in most aspects of organelle gene expression (Table 1). PLS-class PPR proteins have three different types of PPR motifs, which vary in length; P (35 amino acids), L (long, 35–36 amino acids) and S (short, ~31 amino acids), and members of this class are thought to mainly function in RNA editing [4]. Not surprisingly, their presence in particular organisms appears to correlate with the prevalence of RNA editing [45,46]. Subtypes of the PLS class are categorised based on the additional C-terminal domains they possess. In addition to the PLS-type PPR motifs, members of the E subclass contain an E (extended) domain. The E domain is not catalytic, but is predicted to be a protein–protein interaction motif, which acts by recruiting the editing enzyme [47,48]. A second PLS subclass not only contains PPR motifs and an E domain, but also has a DYW domain, named for its highly conserved aspartate–tyrosine–tryptophan (DYW) tripeptide located at the C-terminal end [11]. The DYW domain has been proposed to be an editing domain, as it contains the same conserved catalytic residues found in cytidine deaminase enzymes that mediate C–U conversions [45,46,49]. Similarly to cytidine deaminases, the conserved deaminase residues in the DYW domain possess zinc-binding capabilities and are essential for the editing function of these proteins, providing support for the catalytic role of DYW proteins in editing [50,51]. The PLS-class was originally proposed to be specific to plants, however DYW-containing PPR proteins were subsequently identified in the heterolobosean protist *Naegleria gruberi* and later in several other non-plant organisms [9,52]. Members of the DYW subclass appear to be present in specific eukaryotic lineages but absent in others, suggesting a complex distribution [9]. This punctate phylogenetic distribution indicates that DYW-type PPR proteins in non-plant organisms were acquired by specific lineages from plants via horizontal gene transfer events [9].

While a major group of P-class PPR proteins lack additional domains, multiple subclasses within this group exist that possess catalytic C-terminal domains. In addition to their N-terminal PPR motifs, members of the proteinaceous RNase P (PRORP) subclass possess a C-terminal metallonuclease domain (Fig. 2) [53,54]. These domains mediate the phosphodiester bond cleavage required to release the 5′ end of mitochondrial tRNAs from their polycistronic precursors, and is therefore the catalytic component of the mitochondrial RNase P [53,55,56]. The PPR motifs of PRORP1 in *A. thaliana* are required for recognition of nucleotides C56 and A57 in the TψC loop of precursor tRNAs, which is also likely the case for other PRORP proteins [57]. Unlike most PPR proteins, some members of this subclass do not appear to be confined to organelles, with some PRORP proteins functioning in the nucleus of *A. thaliana* and *T. brucei* [58,59]. Interestingly, *T. brucei* also possesses a mitochondrially targeted PRORP, which can catalyse 5′ tRNA processing *in vitro*, even though the *T. brucei* mitochondrial genome does not encode tRNAs [59]. Therefore, investigations into this protein in tRNA and possibly non-tRNA processing will provide further insight into the functions of the PRORP family [54].

PRORP is highly conserved with homologs present in plants, algae, heterokonts and animals [55,56]. However, these proteins are lacking in distinct lineages, such as cellular slime moulds like *D. discoideum*, despite the fact the mitochondrial RNAs of this amoeba undergo tRNA processing [28,56,60]. It will therefore be of interest to investigate why specific eukaryotic lineages appear to have developed non-PRORP mediated mitochondrial tRNA processing mechanisms.

Although not typically considered a PPR subclass, the mitochondrial RNA polymerases of mammals and yeast also contain PPR motifs (Fig. 2) [61–63]. Mitochondrial RNA polymerases possess a unique N-terminal extension, which is absent from their bacteriophage homologs. The N-terminus contains the PPR motifs, which are required for promoter-specific transcription initiation and the coupling of this process to post-transcriptional regulation of mitochondrial RNAs [18,63,64]. Interestingly, PPR motifs appear to be more difficult to detect in mitochondrial RNA polymerases. In fact, it was only after the development of a Markov hidden algorithm for the identification of PPR motifs in yeast that they were identified in yeast mitochondrial RNA polymerases [61]. It is therefore possible that mitochondrial RNA polymerases from other organisms such as *D. discoideum* possess PPR tracts which remain unidentified. This may be the result of the heavy bias toward *in silico* detection of plant PPR motifs due to the significantly greater number of plant PPR sequences available compared to those from non-plant organisms.

LAGLIDADG motifs are sequences required for splicing and are regularly found in maturases and homing endonucleases (Fig. 2)

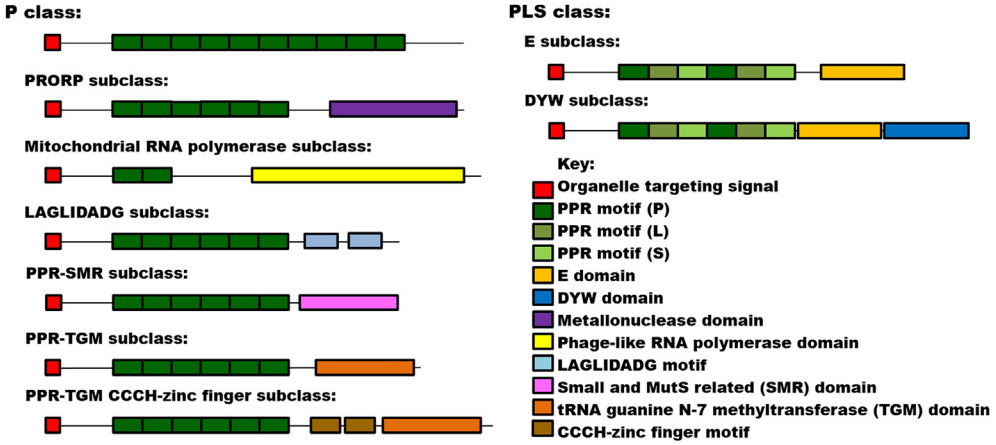

Fig. 2. Basic domain architecture of PPR subfamilies identified to date. Number of PPR motifs displayed in each protein is not characteristic of all members of that respective subfamily and schematics are not to scale.

[65,66]. However, LAGLIDADG motifs are also present in some plant PPR proteins [49,67]. It is therefore not surprising that at least one of these PPR-LAGLIDADG proteins (OTP51) in *A. thaliana* function in transcript splicing [67]. These proteins were predicted to have arisen when an organelle-encoded maturase gene was transferred to the nucleus and fused to a pre-existing PPR-encoding gene [67]. The lack of PPR-LAGLIDADG proteins in other eukaryotic lineages suggests their function in splicing may be catalysed by other proteins in non-plant organisms. PPR-LAGLIDADG proteins remain a largely unexplored subclass of PPR proteins and therefore their significance remains to be investigated.

Only recently discovered was the PPR-SMR group of proteins, whose members contain P-class type PPR motifs with a C-terminal small MutS-related (SMR) domain (Fig. 2) [68–70]. The SMR domain shares homology with bacterial SMR proteins, which have roles in DNA recombination and repair [71]. With the exception of the bacterial pathogen *Legionella longbeachae*, PPR-SMR proteins appear to be restricted to organisms with chloroplasts and have functions in RNA endonucleolytic cleavage and translation [6,68–70]. It is thought that the DNA endonucleolytic activity of the SMR domain, in combination with the RNA-binding ability of the PPR motifs, provides an RNA cleavage function [68]. Interestingly, while the phylogenetic distribution of PPR-SMR proteins would be indicative of functions in chloroplast RNA metabolism, some appear to localise to mitochondria [68,72]. Further research is required to better understand the evolutionary origin, as well as the mechanism of function of PPR-SMR proteins not only in plants but in *L. longbeachae*.

The most recent member of the PPR family identified was the PPR-TGM subfamily. In addition to the PPR tract, these proteins contain a C-terminal tRNA guanine-N7 methyltransferase domain (Fig. 2). The C-terminal domain of the PPR-TGM proteins originated from a chlamydial TGM-encoding gene, acquired via horizontal gene transfer, which evolved PPR motifs following its eukaryotic acquisition [73]. While these proteins are yet to be characterised, their domain architecture suggests they are predicted to function in tRNA metabolism including the *S*-adenosylmethionine-dependent methylation of tRNA [73]. Individual PPR-TGM proteins were found to contain mitochondrial, chloroplast or no targeting signals, indicating that they function in various locations of the cell. PPR-TGM proteins are only present in eukaryotic microbes including cellular slime moulds, entamoebae, algae and diatoms [28,73]. The absence of PPR-TGM proteins in plants provides a useful tool to investigate the evolution of the PPR family in non-plant lineages. In green algae, the PPR-TGM-encoding gene has undergone a duplication event, resulting not only in one traditional PPR-TGM protein, but also in a second subtype with a CCCH-zinc finger motif, which also possess RNA-binding capabilities. The function of these proteins acquiring a second RNA-binding motif remains unclear [73].

Among the eukaryotic microbes possessing PPR-TGM proteins are the oyster pathogen *Perkinsus marinus* and the clinically-relevant human pathogen *Entamoeba histolytica* [73]. Given *in silico* analyses suggest PPR-TGM proteins play important roles in tRNA function (and consequently translation), as well as the deleterious phenotypes displayed by other PPR mutants, it is likely these proteins are important to these pathogens. Additionally, drugs that inhibit RNA methylation such as sinefungin, an *S*-adenosylmethionine analogue, have been developed previously [74,75]. Taken together, PPR-TGM proteins represent new putative targets for the development of drugs that can specifically inhibit their activity in pathogenic eukaryotes.

Interestingly, while PPR-TGM proteins have been found in a diverse range of eukaryotic species, they do not seem to be present in organisms that have DYW-type PPR proteins (Table 2) [9,73]. This may imply functional redundancy between these subfamilies of PPR proteins. However, the significance of these opposing phylogenetic distributions remains to be investigated.

Table 2
Distribution of PPR-TGM and DYW-type PPR proteins in eukaryotes. The presence or absence of a particular PPR subgroup is denoted by '+' or '−', respectively. DYW-type PPR proteins were identified by Schallenberg-Rüdinger et al. [9] PPR-TGM proteins were identified by Manna and Barth [73]. *Arabidopsis thaliana* is representative of all plants.

Organism	DYW-type PPR proteins	PPR-TGM proteins
Amoebozoa		
Dictyostelium spp.	−	+
Polysphondylium pallidum	−	+
Physarum polycephalum	+	−
Acanthamoeba castellanii	+	−
Entamoeba spp.	−	+
Excavata		
Naegleria gruberi	+	−
Malawimonas jakobiformis	+	−
Alveolata		
Perkinsus marinus	−	+
Opisthokonta		
Adineta ricciae	+	−
Philodina roseola	+	−
Chlorophyta		
Ostreococcus spp.	−	+
Micromonas spp.	−	+
Bathycoccus prasinos	−	+
Cryptophyta		
Guillardia theta	−	+
Haptophyta		
Emiliania huxleyi	−	+
Streptophyta		
Nitella spp.	+	−
Embryophyta		
Physcomitrella patens	+	−
Arabidopsis thaliana	+	−
Stramenopila		
Thalassiosira oceanica	−	+
Phaeodactylum tricornutum	−	+

6. Conclusion

This review has provided an overview of the recent findings in the functional and evolutionary investigations into the PPR family. An emerging feature is the distinct presence and/or absence of specific PPR types in particular lineages. Horizontal gene transfer has likely played an important role in the unique phylogenetic distribution of different PPR subclasses. This includes the rare but nonetheless existence of PPR proteins in prokaryotes, the biological role for which remain to be investigated. It is also clear that non-plant PPR proteins are often overlooked and one major opportunity in non-plant PPR research is the characterisation of PPR-TGM proteins, particularly in pathogenic protists and their potential as putative drug targets. Like the PPR-TGM proteins, investigations into the use of synthetic biology to create designer PPR proteins has several applications in agriculture and medicine which remain to be explored. Future studies providing functional and evolutionary insights into the different PPR subfamilies and the purpose of their phylogenetic distributions will facilitate the understanding of one of the largest eukaryotic protein families identified to date.

Conflicts of interest

There is no conflict of interest.

Acknowledgements

I wish to thank Christian Barth, Ashley Franks, Sanja Aracic and Ashley Harman for their suggestions on the manuscript. This work was funded by La Trobe University and an Australian Postgraduate Award.

References

[1] I.D. Small, O. Rackham, A. Filipovska, Organelle transcriptomes: products of a deconstructed genome, Curr. Opin. Microbiol. 16 (2013) 652–658.
[2] S. Aubourg, N. Boudet, M. Kreis, A. Lecharny, In *Arabidopsis thaliana*, 1% of the genome codes for a novel protein family unique to plants, Plant Mol. Biol. 42 (2000) 603–613.
[3] I.D. Small, N. Peeters, The PPR motif – a TPR-related motif prevalent in plant organellar proteins, Trends Biochem. Sci. 25 (2000) 46–47.
[4] C. Schmitz-Linneweber, I. Small, Pentatricopeptide repeat proteins: a socket set for organelle gene expression, Trends Plant Sci. 13 (2008) 663–670.
[5] A. Barkan, I. Small, Pentatricopeptide repeat proteins in plants, Annu. Rev. Plant Biol. 65 (2014) 415–442.
[6] C. Cazalet, L. Gomez-Valero, C. Rusniok, M. Lomma, D. Dervins-Ravault, H.J. Newton, et al., Analysis of the *Legionella longbeachae* genome and transcriptome uncovers unique strategies to cause Legionnaires' disease, PLoS Genet. 6 (2010) e1000851.
[7] M. Choudhary, X. Zanhua, Y. Fu, S. Kaplan, Genome analyses of three strains of *Rhodobacter sphaeroides*: evidence of rapid evolution of chromosome II, J. Bacteriol. 189 (2007) 1914–1921.
[8] S.J. Hallam, N. Putnam, C.M. Preston, J.C. Detter, D. Rokhsar, P.M. Richardson, et al., Reverse methanogenesis: testing the hypothesis with environmental genomics, Science 305 (2004) 1457–1462.
[9] M. Schallenberg-Rüdinger, H. Lenz, M. Polsakiewicz, J.M. Gott, V. Knoop, A survey of PPR proteins identifies DYW domains like those of land plant RNA editing factors in diverse eukaryotes, RNA Biol. 10 (2013) 1549–1556.
[10] E. Delannoy, W.A. Stanley, C.S. Bond, I.D. Small, Pentatricopeptide repeat (PPR) proteins as sequence-specificity factors in post-transcriptional processes in organelles, Biochem. Soc. Trans. 35 (2007) 1643–1647.
[11] C. Lurin, C. Andres, S. Aubourg, M. Bellaoui, F. Bitton, C. Bruyère, et al., Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis, Plant Cell 16 (2004) 2089–2103.
[12] K. Kobayashi, M. Kawabata, K. Hisano, T. Kazama, K. Matsuoka, M. Sugita, et al., Identification and characterization of the RNA binding surface of the pentatricopeptide repeat protein, Nucl. Acids Res. 40 (2012) 2712–2723.
[13] K. Meierhoff, S. Felder, T. Nakamura, N. Bechtold, G. Schuster, HCF152, an *Arabidopsis* RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast *psbB–psbT–psbH–petB–petD* RNAs, Plant Cell 15 (2003) 1480–1495.
[14] S. Mili, S. Piñol-Roma, LRP130, a pentatricopeptide motif protein with a noncanonical RNA-binding domain, is bound *in vivo* to mitochondrial and nuclear RNAs, Mol. Cell. Biol. 23 (2003) 4972–4982.
[15] T. Nakamura, K. Meierhoff, P. Westhoff, G. Schuster, RNA-binding properties of HCF152, an *Arabidopsis* PPR protein involved in the processing of chloroplast RNA, Eur. J. Biochem. 270 (2003) 4070–4081.
[16] H. Ruwe, C. Schmitz-Linneweber, Short non-coding RNA fragments accumulating in chloroplasts: footprints of RNA binding proteins? Nucl. Acids Res. 40 (2011) 3106–3116.
[17] O. Rackham, A. Filipovska, The role of mammalian PPR domain proteins in the regulation of mitochondrial gene expression, Biochim. Biophys. Acta 1819 (2011) 1008–1016.
[18] R. Ringel, M. Sologub, Y.I. Morozov, D. Litonin, P. Cramer, D. Temiakov, Structure of human mitochondrial RNA polymerase, Nature 478 (2011) 269–273.
[19] M.J. Howard, W.H. Lim, C.A. Fierke, M. Koutmos, 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 (2012) 16149–16154.
[20] J. Ke, R.-Z. Chen, T. Ban, X.E. Zhou, X. Gu, M.E. Tan, et al., Structural basis for RNA recognition by a dimeric PPR–protein complex, Nat. Struct. Mol. Biol. 20 (2013) 1377–1382.
[21] P. Yin, Q. Li, C. Yan, Y. Liu, J. Liu, F. Yu, et al., Structural basis for the modular recognition of single-stranded RNA by PPR proteins, Nature 504 (2013) 168–171.
[22] A. Barkan, M. Rojas, S. Fujii, A. Yap, Y.S. Chong, C.S. Bond, I. Small, A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins, PLoS Genet. 8 (2012) e1002910.
[23] Y. Yagi, S. Hayashi, K. Kobayashi, T. Hirayama, T. Nakamura, Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants, PLoS ONE 8 (2013) e57286.
[24] S. Coquille, A. Filipovska, T. Chia, L. Rajappa, J.P. Lingford, M.F. Razif, et al., An artificial PPR scaffold for programmable RNA recognition, Nat. Commun. 5 (2014) 5279.
[25] I. Small, H. Puchta, Emerging tools for synthetic biology in plants, Plant J. 78 (2014) 725–726.
[26] Y. Yagi, T. Nakamura, I. Small, The potential for manipulating RNA with pentatricopeptide repeat proteins, Plant J. 78 (2014) 772–782.
[27] J. Colcombet, M. Lopez-Obando, L. Heurtevin, C. Bernard, K. Martin, R. Berthomé, Systematic study of subcellular localization of *Arabidopsis* PPR proteins confirms a massive targeting to organelles, RNA Biol. 10 (2013) 1557–1575.
[28] S. Manna, J. Brewster, C. Barth, Identification of pentatricopeptide repeat proteins in the model organism *Dictyostelium discoideum*, Int. J. Genom. 2013 (2013) 586498.
[29] M. Pusnik, I. Small, L.K. Read, T. Fabbro, A. Schneider, Pentatricopeptide repeat proteins in *Trypanosoma brucei* function in mitochondrial ribosomes? Mol. Cell. Biol. 27 (2007) 6876–6888.
[30] S. Bentolila, A.A. Alfonso, M.R. Hanson, A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 10887–10892.
[31] D.A. Cushing, N.R. Forsthoefel, D.R. Gestaut, D.M. Vernon, *Arabidopsis emb175* and other *ppr* knockout mutants reveal essential roles for pentatricopeptide repeat (PPR) proteins in plant embryogenesis, Planta 221 (2005) 424–436.
[32] T. Kocábek, J. Řepková, M. Dudová, K. Hoyerová, L. Vrba, Isolation and characterization of a novel semi-lethal *Arabidopsis thaliana* mutant of gene for pentatricopeptide (PPR) repeat-containing protein, Genetica 128 (2006) 395–407.
[33] N. Koizuka, R. Imai, H. Fujimoto, T. Hayakawa, Y. Kimura, J. Kohno-Murase, et al., Genetic characterization of a pentatricopeptide repeat protein gene, *orf687*, that restores fertility in the cytoplasmic male-sterile Kosena radish, Plant J. 34 (2003) 407–415.
[34] I. Tzafrir, R. Pena-Muralla, A. Dickerman, M. Berg, R. Rogers, S. Hutchens, et al., Identification of genes required for embryo development in *Arabidopsis*, Plant Physiol. 135 (2004) 1206–1220.
[35] Z. Wang, Y. Zou, X. Li, Q. Zhang, L. Chen, H. Wu, 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 (2006) 676–687.
[36] C.D. Chase, Cytoplasmic male sterility: a window to the world of plant mitochondrial–nuclear interactions, Trends Genet. 23 (2007) 81–90.
[37] J. Dahan, H. Mireau, The Rf and Rf-like PPR in higher plants, a fast-evolving subclass of PPR genes, RNA Biol. 10 (2013) 1469–1476.
[38] N. Bryant, J. Lloyd, C. Sweeney, F. Myouga, D. Meinke, Identification of nuclear genes encoding chloroplast-localized proteins required for embryo development in *Arabidopsis*, Plant Physiol. 155 (2011) 1678–1689.
[39] V.K. Mootha, P. Lepage, K. Miller, J. Bunkenborg, M. Reich, M. Hjerrild, et al., Identification of a gene causing human cytochrome c oxidase deficiency by integrative genomics, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 605–610.
[40] F. Sasarman, C. Brunel-Guitton, H. Antonicka, T. Wai, E.A. Shoubridge, LRPPRC and SLIRP interact in a ribonucleoprotein complex that regulates post-transcriptional gene expression in mitochondria, Mol. Biol. Cell 21 (2010) 1315–1323.
[41] B. Ruzzenente, M.D. Metodiev, A. Wredenberg, A. Bratic, C.B. Park, Y. Cámara, et al., LRPPRC is necessary for polyadenylation and coordination of translation of mitochondrial mRNAs, EMBO J. 31 (2012) 443–456.
[42] T. Tian, J. Ikeda, Y. Wang, S. Mamat, W. Luo, K. Aozasa, et al., Role of leucine-rich pentatricopeptide repeat motif-containing protein (LRPPRC) for anti-apoptosis and tumourigenesis in cancers, Eur. J. Cancer 48 (2012) 2462–2473.
[43] J. Harmel, B. Ruzzenente, M. Terzioglu, H. Spahr, M. Falkenberg, N.G. Larsson, The leucine-rich pentatricopeptide repeat containing (LRPPRC) protein does not activate transcription in mammalian mitochondria, J. Biol. Chem. 288 (2013) 15510–15519.
[44] L. Liu, M. Sanosaka, S. Lei, M.L. Bestwick, J.H. Frey, Y.V. Surovtseva, et al., LRP130 protein remodels mitochondria and stimulates fatty acid oxidation, J. Biol. Chem. 286 (2011) 41253–41264.
[45] M. Rüdinger, M. Polsakiewicz, V. Knoop, Organellar RNA editing and plant-specific extensions of pentatricopeptide repeat proteins in jungermanniid but not in marchantiid liverworts, Mol. Biol. Evol. 25 (2008) 1405–1414.
[46] V. Salone, M. Rüdinger, M. Polsakiewicz, B. Hoffmann, M. Groth-Malonek, B. Szurek, et al., A hypothesis on the identification of the editing enzyme in plant organelles, FEBS Lett. 581 (2007) 4132–4138.
[47] K. Okuda, F. Myouga, R. Motohashi, K. Shinozaki, T. Shikanai, Conserved domain structure of pentatricopeptide repeat proteins involved in chloroplast RNA editing, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 8178–8183.
[48] T. Shikanai, RNA editing in plant organelles: machinery, physiological function and evolution, Cell. Mol. Life Sci. 63 (2006) 698–708.
[49] N. O'Toole, M. Hattori, C. Andres, K. Iida, C. Lurin, C. Schmitz-Linneweber, et al., On the expansion of the pentatricopeptide repeat gene family in plants, Mol. Biol. Evol. 25 (2008) 1120–1128.
[50] C. Boussardon, A. Avon, P. Kindgren, C.S. Bond, M. Challenor, C. Lurin, I. Small, The cytidine deaminase signature HxE (x) nCxxC of DYW1 binds zinc and is necessary for RNA editing of *ndhD-1*, New Phytol. 203 (2014) 1090–1095.
[51] M.L. Hayes, K. Giang, B. Berhane, R.M. Mulligan, Identification of two pentatricopeptide repeat genes required for RNA editing and zinc binding by C-terminal cytidine deaminase-like domains, J. Biol. Chem. 288 (2013) 36519–36529.

[52] V. Knoop, M. Rüdinger, DYW-type PPR proteins in a heterolobosean protist: plant RNA editing factors involved in an ancient horizontal gene transfer? FEBS Lett. 584 (2010) 4287–4291.
[53] J. Holzmann, P. Frank, E. Löffler, K.L. Bennett, C. Gerner, W. Rossmanith, RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme, Cell 135 (2008) 462–474.
[54] F. Pinker, G. Bonnard, A. Gobert, B. Gutmann, K. Hammani, C. Sauter, et al., PPR proteins shed a new light on RNase P biology, RNA Biol. 10 (2013) 1457–1468.
[55] A. Gobert, B. Gutmann, A. Taschner, M. Gößringer, J. Holzmann, R.K. Hartmann, et al., A single *Arabidopsis* organellar protein has RNase P activity, Nat. Struct. Mol. Biol. 17 (2010) 740–744.
[56] W. Rossmanith, Of P and Z: mitochondrial tRNA processing enzymes, Biochim. Biophys. Acta 1819 (2012) 1017–1026.
[57] T. Imai, T. Nakamura, T. Maeda, K. Nakayama, X. Gao, T. Nakashima, et al., Pentatricopeptide repeat motifs in the processing enzyme PRORP1 in *Arabidopsis thaliana* play a crucial role in recognition of nucleotide bases at TψC loop in precursor tRNAs, Biochem. Biophys. Res. Commun. 450 (2014) 1541–1546.
[58] B. Gutmann, A. Gobert, P. Giegé, PRORP proteins support RNase P activity in both organelles and the nucleus in *Arabidopsis*, Genes Dev. 26 (2012) 1022–1027.
[59] A. Taschner, C. Weber, A. Buzet, R.K. Hartmann, A. Hartig, W. Rossmanith, Nuclear RNase P of *Trypanosoma brucei*: a single protein in place of the multicomponent RNA–protein complex, Cell Rep. 2 (2012) 19–21.
[60] C. Barth, U. Greferath, M. Kotsifas, Y. Tanaka, S. Alexander, H. Alexander, et al., Transcript mapping and processing of mitochondrial RNA in *Dictyostelium discoideum*, Curr. Genet. 39 (2001) 355–364.
[61] K.A. Lipinski, O. Puchta, V. Surendranath, M. Kudla, P. Golik, Revisiting the yeast PPR proteins – application of an iterative hidden Markov model algorithm reveals new members of the rapidly evolving family, Mol. Biol. Evol. 28 (2011) 2935–2948.
[62] I. Kühl, L. Dujeancourt, M. Gaisne, C.J. Herbert, N. Bonnefoy, A genome wide study in fission yeast reveals nine PPR proteins that regulate mitochondrial gene expression, Nucl. Acids Res. 39 (2011) 8029–8041.
[63] G.S. Shadel, Coupling the mitochondrial transcription machinery to human disease, Trends Genet. 20 (2004) 513–519.
[64] V. Posse, E. Hoberg, A. Dierckx, S. Shahzad, C. Koolmeister, N.-G. Larsson, et al., The amino terminal extension of mammalian mitochondrial RNA polymerase ensures promoter specific transcription initiation, Nucl. Acids Res. 42 (2014) 3638–3647.
[65] M. Belfort, R.J. Roberts, Homing endonucleases: keeping the house in order, Nucl. Acids Res. 25 (1997) 3379–3388.
[66] P. Lucas, C. Otis, J.-P. Mercier, M. Turmel, C. Lemieux, Rapid evolution of the DNA-binding site in LAGLIDADG homing endonucleases, Nucl. Acids Res. 29 (2001) 960–969.
[67] A.F. de Longevialle, A. Falcon, L. Hendrickson, N.L. Taylor, E. Delannoy, C. Lurin, et al., The pentatricopeptide repeat gene OTP51 with two LAGLIDADG motifs is required for the *cis*-splicing of plastid *ycf3* intron 2 in *Arabidopsis thaliana*, Plant J. 56 (2008) 157–168.
[68] S. Liu, J. Melonek, L.M. Boykin, I. Small, K.A. Howell, PPR-SMRs: ancient proteins with enigmatic functions, RNA Biol. 10 (2013) 80–89.
[69] R. Zoschke, T. Kroeger, S. Belcher, M.A. Schöttler, A. Barkan, C. Schmitz-Linneweber, The pentatricopeptide repeat-SMR protein ATP4 promotes translation of the chloroplast *atpB/E* mRNA, Plant J. 72 (2012) 547–558.
[70] R. Zoschke, Y. Qu, Y.O. Zubo, T. Börner, C. Schmitz-Linneweber, Mutation of the pentatricopeptide repeat-SMR protein SVR7 impairs accumulation and translation of chloroplast ATP synthase subunits in *Arabidopsis thaliana*, J. Plant Res. 126 (2013) 403–414.
[71] D. Moreira, H. Philippe, Smr: a bacterial and eukaryotic homologue of the C-terminal region of the MutS2 family, Trends Biochem. Sci. 24 (1999) 298–300.
[72] R. Narsai, S.R. Law, C. Carrie, L. Xu, J. Whelan, In-depth temporal transcriptome profiling reveals a crucial developmental switch with roles for RNA processing and organelle metabolism that are essential for germination in *Arabidopsis*, Plant Physiol. 157 (2011) 1342–1362.
[73] S. Manna, C. Barth, Identification of a novel pentatricopeptide repeat subfamily with a C-terminal domain of bacterial origin acquired via ancient horizontal gene transfer, BMC Res. Notes 6 (2013) 525.
[74] C. Pugh, R.T. Borchardt, H. Stone, Sinefungin, a potent inhibitor of virion mRNA (guanine-7-)-methyltransferase, mRNA (nucleoside-2′-)-methyltransferase, and viral multiplication, J. Biol. Chem. 253 (1978) 4075–4077.
[75] M. Vedel, F. Lawrence, M. Robert-Gero, E. Lederer, The antifungal antibiotic sinefungin as a very active inhibitor of methyltransferases and of the transformation of chick embryo fibroblasts by rous sarcoma virus, Biochem. Biophys. Res. Commun. 85 (1978) 371–376.
[76] M.I.G. Lopez Sanchez, T.R. Mercer, S.M.K. Davies, A.M.J. Shearwood, K.K.A. Nygård, T.R. Richman, et al., RNA processing in human mitochondria, Cell Cycle 10 (2011) 2904–2916.
[77] F. Xu, C. Ackerley, M. Maj, J. Addis, V. Levandovskiy, J. Lee, et al., Disruption of a mitochondrial RNA-binding protein gene results in decreased cytochrome b expression and a marked reduction in ubiquinol-cytochrome c reductase activity in mouse heart mitochondria, Biochem. J. 416 (2008) 15–26.
[78] A.L. Greenleaf, J.L. Kelly, I. Lehman, Yeast RPO41 gene product is required for transcription and maintenance of the mitochondrial genome, Proc. Natl. Acad. Sci. U. S. A. 83 (1986) 3391–3394.
[79] J.I. Moreno, K.S. Buie, R.E. Price, M.A. Piva, Ccm1p/Ygr150cp, a pentatricopeptide repeat protein, is essential to remove the fourth intron of both *COB* and *COX1* pre-mRNAs in *Saccharomyces cerevisiae*, Curr. Genet. 55 (2009) 475–484.
[80] G.M. Manthey, J.E. McEwen, The product of the nuclear gene PET309 is required for translation of mature mRNA and stability or production of intron-containing RNAs derived from the mitochondrial *COX1* locus of *Saccharomyces cerevisiae*, EMBO J. 14 (1995) 4031–4043.
[81] J.W. Coffin, R. Dhillon, R.G. Ritzel, F.E. Nargang, The *Neurospora crassa* cya-5 nuclear gene encodes a protein with a region of homology to the *Saccharomyces cerevisiae* PET309 protein and is required in a post-transcriptional step for the expression of the mitochondrially encoded COXI protein, Curr. Genet. 32 (1997) 273–280.
[82] I. Aphasizheva, D. Maslov, X. Wang, L. Huang, R. Aphasizhev, Pentatricopeptide repeat proteins stimulate mRNA adenylation/uridylation to activate mitochondrial translation in Trypanosomes, Mol. Cell 42 (2011) 106–117.
[83] C. Loiselay, N.J. Gumpel, J. Girard-Bascou, A.T. Watson, S. Purton, F.A. Wollman, et al., Molecular identification and function of *cis*- and *trans*-acting determinants for *petA* transcript stability in *Chlamydomonas reinhardtii* chloroplasts, Mol. Cell. Biol. 28 (2008) 5529–5542.
[84] E. Kotera, M. Tasaka, T. Shikanai, A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts, Nature 433 (2005) 326–330.
[85] J. Pfalz, K. Liere, A. Kandlbinder, K.-J. Dietz, R. Oelmüller, pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression, Plant Cell 18 (2006) 176–197.
[86] N. Haïli, N. Arnal, M. Quadrado, S. Amiar, G. Tcherkez, J. Dahan, et al., The pentatricopeptide repeat MTSF1 protein stabilizes the *nad4* mRNA in *Arabidopsis* mitochondria, Nucl. Acids Res. 41 (2013) 6650–6663.
[87] D.G. Fisk, M.B. Walker, A. Barkan, Molecular cloning of the maize gene *crp1* reveals similarity between regulators of mitochondrial and chloroplast gene expression, EMBO J. 18 (1999) 2621–2630.
[88] D. Sosso, S. Mbelo, V. Vernoud, G. Gendrot, A. Dedieu, P. Chambrier, et al., PPR2263, a DYW-subgroup pentatricopeptide repeat protein, is required for mitochondrial *nad5* and *cob* transcript editing, mitochondrion biogenesis, and maize growth, Plant Cell 24 (2012) 676–691.
[89] M. Uchida, S. Ohtani, M. Ichinose, C. Sugita, M. Sugita, The PPR-DYW proteins are required for RNA editing of *rps14*, *cox1* and *nad5* transcripts in *Physcomitrella patens* mitochondria, FEBS Lett. 585 (2011) 2367–2371.
[90] M. Hattori, H. Miyake, M. Sugita, A pentatricopeptide repeat protein is required for RNA processing of *clpP* pre-mRNA in moss chloroplasts, J. Biol. Chem. 282 (2007) 10773–10782.