Pentatricopeptide Repeat Proteins in Plants

Alice Barkan[1] and Ian Small[2]

[1]Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97405; email: abarkan@uoregon.edu

[2]Australian Research Council Centre of Excellence in Plant Energy Biology, University of Western Australia, Crawley 6009, Australia; email: ian.small@uwa.edu.au

Annu. Rev. Plant Biol. 2014. 65:415–42

First published online as a Review in Advance on January 27, 2014

The *Annual Review of Plant Biology* is online at plant.annualreviews.org

This article's doi: 10.1146/annurev-arplant-050213-040159

Keywords

chloroplast, mitochondria, RNA-binding protein, alpha solenoid, helical repeat protein, posttranscriptional control

Abstract

Pentatricopeptide repeat (PPR) proteins constitute one of the largest protein families in land plants, with more than 400 members in most species. Over the past decade, much has been learned about the molecular functions of these proteins, where they act in the cell, and what physiological roles they play during plant growth and development. A typical PPR protein is targeted to mitochondria or chloroplasts, binds one or several organellar transcripts, and influences their expression by altering RNA sequence, turnover, processing, or translation. Their combined action has profound effects on organelle biogenesis and function and, consequently, on photosynthesis, respiration, plant development, and environmental responses. Recent breakthroughs in understanding how PPR proteins recognize RNA sequences through modular base-specific contacts will help match proteins to potential binding sites and provide a pathway toward designing synthetic RNA-binding proteins aimed at desired targets.

Contents

INTRODUCTION

It has been 14 years since the pentatricopeptide repeat (PPR) motif was first recognized to be a distinct class of helical repeat motif defining a family of remarkable size in plants (3, 145). At that time, a smattering of genetic data linking PPR proteins to organelle RNA metabolism provided the only solid information concerning the functions of the PPR motif and the proteins that harbor it. These data, in conjunction with predictions that the majority of PPR proteins localize to organelles and a structural model suggesting that PPR tracts bind extended acidic ligands, led to the hypotheses that (*a*) PPR tracts bind RNA and (*b*) the expansion of the PPR family in plants had an evolutionary link with the complex RNA metabolism characteristic of plant organelles (145). In the context of the time, these were bold suggestions, as the known helical repeat motifs, including the closely related tetratricopeptide repeat (TPR), were believed to form only protein-binding surfaces (94). These hypotheses have since been borne out by a wealth of data, with far-reaching implications for mechanisms of genome evolution and organellar gene expression, and with unforeseen applications in agriculture and biotechnology.

PPR: pentatricopeptide repeat

Today, PPR proteins are recognized to be members of the alpha-solenoid superfamily, which comprises proteins with short helical repeats that stack to form extended surfaces (94). Most alpha-solenoid proteins bind protein ligands, but examples that bind nucleic acids are increasing (46, 132). The best-characterized examples—the DNA-binding transcription activator–like (TAL) motifs and the RNA-binding PUF and PPR motifs—share a key property: modular one-repeat:one-nucleotide binding, whereby base specificity is dictated by amino acids at a few specific positions in the repeating units. This feature offers unprecedented promise for the engineering of nucleic acid–binding proteins with desired sequence specificities. As discussed below, the PPR scaffold may be particularly amenable for the engineering of designer RNA-binding domains.

PPR proteins are found in all eucaryotes and function universally in organellar gene expression, but the PPR family in plants is notable for its enormous size. With more than 400 members in land plants (108), the expansion of the PPR family has had a major impact on the evolutionary trajectory of plant organellar genomes. The essential physiological functions of organellar genomes have, in turn, provided selective pressure to diversify the plant PPR family. We now know that the expansion of the PPR family in plants is closely tied to the evolution of the complex RNA metabolism that is characteristic of plant organelles (12, 53). The unusual features of PPR proteins and the genes that encode them also underpin the frequent emergence of cytoplasmic male sterility (CMS) and cognate nuclear restorer loci in plants (24, 36).

In this review, we summarize the major themes that have emerged from the roughly 250 papers published on PPR proteins to date. We focus on the PPR family in plants, where the deepest insights into mechanisms and evolution have been obtained and where the impact of the PPR family on genomes, gene expression, and physiology are particularly profound. We aim to highlight the enormous progress that has been made as well as the many fundamental questions that remain to be answered.

EVOLUTION OF THE PPR FAMILY

The PPR Motif Likely Evolved from a TPR Progenitor Early in the Evolution of Eucaryotes

The PPR motif was originally defined (145) as a 35-amino-acid repeat (hence the name pentatricopeptide) predicted to fold into a pair of antiparallel alpha helices based on faint but compelling sequence similarity to the TPR motif. These predictions have recently been confirmed by crystal structures of PPR repeats in several contexts (5, 73, 85, 130, 172) (**Figure 1*a***). Arrays of tandem PPR motifs form a superhelical ribbon-like sheet (5, 49, 73, 85, 172) (**Figure 1*b*,*c***). TPR motifs have a broader phylogenetic distribution (present throughout procaryotes and eucaryotes) and are embedded into a greater diversity of protein architectures than are PPR motifs, suggesting that the PPR motif arose through divergence of a TPR motif rather than vice versa. The two motifs are generally easy to distinguish [e.g., by tools such as TPRpred (82)] based on characteristic differences in the conserved residues in each motif (145).

Based on the distribution of PPR proteins in extant organisms, it seems likely that they arose very early in eucaryotic evolution or in the immediate ancestors of the first eucaryotic cell. Essentially all PPR proteins are predicted or established to localize to mitochondria or chloroplasts (108). They are ubiquitous in eucaryotes with a mitochondrial genome but absent from most procaryotes (110). The few bacteria that contain PPR proteins are all pathogens or symbionts of eucaryotes (23, 140), suggesting that they may have acquired PPR sequences via horizontal transfer from their eucaryotic hosts.

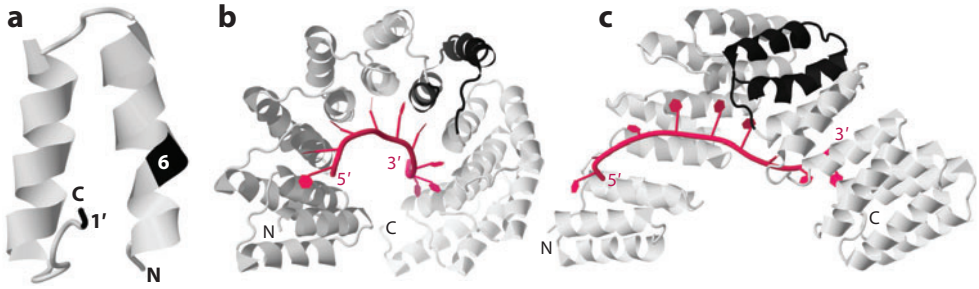

Figure 1

Structure of a PPR motif and model of a PPR tract bound to RNA. Each repeat consists of a pair of antiparallel helices connected by a short loop, with the N-terminal helix of each motif forming the RNA-binding face and the C-terminal helix forming the external surface of the protein. (*a*) A single PPR motif from the *Arabidopsis* organellar RNase P (PRORP1) (73). Residues 210–245 from structure 4G23 are shown, with the 6 and 1′ positions that have been proposed to determine the nucleotide-binding specificity of PPR motifs (6, 152, 169) shaded in black. (*b*) A model of 10 consensus PPR repeats (*gray*) bound to 9 nucleotides of poly(U) RNA (*magenta*), oriented to show that the bases are predicted to interdigitate with the PPR helices. The sixth PPR motif is shaded in black to highlight a single repeat. The model was produced as described in Reference 49 and visualized with Jmol (**http://www.jmol.org**). A recent crystal structure of a PPR–RNA complex confirmed the general features of this model (172). (*c*) The same model viewed from the side, to show that the bases are predicted to interact with the edges of the solenoid array (loop/beginning of helix).

The PPR Family in Land Plants

Most eucaryotes contain a small multigene family encoding ~5–30 PPR proteins, but in the land-plant lineage the family has greatly expanded (118). This expansion probably occurred early in the colonization of land, as all land-plant genomes that have been examined contain many more PPR genes (from 400 to more than 1,000) than any algal genome sequenced to date. The PPR family in plants consists of two major subfamilies, denoted P and PLS (108) (**Figure 2**). Considering the canonical PPR motif as the archetypal P motif, P-class proteins contain arrays of only P motifs; PLS-class proteins, originally termed plant combinatorial modular proteins (3), contain characteristic triplets of P, L, and S motifs, where L motifs are 35–36 amino acids and S motifs are 31 amino acids, each with specific patterns of amino acid conservation that distinguish them (108, 131) (**Figure 2**).

Most P-class proteins consist of one or more PPR tracts, an N-terminal organelle-targeting sequence, and little else. However, additional domains are appended in some cases. Of particular interest are eight proteins conserved across angiosperms that harbor an array of P motifs followed by a small MutS-related (SMR) domain (101) (**Figure 2**). In other proteins, this domain confers RNA or DNA endonuclease activity (51), but its role in these PPR proteins is unknown.

PLS proteins almost always possess C-terminal domains denoted E or DYW (**Figure 2*a***), which are unique to this class of proteins (108). These domains are implicated in RNA editing in plant organelles (26, 119, 120), a deamination reaction converting specific cytidines to uridine (30). The DYW domain is related to nucleotide deaminase domains in other proteins (77) and contains a conserved signature similar to the active-site signature of known cytidine deaminases (138). The expansion of the PLS clade paralleled an increase in RNA editing in organelles, and the diversity of PLS-class PPR proteins in each species correlates with the number of organellar editing sites (50, 136). The driving forces behind these spectacular changes in gene number (from zero in *Marchantia* to more than 1,000 in *Selaginella*) (reviewed in 49) are unclear. One proposal ties this to increased mutation pressure on organellar genomes following the transition to dry land, which would have brought higher light and UV levels and coincided with a significant change in

RNA editing: posttranscriptional change of specific cytidine residues in RNA to uridine

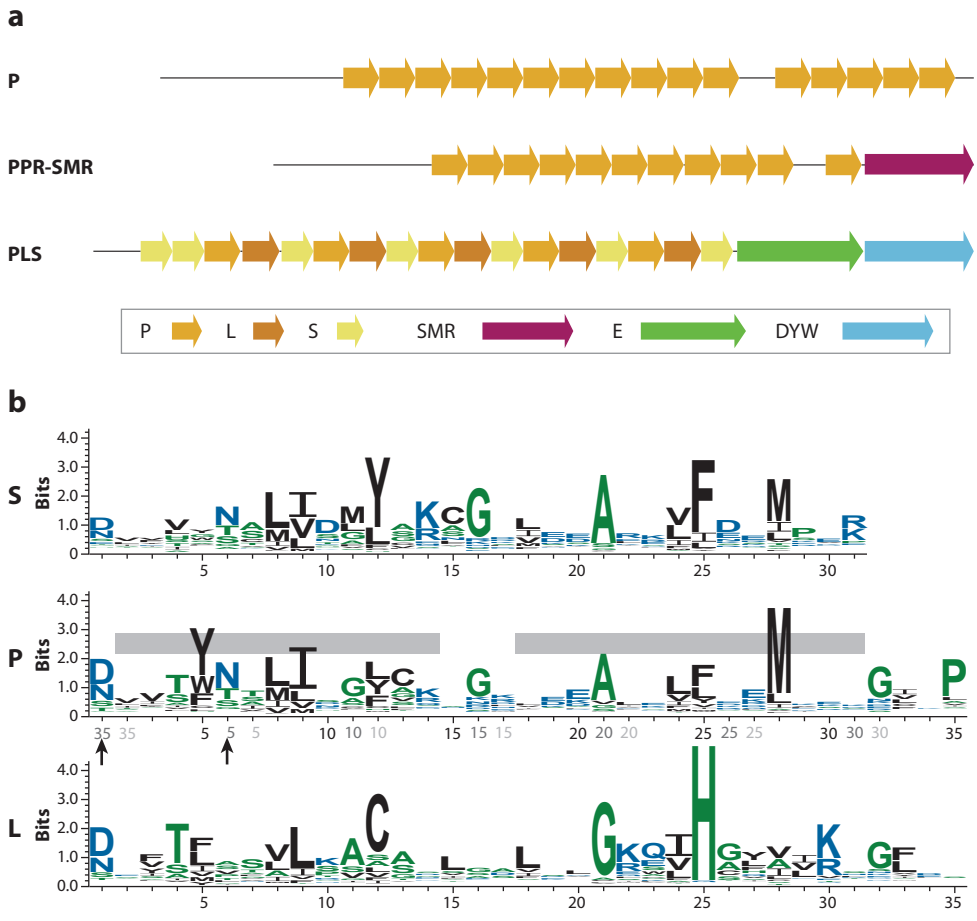

Figure 2

Differences between P- and PLS-class proteins and between P, L, and S motifs. (*a*) The P and PLS subfamilies. P-class proteins consist of one or more tandem arrays of P motifs, whereas PLS-class proteins consist of triplets of P, L, and S motifs, occasionally interspersed with additional S motifs. A few P-class proteins contain additional motifs, such as the PPR–small MutS-related (SMR) example shown here. Almost all PLS-class proteins contain C-terminal E motifs, and many contain, in addition, a DYW motif. All motifs are drawn approximately to scale. (*b*) Sequence logos for P, L, and S motifs. P and S motifs differ primarily at the C terminus, where the S motif is four residues shorter. L motifs differ from P and S motifs throughout the central loop and second helix. The amino acid positions in the P logo are numbered according to three alternative schemes: The black numbers are as in References 6 and 108, the dark gray numbers describe the scheme proposed in Reference 172 based on structural data, and the light gray numbers show the scheme used by Pfam (PF01535). The two positions shown to be the primary determinants of RNA-binding specificity are indicated by arrows (6, 152, 169). The boundaries of the two alpha helices in the P consensus are indicated by gray shading. The secondary structure formed by L and S motifs has not been experimentally validated; although a recent crystal structure was reported to be that of a PLS protein (5), that protein is more accurately viewed as a P-type protein. Sequence logos were constructed with WebLogo 3 (**http://weblogo.berkeley.edu**).

the base composition of the DNA (50). We discuss various explanations for the retention of such huge numbers of PPR genes at the end of this review.

It was long thought that PLS-class proteins were specific to land plants. It is becoming apparent, however, that they are more widely distributed than originally thought. The protist *Naegleria* contains a handful of PLS-class proteins (91) that correlate with the presence of RNA editing in this organism (134). Small numbers of PLS proteins have since been discovered in other protists (and, more surprisingly, in rotifers), but these are likely due to horizontal transfer from plant sequences (140). However, the discovery of DYW-containing PLS proteins in the alga *Nitella* may lead to an understanding of when and where the plant proteins arose (140). It will be interesting to see whether in each case these PLS proteins are associated with RNA-editing events.

Coevolution of PPR Proteins with Organellar Genomes

In general, PPR gene content is similar across angiosperms; for example, more than 80% of rice and *Arabidopsis* proteins form orthologous pairs in phylogenetic trees (118). This implies conservation of both protein activity and target sequence in the organelle genome. Evidence in favor includes editing factors shown to edit heterologous transcripts (14, 33, 34), even sites that no longer exist in their own organelle transcriptome (33, 81, 158). P-class proteins and their target sites are at least as well conserved; for example, the maize MPPR6 protein can complement loss-of-function *Arabidopsis* mutants lacking the orthologous protein (111). For such long-term conservation of function to be possible, the target sequences must be highly constrained (69, 70, 137, 177).

One particular clade of PPR proteins merits a particular mention in a coevolution context. The *restorer-to-fertility-like* (*RFL*) genes (36) form a small subclade of P-class genes that are entirely typical in terms of their motif structure and sequence but entirely atypical in terms of their evolutionary behavior (49). *RFL* genes are highly diverse among species and even among different strains of the same species, showing strong signals of diversifying selection (49, 52). They are generally present in chromosomal clusters (52) and show evolutionary features reminiscent of *R* genes implicated in pathogen resistance (36, 71). At least part of the explanation for their diversity is their involvement in suppressing CMS, a mitochondrially encoded trait that prevents the plant from making fertile pollen (24). The coevolution of fertility-restoring PPR genes with CMS-inducing mitochondrial genes has been described as an arms race between the mitochondrial and nuclear genomes, with the *RFL* genes being the weapons employed by the nuclear genome (159).

EFFECTS OF PPR PROTEINS ON PLANT PHYSIOLOGY

Despite the enormous size of the PPR family, there appears to be little redundancy between different family members, as reflected by the high proportion of mutants with distinct molecular and physiological phenotypes. PPR mutants have been discovered in classical genetic screens searching for many varied phenotypes: photosynthetic defects (7, 20, 47, 65, 78, 86, 98, 112, 119, 120, 171), aberrant leaf development (123), changes in leaf pigmentation (28, 102, 160, 173, 178), restricted growth (149, 174, 179), defective seed or embryo development (19, 38, 58, 87, 104, 111, 146, 147, 174), hypersensitivity to abiotic stress (183), hypersensitivity to abscisic acid (103, 115), tolerance to inhibitors of different biosynthetic pathways (92, 97), and restoration of pollen fertility (1, 9, 18, 37, 74, 84, 88, 95, 96, 166). As we shall see, these diverse phenotypes hide the simple truth that most PPR proteins have similar molecular functions, in that they serve as sequence-specific RNA-binding proteins inside organelles.

Many of the most frequent phenotypes associated with mutations in PPR genes (e.g., seedling lethality or slow growth) result from deficits in energy supply. These arise from a defect either in photosynthesis or in oxidative phosphorylation, correlated with whether the protein acts in chloroplasts or mitochondria, respectively. Pleiotropic effects of energy deficits on plant metabolism and development include alterations to signaling pathways involving sugars and abscisic acid (35, 103) and can explain many of the phenotypes listed above. Although some PPR proteins have been reported to be present outside organelles (38, 63), no physiological phenotypes clearly associated with an extraorganellar function have been demonstrated.

Defects in fertility, embryo, and seed development are also commonly associated with mutations in PPR-encoding genes (19, 42, 108, 111, 146, 147). For mitochondria-localized PPR proteins, these effects likely result from the disruption of respiration. However, photosynthesis per se is not required for seed development, as illustrated by the numerous nonphotosynthetic mutants that germinate and survive as seedlings until seed reserves are exhausted. Therefore, embryo-lethal phenotypes caused by plastid defects must have another cause. These are generally associated with disruption of the plastid translation system, thus preventing expression of factors beyond those required solely for photosynthesis (8, 19, 86, 106, 143, 167). There are interesting differences in the phenotypes conditioned by such mutations among different species: In *Arabidopsis*, mutations leading to loss of plastid translation are lethal early in embryo development (19), whereas in cereals such mutations are generally lethal only at the seedling stage (e.g., 8, 86, 106, 143, 167). Even in maize, however, the stage at which lethality occurs varies depending on the genetic background (176). Possible mechanisms underlying these differences have been discussed elsewhere (19, 176).

Almost all phenotypes exhibited by PPR mutants can be explained as resulting from the loss of one or several mitochondrial or chloroplast gene products. A select few, however, cannot easily be explained in this way. For example, *genomes uncoupled 1* (*gun1*) is disrupted in a retrograde signaling pathway that links the expression of specific nuclear genes to various aspects of chloroplast development (35, 99, 133). A possibly analogous defect is shown by the mutant *flavodentata*, also known as *defectively organized tributaries 4* (123), which shows defects in leaf morphology owing to aberrant expression of transcription factors that determine cell fate in early leaf development (154). We favor the hypothesis that these two PPR proteins have molecular functions similar to those of the others, but that their targets within the plastid must encode gene products that help provide a retrograde signal (157). These are particularly interesting targets for future study.

MOLECULAR FUNCTIONS OF PPR PROTEINS

An intriguing feature of the PPR family is that it comprises a relatively small range of protein architectures yet exhibits a diverse repertoire of molecular functions. Genetic data have implicated PPR proteins in every step of organellar gene expression: transcription, RNA stabilization, RNA cleavage, RNA splicing, RNA editing, and translation (reviewed in 12, 36, 144). Because it can be challenging to distinguish direct from indirect effects of a mutation, evidence for a physical association between a PPR protein and a particular RNA has been instrumental in demonstrating direct molecular functions. Some PPR–RNA partnerships have been validated by coimmunoprecipitation from organelle extracts (8, 16, 54, 86, 124, 142, 143, 161, 180) and some with in vitro binding assays involving recombinant proteins (6, 21, 60, 62, 67, 83, 121, 122, 124, 168). The recent elucidation of an amino acid code by which PPR tracts recognize specific nucleotides offers a computational approach for evaluating the likelihood that a PPR protein binds directly to a particular RNA (6, 68, 152, 169).

ORF: open reading frame

Some readouts of PPR function are relatively easy to spot, such as changes in transcripts encoding proteins that are missing in a particular mutant. Others are more elusive, such as changes in the translation of open reading frames (ORFs) within polycistronic mRNAs. Genome-wide assays facilitate the discovery of the less obvious (but not necessarily less important) functions. The current tool kit includes RNA immunoprecipitation-on-microarray assays (RIP-chip) to detect native PPR–RNA interactions (142), quantitative real-time reverse-transcription polymerase chain reaction (qRT-PCR) analysis of organelle transcriptomes (41, 97), circular RT-PCR (cRT-PCR) profiling of mitochondrial mRNA termini (48), large-scale RNA-editing assays (11, 29, 150), and a ribosome-profiling method that provides a genome-wide readout of chloroplast RNA abundance and ribosome occupancy (182). This last method revealed additional functions for several PPR proteins that had already been characterized in considerable depth (182), suggesting that deeper analysis of other characterized PPR proteins will often reveal functions beyond those that have been reported.

P-Type PPR Proteins Can Stabilize Specific RNAs and Position Processed RNA Termini

A major function of P-type PPR proteins in chloroplasts is the stabilization of specific RNAs. PPR proteins can stabilize 5′ or 3′ termini at the ends of transcription units as well as termini that result from processing between ORFs in polycistronic transcripts. All of these effects reflect the same mechanism: A long PPR tract bound to a high-affinity RNA ligand serves as an effective barrier to exoribonucleases (**Figure 3*a***). The bound protein determines the position of the processed ends while also protecting the adjacent RNA from destruction.

Elucidation of this simple mechanism occurred in steps that spanned more than a decade. Early genetic data showed that maize CRP1 and *Arabidopsis* HCF152 are required for the accumulation of processed RNAs with 5′ or 3′ termini in the *petB-petD* and *psbH-petB* intergenic regions, respectively (7, 47, 112), but it was initially unclear whether they promoted RNA processing or RNA stabilization. Elegant genetic experiments in *Chlamydomonas* later showed that the PPR protein MCA1 protects the chloroplast *petA* mRNA from degradation by a 5′→3′ exonuclease (105). Subsequent analysis of the maize protein PPR10 revealed that the "intercistronic processing" activities of proteins like CRP1 and HCF152 and the 5′ protective function of proteins like MCA1 are fundamentally the same (124, 127). PPR10 is required for the accumulation of processed RNAs with a 5′ or 3′ terminus in the *atpI-atpH* or *psaJ-rpl33* intercistronic region. Mapping the PPR10 binding sites and PPR10-dependent RNA termini to high resolution revealed the mechanism: PPR10 binds similar sequences in each region and blocks exoribonucleases intruding from either the 5′ or 3′ direction (124, 127). This results in a characteristic overlap between processed RNAs with 5′ or 3′ termini in each intercistronic region, with the shared sequences harboring the PPR-binding site (**Figure 3*a***). Clarification of the positions of the HCF152-dependent RNA termini and HCF152 binding site (124, 177) showed that the same mechanism promotes the accumulation of processed RNAs with termini in the *psbH-petB* intercistronic region. Processed RNA termini mapping to many other intercistronic regions in chloroplasts have an analogous overlapping arrangement (137, 177), suggesting that they are typically the result of stabilization by a PPR roadblock rather than the product of site-specific endonucleolytic cleavage.

Plant mitochondria contain an abundance of P-type PPR proteins with architectures similar to those that stabilize chloroplast RNAs (108). Interestingly, most such proteins studied thus far promote RNA cleavage rather than RNA stabilization (see below). An exception is MTSF1 in *Arabidopsis*, which stabilizes the 3′ end of the mitochondrial *nad4* mRNA (60). It will be interesting to discover whether this is a typical mechanism for stabilizing the 3′ termini of RNAs in plant mitochondria, and whether 5′ termini can be stabilized in a similar manner.

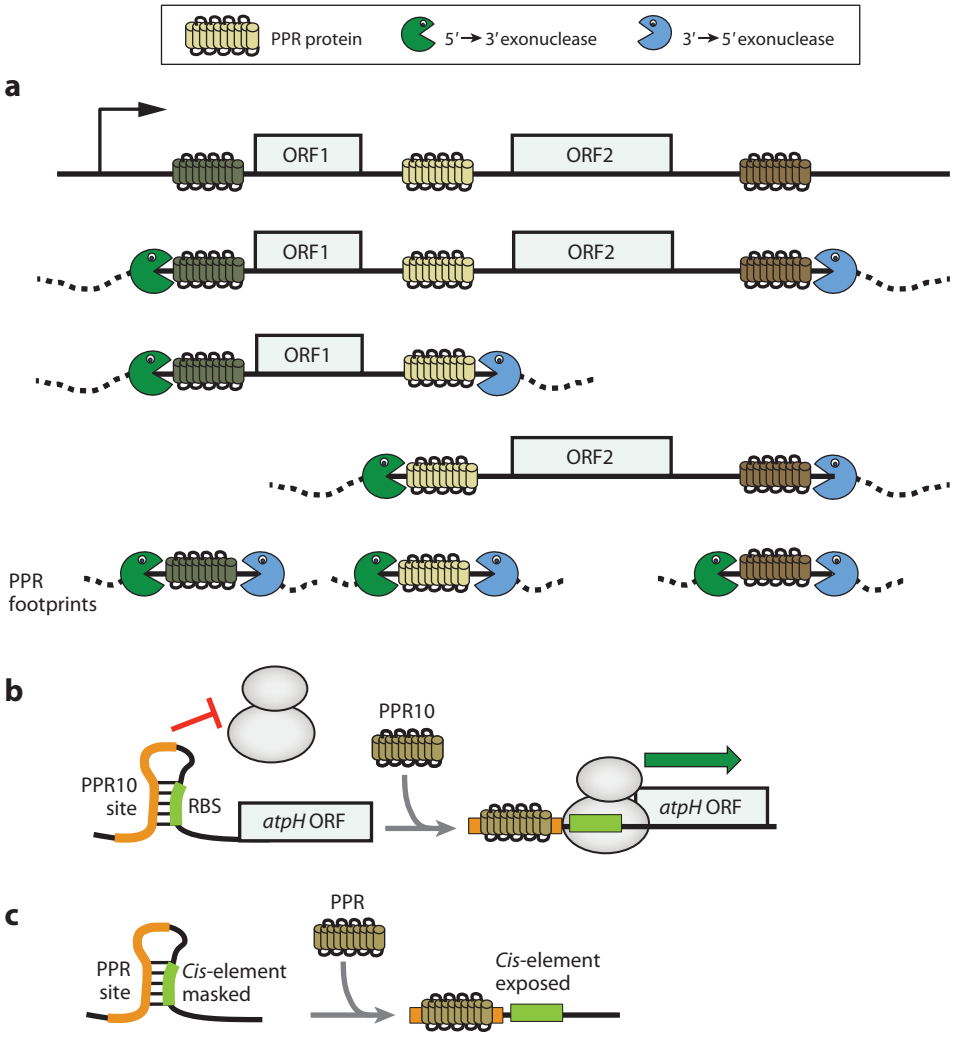

Figure 3

Sequestration of a long RNA segment can account for many activities reported for P-type PPR proteins. P-type PPR tracts have the unusual ability to bind with high affinity to long RNA segments, thereby preventing the bound RNA from interacting with other proteins or RNA. These interactions can affect gene expression in various ways depending on their position with respect to *cis*-elements and open reading frames (ORFs). This mechanism mimics that of many small RNAs in bacteria. (*a*) PPR proteins can stabilize RNAs and define the positions of processed RNA termini by blocking exoribonucleases. Examples of PPR proteins with this activity in chloroplasts include PGR3 and MCA1 (5′ stabilization) (21, 105, 171), PPR10 and HCF152 (intercistronic stabilization) (112, 124, 127), and unidentified PPR proteins that protect 3′ ends of *petD* and *rps16* RNAs (137, 177). MTSF1 stabilizes the 3′ end of a mitochondrial mRNA in this manner (60), but whether this is a general mechanism for RNA stabilization in plant mitochondria is not known. "PPR footprints" accumulate as metastable degradation intermediates when exonucleases converge on a tightly bound PPR protein from both directions (124, 137, 177). Different PPR proteins are shown in different colors. (*b*) Mechanism of translational activation by PPR10. PPR10 prevents the formation of an RNA structure that masks the *atpH* ribosome-binding site (RBS) (124, 127). Binding of PPR10 to this site also stabilizes *atpH* RNA by blocking 5′→3′ degradation. (*c*) RNA remodeling activity of a generic PPR protein could potentially expose (or mask) *cis*-elements for translation, RNA splicing, RNA cleavage, or RNA editing.

PPR-LIKE PROTEINS WITH DISTINCT REPEAT ARCHITECTURES

Although this discussion focuses on PPR proteins, it is important to recognize that several other classes of helical repeat protein play similar roles in plant and algal organelles. The HAT repeat (126) is a variant TPR motif found in several proteins that stabilize and/or activate the translation of specific chloroplast RNAs (15, 45, 139, 163). Mechanisms underlying the effects of the chloroplast HAT protein HCF107 are very similar to those of certain PPR proteins (61), as discussed in the main text. The octatricopeptide repeat (OPR) has been found in proteins involved in the translation or splicing of specific chloroplast mRNAs in *Chlamydomonas* (4, 40, 113, 128). Whereas PPR proteins are few in number in *Chlamydomonas*, the OPR family is quite large (40, 128). Little information is available about specific features of OPR–RNA interactions, but the strong parallels between the architectures of PPR and OPR proteins and the functional niches they have acquired suggest similar modes of action. The mTERF family is another family of helical repeat proteins that bind nucleic acids and function predominantly in organelles (reviewed in 89, 132). This family is particularly large in plants (in comparison to its size in metazoa), and its members may also have mechanistic parallels to PPR proteins.

PPR Footprints Reveal a Global Role for PPR-Mediated RNA Stabilization in Land-Plant Chloroplasts

Supplemental Material

Seven PPR proteins have been shown to stabilize specific mRNA termini in land-plant chloroplasts (7, 21, 47, 66, 67, 78, 112, 124, 127, 171, 180, 182) (**Supplemental Table 1**; follow the **Supplemental Material link** from the Annual Reviews home page at **http://www.annualreviews.org**). However, there is evidence that the impact of this mechanism goes far beyond these examples. The roadblock mechanism predicts that exonucleases should sometimes converge on a bound PPR protein from both directions, producing a small RNA corresponding to the protein's "footprint" (**Figure 3*a***). In fact, the binding sites of PPR10 and HCF152 are represented in small RNA data sets (124), and the boundaries of these small RNAs match the RNA termini that are protected by these proteins in vivo (**Figure 3*a***). Mining of transcriptome data from barley, rice, and *Arabidopsis* identified roughly 40 small chloroplast RNAs with hallmarks of PPR footprints (137, 177); these map to the 5′-terminal ~20 nucleotides of almost all processed mRNAs and to many 3′ termini as well, implying that these termini are protected by PPR or PPR-like proteins (see sidebar, PPR-Like Proteins with Distinct Repeat Architectures). Most of these are orphan footprints, whose protective protein is not known. Nonetheless, these results provide strong evidence that ancestral RNA-based stabilization mechanisms such as 3′ RNA hairpins are, over time, being replaced by protective PPR (or PPR-like) proteins. Furthermore, these findings imply a major role for the PPR family in shaping the complex transcript populations that are characteristic of land-plant chloroplasts (53).

PPR footprint: a small RNA segment that is protected by a bound PPR protein from ribonucleases in vivo

P-Type PPR Proteins Can Activate and Repress the Translation of Specific mRNAs

Four PPR proteins have been shown to stimulate the translation of specific chloroplast ORFs, each of which also stabilizes specific RNAs. PPR10 and PGR3 activate the translation of the same RNAs they stabilize (*atpH* and *petL*, respectively) (21, 124, 127, 171, 182), ATP4 enhances *atpB* translation but stabilizes *atpF* and *rpl14* mRNAs (180, 182), and CRP1 has dual effects at *petD* but serves solely as a translational activator at *psaC* and *petA* (7, 182). All of these proteins have been shown to interact with the 5′ untranslated region (UTR) of the ORFs whose translation they activate (21, 124, 142, 180). Quantitative profiling of chloroplast ribosome positions

in *ppr10* and *crp1* mutants (182) showed that translational activation contributes more than RNA stabilization to enhancement of target gene expression by PPR10 and CRP1. Translational activation may be a much more common feature of PPR function than is currently appreciated: The paucity of examples may simply reflect the technical challenge of detecting translation-level effects.

Study of *atpH* translation and its activation by PPR10 provided insight into potential mechanisms of PPR-mediated translational activation. The notion that the monocistronic *atpH* RNA isoform that is stabilized by PPR10 is intrinsically "more translatable" than its precursors is not supported by the results of in vitro translation experiments (175). Instead, current data (127) support the following view (**Figure 3*b***): In the absence of PPR10, an RNA hairpin occludes the *atpH* ribosome-binding site. PPR10 binding prevents the formation of this inhibitory structure by occupying the 5′ segment of the hairpin. PPR10 bound to this site also blocks a 5′→3′ exoribonuclease, thereby stabilizing the *atpH* ORF. Thus, the translational enhancement and RNA stabilization effects of PPR10 are two sides of the same coin: Both result from the sequestration of the same RNA segment by the long PPR tract (**Figure 3*b***).

There are hints that PPR proteins may stimulate the translation of many other chloroplast ORFs in a similar way. First, the PPR-like half-a-tetratricopeptide (HAT) repeat protein HCF107 employs analogous mechanisms to stabilize the *psbH* ORF and activate *psbH* translation (45, 61, 139). Furthermore, many PPR footprints map very near to (but do not overlap) the anticipated footprint of an initiating ribosome at a start codon (137, 177). This placement of a PPR tract may help to maintain a structure-free ribosome "landing pad" by occupying adjacent RNA that might otherwise pair with the translation initiation region. However, this simple mechanism may not account for all instances of PPR-mediated translational enhancement. For example, RNA coimmunoprecipitation data place the ATP4 binding site hundreds of nucleotides upstream of the *atpB* ORF, whose translation it stimulates (181).

Less is known about the effects of PPR proteins on translation in plant mitochondria. Two PPR proteins in *Arabidopsis* mitochondria have been shown to associate with polysomes (63, 162), but their functions are unknown. These two proteins belong to a large class of relatively short P-class proteins that are abundant in the mitochondrial proteome (90, 156) but apparently absent from plastids, implying mitochondrial-specific functions. Among the more typical PPR proteins, maize MPPR6 has been proposed to stimulate the translation and 5′ processing of the mitochondrial *rps3* mRNA (111), whereas a repressive effect on the translation of sterility-inducing ORFs has been demonstrated for several Rf-type PPR proteins (reviewed in 36). For example, PPR-B from the radish *Rfo* locus (18, 37, 95, 161) prevents the expression of ORF138 without influencing the size or abundance of ORF138 mRNA (100, 161). Rice Rf1a prevents the translation of ORF79 and also triggers the intercistronic cleavage of the *atp6-orf79* mRNA (83, 166), but the mechanistic relationship between these effects is not known. In light of the potent blockade effects of proteins such as PPR10, it seems plausible that some Rf proteins inhibit the translation of CMS ORFs via steric interference with ribosome binding or movement.

Stimulation of RNA Cleavage by P-Type PPR Proteins in Plant Mitochondria

It has been known for some time that several PPR-encoding *Rf* alleles induce the cleavage of sterility-associated mitochondrial RNAs (reviewed in 36). An innovative forward genetic strategy (48, 72) recently demonstrated widespread effects of Rf-like PPR proteins in fertile mitotypes on the processing of mitochondrial RNAs. RPF1 acts on the *nad4* 5′ UTR (72), RPF2 on the *nad9* and *cox3* 5′ UTRs (80), and RPF3 on the *ccmC* 5′ UTR (79). These proteins induce the

formation of adjacent 5′ and 3′ termini, providing strong evidence that they stimulate site-specific endonucleolytic cleavage (72, 79). An analogous effect was recently demonstrated for RPF5, a P-type PPR protein that does not belong to the RFL clade (68). RPF5 stimulates efficient 5′ processing of three different RNAs. The sequences flanking two of the RPF5-dependent sites share striking sequence similarity, including a segment that is an excellent match to the binding site predicted for RPF5 based on the "PPR code" (6). This advance lays the foundation for elucidating how pure PPR proteins promote RNA cleavage. Plausible scenarios include recruitment of an endoribonuclease or a remodeling of RNA structure that unmasks a nuclease-sensitive site (**Figure 3*c***).

PPR code: the combinations of amino acids at two key positions in a PPR motif that specify binding to a particular nucleotide

Chloroplasts harbor many proteins that resemble mitochondrial RFL proteins in their architecture, but none of these have been associated with RNA-cleavage events. However, the DYW and SMR domains appended to some PPR tracts have been suggested to have RNA endonuclease activity (reviewed in 101, 144). CRR2 is a PLS-DYW protein that resembles RNA-editing factors, but it is required for the accumulation of RNAs with 5′ or 3′ termini mapping upstream of *ndhB* (65). It has been proposed that CRR2's DYW domain cleaves the *rps7-ndhB* precursor, and this view is consistent with the RNA-cleavage activity demonstrated for several DYW domains in vitro (117, 119). However, high-resolution mapping of the CRR2-dependent RNA termini (137) revealed an overlapping arrangement that suggests these termini are defined and stabilized via a PPR10-like blockade mechanism. Additional work will be required to reconcile these observations.

An interesting example is provided by ribonuclease P enzymes in plants, which process the 5′ termini of tRNAs (59). Although the corresponding activity in bacteria is performed by a ribozyme, the plant enzymes consist of a polypeptide with several PPR motifs upstream of an endonuclease domain (55). The PPR tract increases the enzyme's affinity for pre-tRNA (73), possibly by interacting with conserved nucleotides typically found in tRNA loops (56).

PPR Proteins Promote the Splicing of Group II Introns in Chloroplasts and Mitochondria

Plant organellar genomes encode numerous introns of the group II type: ribozymes with a complex structure that forms an active site for self-splicing (43). The group II introns in plant organelles have lost the capacity to self-splice and require a plethora of nucleus-encoded proteins for their splicing. PPR proteins feature prominently among these (**Supplemental Table 1**). PPR splicing factors have a variety of architectures: They include short (THA8) and long (PPR4, OTP51, PPR5, and OTP43) P-type proteins both with and without accessory domains (8, 41, 42, 86, 143), as well as PLS-type proteins (OTP70 and Pp_PPR43) (25, 75). Each of these proteins is required for the splicing of just one or two introns, and RIP-chip assays have shown that several of them associate specifically with the introns whose splicing they promote (8, 86, 143).

Supplemental Material

It is likely that many of these proteins stimulate splicing by occupying RNA segments that would otherwise interfere with productive intron folding. PPR4 and OTP43 are special cases in that they promote the *trans*-splicing of independently transcribed intron fragments (42, 143); these could potentially protect the termini of the intron fragments from exonucleases or serve as splints to join intron segments. The most detailed information is available for PPR5, whose binding site has been mapped to a 50-nucleotide region in the maize chloroplast *trnG* intron (8, 168). It was hypothesized that PPR5 prevents the formation of an RNA hairpin that would otherwise occlude a key *cis*-element for splicing. PPR5 bound to this site also protects the *trnG* intron from endonucleolytic cleavage, providing the sole example of PPR employment as an endonuclease blockade.

PLS-Type PPR Proteins Typically Function as Site-Specificity Factors in RNA Editing

RNA editing in plant organelles consists of site-specific deamination of cytidine to uridine, and almost always alters the coding potential of the transcript (reviewed in 30). Many organelle-encoded proteins can be synthesized in a functional form only after their transcripts have been edited. The site specificity of the editing reaction puzzled researchers for many years until the first specificity factor was identified and found to be a PPR protein (98). Subsequently, many others have been identified in classical and reverse genetic screens (**Supplemental Table 1**). So far, all RNA-editing specificity factors that have been discovered are PLS-class PPR proteins (reviewed in 50, 144, 170). In rare cases, P-class proteins have been shown to influence editing (39), presumably by effects on RNA turnover or site accessibility, but none have been shown to act as true specificity factors.

Supplemental Material

The PPR tracts of PLS-class proteins appear to function similarly to those of P-class proteins, i.e., they form a binding surface that recognizes the primary sequence in a single-stranded RNA molecule. Genetic and biochemical evidence showed that the *cis*-elements targeted by editing specificity factors are positioned approximately 5–20 nucleotides upstream of the edited cytidine (e.g., 13, 27, 31, 44, 151), and direct binding of several PPR editing factors to such sites has been demonstrated (93, 114, 121, 122, 155). Current models align the PPR motifs to the RNA sequence such that the terminal S motif is positioned in contact with the nucleotide at the −4 position (6, 152, 169) (**Figure 4**). This position allows the cytidine deaminase activity to specifically act on the cytidine to be edited, as shown by the fact that adjacent cytidine residues on either side are not edited (2). The editing reaction itself requires the C-terminal domains of PLS-class proteins (26, 119, 120). The E domain is always essential, whereas the terminal DYW domain is often facultative, although this may be because it can be recruited from another protein (17). However, although the sequence-specific RNA-binding activity of editing factors is well documented (121, 122, 135, 155), the editing activity itself has not been reconstituted with purified factors (117, 119), leading to expectations that other proteins must be involved. Indeed, RNA-editing interacting factor protein (RIP)/multiple-organellar RNA-editing factor (MORF) proteins are required for editing at many sites in angiosperm organelles (10, 11, 148, 153), although their function is unknown, and they appear to be absent from bryophytes and lycophytes, where the RNA-editing factors are otherwise very similar.

Functions of PPR Tracts That Do Not Rely on an RNA-Binding Activity

The activities of characterized PPR proteins are consistent with the view that the PPR tract contributes an RNA-binding function. However, there are instances in which interactions of a PPR tract with protein or DNA may be physiologically relevant. For example, the wheat mitochondrial protein p63 binds DNA and stimulates transcription by mitochondrial extracts in vitro (76), and the chloroplast protein PTAC2 copurifies with a plastid transcription complex (125). Whether these proteins interact in vivo with DNA, RNA, or both is not known.

Some PPR tracts have the capacity to bind proteins. For example, PPR10 and HCF152 form homodimers in vitro (6, 116). However, PPR10 binds RNA only in monomeric form (6), so the functional significance of its homodimerization is unclear. PNM1 associates with nuclear chromatin and interacts with several chromatin proteins in a yeast two-hybrid assay (63), but the physiological relevance of these interactions is unknown. Functionally relevant protein binding by a PPR tract has been shown for the *Chlamydomonas* protein MCA1. MCA1 collaborates with TCA1 to stabilize the chloroplast *petA* RNA and activate *petA* translation (129) via adjacent

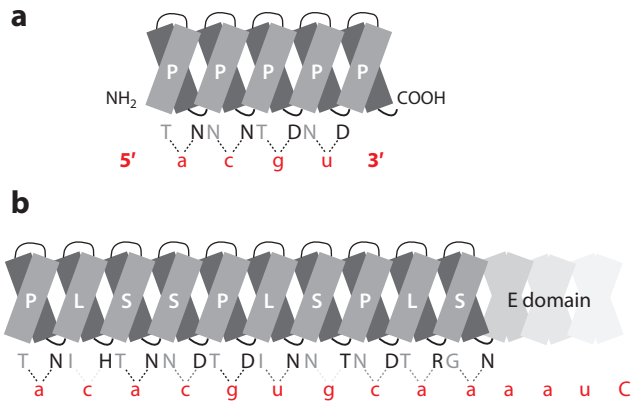

Figure 4

Alignments of PPR proteins on RNA targets. The PPR code (**Table 1**) facilitates the alignment of plant PPR proteins with their targets, but all known natural examples include mismatches and/or gaps in the alignment. (*a*) Segment of a hypothetical P-class PPR protein bound to the sequence ACGU. Residues at position 6 (*gray*) in the front helix and position 1′ (*black*) in the loop connecting adjacent helices specify which nucleotide is bound (*dotted lines*). The protein and RNA align in parallel (N terminus aligned with 5′ end). (*b*) A real-life example showing some of the possible complexities. The *Arabidopsis* editing factor CLB19 (28) is shown aligned to the *cis*-element upstream of the editing site in the *rpoA* transcript. Ten PPR motifs interact with a ten-nucleotide binding site depicted as in panel *a*. Interactions matching the code (**Table 1**) are indicated in darker lines than those that appear to be "mismatches" or that include uncommon combinations of amino acids whose preference has not been studied. As in almost all alignments of editing factors on their target sites, the final S motif aligns with the nucleotide at the −4 position with respect to the edited C (*red uppercase*) (6, 152, 169). The C-terminal E domain is faded out, as how it participates in RNA binding or in constituting the editing complex is not known.

cis-elements in the *petA* 5′ UTR (105). MCA1 also interacts with TCA1 in a yeast two-hybrid assay, and a PPR-containing region of the protein is required for this interaction (16). A second example involves the interaction between RNA-editing factors and their RIP/MORF partners (10, 153), one of which was shown to be mediated by the PPR tract in a yeast two-hybrid assay (10). Still, it remains to be seen whether all PPR-mediated functions have, at their core, a direct interaction between the PPR tract and RNA, or whether this presumably ancestral activity has in some cases been discarded entirely.

RNA RECOGNITION AND THE PPR CODE

The unusual evolutionary history and functional repertoire of the PPR family are likely to be tied to the unusual way in which PPR tracts bind RNA. In comparison with classic RNA-binding domains such as the RNA recognition motif (RRM), K homology (KH), and zinc-finger domains (32), *cis*-elements bound by typical PPR tracts are extremely long: PPR binding sites commonly span more than 12 nucleotides (and can reach up to ~29 nucleotides), and orthologous sites are often highly conserved across their length (e.g., 64, 70, 105, 127, 137, 177). Even the structurally analogous PUF domains recognize only 8–9 nucleotides (reviewed in 46). All PUF proteins, however, recognize related sequences, whereas there is no relation between the target sequences of different PPR proteins. Among the protein families capable of recognizing primary RNA sequence with high specificity, few, if any, have such a diverse range of natural target sequences. These unusual features have made the molecular basis for sequence recognition by PPR tracts an eagerly sought goal.

Table 1 Combinations of amino acids (AAs) in PPR motifs that specify binding to specific bases

Motif type	AA 6	AA 1′	Base	
			Correlation[a]	Experimental[b]
P, S	T	N	A	**A** >>> G,C,U
P, S	S	N	A	
P	S	S	A	
P	N	N	C	**C = U** >>> G,A
P	N	S	C	**C** > **U** >>> G,A
P, S	T	D	G	**G** >>> A,C,U
S	S	D	G	
P, S	N	D	U	**U** > **C** >>> G,A
L	P	D	U	

[a]Base most significantly correlated with this amino acid combination among PPR–RNA partners described in References 6, 152, and 169.

[b]Relative affinities based on affinities of PPR10 variants for modified RNAs in vitro (6). Nucleotides in bold bind with high affinity, and those not in bold bind negligibly.

Sequence-Specific RNA Binding by PPR Tracts Involves Modular One-Repeat:One-Nucleotide Recognition and the Combinatorial Effects of Two Amino Acids in Each Repeat

A structure of a PUF domain bound to RNA (164) provided a first look at the basis for RNA recognition by an alpha-solenoid protein. This structure revealed an elegant one-repeat:one-nucleotide binding mode in which the identities of amino acids at two positions in each repeat specify the bound nucleotide via hydrogen-bonding interactions. With this precedent, it seemed likely that PPR tracts bind RNA in a similar fashion, but experimental proof took a decade to obtain. As it turns out, PPR–RNA recognition has similarities to PUF–RNA recognition but also important differences (6).

Support for a one-repeat:one-nucleotide binding mode by PPR tracts initially came from correlations between the number of PPR motifs and the number of nucleotides in the binding sites of two particularly well-characterized proteins (105, 127) and from modeling of PPR structures based on contact-site prediction (49). Patterns of diversity within PPR motif sequences hinted at which residues might be involved in determining sequence specificity (49), and compelling correlations were subsequently found between these residues and aligned RNA bases in several known or inferred PPR–RNA partners (6, 152, 169). The major determinant is the amino acid at position 6 (**Figures 1** and **4**, **Table 1**). Asparagine at this position correlates strongly with pyrimidine at the corresponding position in the RNA, whereas serine or threonine correlates with purine. The second major determinant is at position 1 of the following motif (referred to here as position 1′). Aspartate at this position correlates with uridine or guanosine, whereas asparagine correlates with cytidine or adenosine. In this way, a simple code can be described whereby each of the four bases in the RNA can be recognized by a particular pair of amino acids at these two sites (**Figure 4**, **Table 1**). Amino acids 6 and 1′ are close to one another in three-dimensional space (**Figure 1**), such that they are well positioned to contact the same base. All of the amino acids reported to be involved in nucleotide specification have side chains that are avid hydrogen bond donors or acceptors, suggesting that base recognition is similar to Watson–Crick pairing in nucleic acid duplexes. These features are reminiscent of the two-amino-acid codes for RNA and DNA binding by PUF and TAL effector proteins, respectively (reviewed in 46). However, it is

important to note that PPR–RNA "duplexes" have the opposite polarity to PUF–RNA duplexes (parallel versus antiparallel) and employ a distinct amino acid code.

Evidence confirming basic features of this code has come from two types of experiments. First, in vitro binding assays have shown that altering the sequence of PPR10 alters its RNA recognition as predicted by the code (6). Second, the code has successfully predicted the binding sites of PLS-type editing factors (152, 169); such predictions would have a very low chance of success if the code were not largely correct.

Although it is tempting to refer to a single code, the reality is more complex. For example, the code is degenerate, with multiple combinations of amino acids at positions 6 and 1′ specifying the same nucleotide, and the same combination of amino acids sometimes being equally compatible with more than one nucleotide. In addition, amino acids at positions other than 6 and 1′ are likely to influence binding. Position 3 has been highlighted as a putative RNA-interacting residue (49) and incorporated into a three-position code (169). However, this residue is generally large and hydrophobic, suggesting a distinct mode of base interaction. In fact, a PPR10-*psaJ* RNA cocrystal structure (172) showed that this residue stacks between consecutive bases in a manner that is similar to the intercalation of amino acid side chains between adjacent bases in PUF-RNA complexes (164). The code also differs between different types of PPR motif: Although the same positions are involved, P and S motifs show slightly different amino acid–nucleotide correlations, and L motifs show even more distinct patterns (6, 152) (**Table 1**).

The current understanding of the code is sufficient to explain many but not all PPR–RNA alignments inferred from genetic and biochemical experiments. For example, alignments between P-type PPR proteins and their RNA ligands tend to include a central region containing gaps or mismatches (6). The crystal structure of PPR10 bound to *psaJ* RNA (172) confirmed anticipated modular interactions between the protein and RNA at the two ends of the complex and revealed the absence of RNA–protein contacts in the central region. In a similar fashion, PUF proteins tend to bind their RNA ligands via two arrays of contiguous motif–base interactions flanking a central region where the interactions are less predictable (165). That being said, some caution is warranted in interpreting the reported PPR10-*psaJ* structure because it involved a dimeric complex whose physiological relevance is unclear (6, 172). Additional crystal structures of PPR–RNA complexes will be invaluable for providing deeper insight into PPR–RNA interactions.

Distinctive RNA-Binding Properties of P and PLS Tracts Likely Underlie Their Distinct Functional Repertoires

The division of labor between PLS- and P-class proteins likely reflects a difference in their RNA-binding characteristics, although sequence recognition is very similar in both (6). Presumably, editing factors bind transiently to their targets, as their stable binding within ORFs (where most editing sites are located) would interfere with translation. In contrast, most functions of P-class proteins require them to bind tightly for extended periods of time, and all known P-class binding sites are in noncoding sequences where high-affinity binding would not disrupt translation. The fact that PPR footprints have been detected only for P-type proteins (137, 177) is in accord with this view, as the accumulation of a footprint would demand a slow protein off-rate. The irregularity of PLS motif organization likely mediates lower-affinity RNA interactions compared with P-type PPR tracts of similar length. This is not necessarily reflected in the K_d values for interactions reported in the literature, but such numbers should be viewed with caution because of variations in experimental conditions.

WHY DO PLANTS HAVE SO MANY PPR PROTEINS?

It is intriguing that hundreds of PPR proteins in plants serve the needs of ~150 organellar genes whose bacterial ancestors did without. This would seem to place an undue load of genomic baggage on the maintenance of organelle functions. How did this apparently unfortunate situation arise? Various adaptive drivers have been suggested. For example, a "genome-debugging" hypothesis (109) posits that rapidly evolving PPR proteins provide RNA-level "repair" of organelle genome mutations that accrue owing to the asexual reproduction of organelles. A genome-conflict theory posits an adaptive role for those organelle genome mutations that confer male sterility and the compensating PPR proteins that restore male fertility (22). The large reservoir of PPR proteins could also provide regulatory factors for coordinating nuclear and organellar gene expression (141).

These adaptive hypotheses are not mutually exclusive, and there is support for each of them. It is clear that PPR proteins can compensate for a variety of genome-level defects: For example, some point mutations are corrected by RNA editing, loss of stabilizing RNA hairpins can be corrected by PPR blockades, and the acquisition of RNA structures that mask start codons can be corrected by a PPR protein that prevents the formation of this structure. In accord with the cross-genome coordination hypothesis, the expression of one PPR protein has been shown to be a limiting factor in the expression of its target chloroplast gene (16). However, based on their important role in organellar gene expression, regulatory roles such as this probably emerged only after the fixation of these genes.

A fundamental chicken-and-egg issue remains: Which came first, the genome mutation or the compensating PPR? The genome-debugging analogy and genome-conflict theories place the organelle mutations prior to the compensating PPR. However, many of the organelle genome mutations that are "corrected" by PPR proteins have such serious physiological consequences (e.g., embryo lethality or inability to photosynthesize) that the mutation-first scenario is implausible. In this context, the constructive neutral evolution theory offers an attractive alternative. Constructive neutral evolution provides a nonadaptive means to generate "gratuitous complexity" in biology, examples of which include RNA editing and RNA splicing (57, 107). According to this scenario, the rescuing PPR existed prior to the organelle genome mutation it repairs and, in fact, allows such mutations to be retained. In the first step, opportunistic binding by a promiscuous PPR protein relaxes prior constraints on RNA sequence. This is followed by the accrual of mutations that would not have been tolerated in the absence of the PPR. For example, fortuitous PPR binding in a 3′ UTR could block 3′→5′ exonucleases and thus permit the degeneration of the 3′ RNA hairpin that originally filled this role. In this way, previously superfluous PPR–RNA interactions are fixed, and millions of years later hundreds of PPR "Band-Aids" are indispensable to plant survival.

PPR TRACTS AS A SCAFFOLD FOR THE ENGINEERING OF DESIGNER RNA-BINDING PROTEINS

Recent advances in understanding the functions and mechanisms of PPR proteins have exciting practical implications in the realms of synthetic biology and biotechnology. The ability to design proteins to bind specified RNA sequences has been a long-sought but elusive goal. The most promising results to date exploit the PUF domain, whose eight helical repeats each bind a single nucleotide via a predictable amino acid code (32, 165). The mechanism of RNA recognition by PPR tracts has similar features (although it differs in details), and the sequence specificity of a PPR protein has been changed in a predictable way according to the PPR code (6). The more malleable architecture and extraordinary evolutionary plasticity of the PPR scaffold together with the unusual ability of long PPR tracts to prevent a substantial length of RNA from interacting

with other macromolecules suggest applications beyond those that are readily accessible with PUF domains. That being said, much work remains to understand fundamental aspects of PPR–RNA interactions such that sequence specificity and affinity can be tailored to the desired function.

Many PPR proteins are required for the expression of one or several organellar genes. Therefore, it is not a stretch to imagine using natural and engineered PPR proteins to control the expression of specific organellar genes as a means to modulate photosynthesis, respiration, and other organelle-dependent processes. Robust and flexible regulation of plastid transgenes (110) may be achieved by employing regulated nuclear genes encoding engineered PPR proteins in conjunction with cognate plastid *cis*-elements. Engineered PPR-based RNA-editing factors could potentially be used to modify the amino acid sequences of organelle-encoded proteins, and functional domains (e.g., RNA-modifying/cleaving enzymes or fluorescent proteins) could be targeted to specific organellar RNAs via designer PPR tracts. The development of analogous applications outside of organelles may be feasible eventually, but will likely be more challenging considering the lack of natural examples to date.

SUMMARY POINTS

1. PPR proteins are nucleus-encoded helical repeat proteins that play a multitude of roles in organellar gene expression and organismal physiology. The enormous size and diverse functional repertoire of the PPR family in plants create an illusion of mechanistic complexity. However, many of the effects of PPR proteins distill down to a few themes and principles.
2. PPR proteins impact many aspects of plant physiology by altering the expression of specific organellar genes. The specificity of the effects of PPR proteins on particular genes derives primarily from the sequence-specific binding of PPR tracts to single-stranded RNA.
3. A functional divide separates the PLS and P subfamilies: The former is dedicated largely to RNA editing and the latter to RNA stabilization, RNA cleavage, translational activation, and RNA splicing.
4. PLS proteins specify sites of cytidine-to-uridine RNA editing by binding upstream of the target cytidine; the manner in which their C-terminal domains assemble the cytidine deaminase activity is unknown.
5. Many functions attributed to P-type proteins are a passive consequence of their ability to sequester a long RNA segment. This activity can block exoribonucleases, thereby defining and stabilizing processed 5′ and 3′ RNA termini. It can also influence the RNA-folding landscape, thereby exposing *cis*-elements for translation, splicing, and possibly RNA cleavage.
6. PPR tracts bind RNA via a modular one-repeat:one-nucleotide mechanism in which the identities of two amino acids in each repeat are the primary determinants of the bound nucleotide. Successes in engineering PPR proteins to bind novel sequences and in predicting the binding sites of uncharacterized proteins validate key aspects of this "code," but there are important gaps in current understanding.
7. Some PPR tracts mediate protein–protein interactions, but the degree to which such interactions contribute to PPR activities outside the realm of RNA editing remains to be seen.

8. The evolutionary malleability of the PPR family, the modular RNA-binding mechanism of PPR motifs, and the diverse functional repertoire that is fostered by varied motif architectures make the PPR scaffold an especially promising target for development of designer RNA-binding proteins with applications in biotechnology and synthetic biology.

FUTURE ISSUES

1. Improved understanding of PPR–RNA interactions will be necessary to accurately predict PPR binding sites and the functional consequences of that binding: What are the structures of P- and PLS-type PPR tracts bound to specific RNA ligands? What is the affinity and specificity of each PPR motif variant for each nucleotide, and do these parameters depend on context within a PPR–RNA duplex? To what extent are gaps or mismatches in a PPR–RNA duplex required or tolerated, and how does this vary depending on the position in the duplex and the biological function of the interaction? Do PPR proteins in other organisms follow the same rules governing RNA target recognition, and are there multiple modes of RNA binding by PPR tracts in plants?
2. What are the biochemical functions of the PPR-associated SMR and DYW domains?
3. To what extent does the regulated expression or activity of PPR proteins regulate organellar gene expression in response to environmental and developmental cues?
4. What is the repertoire of molecular mechanisms that underlie the effects of PPR proteins on translation, splicing, RNA cleavage, RNA editing, and transcription?
5. Can PPRs be engineered to bind RNA and modulate gene expression in the nuclear-cytosolic compartment?
6. How can knowledge of PPR–RNA interactions be exploited to develop useful tools?

DISCLOSURE STATEMENT

This review will be declared as an output of US National Science Foundation grant MCB-1243641 to A.B. and Australian Research Council grants CEO561495 and DP120102870 to I.S. The authors are inventors on International Patent Application PCT/AU2013/000387, which deals with recognition of RNA sequences by PPR proteins.

ACKNOWLEDGMENTS

We are grateful to the many authors of the excellent papers we cited and to the many colleagues with whom we have had stimulating discussions. We apologize to the authors of the many other excellent papers we were unable to cite. We thank Professor Charles Bond for the PPR model depicted in **Figure 1**. Research on PPR proteins in the authors' laboratories is supported by US National Science Foundation grants IOS-0922560 and MCB-1243641 to A.B. and Australian Research Council grants CEO561495 and DP120102870 to I.S.

LITERATURE CITED

1. Akagi H, Nakamura A, Yokozeki-Misono Y, Inagaki A, Takahashi H, et al. 2004. Positional cloning of the rice *Rf-1* gene, a restorer of BT-type cytoplasmic male sterility that encodes a mitochondria-targeting PPR protein. *Theor. Appl. Genet.* 108:1449–57
2. Arenas-M A, Takenaka M, Moreno S, Gómez I, Jordana X. 2013. Contiguous RNA editing sites in the mitochondrial *nad1* transcript of *Arabidopsis thaliana* are recognized by different proteins. *FEBS Lett.* 587:887–91
3. Aubourg S, Boudet N, Kreis M, Lecharny A. 2000. In *Arabidopsis thaliana*, 1% of the genome codes for a novel protein family unique to plants. *Plant Mol. Biol.* 42:603–13
4. Auchincloss A, Zerges W, Perron K, Girard-Bascou J, Rochaix JD. 2002. Characterization of Tbc2, a nucleus-encoded factor specifically required for translation of the chloroplast *psbC* mRNA in *Chlamydomonas reinhardtii*. *J. Cell Biol.* 157:953–62
5. Ban T, Ke J, Chen R, Gu X, Tan MH, et al. 2013. Structure of a PLS-class pentatricopeptide repeat protein provides insights into mechanism of RNA recognition. *J. Biol. Chem.* 288:31540–48
6. Barkan A, Rojas M, Fujii S, Yap A, Chong YS, et al. 2012. A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet.* 8:e1002910
7. Barkan A, Walker M, Nolasco M, Johnson D. 1994. A nuclear mutation in maize blocks the processing and translation of several chloroplast mRNAs and provides evidence for the differential translation of alternative mRNA forms. *EMBO J.* 13:3170–81
8. Beick S, Schmitz-Linneweber C, Williams-Carrier R, Jensen B, Barkan A. 2008. The pentatricopeptide repeat protein PPR5 stabilizes a specific tRNA precursor in maize chloroplasts. *Mol. Cell. Biol.* 28:5337–47
9. Bentolila S, Alfonso A, Hanson M. 2002. A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. *Proc. Natl. Acad. Sci. USA* 99:10887–92
10. Bentolila S, Heller WP, Sun T, Babina AM, Friso G, et al. 2012. RIP1, a member of an *Arabidopsis* protein family, interacts with the protein RARE1 and broadly affects RNA editing. *Proc. Natl. Acad. Sci. USA* 109:E1453–61
11. Bentolila S, Oh J, Hanson MR, Bukowski R. 2013. Comprehensive high-resolution analysis of the role of an *Arabidopsis* gene family in RNA editing. *PLoS Genet.* 9:e1003584
12. Binder S, Hölzle A, Jonietz C. 2011. RNA processing and RNA stability in plant mitochondria. In *Plant Mitochondria*, ed. F Kempken, pp. 107–30. New York: Springer
13. Bock R, Hermann M, Kossel H. 1996. In vivo dissection of *cis*-acting determinants for plastid RNA editing. *EMBO J.* 15:5052–59
14. Bolle N, Kempken F. 2006. Mono- and dicotyledonous plant-specific RNA editing sites are correctly edited in both in organello systems. *FEBS Lett.* 580:4443–48
15. Boudreau E, Nickelsen J, Lemaire SD, Ossenbuhl F, Rochaix JD. 2000. The *Nac2* gene of *Chlamydomonas* encodes a chloroplast TPR-like protein involved in *psbD* mRNA stability. *EMBO J.* 19:3366–76
16. Boulouis A, Raynaud C, Bujaldon S, Aznar A, Wollman FA, Choquet Y. 2011. The nucleus-encoded *trans*-acting factor MCA1 plays a critical role in the regulation of cytochrome *f* synthesis in *Chlamydomonas* chloroplasts. *Plant Cell* 23:333–49
17. Boussardon C, Salone V, Avon A, Berthome R, Hammani K, et al. 2012. Two interacting proteins are necessary for the editing of the NdhD-1 site in *Arabidopsis* plastids. *Plant Cell* 24:3684–94
18. Brown GG, Formanova N, Jin H, Wargachuk R, Dendy C, et al. 2003. The radish *Rfo* restorer gene of Ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. *Plant J.* 35:262–72
19. Bryant N, Lloyd J, Sweeney C, Myouga F, Meinke D. 2011. Identification of nuclear genes encoding chloroplast-localized proteins required for embryo development in *Arabidopsis thaliana*. *Plant Physiol.* 155:1678–89
20. Cai W, Ji D, Peng L, Guo J, Ma J, et al. 2009. LPA66 is required for editing *psbF* chloroplast transcripts in Arabidopsis. *Plant Physiol.* 150:1260–71
21. Cai W, Okuda K, Peng L, Shikanai T. 2011. PROTON GRADIENT REGULATION 3 recognizes multiple targets with limited similarity and mediates translation and RNA stabilization in plastids. *Plant J.* 67:318–27

22. Castandet B, Araya A. 2012. The nucleocytoplasmic conflict, a driving force for the emergence of plant organellar RNA editing. *IUBMB Life* 64:120–25
23. Cazalet C, Gomez-Valero L, Rusniok C, Lomma M, Dervins-Ravault D, et al. 2010. Analysis of the *Legionella longbeachae* genome and transcriptome uncovers unique strategies to cause Legionnaires' disease. *PLoS Genet.* 6:e1000851
24. Chase CD. 2007. Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions. *Trends Genet.* 23:81–90
25. Chateigner-Boutin AL, Colas des Francs-Small C, Delannoy E, Kahlau S, Tanz SK, et al. 2011. OTP70 is a pentatricopeptide repeat protein of the E subgroup involved in splicing of the plastid transcript *rpoC1*. *Plant J.* 65:532–42
26. Chateigner-Boutin AL, Colas des Francs-Small C, Fujii S, Okuda K, Tanz S, Small I. 2013. The E domains of pentatricopeptide repeat proteins from different organelles are not functionally equivalent for RNA editing. *Plant J.* 74:935–45
27. Chateigner-Boutin AL, Hanson MR. 2002. Cross-competition in transgenic chloroplasts expressing single editing sites reveals shared *cis* elements. *Mol. Cell. Biol.* 22:8448–56
28. Chateigner-Boutin AL, Ramos-Vega M, Guevara-Garcia A, Andrés C, de la Luz Gutiérrez-Nava M, et al. 2008. CLB19, a pentatricopeptide repeat protein required for editing of *rpoA* and *clpP* chloroplast transcripts. *Plant J.* 56:590–602
29. Chateigner-Boutin AL, Small I. 2007. A rapid high-throughput method for the detection and quantification of RNA editing based on high-resolution melting of amplicons. *Nucleic Acids Res.* 35:e114
30. Chateigner-Boutin AL, Small I. 2010. Plant RNA editing. *RNA Biol.* 7:213–19
31. Chaudhuri S, Maliga P. 1996. Sequences directing C to U editing of the plastid *psbL* mRNA are located within a 22 nucleotide segment spanning the editing site. *EMBO J.* 15:5958–64
32. Chen Y, Varani G. 2013. Engineering RNA-binding proteins for biology. *FEBS J.* 280:3734–54
33. Choury D, Araya A. 2006. RNA editing site recognition in heterologous plant mitochondria. *Curr. Genet.* 50:405–16
34. Choury D, Farré J, Jordana X, Araya A. 2005. Gene expression studies in isolated mitochondria: *Solanum tuberosum rps10* is recognized by cognate potato but not by the transcription, splicing and editing machinery of wheat mitochondria. *Nucleic Acids Res.* 33:7058–65
35. Cottage A, Mott E, Kempster J, Gray J. 2010. The *Arabidopsis* plastid-signalling mutant *gun1* (*genomes uncoupled1*) shows altered sensitivity to sucrose and abscisic acid and alterations in early seedling development. *J. Exp. Bot.* 61:3773–86
36. Dahan J, Mireau H. 2013. The Rf and Rf-like PPR in higher plants, a fast-evolving subclass of PPR genes. *RNA Biol.* 10:1286–93
37. Desloire S, Gherbi H, Laloui W, Marhadour S, Clouet V, et al. 2003. Identification of the fertility restoration locus, *Rfo*, in radish, as a member of the pentatricopeptide-repeat protein family. *EMBO Rep.* 4:588–94
38. Ding YH, Liu NY, Tang ZS, Liu J, Yang WC. 2006. Arabidopsis GLUTAMINE-RICH PROTEIN23 is essential for early embryogenesis and encodes a novel nuclear PPR motif protein that interacts with RNA polymerase II subunit III. *Plant Cell* 18:815–30
39. Doniwa Y, Ueda M, Ueta M, Wada A, Kadowaki K, Tsutsumi N. 2010. The involvement of a PPR protein of the P subfamily in partial RNA editing of an *Arabidopsis* mitochondrial transcript. *Gene* 454:39–46
40. Eberhard S, Loiselay C, Drapier D, Bujaldon S, Girard-Bascou J, et al. 2011. Dual functions of the nucleus-encoded factor TDA1 in trapping and translation activation of *atpA* transcripts in *Chlamydomonas reinhardtii* chloroplasts. *Plant J.* 67:1055–66
41. Falcon de Longevialle AF, Hendrickson L, Taylor N, Delannoy E, Lurin C, et al. 2008. 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.* 56:157–68
42. Falcon de Longevialle AF, Meyer EH, Andrès C, Taylor NL, Lurin C, et al. 2007. The pentatricopeptide repeat gene *OTP43* is required for *trans*-splicing of the mitochondrial *nad1* intron 1 in *Arabidopsis thaliana*. *Plant Cell* 19:3256–65
43. Falcon de Longevialle AF, Small I, Lurin C. 2010. Nuclearly encoded splicing factors implicated in RNA splicing in higher plant organelles. *Mol. Plant* 3:691–705

44. Farré JC, Leon G, Jordana X, Araya A. 2001. *cis* Recognition elements in plant mitochondrion RNA editing. *Mol. Cell. Biol.* 21:6731–37
45. Felder S, Meurer J, Meierhoff K, Klaff P, Bechtold N, Westhoff P. 2001. The nucleus-encoded *HCF107* gene of Arabidopsis provides a link between intercistronic RNA processing and the accumulation of translation-competent *psbH* transcripts in chloroplasts. *Plant Cell* 13:2127–41
46. Filipovska A, Rackham O. 2012. Modular recognition of nucleic acids by PUF, TALE and PPR proteins. *Mol. Biosyst.* 8:699–708
47. Fisk DG, Walker MB, Barkan A. 1999. Molecular cloning of the maize gene *crp1* reveals similarity between regulators of mitochondrial and chloroplast gene expression. *EMBO J.* 18:2621–30
48. Forner J, Hölzle A, Jonietz C, Thuss S, Schwarzlander M, et al. 2008. Mitochondrial mRNA polymorphisms in different Arabidopsis accessions. *Plant Physiol.* 148:1106–16
49. Fujii S, Bond CS, Small I. 2011. Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution. *Proc. Natl. Acad. Sci. USA* 108:1723–28
50. Fujii S, Small I. 2011. The evolution of RNA editing and pentatricopeptide repeat genes. *New Phytol.* 191:37–47
51. Fukui K, Kuramitsu S. 2011. Structure and function of the small MutS-related domain. *Mol. Biol. Int.* 2011:691735
52. Geddy R, Brown GG. 2007. Genes encoding pentatricopeptide repeat (PPR) proteins are not conserved in location in plant genomes and may be subject to diversifying selection. *BMC Genomics* 8:130
53. Germain A, Hotto AM, Barkan A, Stern DB. 2013. RNA processing and decay in plastids. *Wiley Interdiscip. Rev. RNA* 4:295–316
54. Gillman JD, Bentolila S, Hanson MR. 2007. The petunia restorer of fertility protein is part of a large mitochondrial complex that interacts with transcripts of the CMS-associated locus. *Plant J.* 49:217–27
55. Gobert A, Gutmann B, Taschner A, Gossringer M, Holzmann J, et al. 2010. A single *Arabidopsis* organellar protein has RNase P activity. *Nat. Struct. Mol. Biol.* 17:740–44
56. Gobert A, Pinker F, Fuchsbauer O, Gutmann B, Boutin R, et al. 2013. Structural insights into protein-only RNase P complexed with tRNA. *Nat. Commun.* 4:1353
57. Gray MW, Lukes J, Archibald JM, Keeling PJ, Doolittle WF. 2010. Cell biology: irremediable complexity? *Science* 330:920–21
58. Gutierrez-Marcos JF, Dal Pra M, Giulini A, Costa LM, Gavazzi G, et al. 2007. *empty pericarp4* encodes a mitochondrion-targeted pentatricopeptide repeat protein necessary for seed development and plant growth in maize. *Plant Cell* 19:196–210
59. Gutmann B, Gobert A, Giegé P. 2012. PRORP proteins support RNase P activity in both organelles and the nucleus in *Arabidopsis*. *Genes Dev.* 26:1022–27
60. Haili N, Arnal N, Quadrado M, Amiar S, Tcherkez G, et al. 2013. The pentatricopeptide repeat MTSF1 protein stabilizes the *nad4* mRNA in *Arabidopsis* mitochondria. *Nucleic Acids Res.* 41:6650–63
61. Hammani K, Cook W, Barkan A. 2012. RNA binding and RNA remodeling activities of the half-a-tetratricopeptide (HAT) protein HCF107 underlie its effects on gene expression. *Proc. Natl. Acad. Sci. USA* 109:5651–16
62. Hammani K, Colas des Francs-Small C, Takenaka M, Tanz SK, Okuda K, et al. 2011. The pentatricopeptide repeat protein OTP87 is essential for RNA editing of *nad7* and *atp1* transcripts in *Arabidopsis* mitochondria. *J. Biol. Chem.* 286:21361–71
63. Hammani K, Gobert A, Hleibieh K, Choulier L, Small I, Giegé P. 2011. An *Arabidopsis* dual-localized pentatricopeptide repeat protein interacts with nuclear proteins involved in gene expression regulation. *Plant Cell* 23:730–40
64. Hammani K, Okuda K, Tanz SK, Chateigner-Boutin AL, Shikanai T, Small I. 2009. A study of new *Arabidopsis* chloroplast RNA editing mutants reveals general features of editing factors and their target sites. *Plant Cell* 21:3686–99
65. Hashimoto M, Endo T, Peltier G, Tasaka M, Shikanai T. 2003. A nucleus-encoded factor, CRR2, is essential for the expression of chloroplast *ndhB* in *Arabidopsis*. *Plant J.* 36:541–49
66. Hattori M, Miyake H, Sugita M. 2007. A pentatricopeptide repeat protein is required for RNA processing of *clpP* pre-mRNA in moss chloroplasts. *J. Biol. Chem.* 282:10773–82

67. Hattori M, Sugita M. 2009. A moss pentatricopeptide repeat protein binds to the 3′ end of plastid *clpP* pre-mRNA and assists with mRNA maturation. *FEBS J.* 276:5860–69
68. Hauler A, Jonietz C, Stoll B, Stoll K, Braun HP, Binder S. 2013. RNA PROCESSING FACTOR 5 is required for efficient 5′ cleavage at a processing site conserved in RNAs of three different mitochondrial genes in *Arabidopsis thaliana*. *Plant J.* 74:593–604
69. Hayes ML, Hanson M. 2008. High conservation of a 5′ element required for RNA editing of a C target in chloroplast *psbE* transcripts. *J. Mol. Evol.* 67:233–45
70. Hayes ML, Mulligan RM. 2011. Pentatricopeptide repeat proteins constrain genome evolution in chloroplasts. *Mol. Biol. Evol.* 28:2029–39
71. Hernandez Mora JR, Rivals E, Mireau H, Budar F. 2010. Sequence analysis of two alleles reveals that intra-and intergenic recombination played a role in the evolution of the radish fertility restorer (Rfo). *BMC Plant Biol.* 10:35
72. Hölzle A, Jonietz C, Törjek O, Altmann T, Binder S, Forner J. 2011. A RESTORER OF FERTILITY-like PPR gene is required for 5′-end processing of the *nad4* mRNA in mitochondria of *Arabidopsis thaliana*. *Plant J.* 65:737–44
73. Howard MJ, Lim WH, Fierke CA, Koutmos M. 2012. Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5′ processing. *Proc. Natl. Acad. Sci. USA* 109:16149–54
74. Hu J, Wang K, Huang W, Liu G, Gao Y, et al. 2012. 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* 24:109–22
75. Ichinose M, Tasaki E, Sugita C, Sugita M. 2012. A PPR-DYW protein is required for splicing of a group II intron of *cox1* pre-mRNA in *Physcomitrella patens*. *Plant J.* 70:271–78
76. Ikeda T, Gray M. 1999. Characterization of a DNA-binding protein implicated in transcription in wheat mitochondria. *Mol. Cell. Biol.* 19:8113–22
77. Iyer LM, Zhang D, Rogozin IB, Aravind L. 2011. Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. *Nucleic Acids Res.* 39:9473–97
78. Johnson X, Wostrikoff K, Finazzi G, Kuras R, Schwarz C, et al. 2010. MRL1, a conserved pentatricopeptide repeat protein, is required for stabilization of *rbcL* mRNA in *Chlamydomonas* and *Arabidopsis*. *Plant Cell* 22:234–48
79. Jonietz C, Forner J, Hildebrandt T, Binder S. 2011. RNA PROCESSING FACTOR3 is crucial for the accumulation of mature *ccmC* transcripts in mitochondria of Arabidopsis accession Columbia. *Plant Physiol.* 157:1430–39
80. Jonietz C, Forner J, Hölzle A, Thuss S, Binder S. 2010. RNA PROCESSING FACTOR2 is required for 5′ end processing of *nad9* and *cox3* mRNAs in mitochondria of *Arabidopsis thaliana*. *Plant Cell* 22:443–53
81. Karcher D, Kahlau S, Bock R. 2008. Faithful editing of a tomato-specific mRNA editing site in transgenic tobacco chloroplasts. *RNA* 14:217–24
82. Karpenahalli MR, Lupas AN, Soding J. 2007. TPRpred: a tool for prediction of TPR-, PPR- and SEL1-like repeats from protein sequences. *BMC Bioinforma.* 8:2
83. Kazama T, Nakamura T, Watanabe M, Sugita M, Toriyama K. 2008. Suppression mechanism of mitochondrial ORF79 accumulation by Rf1 protein in BT-type cytoplasmic male sterile rice. *Plant J.* 55:619–28
84. Kazama T, Toriyama K. 2003. A pentatricopeptide repeat-containing gene that promotes the processing of aberrant *atp6* RNA of cytoplasmic male-sterile rice. *FEBS Lett.* 544:99–102
85. Ke J, Chen RZ, Ban T, Zhou XE, Gu X, et al. 2013. Structural basis for RNA recognition by a dimeric PPR-protein complex. *Nat. Struct. Mol. Biol.* 20:1377–82
86. Khrouchtchova A, Monde RA, Barkan A. 2012. A short PPR protein required for the splicing of specific group II introns in angiosperm chloroplasts. *RNA* 18:1197–209
87. Kim S, Yang J, Moon S, Ryu C, An K, et al. 2009. Rice *OGR1* encodes a pentatricopeptide repeat–DYW protein and is essential for RNA editing in mitochondria. *Plant J.* 59:738–49

88. Klein RR, Klein PE, Mullet JE, Minx P, Rooney WL, Schertz KF. 2005. Fertility restorer locus *Rf1* of sorghum (*Sorghum bicolor* L.) encodes a pentatricopeptide repeat protein not present in the colinear region of rice chromosome 12. *Theor. Appl. Genet.* 111:994–1012
89. Kleine T. 2012. *Arabidopsis thaliana* mTERF proteins: evolution and functional classification. *Front. Plant Sci.* 3:233
90. Klodmann J, Senkler M, Rode C, Braun HP. 2011. Defining the protein complex proteome of plant mitochondria. *Plant Physiol.* 157:587–98
91. Knoop V, Rüdinger M. 2010. DYW-type PPR proteins in a heterolobosean protist: plant RNA editing factors involved in an ancient horizontal gene transfer? *FEBS Lett.* 584:4287–91
92. Kobayashi K, Suzuki M, Tang J, Nagata N, Ohyama K, et al. 2007. LOVASTATIN INSENSITIVE 1, a novel pentatricopeptide repeat protein, is a potential regulatory factor of isoprenoid biosynthesis in *Arabidopsis*. *Plant Cell Physiol.* 48:322–31
93. Kobayashi Y, Matsuo M, Sakamoto K, Wakasugi T, Yamada K, Obokata J. 2008. Two RNA editing sites with *cis*-acting elements of moderate sequence identity are recognized by an identical site-recognition protein in tobacco chloroplasts. *Nucleic Acids Res.* 36:311–18
94. Kobe B, Kajava AV. 2000. When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. *Trends Biochem. Sci.* 25:509–15
95. Koizuka N, Imai R, Fujimoto H, Hayakawa T, Kimura Y, et al. 2003. Genetic characterization of a pentatricopeptide repeat protein gene, *orf687*, that restores fertility in the cytoplasmic male-sterile Kosena radish. *Plant J.* 34:407–15
96. Komori T, Ohta S, Murai N, Takakura Y, Kuraya Y, et al. 2004. Map-based cloning of a fertility restorer gene, *Rf-1*, in rice (*Oryza sativa* L.). *Plant J.* 37:315–25
97. Koprivova A, Colas des Francs-Small C, Calder G, Mugford ST, Tanz S, et al. 2010. Identification of a pentatricopeptide repeat protein implicated in splicing of intron 1 of mitochondrial *nad7* transcripts. *J. Biol. Chem.* 285:32192–99
98. Kotera E, Tasaka M, Shikanai T. 2005. A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature* 433:326–30
99. Koussevitzky S, Nott A, Mockler TC, Hong F, Sachetto-Martins G, et al. 2007. Signals from chloroplasts converge to regulate nuclear gene expression. *Science* 316:715–19
100. Krishnasamy S, Makaroff CA. 1994. Organ-specific reduction in the abundance of a mitochondrial protein accompanies fertility restoration in cytoplasmic male-sterile radish. *Plant Mol. Biol.* 26:935–46
101. Liu S, Melonek J, Boykin L, Small I, Howell K. 2013. PPR-SMRs: ancient proteins with enigmatic functions. *RNA Biol.* 10:1312–21
102. Liu X, Yu F, Rodermel S. 2010. An Arabidopsis pentatricopeptide repeat protein, SUPPRESSOR OF VARIEGATION7, is required for FtsH-mediated chloroplast biogenesis. *Plant Physiol.* 154:1588–601
103. Liu Y, He J, Chen Z, Ren X, Hong X, Gong Z. 2010. *ABA overly-sensitive 5* (*ABO5*), encoding a pentatricopeptide repeat protein required for *cis*-splicing of mitochondrial *nad2* intron 3, is involved in the abscisic acid response in Arabidopsis. *Plant J.* 63:749–65
104. Liu YJ, Xiu ZH, Meeley R, Tan BC. 2013. *Empty pericarp5* encodes a pentatricopeptide repeat protein that is required for mitochondrial RNA editing and seed development in maize. *Plant Cell* 25:868–83
105. Loiselay C, Gumpel NJ, Girard-Bascou J, Watson AT, Purton S, et al. 2008. Molecular identification and function of *cis*- and *trans*-acting determinants for *petA* transcript stability in *Chlamydomonas reinhardtii* chloroplasts. *Mol. Cell. Biol.* 28:5529–42
106. Lu Y, Li C, Wang H, Chen H, Berg H, Xia Y. 2011. AtPPR2, an Arabidopsis pentatricopeptide repeat protein, binds to plastid 23S rRNA and plays an important role in the first mitotic division during gametogenesis and in cell proliferation during embryogenesis. *Plant J.* 67:13–25
107. Lukes J, Archibald JM, Keeling PJ, Doolittle WF, Gray MW. 2011. How a neutral evolutionary ratchet can build cellular complexity. *IUBMB Life* 63:528–37
108. Lurin C, Andrés C, Aubourg S, Bellaoui M, Bitton F, et al. 2004. Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* 16:2089–103
109. Maier UG, Bozarth A, Funk HT, Zauner S, Rensing SA, et al. 2008. Complex chloroplast RNA metabolism: just debugging the genetic programme? *BMC Biol.* 6:36

110. Maliga P, Bock R. 2011. Plastid biotechnology: food, fuel, and medicine for the 21st century. *Plant Physiol.* 155:1501–10
111. Manavski N, Guyon V, Meurer J, Wienand U, Brettschneider R. 2012. An essential pentatricopeptide repeat protein facilitates 5′ maturation and translation initiation of *rps3* mRNA in maize mitochondria. *Plant Cell* 24:3087–105
112. Meierhoff K, Felder S, Nakamura T, Bechtold N, Schuster G. 2003. HCF152, an Arabidopsis RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast *psbB-psbT-psbH-petB-petD* RNAs. *Plant Cell* 15:1480–95
113. Merendino L, Perron K, Rahire M, Howald I, Rochaix JD, Goldschmidt-Clermont M. 2006. A novel multifunctional factor involved in *trans*-splicing of chloroplast introns in *Chlamydomonas*. *Nucleic Acids Res.* 34:262–74
114. Miyamoto T, Obokata J, Sugiura M. 2004. A site-specific factor interacts directly with its cognate RNA editing site in chloroplast transcripts. *Proc. Natl. Acad. Sci. USA* 101:48–52
115. Murayama M, Hayashi S, Nishimura N, Ishide M, Kobayashi K, et al. 2012. Isolation of *Arabidopsis abg11*, a weak ABA hypersensitive mutant defective in *nad4* RNA editing. *J. Exp. Bot.* 63:5301–10
116. Nakamura T, Meierhoff K, Westhoff P, Schuster G. 2003. RNA-binding properties of HCF152, an *Arabidopsis* PPR protein involved in the processing of chloroplast RNA. *Eur. J. Biochem.* 270:4070–81
117. Nakamura T, Sugita M. 2008. A conserved DYW domain of the pentatricopeptide repeat protein possesses a novel endoribonuclease activity. *FEBS Lett.* 582:4163–68
118. O'Toole N, Hattori M, Andrés C, Iida K, Lurin C, et al. 2008. On the expansion of the pentatricopeptide repeat gene family in plants. *Mol. Biol. Evol.* 25:1120–28
119. Okuda K, Chateigner-Boutin AL, Nakamura T, Delannoy E, Sugita M, et al. 2009. Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in *Arabidopsis* chloroplasts. *Plant Cell* 21:146–56
120. Okuda K, Myouga F, Motohashi R, Shinozaki K, Shikanai T. 2007. Conserved domain structure of pentatricopeptide repeat proteins involved in chloroplast RNA editing. *Proc. Natl. Acad. Sci. USA* 104:8178–83
121. Okuda K, Nakamura T, Sugita M, Shimizu T, Shikanai T. 2006. A pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing. *J. Biol. Chem.* 281:37661–67
122. Okuda K, Shikanai T. 2012. A pentatricopeptide repeat protein acts as a site-specificity factor at multiple RNA editing sites with unrelated *cis*-acting elements in plastids. *Nucleic Acids Res.* 40:5052–64
123. Petricka J, Clay N, Nelson T. 2008. Vein patterning screens and the defectively organized tributaries mutants in *Arabidopsis thaliana*. *Plant J.* 56:251–63
124. Pfalz J, Bayraktar O, Prikryl J, Barkan A. 2009. Site-specific binding of a PPR protein defines and stabilizes 5′ and 3′ mRNA termini in chloroplasts. *EMBO J.* 28:2042–52
125. Pfalz J, Liere K, Kandlbinder A, Dietz KJ, Oelmüller R. 2006. pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression. *Plant Cell* 18:176–97
126. Preker PJ, Keller W. 1998. The HAT helix, a repetitive motif implicated in RNA processing. *Trends Biochem. Sci.* 23:15–16
127. Prikryl J, Rojas M, Schuster G, Barkan A. 2011. Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. *Proc. Natl. Acad. Sci. USA* 108:415–20
128. Rahire M, Laroche F, Cerutti L, Rochaix JD. 2012. Identification of an OPR protein involved in the translation initiation of the PsaB subunit of photosystem I. *Plant J.* 72:652–61
129. Raynaud C, Loiselay C, Wostrikoff K, Kuras R, Girard-Bascou J, et al. 2007. Evidence for regulatory function of nucleus-encoded factors on mRNA stabilization and translation in the chloroplast. *Proc. Natl. Acad. Sci. USA* 104:9093–98
130. Ringel R, Sologub M, Morozov YI, Litonin D, Cramer P, Temiakov D. 2011. Structure of human mitochondrial RNA polymerase. *Nature* 478:269–73
131. Rivals E, Bruyére C, Toffano-Nioche C, Lecharny A. 2006. Formation of the Arabidopsis pentatricopeptide repeat family. *Plant Physiol.* 141:825–39
132. Rubinson EH, Eichman BF. 2012. Nucleic acid recognition by tandem helical repeats. *Curr. Opin. Struct. Biol.* 22:101–9

133. Ruckle M, Larkin R. 2009. Plastid signals that affect photomorphogenesis in *Arabidopsis thaliana* are dependent on GENOMES UNCOUPLED 1 and cryptochrome 1. *New Phytol.* 182:367–79
134. Rüdinger M, Fritz-Laylin L, Polsakiewicz M, Knoop V. 2011. Plant-type mitochondrial RNA editing in the protist *Naegleria gruberi*. *RNA* 17:2058–62
135. Rüdinger M, Kindgren P, Zehrmann A, Small I, Knoop V. 2013. A DYW-protein knockout in *Physcomitrella* affects two closely spaced mitochondrial editing sites and causes a severe developmental phenotype. *Plant J.* 76:420–32
136. Rüdinger M, Volkmar U, Lenz H, Groth-Malonek M, Knoop V. 2012. Nuclear DYW-type PPR gene families diversify with increasing RNA editing frequencies in liverwort and moss mitochondria. *J. Mol. Evol.* 74:37–51
137. Ruwe H, Schmitz-Linneweber C. 2012. Short non-coding RNA fragments accumulating in chloroplasts: footprints of RNA binding proteins? *Nucleic Acids Res.* 40:3106–16
138. Salone V, Rüdinger M, Polsakiewicz M, Hoffmann B, Groth-Malonek M, et al. 2007. A hypothesis on the identification of the editing enzyme in plant organelles. *FEBS Lett.* 581:4132–38
139. Sane AP, Stein B, Westhoff P. 2005. The nuclear gene HCF107 encodes a membrane-associated R-TPR (RNA tetratricopeptide repeat)-containing protein involved in expression of the plastidial *psbH* gene in Arabidopsis. *Plant J.* 42:720–30
140. Schallenberg-Rüdinger M, Lenz H, Polsakiewicz M, Gott J, Knoop V. 2013. A survey of PPR proteins identifies DYW domains like those of land plant RNA editing factors in diverse eukaryotes. *RNA Biol.* 10:1343–50
141. Schmitz-Linneweber C, Small I. 2008. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends Plant Sci.* 13:663–70
142. Schmitz-Linneweber C, Williams-Carrier RE, Barkan A. 2005. RNA immunoprecipitation and microarray analysis show a chloroplast pentatricopeptide repeat protein to be associated with the 5′-region of mRNAs whose translation it activates. *Plant Cell* 17:2791–804
143. Schmitz-Linneweber C, Williams-Carrier RE, Williams-Voelker PM, Kroeger TS, Vichas A, Barkan A. 2006. A pentatricopeptide repeat protein facilitates the *trans*-splicing of the maize chloroplast rps12 pre-mRNA. *Plant Cell* 18:2650–63
144. Shikanai T, Fujii S. 2013. Function of PPR proteins in plastid gene expression. *RNA Biol.* 10:1263–73
145. Small I, Peeters N. 2000. The PPR motif—a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem. Sci.* 25:46–47
146. Sosso D, Canut M, Gendrot G, Dedieu A, Chambrier P, et al. 2012. *PPR8522* encodes a chloroplast-targeted pentatricopeptide repeat protein necessary for maize embryogenesis and vegetative development. *J. Exp. Bot.* 63:5843–57
147. Sosso D, Mbelo S, Vernoud V, Gendrot G, Dedieu A, et al. 2012. PPR2263, a DYW-subgroup pentatricopeptide repeat protein, is required for mitochondrial *nad5* and *cob* transcript editing, mitochondrion biogenesis, and maize growth. *Plant Cell* 24:676–91
148. Sun T, Germain A, Giloteaux L, Hammani K, Barkan A, et al. 2013. An RNA recognition motif-containing protein is required for plastid RNA editing in *Arabidopsis* and maize. *Proc. Natl. Acad. Sci. USA* 110:E1169–78
149. Sung T, Tseng C, Hsieh M. 2010. The SLO1 PPR protein is required for RNA editing at multiple sites with similar upstream sequences in Arabidopsis mitochondria. *Plant J.* 63:499–511
150. Takenaka M, Brennicke A. 2012. Using multiplex single-base extension typing to screen for mutants defective in RNA editing. *Nat. Protoc.* 7:1931–45
151. Takenaka M, Neuwirt J, Brennicke A. 2004. Complex *cis*-elements determine an RNA editing site in pea mitochondria. *Nucleic Acids Res.* 32:4137–44
152. Takenaka M, Zehrmann A, Brennicke A, Graichen K. 2013. Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. *PLoS ONE* 8:e65343
153. Takenaka M, Zehrmann A, Verbitskiy D, Kugelmann M, Hartel B, Brennicke A. 2012. Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. *Proc. Natl. Acad. Sci. USA* 109:5104–9

154. Tameshige T, Fujita H, Watanabe K, Toyokura K, Kondo M, et al. 2013. Pattern dynamics in adaxial-abaxial specific gene expression are modulated by a plastid retrograde signal during *Arabidopsis thaliana* leaf development. *PLoS Genet.* 9:e1003655
155. Tasaki E, Hattori M, Sugita M. 2010. The moss pentatricopeptide repeat protein with a DYW domain is responsible for RNA editing of mitochondrial *ccmFc* transcript. *Plant J.* 62:560–70
156. Taylor NL, Heazlewood JL, Millar AH. 2011. The *Arabidopsis thaliana* 2-D gel mitochondrial proteome: refining the value of reference maps for assessing protein abundance, contaminants and post-translational modifications. *Proteomics* 11:1720–33
157. Terry M, Smith A. 2013. A model for tetrapyrrole synthesis as the primary mechanism for plastid-to-nucleus signaling during chloroplast biogenesis. *Front. Plant Sci.* 4:14
158. Tillich M, Poltnigg P, Kushnir S, Schmitz-Linneweber C. 2006. Maintenance of plastid RNA editing activities independently of their target sites. *EMBO Rep.* 7:308–13
159. Touzet P, Budar F. 2004. Unveiling the molecular arms race between two conflicting genomes in cytoplasmic male sterility? *Trends Plant Sci.* 9:568–70
160. Tseng C, Sung T, Li Y, Hsu S, Lin C, Hsieh M. 2010. Editing of *accD* and *ndhF* chloroplast transcripts is partially affected in the *Arabidopsis vanilla cream1* mutant. *Plant Mol. Biol.* 73:309–23
161. Uyttewaal M, Arnal N, Quadrado M, Martin-Canadell A, Vrielynck N, et al. 2008. Characterization of *Raphanus sativus* pentatricopeptide repeat proteins encoded by the fertility restorer locus for Ogura cytoplasmic male sterility. *Plant Cell* 20:3331–45
162. Uyttewaal M, Mireau H, Rurek M, Hammani K, Arnal N, et al. 2008. PPR336 is associated with polysomes in plant mitochondria. *J. Mol. Biol.* 375:626–36
163. Vaistij FE, Boudreau E, Lemaire SD, Goldschmidt-Clermont M, Rochaix JD. 2000. Characterization of Mbb1, a nucleus-encoded tetratricopeptide-like repeat protein required for expression of the chloroplast *psbB/psbT/psbH* gene cluster in *Chlamydomonas reinhardtii*. *Proc. Natl. Acad. Sci. USA* 97:14813–18
164. Wang X, McLachlan J, Zamore PD, Hall TM. 2002. Modular recognition of RNA by a human pumilio-homology domain. *Cell* 110:501–12
165. Wang Y, Wang Z, Hall TM. 2013. Engineered proteins with Pumilio/*fem-3* mRNA binding scaffold to manipulate RNA metabolism. *FEBS J.* 280:3755–67
166. Wang Z, Zou Y, Li X, Zhang Q, Chen L, et al. 2006. 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* 18:676–87
167. Williams P, Barkan A. 2003. A chloroplast-localized PPR protein required for plastid ribosome accumulation. *Plant J.* 36:675–86
168. Williams-Carrier R, Kroeger T, Barkan A. 2008. Sequence-specific binding of a chloroplast pentatricopeptide repeat protein to its native group II intron ligand. *RNA* 14:1930–41
169. Yagi Y, Hayashi S, Kobayashi K, Hirayama T, Nakamura T. 2013. Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS ONE* 8:e57286
170. Yagi Y, Tachikawa M, Noguchi H, Satoh S, Obokata J, Nakamura T. 2013. Pentatricopeptide repeat proteins involved in plant organellar RNA editing. *RNA Biol.* 10:1236–42
171. Yamazaki H, Tasaka M, Shikanai T. 2004. PPR motifs of the nucleus-encoded factor, PGR3, function in the selective and distinct steps of chloroplast gene expression in *Arabidopsis*. *Plant J.* 38:152–63
172. Yin P, Li Q, Yan C, Liu Y, Liu J, et al. 2013. Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* 504:168–71
173. Yu Q, Jiang Y, Chong K, Yang Z. 2009. AtECB2, a pentatricopeptide repeat protein, is required for chloroplast transcript *accD* RNA editing and early chloroplast biogenesis in *Arabidopsis thaliana*. *Plant J.* 59:1011–23
174. Yuan H, Liu D. 2012. Functional disruption of the pentatricopeptide protein SLG1 affects mitochondrial RNA editing, plant development, and responses to abiotic stresses in Arabidopsis. *Plant J.* 70:432–44
175. Yukawa M, Kuroda H, Sugiura M. 2007. A new in vitro translation system for non-radioactive assay from tobacco chloroplasts: effect of pre-mRNA processing on translation in vitro. *Plant J.* 49:367–76
176. Zhang YF, Hou MM, Tan BC. 2013. The requirement of WHIRLY1 for embryogenesis is dependent on genetic background in maize. *PLoS ONE* 8:e67369

177. Zhelyazkova P, Hammani K, Rojas M, Voelker R, Vargas-Suárez M, et al. 2012. Protein-mediated protection as the predominant mechanism for defining processed mRNA termini in land plant chloroplasts. *Nucleic Acids Res.* 40:3092–105
178. Zhou W, Cheng Y, Yap A, Chateigner-Boutin A, Delannoy E, et al. 2009. The Arabidopsis gene *YS1* encoding a DYW protein is required for editing of *rpoB* transcripts and the rapid development of chloroplasts during early growth. *Plant J.* 58:82–96
179. Zhu Q, Dugardeyn J, Zhang C, Takenaka M, Kuhn K, et al. 2012. SLO2, a mitochondrial pentatricopeptide repeat protein affecting several RNA editing sites, is required for energy metabolism. *Plant J.* 71:836–49
180. Zoschke R, Kroeger T, Belcher S, Schottler MA, Barkan A, Schmitz-Linneweber C. 2012. The pentatricopeptide repeat-SMR protein ATP4 promotes translation of the chloroplast *atpB/E* mRNA. *Plant J.* 72:547–58
181. Zoschke R, Qu Y, Zubo YO, Börner T, Schmitz-Linneweber C. 2013. Mutation of the pentatricopeptide repeat-SMR protein SVR7 impairs accumulation and translation of chloroplast ATP synthase subunits in *Arabidopsis thaliana*. *J. Plant Res.* 126:403–14
182. Zoschke R, Watkins K, Barkan A. 2013. A rapid microarray-based ribosome profiling method elucidates chloroplast ribosome behavior in vivo. *Plant Cell* 25:2265–75
183. Zsigmond L, Szepesi A, Tari I, Rigó G, Király A, Szabados L. 2012. Overexpression of the mitochondrial *PPR40* gene improves salt tolerance in Arabidopsis. *Plant Sci.* 182:87–93