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].

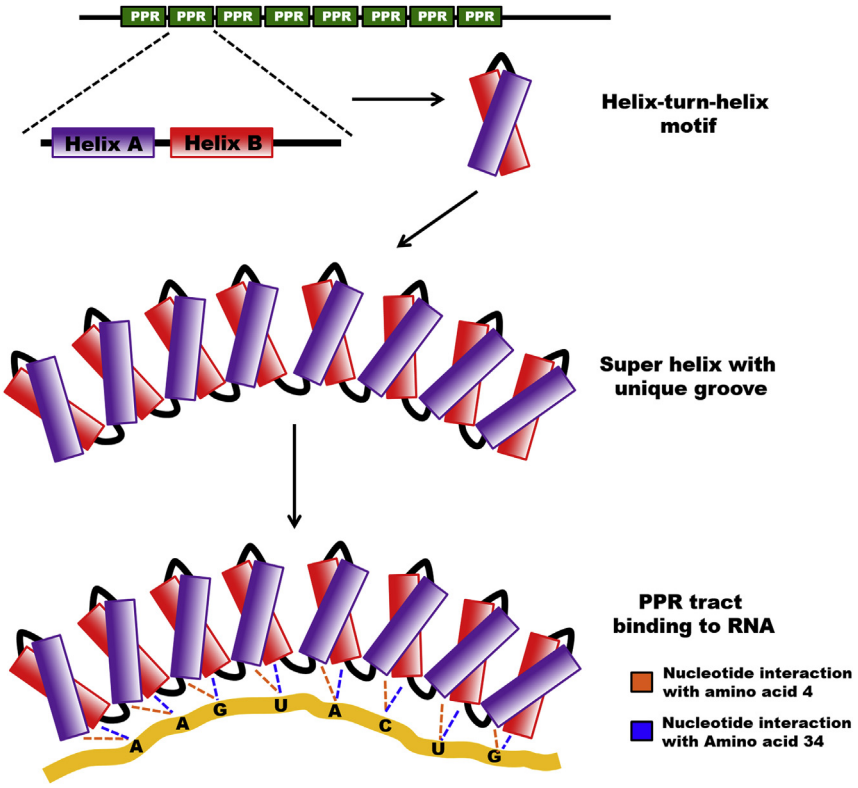

Fig. 1. PPR protein structure and the mechanism of transcript recognition. Each PPR motif forms two α-helices, which interact to form a helix turn helix motif. The series of helix turn helix motifs throughout the protein are stacked together to form a superhelix with an RNA binding groove. Modular recognition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