Pentatricopeptide repeat (PPR) proteins as sequence-specificity factors in post-transcriptional processes in organelles

E. Delannoy*, W.A. Stanley*, C.S. Bond† and I.D. Small*1

1ARC Centre of Excellence in Plant Energy Biology, University of Western Australia, 35 Stirling Highway, Crawley, Perth 6009, WA, Australia, and †School of Biomedical, Biomolecular and Chemical Sciences, Molecular and Chemical Sciences Building (M316), University of Western Australia, 35 Stirling Highway, Crawley, Perth 6009, WA, Australia

Abstract
PPR (pentatricopeptide repeat) genes form a large family particularly prevalent in higher plants and targeted to organelles. They are involved in many post-transcriptional processes such as splicing, editing, processing and translation. Current data suggest that PPR proteins are involved in targeting effectors to the correct sites on the correct transcripts but the molecular mechanisms for RNA binding and effector recruitment by PPR proteins are not understood yet.

Introduction
Transcription in plant organelles often yields multiple primary transcripts with long 5′- and 3′-untranslated extremities, often containing several coding sequences and introns [1–3]. These primary transcripts must undergo extensive processing before they can be efficiently translated. These post-transcriptional modifications include 5′- and 3′-maturation, cis- and trans-splicing, cleavage [4] and editing [5,6]. Furthermore, these transcripts tend to accumulate to high steady-state levels in apparent excess over requirements [7], so gene expression in plant organelles is mostly regulated at the post-transcriptional level or translational level [8,9]. All of these steps require numerous nucleus-encoded specificity factors [10,11].

Recently, the PPR (pentatricopeptide repeat) gene family was described to contain approx. 450 members in the thale cress, Arabidopsis thaliana [12,13], three-quarters of which are predicted to be targeted to organelles. This proportion may even be an underestimate given the uncertainties of targeting predictions and that the limited experimental testing done so far showed exclusively organelar targeting. This proportion may be the result of different members in the family, mutations in PPR genes result in a wide range of phenotypes including embryo-lethality [14], cytoplasmic male sterility [15,16], defects in seed maturation [17] and impaired photosynthesis [18], suggesting that PPR proteins are implicated in many different physiological processes. Further studies showed that they are associated with diverse molecular events, mostly post-transcriptional (Table 1). CRR (chlororespiratory reduction) 4 and CRR21 are involved in editing events [18,19], CRR2 [20], HCF152 (high chlorophyll fluorescence 152) [21], RF1 (ribosome release factor 1) [22] are involved in processing, PET309 (petite 309) [23], RFO (resistance to Fusarium oxysporum) [24], CRP1 (chloroplast RNA processing 1) [25] in translation, RF1β [22], MCA1 [26], AEP3 [27] and BSF (bicoid stability factor) [28] in RNA stability and PPR4 [29] and PPR31-11 [30] in splicing. Furthermore, as mutations in PPR proteins often result in severe physiological defects, they cannot complement each other [12], which suggests that PPR proteins act in a very specific way. In fact, PPR inactivation is most likely to result in a unique primary defect restricted to a single transcript, as shown in Table 1. The best illustration so far comes from the description of two PPR proteins involved in the edition of a single chloroplastic transcript, ndhD, but targeting two different editing sites on that transcript [18,31]. Collectively, these results suggest that PPR proteins are most probably not direct effectors of the post-transcriptional processes in organelles but rather sequence-specificity factors targeting different effectors to the correct sites.

In the present review, we shall discuss the evidence for and against this view and point out some of the uncertainties that need to be worked on to understand the real functions of this large and still little-understood family of proteins.

Sequence-specific RNA binding
PPR motifs comprise a set of three related degenerate motifs of 31–36 amino acids [12]. They occur as tandem arrays of 2–26 motifs per protein with an average of 12 in plants [12]. PPR motifs are related to TPR (tetra-tricopeptide repeat) motifs, which mediate protein–protein interactions [32]. TPR motifs comprise a pair of antiparallel α-helices, and tandem repeats form a superhelix enclosing a groove, which is...
Only a few PPR proteins contain other recognizable RNA-binding proteins, responsible for the RNA-binding activity of PPR proteins. The balance of the evidence is in favor of PPR motifs being RNA-binding motifs, such as PPR4, which contains an RRM (RNA-recognition motif) domain [29].

The specificity of the RNA-binding activity displayed by PPR proteins has proved relatively difficult to ascertain. RIP-chip (RNA immunoprecipitation-chip) experiments showed that CRP1 binds in vivo in the region of a 69 nt motif in the 5′-UTR (5′-untranslated region) of psaC and petA [25]. However, this type of assay does not allow the delineation of the exact binding site. Attempts to discover the binding site(s) for HCF152 on the petB transcript were also unable to show a clear, unique intersection site [40]. Much clearer results have come from studies of the editing factors CRR4 and CRR21. The binding site of CCR4 was mapped in vitro to lie within the region from 25 nt upstream to 10 nt downstream of the target site using recombinant CRR4 proteins in EMSAs (electrophoretic mobility-shift assays). Domain swapping experiments between CRR4 and CRR21 clearly demonstrated in vivo that PPR motifs are necessary and sufficient for precise recognition of the base to be edited [19]. More studies are still required, but the data already available strongly suggest that PPR motifs are RNA-binding motifs that act collectively to achieve high affinity and sequence specificity. It cannot yet be formally ruled out that one or more other proteins are required to achieve full specificity, as most of the convincing data come from in vivo experiments.

Little, if anything, is known about the molecular basis of RNA binding by PPR proteins. Which amino acids are involved in the interaction? What are the determinants of the sequence specificity? A number of other sequence-specific RNA-binding proteins have been studied and their comparison can give some leads to understanding the binding site for the protein ligand [33]. Similarly, structural predictions suggest that PPR motifs consist of two antiparallel α-helices and presumably tandem arrays would form a superhelix with an internal binding face similar to that of TPR proteins (Figure 1) [34–37]. However, as opposed to TPR motifs, the predicted residues projecting into the central groove are almost exclusively hydrophilic and the bottom of the groove is positively charged, suggesting that the ligand is likely to be hydrophilic and acidic. This led to an initial suggestion that PPR motifs could be RNA-binding motifs [37]. The Puf domain family of sequence-specific RNA-binding proteins consists of tandem arrays of three-helix repeats [36] and may provide a guide as to how PPR proteins might contact RNA (Figure 1).

The RNA-binding property of PPR proteins was confirmed by several in vitro experiments using gel shift [38], UV cross-linking [21,39,40] or affinity [12,38,39] assays. It was also shown that P63 from wheat binds DNA [41], suggesting that a subset of PPR proteins may bind DNA rather than, or as well as, RNA. Nakamura et al. [40] showed that PPR motifs are collectively required for RNA binding of HCF152: proteins with only two PPR motifs showed very low affinity, whereas higher affinity was achieved with the 12 PPR motifs present in HCF152. A contradictory result came from LRPPRC. (leucine-rich PPR-motif containing protein) [39] where the binding domain was localized at the C-terminal part of the protein including only 2 of the 11 PPR motifs present in this PPR protein. Nevertheless, the balance of the evidence is in favor of PPR motifs being responsible for the RNA-binding activity of PPR proteins. Only a few PPR proteins contain other recognizable RNA-binding motifs, such as PPR4, which contains an RRM (RNA-recognition motif) domain [29].

### Table 1 | PPR proteins with elucidated target and mode of action

<table>
<thead>
<tr>
<th>Gene</th>
<th>Organism</th>
<th>Molecular process</th>
<th>Target Location</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rf1</td>
<td>Raphanus sativus</td>
<td>Translation</td>
<td>orf125</td>
<td>Mt [53]</td>
</tr>
<tr>
<td>Rf1b</td>
<td>Oryza sativa</td>
<td>RNA degradation</td>
<td>atp6-orf79</td>
<td>Mt [22]</td>
</tr>
<tr>
<td>Rf1</td>
<td>O. sativa</td>
<td>RNA cleavage</td>
<td>atp6-orf79</td>
<td>Mt [16,22]</td>
</tr>
<tr>
<td>CRP1</td>
<td>Zea mays</td>
<td>Translation</td>
<td>petB</td>
<td>Cp [21]</td>
</tr>
<tr>
<td>CRR4</td>
<td>A. thaliana</td>
<td>RNA cleavage</td>
<td>ndhB</td>
<td>Cp [20]</td>
</tr>
<tr>
<td>CRR2</td>
<td>A. thaliana</td>
<td>RNA cleavage</td>
<td>ndhB</td>
<td>Cp [21]</td>
</tr>
<tr>
<td>PGR3</td>
<td>A. thaliana</td>
<td>DNA stabilization</td>
<td>petB</td>
<td>Cp [21]</td>
</tr>
<tr>
<td>HCF152</td>
<td>A. thaliana</td>
<td>RNA splicing and maturation</td>
<td>petA, psoC</td>
<td>Cp [25]</td>
</tr>
<tr>
<td>MCA1</td>
<td>Chlamydomonas reinhardti</td>
<td>RNA stabilization</td>
<td>petB</td>
<td>Cp [21]</td>
</tr>
<tr>
<td>LRPPRC</td>
<td>Neurospora crassa</td>
<td>Splicing</td>
<td>ndhB</td>
<td>Cp [20]</td>
</tr>
<tr>
<td>PPR531-11</td>
<td>Physcomitrella potens</td>
<td>RNA splicing and maturation</td>
<td>ndhB</td>
<td>Cp [19]</td>
</tr>
<tr>
<td>CPF1</td>
<td>Human</td>
<td>RNA stabilization</td>
<td>cis, coxl</td>
<td>Mt [39,51]</td>
</tr>
<tr>
<td>Cya5</td>
<td>Saccharomyces cerevisiae</td>
<td>RNA stabilization</td>
<td>coxl</td>
<td>Mt [55]</td>
</tr>
<tr>
<td>Pet309</td>
<td>S. cerevisiae</td>
<td>RNA stabilization</td>
<td>atp6 atp8</td>
<td>Mt [27]</td>
</tr>
<tr>
<td>BSF</td>
<td>Drosophila melanogaster</td>
<td>RNA stabilization</td>
<td>5′-UTR bicoid</td>
<td>C [28]</td>
</tr>
<tr>
<td>P63</td>
<td>Triticum aestivum</td>
<td>Transcription</td>
<td>coxi</td>
<td>Mt [41]</td>
</tr>
</tbody>
</table>

Abbreviations: Mt, mitochondria; Cp, chloroplasts; C, cytoplasm.
Putative structural features of PPR proteins

RNA binding by PPR proteins [42]. Natural mutants show that point mutations in PPR motifs resulting in amino acid changes can abolish the function of PGR3 (proton gradient regulation 3) [43] and RF1 [22], suggesting that although PPR motifs are collectively required for RNA binding, some amino acids are essential for this interaction. The discovery of more PPR-binding sites, along with site-directed mutagenesis and structural studies of PPR motifs, should allow progress towards the elucidation of this binding ‘code’.

Figure 1 | Putative structural features of PPR proteins

Ten tandemly repeated TPR motifs from O-linked GlcNAc transferase [34] (bottom left) were manually aligned with the ten PPR motifs of At1g05750, a typical PLS subfamily protein. This allowed threading of the PPRs around the TPR array to produce a tentative superhelical model (bottom right), produced in Swiss-Model [35]. Ribbon models and electrostatic surfaces were prepared in PyMol (http://www.pymol.org). While the concave inner face of the O-linked GlcNAc transferase superhelix is rather acidic, the inner face of the protein coded by At1g05750 is overwhelmingly basic (as for many other PPR proteins), which may provide a suitable interface for interaction with the phosphate backbone of RNA ligands. Given the tentative nature of the model, it is conceivable that the PPR array may in reality adopt a superhelix of considerably different pitch and inner cavity dimensions, or a looser non-superhelical conformation. The well-characterized single-stranded RNA-binding protein Pumilio1 adopts an arched conformation of tandemly repeated helical PUF motifs [36]. This conformation forms an electrostatic network in which a long acidic surface patch runs parallel to a basic patch (top left). The basic patch interacts with the phosphate backbone of single-stranded RNA and the RNA bases point towards the acidic patch, demonstrating a mechanistic basis for sequence-specific RNA recognition (top right). If PPR proteins bind RNA in a way that resembles Pumilio–RNA interactions, approx. one RNA base would contact one PPR motif.

PPR as site-recognition factors for effector proteins

As most PPR proteins lack any known catalytic sites, the RNA processing events triggered by them must be carried out by other proteins that in some way associate with the PPR–RNA complex. Indeed, several PPR proteins have been shown to be components of complexes. PPR proteins were identified in association with the small ribosomal subunit of mitochondria from mammals [44] and Leishmania tarentolae [45] and as part of yeast mitochondrial transcription and translation complexes [46]. In plants, the Petunia hybrida restorer protein RF was shown to be associated with a soluble, mitochondrial, RNase-sensitive, high-molecular-mass protein complex attached to the inner membrane [47]. This complex is likely to be at least part of the translation machinery, and similarly, PPR4 and PPR2 from maize chloroplasts associate with RNase-sensitive high-molecular-mass complexes that co-sediment with ribosomes [29,47,48]. In the same vein, Arabidopsis pTAC2 (plastid transcriptionally active chromosome 2) [49] and wheat P63 [41] co-purify with transcription complexes from mitochondria and chloroplasts respectively. However, very few direct protein partners of PPR proteins have been identified. GRP23 (glutamine-rich protein 23) was reported to interact in a yeast two-hybrid assay and in planta with the nRNA polymerase II subunit 3 [50], and several interacting partners of the SEC1 domain of LRPPRC were identified through yeast two-hybrid screening [51]. However, to date, no direct interacting partner of mitochondrial- or chloroplast-targeted PPR proteins has been described in plants. Obviously, a better knowledge of the interacting partners of PPR proteins is required to understand their functions in organelles and to allow the study of the molecular basis of their roles in post-transcriptional processing.

Ding et al. [50] showed that the interaction of GRP23 with RNA polymerase II was mediated by the C-terminal part of the protein, which contains multiple copies of a novel glutamine-rich motif. The SMR (small mutS-related) domain is found at the C-terminus of eight PPR proteins in A. thaliana, including pTAC2 and GUN1; in the latter, it has been shown to provide DNA-binding activity [52], which tallies with the observed association of pTAC2 with chloroplast DNA [49]. Similarly, domain-swapping experiments with CRR4 and CRR21 showed that the C-terminal extremity of each protein was required for the edition of the target site, but not for target specificity, demonstrating that physical recruitment of the editing machinery requires these non-PPR domains [19]. Extrapolating from these results, and considering that the PPR motifs are involved in RNA–protein interactions, we can postulate that the diverse C-terminal domains condition protein–protein (or protein–DNA) interactions of PPR proteins. The most common C-terminal domains are the so-called E and E+ domains found in almost 150 of the Arabidopsis PPR proteins, including the editing factors CRR4 and CRR21. In more than half of these proteins, these domains are associated with a further C-terminal
domain, the DYW domain. The only DYW-containing PPR protein for which the molecular function is clear is CRR2, linked to a site-specific cleavage of the ndhB transcript.

It is tempting to think that these various C-terminal domains determine which effector proteins or complexes are recruited to the target RNA by the PPR protein. So far the models proposed for the action of PPR proteins as specific molecular adaptors fit the data available. However, the experimental evidence is still limited to only a few examples and many PPR proteins lack obvious C-terminal domains for recruitment of effector complexes. Hence interaction studies urgently need to be extended to other types of PPR proteins in order to define their protein–protein interaction domains.

Future directions

Although growing evidence is now available to show that PPR proteins play an important role as sequence-specificity factors in post-transcriptional processes in organelles, numerous questions remain. Future studies will probably be aimed at understanding the molecular basis of RNA–protein and protein–protein interactions, as suggested above, but it is also important to try to integrate PPR proteins into the whole context of cellular metabolism. It is highly likely that some, or even many, PPR proteins are not only essential for the processes that they are involved in but also regulate them in accordance with cellular requirements. The first evidence of PPR protein as a true regulator of organelle metabolism came from *Chlamydomonas reinhardtii*, where MCA1 was shown to be the limiting factor for the accumulation of *petA* transcripts while its abundance varied with physiological conditions affecting *petA* gene expression [26]. Similar experiments have not yet been carried out with equivalent land plant proteins, so there is little direct evidence yet that PPR proteins are involved in the regulation of organelle metabolism in response to developmental or environmental changes. In addition, PPR proteins may be involved in retrograde signalling from organelles to the nucleus; GUN1 is a chloroplast PPR protein that somehow represses nuclear gene expression for a wide range of genes encoding chloroplast proteins if plastid function is blocked [52] and LOI1 (lovastatin insensitive 1), a mitochondrial PPR protein, is implicated in the regulation of cytosolic isoprenoid biosynthesis pathways [38]. It is not yet clear whether these effects are indirect (via metabolic feedback), or whether they involve direct involvement of the proteins in signalling pathways. It cannot be ruled out that some PPR proteins function partially or wholly in the nucleus (as reported for GRP23). So far, the more we discover about this large family of proteins, the more questions are raised and the more interesting their functions appear to become.

References


Received 2 July 2007
doi:10.1042/BST0351643