A Pentatricopeptide Repeat Protein in the *Plasmodium* apicoplast is essential and shows sequence-specific RNA binding

Joanna L. Hicks[1], Imen Lassadi, Emma Carpenter, Madeleine Eno, Alexandros Vardakis, Ross F. Waller, Christopher J. Howe*, R. Ellen R. Nisbet

Department of Biochemistry, Hopkins Building, Downing Site, University of Cambridge CB2 1QW Cambridge, United Kingdom

[1] now at Faculty of Science, Waikato University, Hamilton, New Zealand

*Corresponding author

ch26@cam.ac.uk

Keywords: RNA processing, apicoplast, plastid, apicomplexa, *Plasmodium*, *Toxoplasma*

Short title A single *Plasmodium* apicoplast PPR

ABSTRACT

The malaria parasite *Plasmodium* and other apicomplexans such as *Toxoplasma* evolved from photosynthetic organisms and contain an essential, remnant plastid termed the apicoplast. Transcription of the apicoplast genome is polycistronic with extensive RNA processing. Little is known about the mechanism of post-transcriptional processing. In plant chloroplasts, post-transcriptional RNA processing is controlled by multiple pentatricopeptide repeat (PPR) proteins. Here, we present the biochemical characterisation of the single apicoplast-targeted PPR protein. Apicoplast PPR1 is essential, and binds specific RNA sequences corresponding with previously characterized RNA processing sites. We identify the specific binding motif of PPR1. In RNAse protection assays, PPR1 shields apicoplast transcripts from ribonuclease degradation. Our results show that apicoplast RNA processing is under the control of a single protein, thus presenting an Achilles' heel for the development of new anti-malarial drugs.

INTRODUCTION

The malaria parasite *Plasmodium falciparum* and related apicomplexan parasites such as *Toxoplasma* evolved from photosynthetic organisms and contain a remnant plastid known as an apicoplast (Gardner, Williamson, & Wilson, 1991; Howe, 1992; McFadden, Reith, Munholland, & Lang-Unnasch, 1996). The ability to photosynthesise has been lost yet the apicoplast remains essential for parasite survival. The apicoplast genome encodes 30 proteins, two rRNAs and 25 tRNAs (Wilson et al., 1996). Primary RNA transcripts are polycistronic and there must be extensive RNA processing to produce individual tRNA, rRNA and mRNA molecules. Many mRNA transcripts show discrete ends coinciding with tRNA sequences, while others show heterogeneous ends within protein coding sequences (R. Ellen R. Nisbet, Kurniawan, Bowers, & Howe, 2016; R. E. R. Nisbet & McKenzie, 2016). The processing of polycistronic transcripts to individual RNAs requires both site recognition and RNA cleavage.

In plants, the primary agents through which the nucleus exerts control on organelle gene expression are pentatricopeptide repeat (PPR) proteins. PPR proteins are encoded in the nuclear genome and are targeted to the mitochondrion or plastid (Barkan & Small, 2014). Plants contain many hundreds of PPRs (Lurin et al., 2004). By contrast, genomes of algae and non-photosynthetic eukaryotes encode relatively few PPR proteins (Manna, 2015; Tourasse, Choquet, & Vallon, 2013). PPR proteins are involved in all aspects of organelle RNA biology, including splicing, editing, transcript stability and translation. In plants, chloroplast PPR mutants show defects in fertility and embryo and seed development (Bryant, Lloyd, Sweeney, Myouga, & Meinke, 2011; Lurin et al., 2004; Prikryl, Rojas, Schuster, & Barkan, 2011; Sosso, Canut, et al., 2012; Sosso, Mbelo, et al., 2012). Additionally, a

mitochondrial mutation which causes cytoplasmic male sterility (pollen sterility) is suppressed by a PPR protein (Brown et al., 2003).

PPR proteins are sequence-specific RNA-binding proteins. They contain 2-30 tandem repeats, each repeat comprising a 35-amino acid motif (Manna, 2015; Prikryl et al., 2011). The PPR protein family can be divided into two classes: (1) the P-class which contain only the 35 amino acid repeats which stabilize specific RNAs by binding to the 5' or 3' termini of RNA transcripts and providing a barrier to exonuclease digestion (Lurin et al., 2004; Pfalz, Bayraktar, Prikryl, & Barkan, 2009; Prikryl et al., 2011; Zhelyazkova et al., 2012); (2) the PLS-class which contains repeats of 31-36 amino acids together with further domains at the carboxyl terminus of the protein. The PLS class can be further divided into E and DYW subclasses, both implicated in RNA editing, carrying out a deamination reaction converting specific cytidines to uridine in plant organelles.

Very little is known about the molecular mechanisms of post-transcriptional processing in the apicoplast. A number of nucleus-encoded, apicoplast-targeted proteins have been identified which may function in RNA processing. Only one RNA-binding protein (*P. vivax PVX_084415*) has been partially characterized, although the stability of the heterologously expressed protein was such that it was not possible to carry out functional assays, though it did bind to uridine rich RNA (García-Mauriño et al., 2018).

Here, we report the identification of a single apicoplast PPR protein. We show that this protein, designated PPR1, is localized within the apicoplast of both *Plasmodium falciparum*

and *Toxoplasma gondii*. It is essential for growth of *T. gondii*. Biochemical characterisation of the *P. falciparum* PPR protein shows it binds to a specific RNA sequence and protects RNA transcripts from degradation by ribonucleases. Although the presence of a PPR proteins in the apicoplast is not unexpected, the dependence of a plastid on just a single PPR protein is unique and identifies PPR1 as a good potential drug target.

RESULTS

A single apicoplast PPR protein present in both Plasmodium and Toxoplasma

Searches of the *P. falciparum* genome for genes encoding PPR proteins identified only two genes, *Pf*PPR1 (PF3D7_1406400 (PF14_0061)) and *Pf*PPR2 (PF3D7_1233300 (PFL1605W)). Both genes encode proteins with 10 PPR motifs, as predicted by TPRpred (Karpenahalli, Lupas, & Söding, 2007). *Pf*PPR1 belongs to the P-class of PPR proteins, as the repeats all comprise 35 amino acids, and the final PPR motif is situated at the C-terminus of the protein. *Pf*PPR2 may belong to the PLS class, as its PPR elements are not located at its C-terminus. Orthologues of both *Pf*PPR1 and *Pf*PPR2 were found in all *Plasmodium* species with no evidence of paralogues created by lineage-specific gene duplications.

For a protein to be targeted to the apicoplast it must contain both a signal peptide and a plastid-targeting sequence. *Pf*PPR1 and *Pf*PPR2 were analyzed by PlasmoAP and PlasMit for putative apicoplast and/or mitochondrial localization signals (Bender, van Dooren, Ralph, McFadden, & Schneider, 2003; Foth et al., 2003). *Pf*PPR1 analysis by PlasmoAP resulted in 3/4 positive tests for a signal peptide and 5/5 positive tests for an apicoplast targeting peptide, while PlasMit gave a prediction of 99% for not being mitochondrial. This is consistent with an overall strong prediction that *Pf*PPR1 traffics to the apicoplast. PlasmoAP analyses of PPR1 sequences from other *Plasmodium* species similarly predicted that most have an apicoplast localization (Supplementary Table S1), the exceptions being those encoded on genomes with a high GC content, where PlasmoAP is less accurate (Foth et al., 2003). Alignments of PPR1 show that the protein is well conserved across *Plasmodium* species (Figure S1). *Pf*PPR2 was predicted to lack both a signal peptide and a mitochondrial targeting sequence so its location is unknown, but is unlikely to be apicoplast targeted.

To test for the presence of PPRs more broadly in the Apicomplexa we searched for homologues in *Toxoplasma* and *Cryptosporidium.* Analysis of the *Toxoplasma* genome using BLAST and TPRPred, identified five PPR proteins. Only one protein (TGGT1_244050, *Tg*PPR1) contained a predicted signal peptide followed by a plastid-targeting sequence as analyzed by SignalP and iPSORT (Bannai, Tamada, Maruyama, Nakai, & Miyano, 2002; Nielsen, 2017). None of the other four proteins was predicted to include a signal peptide. This indicates that there is only one apicoplast targeted PPR protein in *Toxoplasma*, as is the case in *Plasmodium* spp. The apicomplexan *Cryptosporidium*, which has lost the apicoplast, did not contain any genes encoding PPR proteins.

To test for the localization of the *Pf*PPR1 protein, we expressed recombinant *Pf*PPR1 in *Escherichia coli*, raised poly-clonal antisera in rabbits and used immunofluorescence microscopy to locate this protein. The signal from cells stained with anti-*Pf*PPR1 co-localized with apicoplast-located GFP in the *P. falciparum* D10-ACP_L parasite line (Waller, Reed, Cowman, & McFadden, 2000) (Figure 1). This confirmed the predicted apicoplast-localization of *Pf*PPR1. A Western blot of *P. falciparum* lysate probed with the anti-*Pf*PPR1 antibody showed no detectable band, presumably due to low expression of the endogenous protein.

To test for the localization of the *Toxoplasma Tg*PPR1, a 3'-PPR1-mCherry fusion construct was created to tag the endogenous *Tg*PPR gene. A Western blot probed with antibodies to the mCherry reporter protein showed a faint band, corresponding to low expression levels of the mature fusion protein (Figure 2A, *Tg*PPR1-mCherry), but no signal was apparent by fluorescence microscopy. When the endogenous PPR promoter was replaced by the inducible t7s4 promoter, in cells designated i△*Tg*PPR1-mCherry, a higher expression level of this PPR-fusion was seen by Western blot. Furthermore, two bands were present, of apparent sizes consistent with the mature protein and a preprocessed PPR targeting intermediate still

bearing the predicted apicoplast targeting peptide (Figure 2A). The presence of these two bands is characteristic for many apicoplast-targeted proteins (Waller et al., 2000). Given the presence of only the shorter, processed band when PPR1-mCherry fusion was expressed from the native promoter, this would indicate near-complete processing under normal expression levels.

To confirm apicoplast localization, we performed an immunofluorescence assay on the i△PPR1-mCherry cells (green) co-stained with streptavidin-594 (red) which serves as an apicoplast marker due to endogenous biotinylated apicoplast proteins (Chen et al., 2015). Using this cell line we could detect *Tg*PPR1-mCherry location and observed it co-locating with the apicoplast steptavidin marker (Figure 2B). We conclude that the PPR protein is localized to the apicoplast, and is normally expressed at a very low level.

PPR1 is essential for normal growth

As PPR proteins in plants are known to be essential for chloroplast function, we tested if the apicoplast PPR1 was also important for parasite growth. Knock-down of *Tg*PPR1 in the i△*Tg*PPR1-mCherry line is induced by the addition of ATc, which represses the t7s4 promoter required for PPR1 expression. ATc treatment of i△*Tg*PPR1-mCherry showed rapid depletion of *Tg*PPR1-mCherry with the preprocessed protein undetectable within 12 hours of treatment, and no protein detected after 48 hours by Western blot (Figure 2C). To test for a growth phenotype with PPR1 depletion we used a i△*Tg*PPR1 cell line (i.e. t7s4 promoter and no mCherry fusion). Without ATc-induced depletion these cells showed normal growth by an eight-day plaque assay. With ATc treatment no plaques were observed indicating a strong growth inhibition phenotype in cells depleted of *Tg*PPR1 (Figure 2D). This same growth inhibition phenotype was also seen for i△*Tg*PPR-mCherry with ATc treatment (not shown).

These results are consistent with *Tg*PPR1 gene disruption being reported to have a negative growth phenotype in a genome-wide CRISPR knockout screen (Sidik et al., 2016). Similarly, in *Plasmodium berghei* a recent genetic screen showed that the PPR1 orthologue (PBANKA_1035800) is also essential to blood-stage growth (Bushell et al., 2017). Together these data suggest that the apicoplast PPR1 is broadly essential to apicomplexan parasites.

PfPPR1 binds* in vitro *transcribed apicoplast transcripts

We then tested if *Pf*PPR1 would bind apicoplast RNA transcripts. The recombinant *Pf*PPR1 was assessed for folding by circular dichroism which revealed a folded, alpha helical protein consistent with the alpha helical nature of PPR proteins (Supplementary Figure S2). The protein eluted as a dimer from a gel filtration column (see below), and this dimerization was confirmed by analytical ultracentrifugation following cleavage of the TRX-His_6 tag by HRV 3C protease (Supplementary Figures S3 and S4). The observed folding and dimerization is consistent with other reported plant PPR proteins (Barkan et al., 2012; Ke et al., 2013) and indicates an appropriate conformation of our purified *Pf*PPR1.

Based on the results of Nisbet et al (R. Ellen R. Nisbet et al., 2016), we generated apicoplast RNA *in vitro* transcripts spanning the *tufA* to *clpC* region of the apicoplast genome as this shows two clearly defined processing sites (at the tRNA-Phe and tRNA-Trp genes), and a second RNA transcript spanning the *LSUrRNA* to *rpoB* region that shows a processing site at tRNA-Thr. To test for *Pf*PPR1-binding we biotinylated the 3' end of each transcript and performed pull-down experiments against the *Pf*PPR1 protein. *Pf*PPR1 was observed to bind to both transcripts, Figure 3A. As a control, PPR protein was replaced by a DNA-binding

protein from *Mycobacterium smegmatis* (AmtR), which had been expressed and isolated using the same procedure as *Pf*PPR1 (Petridis et al.). No binding to AmtR was seen.

PfPPR1 shows RNA Sequence-Specific Binding

We next sought to determine if *Pf*PPR1 had a sequence-specific preference for RNA-binding. Analysis of the *Pf*PPR1 amino acid sequence using programs designed for the prediction of plant PPR RNA-binding sequences (Barkan et al., 2012; Takenaka, Zehrmann, Brennicke, & Graichen, 2013) did not result in any sequence predictions, presumably due to low sequence identity between plant and apicoplast PPR proteins. To determine any RNA sequence specificity of *Pf*PPR1, we performed SELEX (Systematic Evolution of Ligands by EXponential enrichment) (Manley, 2013). We constructed a SELEX library with a random 25 nucleotide sequence (N25) in the middle of a 150 nucleotide RNA sequence (Manley, 2013). After four rounds of selection using recombinant His_6-TRX-*Pf*PPR1, sequences containing the motif UUAU were identified in 20 of the 50 final round clones (Figure 3B), with little sequence homology amongst the other 30 final round clones, suggesting a binding motif preferred by *Pf*PPR1. This sequence motif identified by SELEX is very similar to the UUAU apicoplast RNA processing site previously identified (R. Ellen R. Nisbet et al., 2016).

To confirm that *Pf*PPR1 binds the RNA molecules containing the identified sequence motif, we performed PPR pull-down assays using a range of biotinylated 150 nucleotide RNA molecules as ‘bait’ for protein binding. The RNAs were obtained from *in vitro* transcription from five clones isolated in the final round of the SELEX experiment above. RNAs 1, 4 and 5 contained either one or two predicted PPR binding sites while RNA 2 contained a variation of the binding site and RNA 3 lacked the binding site, as shown in Figure 3C. The biotin-labelled transcripts were bound to streptavidin magnetic resin and incubated with

recombinant His_6-TRX-*Pf*PPR1 in a 2:1 protein:RNA molar ratio (as it has been reported to act as a dimer in plants). Western blots were used to detect PPR protein bound to the 'bait' RNA. RNAs 4 and 5 showed strongest bound *Pf*PPR1 and each contained two of the consensus binding motifs (Figure 3C). RNAs 1 and 2 that contain slight variations of the consensus sites, UUAA and UUAC respectively, and showed less bound *Pf*PPR1. RNA 3 did not contain the consensus site and exhibited very weak binding to PfPPR1 (Figure 3C). Thus, PfPPR1 binding correlated with presence and number of the consensus binding motif. As a control, the PPR protein was replaced by the *M. smegmatis* DNA binding protein (AmtR (Petridis et al.)) which showed no binding to the RNA transcripts (Figure 3C). We also repeated the experiment with the *Pf*PPR1 protein minus the His_6-TRX tag and with an MBP tag instead of TRX to ensure that the tag on the PPR protein did not interfere with RNA binding, and no difference in the results was seen (data not shown).

To further test the specificity of *Pf*PPR1 for the UUAU consensus motif, and not elsewhere on the RNA molecules we synthesized three 19 nucleotide RNA oligonucleotides with identical flanking sequences but differing at the potential binding motif. RNA oligos 1 and 2 contained the consensus binding sequence, followed by either an AA or a AU, while oligo 3 did not contain the binding sequence (RNA oligo 1 *UUAUAA*, RNA oligo 2 *UUAUAU*, RNA oligo 3, *UGACGA*). Each RNA oligonucleotide was biotinylated at the 3' end and used to pull-down *Pf*PPR1 or *M. smegmatis* AmtR (control), as above. Western blot analysis showed that *Pf*PPR1 was recovered from RNA oligos 1 and 2 binding assays, but not using RNA oligo 3. None of the RNA oligos bound to AmtR (Figure 3D). Together, these results demonstrate *Pf*PPR1 shows a strong preference for binding RNA at UUAU.

PfPPR1 binds RNA as a dimer

PPR proteins THA8 from *Brachypodium distachyon* (Ke et al., 2013), *Arabidopsis thaliana* HCF152 and *Zea mays* PPR10 bind RNA as a dimer (Li et al., 2014; Meierhoff, Felder, Nakamura, Bechtold, & Schuster, 2003; Yin et al., 2013), whereas *Z. mays* PPR4 and PPR5 exist as monomers. We therefore sought to determine if PPR1 bound RNA as a monomer or as a dimer. Analysis by gel filtration chromatography showed the *Pf*PPR1 protein eluted from an analytical gel filtration column at 12.73 ml with a predicted molecular weight of 144 kDa corresponding to a PPR dimer in the absence of any RNA (Figure 4A) and consistent with AUC analysis (Supp Figure 4). To test if this dimer conformation is maintained upon binding to RNA, *Pf*PPR1 was incubated with each of three 150 nt baits used above: RNA 4 and RNA 5 both containing two binding motifs; and RNA 3 that lacks the motif. Incubations were performed in a 1:1 molar ratio at room temperature for 15 minutes. Compared to the no RNA control, RNA 4- and RNA 5-incubated *Pf*PPR1 eluted earlier from the column at 10.97 ml consistent with RNA bound to the *Pf*PPR1 dimer (Figure 4A) (If *Pf*PPR1 bound RNA as a monomer we would expect the elution volume to be greater than 12.73 ml). The presence of the *Pf*PPR1 protein in the elution fraction was confirmed by SDS-PAGE (Figure 4B). The protein fraction was also treated with Proteinase K and RNA extracted by phenol/chloroform treatment. When analyzed via agarose gel electrophoresis, RNA was visible (Figure 4C). These data confirm binding of *Pf*PPR1 to the consensus RNA motif, and that this binding occurs as a PPR dimer.

PfPPR1 protects transcripts from ribonuclease activity

As the RNA consensus motif is also associated with known transcript cleavage sites, the binding of *Pf*PPR1 was predicted to be able to protect RNA from degradation by

ribonucleases. To test for *Pf*PPR1 protected footprints, we performed RNase protection assays. Three 150 nt RNA molecules with either one or two consensus binding sites (RNAs 1, 4 and 5, as before) were pre-incubated with *Pf*PPR1 protein and then incubated with the RNA endonuclease RNase A. In the absence of *Pf*PPR1 the transcripts were completely degraded by RNase A. However, with pre-incubation with *Pf*PPR1, a small RNA fragment (less than 50 nucleotides) remained after RNase A treatment in each (Figure 5A). We similarly tested for protection of three 19 nucleotide RNA oligonucleotides (RNA oligos 1, 2, and 3, as before). No degradation was evident when the RNA oligonucleotides 1 and 2 were incubated with *Pf*PPR (Figure 5B). In contrast, RNA oligonucleotide 3, which does not contain the binding sequence, was completely degraded by RNase A in the presence of *Pf*PPR1 (Figure 5B). These data show that *Pf*PPR1 protects an RNA region in a consensus binding motif-specific manner.

DISCUSSION

The discovery of a genome of plastid origin in the malaria parasite *Plasmodium* twenty-five years ago was a great surprise. Further investigation revealed the presence of small, but essential organelle subsequently called the apicoplast. We now know a considerable amount about the biochemistry and evolutionary history of this organelle. However, very little is known about how the apicoplast genome itself is transcribed, or how post-transcriptional processing is regulated. Here, we present the characterization of the single apicoplast pentatricopeptide repeat (PPR1) protein and show that it is key to the regulation of post-transcriptional RNA processing.

PPR1 is a nucleus-encoded RNA-binding protein. We confirm that it is targeted to the apicoplast in both *Plasmodium* and *Toxoplasma.* It is essential for normal growth in *Toxoplasma*, and is highly likely also to be essential in *Plasmodium* (Bushell et al., 2017). PPR1 is a P-class PPR protein as it has no additional C-terminal domains and thus no catalytic function. It is predominantly alpha helical in structure, contains 10 PPR motifs, and forms a dimer.

We show that the *Plasmodium* apicoplast PPR1 binds RNA at a UUAU-type motif. This motif is found at known cleavage sites for *Plasmodium* apicoplast transcripts (R. Ellen R. Nisbet et al., 2016). We show that the protein binds *in vitro* to apicoplast RNA transcripts containing known cleavage sites, and can protect RNA from degradation by RNase A. We have previously shown that long, polycistronic RNA molecules are the primary form of transcription in the *Plasmodium* apicoplast (R. Ellen R. Nisbet et al., 2016). These transcripts are processed into individual mRNA, tRNA and rRNA molecules often involving cleavage at

the same UUAU motif (R. Ellen R. Nisbet et al., 2016; R. E. R. Nisbet & McKenzie, 2016). Together, these results suggest that the PPR1 protein is involved in the selection of RNA transcript cleavage sites in the apicoplast.

The abundance of apicoplast transcripts over the *Plasmodium* erythrocytic life cycle is coordinated with the abundance of *Pf*PPR1 transcripts, with levels of *Pf*PPR1 transcript peaking at the trophozoite stage that immediately precedes the apicoplast transcripts peak at the schizont stage (Bozdech et al., 2003; Le Roch et al., 2003; Llinás, Bozdech, Wong, Adai, & DeRisi, 2006). This indicates that presence of PPR1 is required at the point of increase in apicoplast transcription, and consistent with a role in transcript maturation. Given that *Pf*PPR1 lacks any catalytic domain, and thus is similar to other P-class PPRs, apicoplast PPR1 binding might protect and define mature transcript ends from RNA exonucleases, facilitating their maturation. Alternatively, it could be that *Pf*PPR1 is involved in recruiting the endonuclease activity required to cleave RNA at the specific sites relevant to functional transcript production. In either case, the coordination of PPR1 expression and apicoplast transcription is likely necessary for the spike in apicoplast biosynthetic activity ahead of organelle segregation into schizonts to form the next generation of infective merozoites.

Apicomplexan PPR1 is the first plastid PPR protein to be characterized outside the green chloroplast lineage. Unlike other plant and algal taxa, there is only a single plastid PPR protein in apicomplexan plastids. Apicoplast PPR1, like PPR proteins in plants and green algae, binds at specific RNA sites, known to be the sites of RNA processing. A single and essential apicoplast PPR protein in *Plasmodium* represents a very attractive target for the design of novel antimalarial agents.

Materials and Methods

PPR alignment

PPR sequences from *Plasmodium* species were obtained by pBLAST analysis. The PF14_0061 protein alignment was generated in Geneious using a ClustalW algorithm with a BLOSUM cost matrix, a gap open cost of 10 and a gap extend cost of 0.1.

P. falciparum culture

Blood stage *P. falciparum* D10 ACP-GFP (MRA-568) was cultured according to (Tarr, Nisbet, & Howe, 2011). All work was carried out in accordance with the UK Human Tissue Act 2004.

Expression of recombinant PfPPR1

PF14_0061 (*Pf*PPR1) was codon optimized for *E. coli* and synthesized by GeneArt and cloned into the pOPIN vector system using InFusion (Takara Biotech) (Berrow et al., 2007). Expression was carried out in BL21(DE3)pLysS in ZY-5052 auto-induction medium (100 µg/ml ampicillin and 34 µg/ml chloramphenicol. PPR1 was putified via HisTrap column on an AKTA FPLC (GE Healthcare). The protein was eluted from the column using an imidazole gradient. Fractions containing His_6-TRX-PfPPR1 were pooled and concentrated using a Vivaspin concentrator (10 000 MWCO) before gel filtration chromatography using a S200 10/300 analytical size exclusion column. Fractions containing protein were analyzed by SDS-PAGE, and confirmed by MALDI-TOF MS analysis. The His_6-TRX tag was cleaved from *Pf*PPR1 by incubation of the His_6-TRX-PfPPR1 fusion protein with 1% (v/v) recombinant HRV 3C protease followed by by gel filtration chromatography. Full details are given in Supplementary methods.

PfPPR1 antibody production and purification

Recombinant *Pf*PPR1 minus the His_6-TRX tag was used to generate *Pf*PPR1 antibodies in two rabbits by Pacific Immunology. Pre-immune serum was taken followed by injection of recombinant *Pf*PPR1 plus adjuvant. Four production bleeds were taken at 2 week intervals followed by a final bleed after 3 months.

To purify the anti-*Pf*PPR1 antibody, 150 µl His affinity resin was washed with 1 ml water and three times with 1 x TBS pH 7.6. 250 µl of His_6-TRX-*Pf*PPR1 (1.7 mg/ml) was added and incubated at room temperature for 15 minutes with agitation. The supernatant was removed and the resin washed three times in 1 x TBS. 1 ml of antibody serum was added and incubated for one hour at room temperature with gentle mixing. The supernatant was removed and the resin washed four times in 1 x TBS. The bound anti-*Pf*PPR1 antibody was eluted by addition of 200 µl 0.1 M glycine pH 2.5. The supernatant was neutralized by the addition of 20 µl 1 M Tris.HCl pH 8.5 to produce purified antibody.

Immunofluorescence Microscopy for Localization of PfPPR1

Asynchronous *P. falciparum* D10 ACP_L-GFP cultures were used for immunofluorescence microscopy experiments, essentially following (Tonkin et al., 2004). Purified anti-*Pf*PPR1 antibody was diluted 1:1000 with blocking solution and AlexaFluor-568 Donkey anti-rabbit IgG was diluted 1:2000 with blocking solution. Slides were visualized using an Olympus IX81 confocal microscope at 60 x magnification. Two channels, one to detect GFP fluorescence (eGFP) and the other to detect AlexaFluor-568 fluorescence (Cy3) plus bright field were used to image slides. Images were overlaid using Fluoview version 5.0 microscopy software.

T. gondii cell culture and generation of cell lines

T. gondii RH Δku80/TATi tachyzoites were grown by inoculation in confluent human foreskin

fibroblast (HFF) cells as previously described (Striepen, 2007). Endogenous promoter replacement with the t7s4 promoter was induced by Cas9-mediated cleavage at the 5' end of the *ppr* locus. Plasmid pCRISPR/Cas9-GFP_PPR-sgRNA (see supplemental data) was assembled using the Golden Gate assembly method (Engler et al., 2014). A linear donor molecule including the t7s4 promoter and DHFR resistance gene was amplified from plasmid pPR2-HA3 (Katris et al., 2014) with primers KDPPR-Fwd and KDPPR-Rev that included flanking sequences directed to the 5' end of the *ppr* locus on either side of the Cas9 cleavage site. pCRISPR/Cas9-GFP_PPR-sgRNA and this linear donor were co-transfected into *T. gondii* TATiDku80 parasites (a kind gift from Lilach Sheiner and Boris Striepen, U. Georgia; (Sheiner et al., 2011)) and transformants selected on pyrimethamine and cloned by limiting dilution(Katris et al., 2014; Striepen, 2007). Successful promoter replacement was verified by PCR from genomic DNA. Endogenous in-frame 5' tagging of the *ppr* locus with reporter protein gene *mCherry* was achieved using plasmid pPPR-mCherry_CAT (see supplemental data) assembled using the Golden Gate method. Prior to transfection of parasites this plasmid was linearized with *Bam*HI, and transformants were selected with chloramphenicol and cloned by limiting dilutions (Striepen, 2007).

Toxoplasma PPR assays

Western blot detection of SDS-PAGE resolved whole cell lysates was performed using a rabbit anti-mCherry (1/1000 dilution) (Abcam) and anti-TOM40 as a control (Katris et al., 2014). Immunofluorescence microscopy was performed on intracellular tachyzoites using anti-mCherry (1/1000) with secondary antibody AlexaFluor-488 Goat anti-rabbit IgG (Life technology). Apicoplasts were co-stained with AlexaFluor-594 anti-Steptavidine (Life technology)(Chen et al., 2015). Samples were mounted with ProLong Diamond antifade mountant with DAPI (Invitrogen) and sealed with nail polish. Cells were imaged using an Inverted Nikon Eclipse Ti microscope, a Nikon objective lens (Plan APO, 100x/1.45 oil), and a Hamamatsu C11440, ORCA Flash 4.0 camera.

PPR knockdown was induced by addition of anhydrotetracycline (ATc) (0.5 µg/ml) to the growth medium upon parasite inoculation of HFF cells. For growth assays extracellular parasites were filtered, counted by haemocytometer, and 500 parasites added to 25 cm^2 tissue culture flasks containing a confluent monolayer of HFF cells. To visualize plaque sizes in the presence or absence of ATc, flasks were aspirated, fixed with 5 ml 100% ethanol (5 minutes), stained with 5 ml crystal violet solution (15 minutes) then washed once with phosphate-buffered saline (PBS) and dried before imaging.

SELEX for determination of PfPPR1 RNA sequence specificity

A SELEX library was constructed as in (Manley, 2013). SELEX using recombinant His_6-TRX-*Pf*PPR1 was carried out as per the protocol in (Manley, 2013). Four rounds of selection were performed, following which PCR products were cloned into pGEM-T easy (Promega) and transformed into chemically competent *E. coli* DH5α and plated onto LB agar (100 µg/ml ampicillin, 0.1 mM IPTG and 40 µg/ml X-gal). Plasmids were extracted from 50 clones and sequenced. In addition, 10 input clones were sequenced from each round to determine enrichment.

In vitro RNA Transcription and 3' End Biotinylation

Apicoplast *P. falciparum* apicoplast PCR products were obtained using the following primers: LSUrRNA Fwd/rpoB Rev and tufA Fwd. clpC Rev. T7 promoter sequences (TAATACGACTCACTATAG) were added in a further round of PCR with the T7 promoter sequence appended to the 5' end of the forward primer. Additionally, PCR products were obtained from SELEX clones (above) The Ambion T7 MEGAScript Kit was used for *in vitro* transcription. For 3' end biotinylation 50 pmol RNA transcript was heated at 85 °C for 3-5 minutes. Once on ice, 3 µl 10 x T4 RNA ligase buffer (NEB), 1 µl rRNasin (Promega), 50 pmol RNA, 1 µl pCp Biotin, 2 µl T4 RNA Ligase (NEB), water to 15 µl and 15 µl 30% PEG

was added. Ligation reactions were incubated overnight at 16 °C. 70 µl water was added followed by 100 µl chloroform:isoamyl alcohol (49:1). Reactions were centrifuged at 13 000 x g for three minutes and the upper phase (aqueous layer) removed and transferred to a new tube. 10 µl 3 M sodium acetate pH 5.2 and 250 µl 100% ethanol were added to precipitate RNA and stored at -20°C. The RNA was pelleted by centrifugation at 13 000 x g, 4 °C for 20 minutes, washed with 70% ethanol and centrifuged again. The pellet was resuspended in 20 µl water and RNA was quantified with a Nanodrop-1000.

Biotinylated RNA-PfPPR1 Pull-downs

25 µl of resuspended streptavidin magnetic beads (Thermo Scientific) were washed in 1 ml water and then three times in binding buffer (10 mM HEPES pH 7.5, 20 mM KCl, 1 mM $MgCl_2$, 1 mM DTT) and resuspended in 20 µl binding buffer. 400 nM biotinylated RNA transcript was added and incubated at 4 °C for 15 minutes. The beads were washed three times in 1 ml binding buffer and resuspended in 20 µl binding buffer. 800 nM His_6-TRX-*Pf*PPR1 was added followed by incubation at room temperature for 15 minutes with gentle agitation. The beads were washed three times with 1 ml binding buffer and resuspended in 25 µl 4 x SDS loading dye, heated to 100 °C for 5 minutes and loaded onto a 4-15% SDS-PAGE gel (BioRad). Following PAGE, the samples were transferred via wetblot to PVDF membrane for western blot analysis with a mouse-anti His antibody and a secondary goat anti-mouse antibody conjugated with HRP. Western blots were visualized using Western Bright Quantum (Advasnsta) chemiluminiscent substrate and exposed using a CCD camera (Genebox).

Gel Filtration Chromatography

His_6-TRX-*Pf*PPR1 was incubated with RNA transcripts in a 1:1 molar ratio in 1 x binding buffer (10 mM HEPES pH 7.5, 20 mM KCl, 1 mM $MgCl_2$ and 1 mM DTT) for 15 minutes at

room temperature. A control without RNA was included in these reactions. Samples were analyzed using a S200 10/300 analytical gel filtration column pre-equilibrated in 50 mM Tris pH 8.0, 200 mM NaCl, 5 mM DTT.

RNase A protection assays

10 µM RNA, 1 x binding buffer buffer (10 mM HEPES pH 7.5, 20 mM KCl, 1 mM $MgCl_2$ and 1 mM DTT) and 10 µM PfPPR1 in a volume of 20 µl were incubated for 15 minutes at room temperature, followed by addition of 0.01% (v/v) 20 mg/ml RNase A and incubated at 37 °C with shaking for 30 minutes. Reactions were stopped by the addition of 10 µl 2 x formamide loading dye (95% formamide, 0.025% w/v bromophenol blue, 0.025% xylene cyanol FF, 5 mM EDTA) and heating to 70 °C for 5 minutes. Assay reactions were analyzed by 12% urea-denaturing PAGE gel, and visualized with SYBR Safe nucleic acid stain. Three controls (no *Pf*PPR1, no RNA, no RNase) were carried out, where the reagent was replaced by water.

ADDITIONAL INFORMATION

We thank the numerous anonymous blood donors. The research project was approved by the Human Biology Research Ethics Committee, School of Biological Sciences, University of Cambridge, project 2007.04. All data are included in this article and its supplementary information files. We have no competing interests. This work was funded by a Wellcome Trust project grant (094249) to CJH and RERN, a MRC grant (MR/M011690/1) to RFW, an MRC Confidence in Concept grant to RERN and CJH and a University of Cambridge Returning Carers Award to RERN. All authors gave final approval for publication.

Author contributions: JLH, IL, EC, ME, AV and RERN performed the experiments. RFW provided expertise in *Toxoplasma*. JLH, RFW, CJH and RERN drafted the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

We thank Dr Katherine Stott at the Biophysics facility in the Department of Biochemistry at the University of Cambridge for assistance with CD and AUC experiments, and Dr. Joanna-Marie Howes for assistance with confocal microscopy. We thank Prof Mark Carrington, Dr Erin Butterfield and Dr Richard Dorrell for helpful discussion.

References

Bannai, H., Tamada, Y., Maruyama, O., Nakai, K., & Miyano, S. (2002). Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics,* **18(2)**, 298-305.

Barkan, A., Rojas, M., Fujii, S., Yap, A., Chong, Y. S., Bond, C. S., & Small, I. (2012). A Combinatorial Amino Acid Code for RNA Recognition by Pentatricopeptide Repeat Proteins. *PLoS Genet,* **8(8**), e1002910.

Barkan, A., & Small, I. (2014). Pentatricopeptide Repeat Proteins in Plants. *Ann Rev Plant Biol,* **65(1)**, 415-442.

Bender, A., van Dooren, G. G., Ralph, S. A., McFadden, G. I., & Schneider, G. (2003). Properties and prediction of mitochondrial transit peptides from Plasmodium falciparum. *Mol Biochem Parasitol,* **132(2)**, 59-66.

Berrow, N. S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., . . . Owens, R. J. (2007). A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res,* **35(6)**, e45-e45.

Bozdech, Z., Llinás, M., Pulliam, B. L., Wong, E. D., Zhu, J., & DeRisi, J. L. (2003). The Transcriptome of the Intraerythrocytic Developmental Cycle of *Plasmodium falciparum*. *PLoS Biol,* **1(1)**, e5.

Brown, G. G., Formanová, N., Jin, H., Wargachuk, R., Dendy, C., Patil, P., et al. (2003). The radish Rfo restorer gene of Ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. *The Plant Journal,* **35(2)**, 262-272.

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. *Plant Physiol,* **155(4)**, 1678-1689.

Bushell, E., Gomes, A. R., Sanderson, T., Anar, B., Girling, G., Herd, C., et al. (2017). Functional Profiling of a Plasmodium Genome Reveals an Abundance of Essential Genes. *Cell,* **170(2)**, 260-272.e268.

Chen, A. L., Kim, E. W., Toh, J. Y., Vashisht, A. A., Rashoff, A. Q., Van, C., et al (2015). Novel Components of the Toxoplasma Inner Membrane Complex Revealed by BioID. *mBio,* **6(1**).

Engler, C., Youles, M., Gruetzner, R., Ehnert, T.-M., Werner, S., Jones, J. D. G. et al (2014). A Golden Gate Modular Cloning Toolbox for Plants. *ACS Synthetic Biology,* **3(11)**, 839-843.

Foth, B. J., Ralph, S. A., Tonkin, C. J., Struck, N. S., Fraunholz, M., Roos, D. S. et al. (2003). Dissecting Apicoplast Targeting in the Malaria Parasite *Plasmodium falciparum*. *Science, 299*(5607), 705-708.

García-Mauriño, S. M., Díaz-Quintana, A., Rivero-Rodríguez, F., Cruz-Gallardo, I., Grüttner, C., Hernández-Vellisca, M., & Díaz-Moreno, I. (2018). A putative RNA binding protein from Plasmodium vivax apicoplast. *FEBS Open Bio,* **8(2)**, 177-188.

Gardner, M. J., Williamson, D. H., & Wilson, R. J. M. (1991). A circular DNA in malaria parasites encodes an RNA polymerase like that of prokaryotes and chloroplasts. *Mol Biochem Parasitol,* **44(1)**, 115-123.

Howe, C. J. (1992). Plastid origin of an extrachromosomal DNA molecule from Plasmodium, the causative agent of malaria. *J. Theor Biol,* **158**, 199-205.

Karpenahalli, M. R., Lupas, A. N., & Söding, J. (2007). TPRpred: a tool for prediction of TPR-, PPR- and SEL1-like repeats from protein sequences. *BMC Bioinformatics,* **8,** 2-2.

Katris, N. J., van Dooren, G. G., McMillan, P. J., Hanssen, E., Tilley, L., & Waller, R. F. (2014). The Apical Complex Provides a Regulated Gateway for Secretion of Invasion Factors in Toxoplasma. *PLOS Pathogens,* **10(4**), e1004074.

Ke, J., Chen, R.-Z., Ban, T., Zhou, X. E., Gu, X., Tan, M. H. E. et al. (2013). Structural basis for RNA recognition by a dimeric PPR-protein complex. *Nat Struct Mol Biol,* **20,** 1377-1382.

Laue, T. M., Shah, B. D., Ridgeway, T. M., & Peletier, S. L. (1992). *Analytical ultracentrifugation in biochemistry and polymer science*. Cambridge, UK: The Royal Society of Chemistry.

509 Le Roch, K. G., Zhou, Y., Blair, P. L., Grainger, M., Moch, J. K., Haynes, J. D. et al (2003). Discovery
510 of Gene Function by Expression Profiling of the Malaria Parasite Life Cycle. *Science,*
511 **301(5639)**, 1503-1508.
512 Li, Q., Yan, C., Xu, H., Wang, Z., Long, J., Li, W., . . . Yan, N. (2014). Examination of the dimerization
513 states of the single-stranded RNA recognition protein pentatricopeptide repeat 10
514 (PPR10). *J Biol Chem,* **289(45)**, 31503-31512.
515 Llinás, M., Bozdech, Z., Wong, E. D., Adai, A. T., & DeRisi, J. L. (2006). Comparative whole genome
516 transcriptome analysis of three *Plasmodium falciparum* strains. *Nucleic Acids Res,* **34(4),**
517 1166-1173.
518 Lurin, C., Andrés, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyère, C., et al (2004). Genome-Wide
519 Analysis of Arabidopsis Pentatricopeptide Repeat Proteins Reveals Their Essential Role
520 in Organelle Biogenesis. *The Plant Cell,* **16(8)**, 2089-2103.
521 Manley, J. L. (2013). SELEX to Identify Protein-Binding Sites on RNA. *Cold Spring Harbor*
522 *protocols,* **2013(2)**, 156-163.
523 Manna, S. (2015). An overview of pentatricopeptide repeat proteins and their applications.
524 *Biochimie,* **113,** 93-99.
525 McFadden, G. I., Reith, M. E., Munholland, J., & Lang-Unnasch, N. (1996). Plastid in human
526 parasites. *Nature,* **381,** 482.
527 Meierhoff, K., Felder, S., Nakamura, T., Bechtold, N., & Schuster, G. (2003). HCF152, an
528 Arabidopsis RNA binding pentatricopeptide repeat protein involved in the processing of
529 chloroplast psbB-psbT-psbH-petB-petD RNAs. *Plant Cell,* **15(6)**, 1480-1495.
530 Nielsen, H. (2017). Predicting Secretory Proteins with SignalP. In D. Kihara (Ed.), *Protein*
531 *Function Prediction: Methods and Protocols* (pp. 59-73). New York, NY: Springer New
532 York.
533 Nisbet, R. E. R., Kurniawan, D. P., Bowers, H. D., & Howe, C. J. (2016). Transcripts in the
534 *Plasmodium* Apicoplast Undergo Cleavage at tRNAs and Editing, and Include Antisense
535 Sequences. *Protist,* **167,** 377-388.
536 Nisbet, R. E. R., & McKenzie, J. L. (2016). Transcription of the apicoplast genome. *Mol Biochem*
537 *Parasitol,* **210(1)**, 5-9. d
538 Petridis, M., Vickers, C., Robson, J., McKenzie, J. L., Bereza, M., Sharrock, A., . . . Cook, G. M. (2016).
539 Structure and function of AmtR in *Mycobacterium smegmatis*: Implications for post-
540 transcriptional regulation of urea metabolism through a small antisense RNA. *J. Mol Biol,*
541 **428(21),** 4315-4329.
542 Pfalz, J., Bayraktar, O. A., Prikryl, J., & Barkan, A. (2009). Site-specific binding of a PPR protein
543 defines and stabilizes 5' and 3' mRNA termini in chloroplasts. *EMBO,* **28(14)**, 2042-2052.
544 Prikryl, J., Rojas, M., Schuster, G., & Barkan, A. (2011). Mechanism of RNA stabilization and
545 translational activation by a pentatricopeptide repeat protein. *Proc Natl Acad Sci,*
546 **108(1)**, 415-420.
547 Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity
548 ultracentrifugation and lamm equation modeling. *Biophys J,* **78(3)**, 1606-1619.
549 Sheiner, L., Demerly, J. L., Poulsen, N., Beatty, W. L., Lucas, O., Behnke, M. S. et al (2011). A
550 Systematic Screen to Discover and Analyze Apicoplast Proteins Identifies a Conserved
551 and Essential Protein Import Factor. *PLOS Pathogens,* **7(12)**, e1002392.
552 Sidik, S. M., Huet, D., Ganesan, S. M., Huynh, M.-H., Wang, T., Nasamu, A. S., et al (2016). A
553 Genome-wide CRISPR Screen in Toxoplasma Identifies Essential Apicomplexan Genes.
554 *Cell,* **166(6)**, 1423-1435.
555 Sosso, D., Canut, M., Gendrot, G., Dedieu, A., Chambrier, P., Barkan, A. et al (2012). PPR8522
556 encodes a chloroplast-targeted pentatricopeptide repeat protein necessary for maize
557 embryogenesis and vegetative development. *J. Exp Bot,* **63(16)**, 5843-5857.
558 Sosso, D., Mbelo, S., Vernoud, V., Gendrot, G., Dedieu, A., Chambrier, P., et al (2012). PPR2263, a
559 DYW-Subgroup Pentatricopeptide Repeat Protein, Is Required for Mitochondrial
560 Transcript Editing, Mitochondrion Biogenesis, and Maize Growth. *The Plant Cell,* **24(2),**
561 676-691.

562 Striepen, B., Soldati, D. (2007). Genetic manipulation of *Toxoplasma gondi*i. In L. M. Weiss, Kim, K.
563 (Ed.), *Toxoplasma gondii. The Model Apicomplexan-Perspectives and Methods* (pp. 391-
564 415). London: Elsevier.
565 Takenaka, M., Zehrmann, A., Brennicke, A., & Graichen, K. (2013). Improved Computational
566 Target Site Prediction for Pentatricopeptide Repeat RNA Editing Factors. *PLoS ONE,* **8(6)**,
567 e65343.
568 Tarr, S. J., Nisbet, R. E. R., & Howe, C. J. (2011). Transcript-level responses of *Plasmodium*
569 *falciparum* to thiostrepton. *Mol Biochem Parasitol,* **179(1),** 37-41.
570 Tonkin, C. J., van Dooren, G. G., Spurck, T. P., Struck, N. S., Good, R. T., Handman, E., . . . McFadden,
571 G. I. (2004). Localization of organellar proteins in *Plasmodium falciparum* using a novel
572 set of transfection vectors and a new immunofluorescence fixation method. *Molecular*
573 *and Biochemical Parasitology,* **137(1),** 13-21.
574 Tourasse, N. J., Choquet, Y., & Vallon, O. (2013). PPR proteins of green algae. *RNA Biology,* **10(9)**,
575 1526-1542.
576 Waller, R. F., Reed, M. B., Cowman, A. F., & McFadden, G. I. (2000). Protein trafficking to the
577 plastid of *Plasmodium falciparum* is via the secretory pathway. *EMBO J, 19*(8), 1794-
578 1802.
579 Wilson, R. J. M., Denny, P. W., Preiser, P. R., Rangachari, K., Roberts, K., Roy, A., et al (1996).
580 Complete Gene Map of the Plastid-like DNA of the Malaria Parasite *Plasmodium*
581 *falciparum*. *J. Mol Biol,* **261(2)**, 155-172.
582 Yin, P., Li, Q., Yan, C., Liu, Y., Liu, J., Yu, F., . . . Yan, N. (2013). Structural basis for the modular
583 recognition of single-stranded RNA by PPR proteins. *Nature,* **504(7478)**, 168-171.
584 Zhelyazkova, P., Hammani, K., Rojas, M., Voelker, R., Vargas-Suárez, M., Börner, T., & Barkan, A.
585 (2012). Protein-mediated protection as the predominant mechanism for defining
586 processed mRNA termini in land plant chloroplasts. *Nucleic Acids Res,* **40(7)**, 3092-3105.

587

1

1

Figure 1

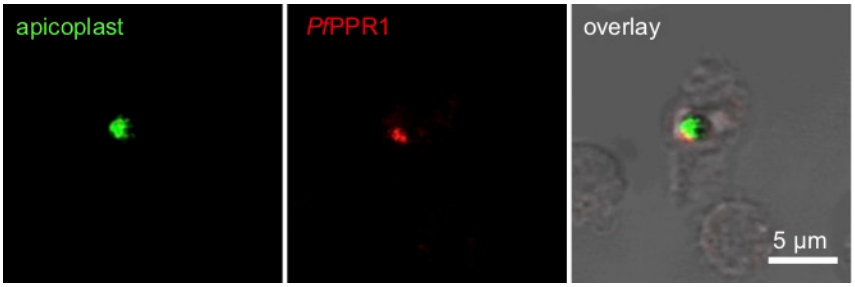


Figure 2

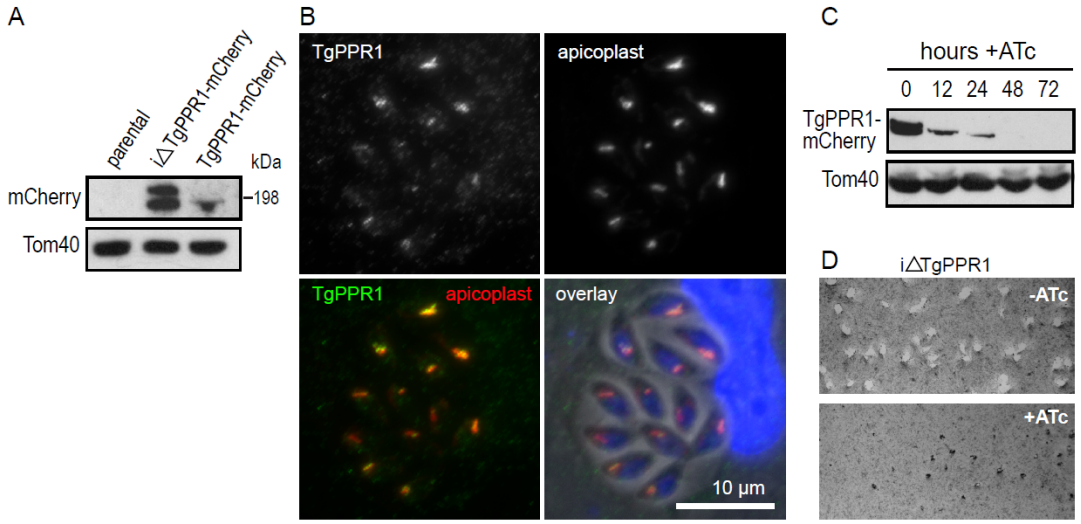

A
parental
iΔTgPPR1-mCherry
TgPPR1-mCherry
kDa
mCherry
198
Tom40
B
TgPPR1
apicoplast
TgPPR1
apicoplast
overlay
10 µm
C
hours +ATc
0 12 24 48 72
TgPPR1-mCherry
Tom40
D
iΔTgPPR1
-ATc
+ATc

Figure 3

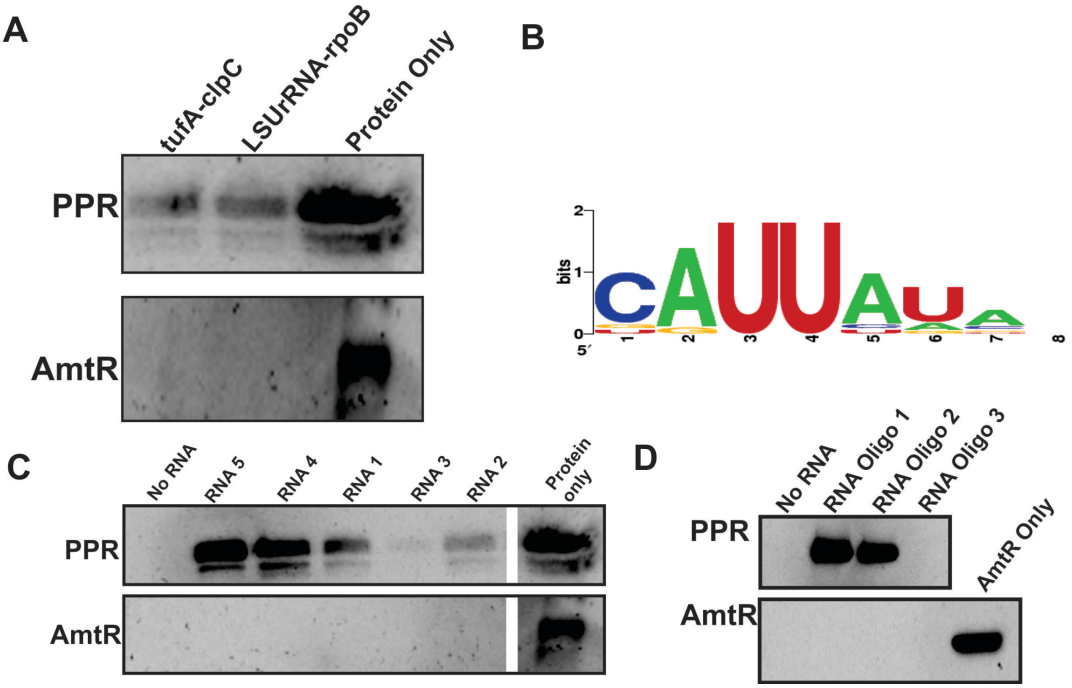

A
tufA-clpC
LSUrRNA-rpoB
Protein Only
PPR
AmtR
B
bits
2
1
0
5'
1
2
3
4
5
6
7
8
C
No RNA
RNA 5
RNA 4
RNA 1
RNA 3
RNA 2
Protein only
PPR
AmtR
D
No RNA
RNA Oligo 1
RNA Oligo 2
RNA Oligo 3
AmtR Only
PPR
AmtR

Figure 4

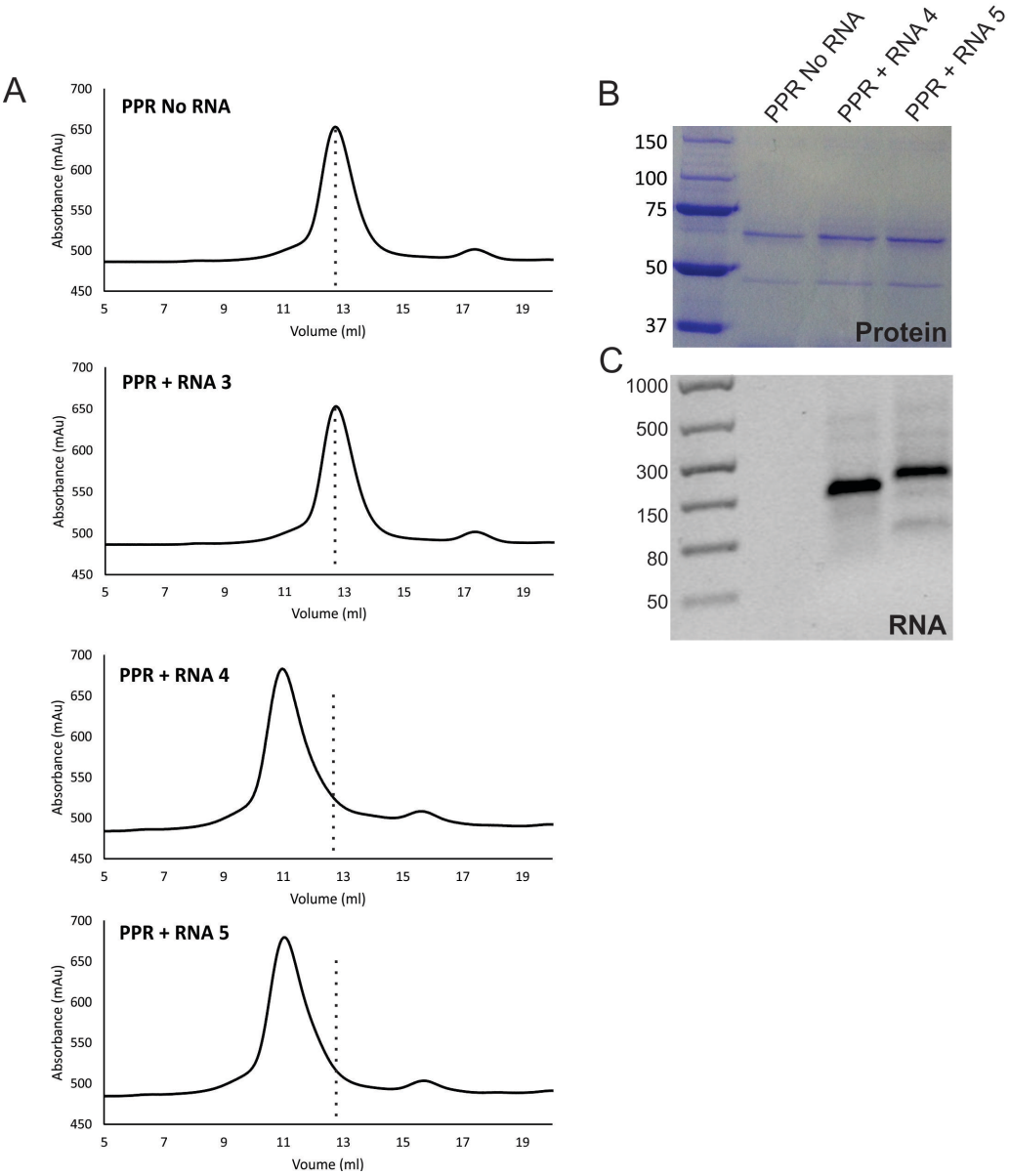
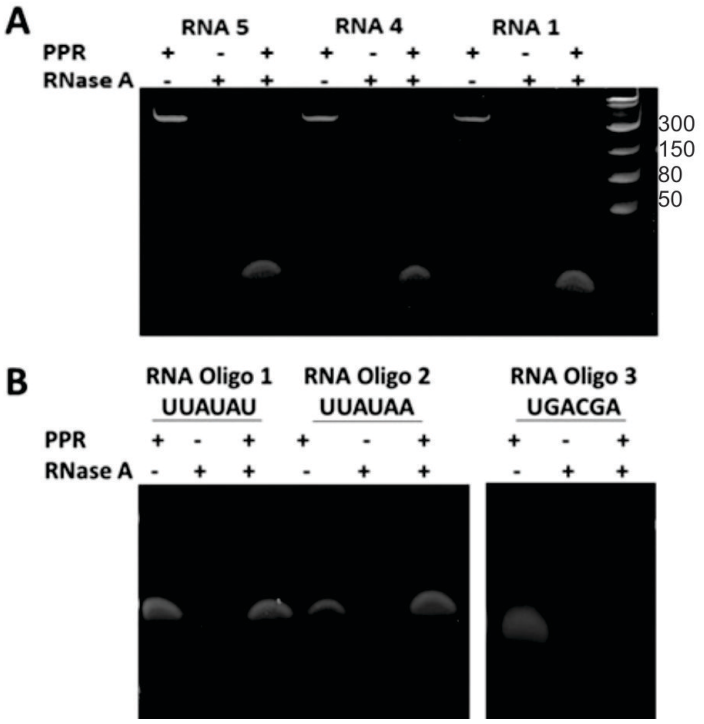

A
PPR No RNA
PPR + RNA 3
PPR + RNA 4
PPR + RNA 5
Absorbance (mAu)
700
650
600
550
500
450
5
7
9
11
13
15
17
19
Volume (ml)
Voume (ml)
B
PPR No RNA
PPR + RNA 4
PPR + RNA 5
150
100
75
50
37
Protein
C
1000
500
300
150
80
50
RNA

Figure 5


A
RNA 5
RNA 4
RNA 1
PPR
RNase A
300
150
80
50
B
RNA Oligo 1
UUAUAU
RNA Oligo 2
UUAUAA
RNA Oligo 3
UGACGA
PPR
RNase A