


# SCIENTIFIC REPORTS



OPEN

## The 27 kDa *Trypanosoma brucei* Pentatricopeptide Repeat Protein is a G-tract Specific RNA Binding Protein

Pakoyo F. Kamba<sup>1,2,4</sup>, David A. Dickson<sup>1,5</sup>, Neil A. White<sup>1,6</sup>, Jennifer L. Ekstrom<sup>1,7</sup>, Donna J. Koslowsky<sup>2,3</sup> & Charles G. Hoogstraten<sup>1</sup> 

Pentatricopeptide repeat (PPR) proteins, a helical repeat family of organellar RNA binding proteins, play essential roles in post-transcriptional RNA processing. In *Trypanosoma brucei*, an expanded family of PPR proteins localize to the parasite's single mitochondrion, where they are believed to perform important roles in both RNA processing and translation. We studied the RNA binding specificity of the simplest *T. brucei* PPR protein (KRIPP11) using electrophoretic mobility shift assays, fluorescence anisotropy, circular dichroism spectroscopy, and *in vitro* selection. We found KRIPP11 to be an RNA binding protein with specificity for sequences of four or more consecutive guanosine residues (G-tracts). Such G-tracts are dramatically enriched in *T. brucei* mitochondrial transcripts that are destined for extensive uridine insertion/deletion editing but are not present in mRNAs following editing. We further found that the quadruplex oligoguanosine RNA conformation is preferentially recognized by KRIPP11 over other conformational forms, and is bound without disruption of the quadruplex structure. In combination with prior data demonstrating association of KRIPP11 with the small ribosomal subunit, these results suggest possible roles for KRIPP11 in bridging mRNA maturation and translation or in facilitating translation of unusual dual-coded open reading frames.

Trypanosomes, including the causative agent of human African trypanosomiasis (HAT) *Trypanosoma brucei*, are characterized by a single, large mitochondrion (kinetoplast) containing a disc-shaped DNA genome, itself made from a concatenated network of large DNA maxicircles and small DNA minicircles<sup>1,2</sup>. This unusual mitochondrion is central to the life cycle and pathogenesis of kinetoplastid parasites. In *T. brucei*, modulation of mitochondrial activity facilitates parasite adaptation from the sugar-deficient tsetse fly gut, where it depends on oxidative phosphorylation, to the sugar-rich environment in human blood, where it depends on glycolysis for energy<sup>3,4</sup>. In these parasites, both nuclear and mitochondrial genes are polycistronic and most regulation of gene expression occurs at the RNA level<sup>5–9</sup>. Strikingly, 12 of the 18 protein coding genes in trypanosome mitochondrial DNA are composed of incomplete open reading frames (ORFs) which are post-transcriptionally converted into translatable mRNAs by a process of systematic uridine insertion and/or deletion called RNA editing<sup>10</sup>. In fact, the primary (pre-edited) transcripts of nine of the *T. brucei* mitochondrial mRNA genes undergo such massive editing (pan-editing) that approximately 50% of the sequence of mature mRNA consists of externally inserted uridines<sup>10,11</sup>. RNA editing by nucleotide insertion and deletion is unique to trypanosomes<sup>12</sup>, and thus serves as a potential target for pharmacological intervention.

RNA editing and other post-transcriptional RNA processing events in trypanosome mitochondria are executed by a variety of RNA binding proteins whose roles and mechanisms of action are only now being

<sup>1</sup>Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan, 48824-1319, USA. <sup>2</sup>Graduate Program in Cell and Molecular Biology, Michigan State University, East Lansing, Michigan, 48824-1319, USA. <sup>3</sup>Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan, 48824-1319, USA. <sup>4</sup>Makerere University, Kampala, Uganda. <sup>5</sup>Present address: Sackler School of Graduate Biomedical Sciences, Tufts University, Medford, Massachusetts, USA. <sup>6</sup>Present address: Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, Connecticut, USA. <sup>7</sup>Present address: Department of Food Science and Human Nutrition, Michigan State University, East Lansing, Michigan, USA. Correspondence and requests for materials should be addressed to C.G.H. (email: [hoogstr3@msu.edu](mailto:hoogstr3@msu.edu))

characterized<sup>10,13–17</sup>. Deciphering the RNA binding characteristics of these important proteins is therefore critical to our understanding of the relationship between specific RNA targeting and biological function. Among trypanosome mitochondrial RNA binding proteins, the pentatricopeptide repeat (PPR) family of sequence-specific RNA binding proteins has proven to be particularly important. First discovered in *Arabidopsis thaliana* in 2000, PPR proteins are characterized by tandem repeats of 35 amino acids, with each motif folding into a pair of antiparallel alpha-helices similar to the peptide-binding tetratricopeptide (34 amino acid) repeat (TPR) motif<sup>18,19</sup>. Over the last decade, these proteins have been linked to a wide variety of stages of post-transcriptional RNA processing in chloroplasts and plant mitochondria<sup>20–24</sup>. Typically, plant genomes code for hundreds of PPR proteins whereas non-plant eukaryotes contain fewer than ten<sup>25</sup>. An exception is found in trypanosomes, where approximately 40 PPR proteins have been reported in *T. brucei*<sup>14,26,27</sup>. These proteins vary dramatically both in the number of PPR repeats and in the presence and nature of accessory non-PPR domains. PPR proteins are highly conserved across trypanosomatids and RNAi knockdowns cause parasite growth retardation, growth arrest, and death, consistent with functional essentiality<sup>16,26–28</sup>. Underlying the severe phenotypes is deterioration in oxidative phosphorylation and thus in mitochondrial function<sup>16,26–29</sup>.

Proteomic and genetic analyses have provided some insights into the functions of *T. brucei* PPR proteins. Most *T. brucei* PPR proteins co-purify with polyadenylation complexes and/or mitochondrial ribosomal subunits, suggesting a role in their respective functions<sup>29–32</sup>. For PPR proteins in which the effect of RNAi on mRNA processing has been reported, namely, KPAF1, KRIPP1 and KRIPP8, compromised poly(A/U) synthesis and translation of mitochondrial mRNA has been observed<sup>29,32</sup>. However, knockdown of different PPR proteins often affects specific post-transcriptional processes and specific gene transcripts<sup>26–29,32</sup>, suggesting their cognate RNA sequences are unique.

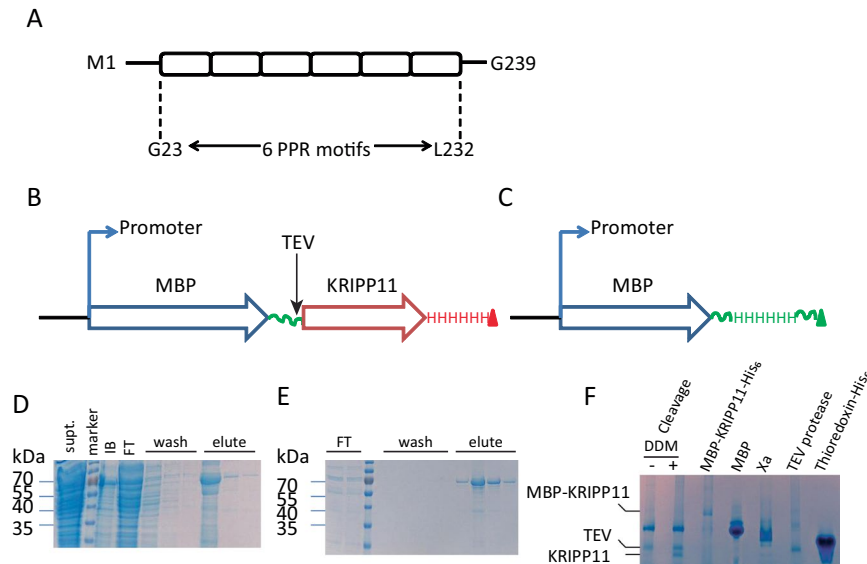
Identifying the specific RNA ligands for PPR proteins is crucial to elucidating their mechanism(s) of RNA binding. Unfortunately, poor protein solubility and difficulty in heterologous expression<sup>25,33</sup> have limited such studies to a restricted subset of plant PPR proteins. These studies have shown that PPR proteins exhibit high sequence specificity in RNA recognition. For example, maize PPR10 protects and defines both the 5' and 3' termini of transcripts from two different loci (*atpH* and *psaI*) by binding to a highly conserved RNA motif located in the intergenic regions downstream of each gene<sup>34,35</sup>. Another PPR protein, CRR4 from *Arabidopsis thaliana*, specifies the site of RNA editing in the chloroplast *ndhD* gene by recognizing a specific motif surrounding the editing site<sup>24,36</sup>. In maize, PPR103 protects the 5'-terminus of processed *rpl16* transcripts from exonucleolytic degradation<sup>37</sup>. Numerous examples of this type have led to a general picture of plant PPR proteins as sequence-specific effectors and modulators of specialized organellar RNA metabolism<sup>23,38</sup>.

For the plant PPR proteins whose specific RNA ligands are known, progress has been made in elucidating their RNA binding mechanisms via biochemical studies with truncated PPR proteins, computational modeling, and x-ray crystallography. Biochemical studies have shown retention of RNA binding in severely truncated proteins, suggesting that the RNA binding activity resides in the PPR modules. Computational studies by phylogenetic analysis of residue conservation across numerous PPR motifs<sup>39</sup> and iterative alignment of residues at each position in the PPR motif with bases in the cognate RNA footprint have suggested a common code of RNA binding in which each PPR repeat recognizes a single nucleotide via RNA contacts from amino acids at positions 1, 3 and 6<sup>40–42</sup>. Experimental studies using engineered PPR proteins have provided strong support for this mechanism of single-stranded RNA recognition<sup>43</sup>. Crystal structures of two plant-derived PPR proteins in complex with RNA targets (maize PPR10 and *Arabidopsis thaliana* THA8) confirmed the recognition of individual nucleotides of single-stranded RNA by individual PPR repeats and elucidated the basis for many aspects of the emerging recognition code<sup>44–46</sup>. These correspondences have been more finely mapped by a set of four crystal structures of designed PPR proteins (dPPRs) in complex with their predicted RNA targets<sup>47</sup>.

In contrast to plants, the specific RNA sequences recognized by individual *T. brucei* PPR proteins have hardly been elucidated, hindering progress in their mechanistic studies and subsequent biotechnological applications. Trypanosomal PPR proteins vary greatly in the number of PPR repeats and in the number and nature of accessory domains. In the initial work reported here, we focused on the 27 kDa *T. brucei* PPR protein KRIPP11 (TriTrypDB Tb927.8.6040; genbank XM\_842341; previously PPR27), which is one of the smallest and simplest members of the protein family. The four PPR repeats originally identified in this protein<sup>26</sup>, together with two additional 35-residue regions predicted to carry strong helical content, yield a probable total of six tandem PPR repeats in 239 amino acids and no identifiable accessory domains (Fig. 1A). Though KRIPP11 was initially thought to be a cytoplasmic protein due to negative mitochondrial targeting by classical computational tools<sup>27</sup>, contemporary computational tools support mitochondrial localization (Supplementary Material, Table S1). Mitochondrial localization is experimentally supported by a strong kinetoplast signal in imaging data from the TrypTag database<sup>48</sup> and by the demonstration of stable RNA-independent association with mitochondrial small subunit ribosomal complexes<sup>31,32</sup>. In this work, we use an integrated approach of biophysical probing of RNA affinity using electrophoretic shift mobility assay (EMSA) and equilibrium fluorescence anisotropy and *in vitro* selection of protein ligands to derive significant insight into the binding specificity and function of this protein. We find that KRIPP11 shows a strong preference for sequences containing G-tracts of four to six residues or more, such as are extensively found in the subset of kinetoplastid pre-mRNAs destined to undergo extensive uridine insertion/deletion editing. Preliminary *in vitro* evidence suggests targeting of G-tracts in quadruplex RNA form.

## Results

**Heterologous expression of soluble, monomeric KRIPP11.** KRIPP11 was expressed as a fusion protein with maltose binding protein (MBP) to promote yield and solubility. Using the constructs in Fig. 1B,C, we obtained a mean soluble yield of 5.0 mg per liter of culture for the MBP-KRIPP11 fusion protein. After tandem affinity chromatography using sequential Ni-NTA and amylose resins, fusion protein of over 90% purity was obtained (Fig. 1D,E). Following removal of the MBP tag with TEV protease, the oligomeric state of mature



**Figure 1.** (A) Organization of PPR motifs in KRIPP11. (B) Scheme of the overexpression construct for the MBP-KRIPP11 fusion protein emphasizing the His<sub>6</sub> tag at the C-terminus of KRIPP11. TEV protease cleavage site, which leaves two additional residues (GS) at the N-terminus of KRIPP11, is indicated. (C) Overexpression construct for free MBP with a His<sub>6</sub> tag at the C-terminus. (D) SDS-PAGE for MBP-KRIPP11-His<sub>6</sub> fractions from the Ni-NTA column. Supt., supernatant; m, PageRuler™ size marker; IB, inclusion bodies; FT, flow through; three lanes of wash and three of eluate. (E) SDS-PAGE of MBP-KRIPP11-His<sub>6</sub> fractions from the amylose column. FT, two lanes flow through; m, marker; four lanes of wash and four of eluate. (F) 6–12% gradient native polyacrylamide gel showing the monomeric state of KRIPP11. TEV cleavage mixture is shown in the absence and presence of DDM detergent (see text). For this and subsequent figures, uncropped originals of all gel images are available as Supplementary Information (Fig. S2).

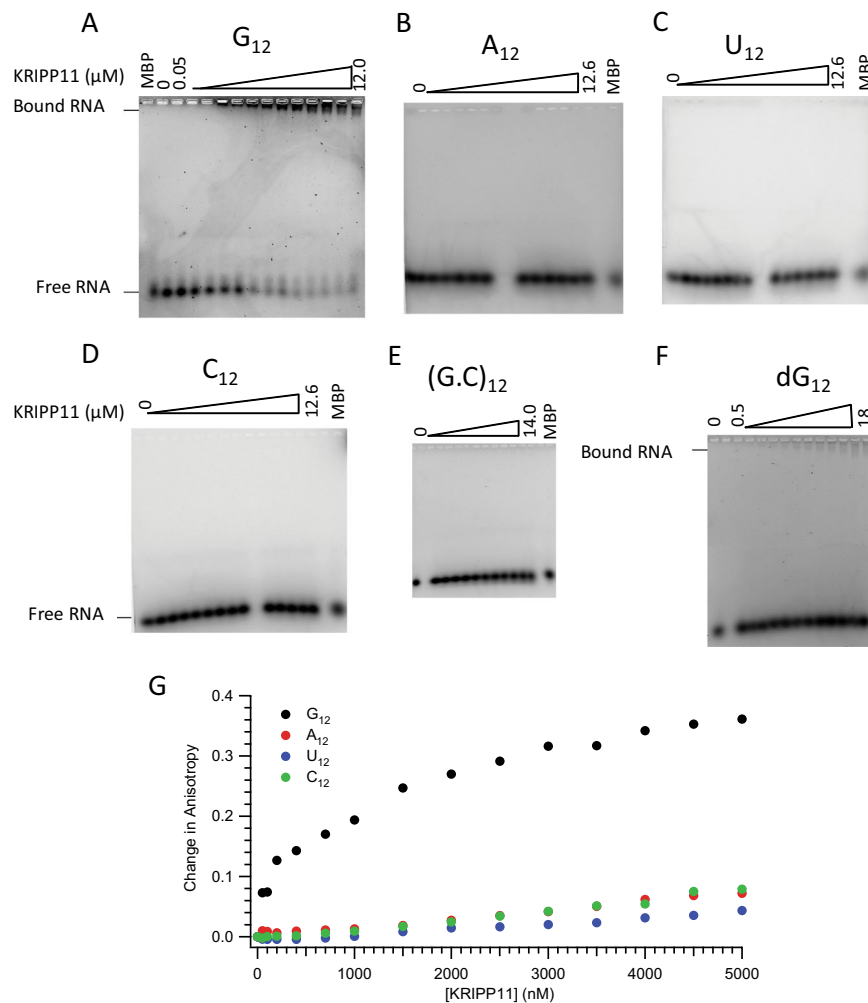
KRIPP11 was verified using native PAGE. The mature KRIPP11 migrated slightly faster than the 27 kDa TEV protease, consistent with a monomeric state of this 24.4 kDa protein in the cleavage mixture (Fig. 1F). Since the cleaved KRIPP11 was found to be vulnerable to aggregation and precipitation upon purification, uncleaved MBP-KRIPP11-His<sub>6</sub> fusion protein was used for the binding assays and selections. Separately expressed MBP-His<sub>6</sub> (Fig. 1C) was used as a control to rule out MBP as the source of observed nucleic-acid binding.

**Binding specificity of KRIPP11 for single-stranded poly(G) RNA.** Preliminary insights into the sequence- and chemical specificity of KRIPP11's nucleic acid ligand were obtained by assessing the protein's affinity for homooligomeric and simple heterooligomeric RNAs and DNA. Among 12-nucleotide homooligomers, EMSA showed an upward shift in the mobility of 12-nucleotide ssRNA for only poly(G) for protein concentrations up to 12 μM (Fig. 2A–D).

We further explored the binding specificity of KRIPP11 for single- versus double-stranded RNA as well as versus the identical DNA sequence. Annealing of G<sub>12</sub><sup>-</sup> with an excess of C<sub>12</sub><sup>-</sup> RNA to form double-stranded (G.C)<sub>12</sub> abolishes the KRIPP11-induced upward shift in poly(G) mobility (Fig. 2E). By contrast, an interaction with G<sub>12</sub> ssDNA is observed, although only at higher concentrations of protein, indicating a reduced binding affinity (Fig. 2F). Finally, we tested the affinity of KRIPP11 for shorter G homooligomers as well as representative G-rich heterooligomers and a (GGU)<sub>4</sub> sequence suggested by the *in vitro* selection results (see below). Of the sequences tested, intense shifted bands on increasing KRIPP11 concentration only appeared with the G<sub>9</sub> homooligomer (Fig. 3). EMSA experiments testing G<sub>12</sub>, G<sub>9</sub>, and G<sub>6</sub> against MBP-His<sub>6</sub> showed no evidence of binding, confirming that the observed high-affinity binding is to the KRIPP11 moiety rather than to the MBP or His<sub>6</sub> fusion partners (“MBP” lanes in Figs 2 and 3). The EMSA results thus emphasize that, among the variety of sequences examined, stable KRIPP11-RNA complexes are only observed for longer poly(G) sequences.

For quantitative determination of equilibrium binding constants  $K_d$ , we applied a fluorescence anisotropy binding assay. In this technique, the reorientation of a fluorophore bound to the nucleic acid oligomer is slowed upon protein complex formation due to slower global tumbling, leading to a greater retention of polarization in fluorescence emission upon excitation with a polarized source. Examination of titrations for the 12-mer homooligomers showed a substantial, saturable shift in anisotropy only for G<sub>12</sub> (Fig. 2G), consistent with the EMSA results (Fig. 2A–D). Upon fitting to a single-site isotherm (Equation 2, Methods), a  $K_d$  of  $(0.66 \pm 0.13) \mu\text{M}$  was obtained for G<sub>12</sub>, increasing to  $(1.88 \pm 0.13) \mu\text{M}$  for the corresponding deoxy sequence (Fig. 4). The binding curve saturated and fit well to the standard isotherm, consistent with a specific interaction. Two independent measures of binding thus confirm the specificity of KRIPP11 for poly(G) among the homooligomers.

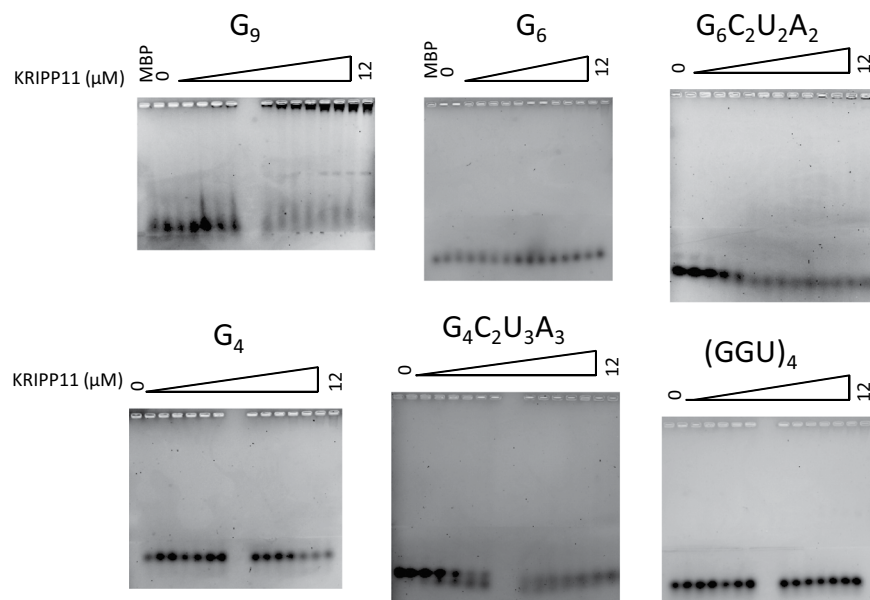
By contrast, poly(G) sequences as short as G<sub>4</sub> and a subset of the heteropolymers tested displayed saturable, specific interactions via fluorescence anisotropy assays, albeit in all cases with reduced affinity compared to G<sub>12</sub>, despite the lack of complex formation in EMSA (Supplementary Material, Fig. S1). Dissociation constants for



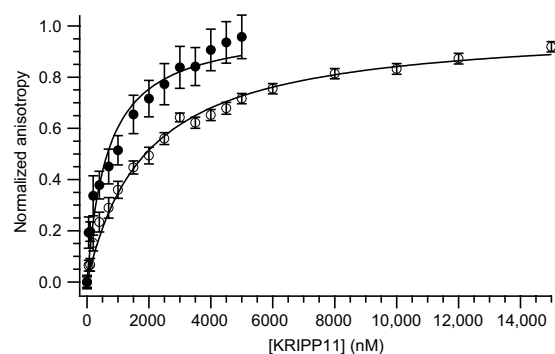
**Figure 2.** Analysis of KRIPP11 binding. (Panels A–D) EMSA analysis on 2% tris-glycine agarose gels showing retardation of G<sub>12</sub> and not the other RNA homooligomers (A<sub>12</sub>, U<sub>12</sub>, and C<sub>12</sub>) as a function of protein concentration; (E) (G.C)<sub>12</sub>; (F) dG<sub>12</sub>. A reaction with MBP-His<sub>6</sub> was used as a negative control for EMSA experiments. (G) Fluorescence anisotropy binding isotherms for KRIPP11 against ssRNA homooligomers.

measured sequences are collected in Table 1. Affinity tracked roughly with the longest stretch of consecutive G residues; for example, (GGU)<sub>4</sub> displayed the weakest affinity of the sequences analyzed despite having a relatively high number (eight) of total G residues. The contrast between anisotropy data and EMSA results may be explained by the nature of EMSA as a non-equilibrium assay, which makes it poorly suited to the detection of interactions with a fast ligand dissociation rate. On the other hand, FA is an equilibrium assay with greater dynamic range of analysis for binding affinity than EMSA<sup>49</sup>. The quantitative results in Table 1 reiterate the nature of KRIPP11 as a nucleic acid binding protein with specificity for nonduplex G-tract RNA sequences.

***In vitro* selection of KRIPP11 binding sequences.** To delineate the binding specificity of KRIPP11 in a more unbiased fashion, we applied the technology of *in vitro* selection to isolate RNA sequences with high affinity to the protein from a pool of over 10<sup>18</sup> RNA sequences randomized at 20 continuous positions. *In vitro* selection of aptamer sequences from fully randomized pools such as this provides an unbiased view of consensus RNA binding preferences for proteins and other molecules<sup>50,51</sup>. During 11 rounds of *in vitro* selection, the RNA pool was progressively enriched with KRIPP11-binding activity as measured by the partitioning of fluorescent RNA into the protein-bound pool (Fig. 5A). Sequencing of 100 clones from the final selection round revealed a bound pool enriched in poly(G) (4 to 8 consecutive G nucleotides) and to a lesser extent poly(U) (4 to 5 consecutive U nucleotides) sequences (Table S2). Little specificity for specific sequences was obtained, as reflected in the relatively low-information consensus binding sequence produced by standard algorithms (Fig. 5B). Assay of binding to a repeating GGU sequence suggested by the apparent consensus sequence yielded weaker binding than a simple G homooligomer (Table 1). This lack of validation tends to argue against the functional significance of the U-containing motif, although effects arising from the somewhat different buffer conditions in use for the two procedures may also account for the discrepancy (see Methods; briefly, fluorescence polarization assays were conducted in 20 mM Tris-HCl pH 7.5, 0.15 M KCl, 5 mM MgCl<sub>2</sub>, 1 mM DTT, whereas the binding buffer for selection experiments



**Figure 3.** EMSA of KRIPP11 with G-tracts varying in length and sequence.



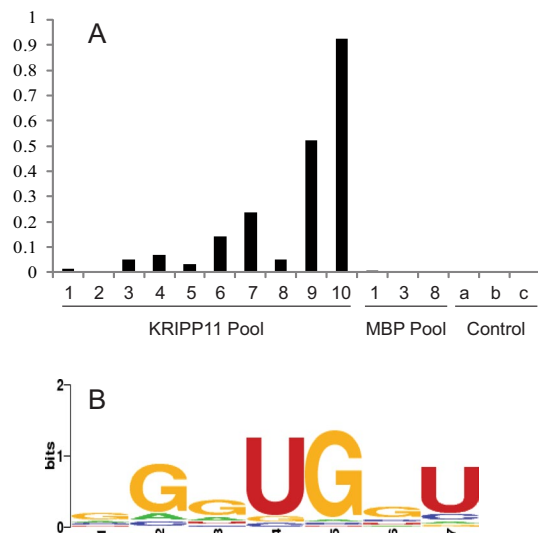
**Figure 4.** Determination of the  $K_d$  for interaction of KRIPP11 with  $G_{12}$  ssRNA (solid circles) and  $G_{12}$  ssDNA (open circles) from fluorescence anisotropy data. Curves represent fits to a single-site binding isotherm (Equation 2).

RNA	$K_d$ ( $\mu\text{M}$ )
$G_{12}$	$0.66 \pm 0.13$
$G_9$	$1.87 \pm 0.03$
$G_6$	$1.07 \pm 0.01$
$G_6C_2U_2A_2$	$0.93 \pm 0.01$
$G_4$	$3.42 \pm 0.19$
$G_4C_2U_3A_3$	$2.21 \pm 0.03$
$(GGU)_4$	$6.55 \pm 0.10$

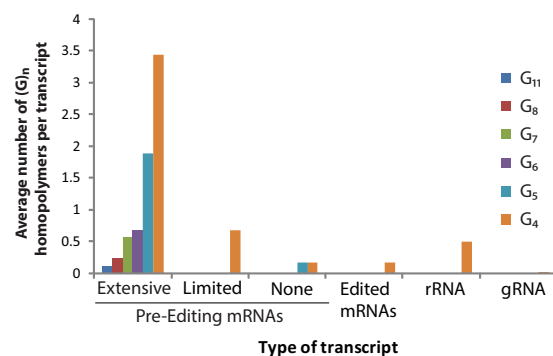
**Table 1.** Affinity of KRIPP11 for RNA G-tracts of varying sequence determined using fluorescence anisotropy.

was 50 mM phosphate pH 7.5, 0.1 M KCl, 1.5 mM  $\text{MgCl}_2$ , 2 mM imidazole). Taken together with the homopolymer binding data above, the selection results establish KRIPP11 as a G-tract specific RNA binding protein.

**Potential functional role of a G-tract binding protein in *T. brucei*.** To investigate the potential biological significance of these results, we examined the *T. brucei* mitochondrial genome for open reading frames coding for G-rich pre-mRNAs. Interestingly, a strong pattern of preferential enrichment of G-tracts in some transcripts was observed, particularly with respect to the U-insertion/deletion editing state of the transcripts. Specifically, we divided pre-edited transcripts into nine that undergo extensive editing, with more than 50% of the final sequence consisting of inserted uridines (“pan-edited” henceforth); three that undergo editing events



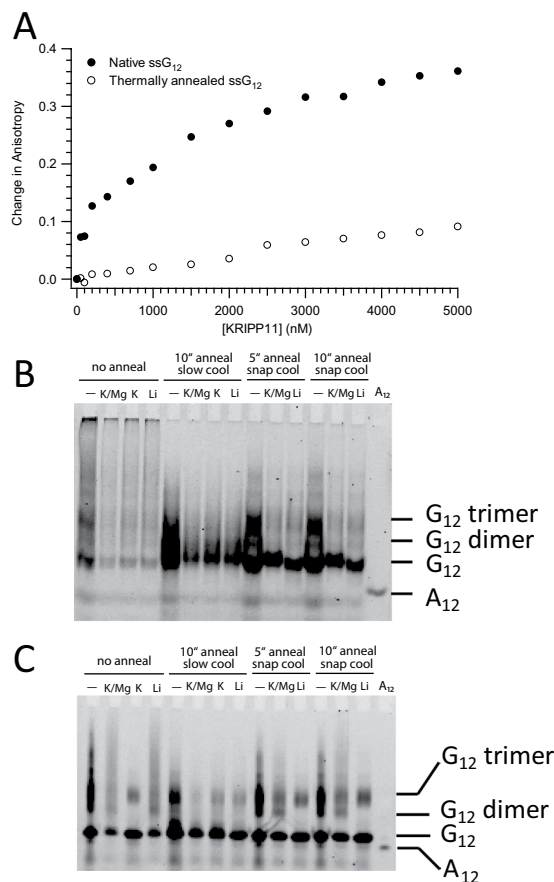
**Figure 5.** (A) RiboGreen assay showing increase in percentage of protein that bound RNA with successive cycles of *in vitro* selection with KRIPP11. After the second round of selection, the amount of RNA-bound protein was undetectable, attributed to an error in the assay. Control assays: a, protein-free beads in presence of RNA pool; b, protein-bound beads in absence of RNA pool; c, protein-free beads in absence of RNA pool. (B) WebLogo of RNA sequences sequenced following the final round of selection with KRIPP11.



**Figure 6.** Prevalence of poly(G) homopolymers of 4, 5, 6, 7, 8, and 11 contiguous guanosine residues. The vertical axis reports the number of homopolymeric G sequences of exactly  $n$  residues found in the indicated family of transcripts divided by the number of transcripts in the family. Pre-editing mRNAs are divided into classes destined to undergo extensive editing, limited editing, or no editing (see text).

to a much less dramatic extent and exclusively to the 5' end (Cyb and Murf II) or correction of an internal frameshift (COII) (“limited” editing); and the six mitochondrial transcripts that undergo no insertion/deletion editing (“never-edited”). The final versions of the twelve pan- and limited-editing transcripts after all editing events are complete are denoted “post-editing”. Of these four classes, as well as mitochondrial rRNA transcripts, the pan-edited transcripts were specifically and dramatically enriched in G-tracts of at least four consecutive guanosines. Indeed, all but one of the pan-edited transcripts contained at least one, and usually several, G-tracts of six or more consecutive guanosines, a pattern with zero occurrences in the other transcript classes (Fig. 6; Supplementary Material, Tables S3–S7). Further, all of the more than 60 G-tracts identified in pan-edited transcripts are destroyed by the U-insertion/deletion editing, and no new ones are created (cf. Tables S3 and S5); in an interesting contrast, the two four-nucleotide G-tracts present in transcripts destined for limited editing survive the editing process intact (cf. Tables S4 and S5). We also searched available libraries of insect- and bloodstream-form gRNA sequences from high-throughput sequencing<sup>52,53</sup>; an overall average of 0.22% of guide RNAs contained a G-tract of four or more nucleotides. Given that G-tracts of four to six or more consecutive nucleotides are necessary for stable KRIPP11 binding, the conclusion is that only pan-edited transcripts are likely to be bound by KRIPP11 *in vivo*, and only prior to the completion of the editing process.

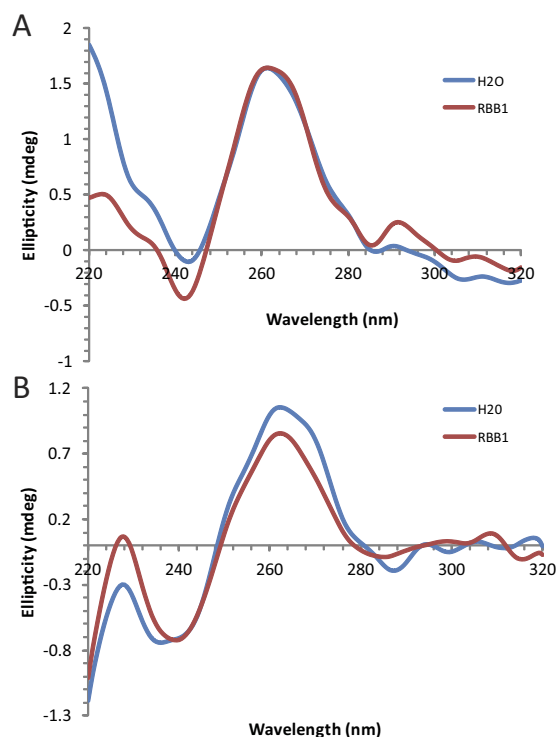
**Preference of KRIPP11 for putative G-quadruplex forming sequences.** RNA sequences containing contiguous runs of guanosine residues can, under appropriate conditions, form intra- and intermolecular G-quadruplex structures<sup>54–56</sup>. Thus, two interesting questions for a G-tract specific binding protein such as



**Figure 7.** (A) Anisotropy titrations of native and thermally denatured G<sub>12</sub> ssRNA with KRIPP11. (B) Native 18% polyacrylamide gel showing the electrophoretic profile of 5'-FLUO-G<sub>12</sub> pre-exposed to different thermal histories and salt conditions. “No anneal” samples were purified under non-denaturing conditions and exchanged to the indicated buffer. For other samples, time of incubation in minutes at 95 °C and mode of cooling is indicated; see main text for details. K/Mg buffer, 20 mM Tris-HCl, pH 7.5, 150 mM KCl, 5.0 mM MgCl<sub>2</sub>, 1.0 mM DTT; K, 20 mM Tris-HCl, pH 7.5, 150 mM KCl, 1.0 mM DTT; Li, 20 mM Tris-HCl, pH 7.5, 150 mM LiCl, 1.0 mM DTT. A<sub>12</sub> ssRNA was dissolved without annealing in buffer TE (10 mM Tris-HCl, pH 7.5, 0.1 mM EDTA). (C) Denaturing (6.8 M urea) 17% polyacrylamide gel with the same lane content as in B.

KRIPP11 are whether it will selectively bind RNA in quadruplex form in solution and, if so, whether that structure is maintained in the protein-bound form. Since the formation of intermolecular G-quadruplexes by short G-rich oligos is likely to be dependent on the thermal history of the sample, we compared the binding of KRIPP11 to RNA homo-oligomers prepared under thermally denaturing and native conditions (see Methods). When RNA previously thermally denatured at 95 °C followed by snap cooling on ice was used in binding assays, the increase in anisotropy with KRIPP11 concentration was greatly diminished relative to native RNA and the binding isotherm did not saturate, consistent with weakened or less-specific binding (Fig. 7A). To assign the structural differences between denatured/renatured and natively-prepared RNA samples, we applied circular dichroism (CD) spectroscopy and native gel analysis. The CD spectra of native G<sub>12</sub> and G<sub>4</sub> ssRNA displayed minima at 245 nm and maxima at 265 nm (Fig. 8), consistent with a parallel G-quadruplex conformation<sup>55,57</sup>. On native PAGE, native G<sub>12</sub> not only migrated as a ladder-like smear of various RNA sizes larger than that of A<sub>12</sub> RNA, most of the RNA was also retained in the well (Fig. 7B), consistent with a multimeric state for these species. Following Tlucikova and coworkers<sup>58</sup>, we speculate that the slowest migrating native G<sub>12</sub> RNA forms and those retained in the well are tetrameric quadruplexes and their aggregates, respectively. A denaturing gel of the G-tracts revealed the resistance of some quadruplexes to denaturation (Fig. 7C).

Interestingly, the intensity of native G<sub>12</sub> bands on non-denaturing gels was severely diminished in the presence of monovalent ions compared to that of thermally treated RNA. Monovalent and divalent ions stabilize quadruplex structure<sup>59,60</sup>, and it is possible that they drive the G-tract RNA to a state in which the fluorescein stacks against the guanosine, thereby reducing its quantum yield. Thermally treated but slowly cooled G<sub>12</sub> RNA migrates as a single elongated band at the speed of the fastest migrating band of native RNA (Fig. 7B). It is likely that the elongated band results from overlapping dimeric quadruplexes and monomeric RNA<sup>58</sup>, of which the former severely bleaches fluorescein in the presence of metal ions. In contrast, when thermally treated RNA is rapidly cooled in the absence of metal ions, up to 20% of G<sub>12</sub> migrated at a size most likely to be a trimer while



**Figure 8.** CD spectra of  $G_{12}$  (A) and  $G_4$  (B) in MQ- $H_2O$  (blue) and K/Mg buffer (brown). The minima at 245 nm and maxima at 265 nm are characteristic of parallel G-quadruplexes.

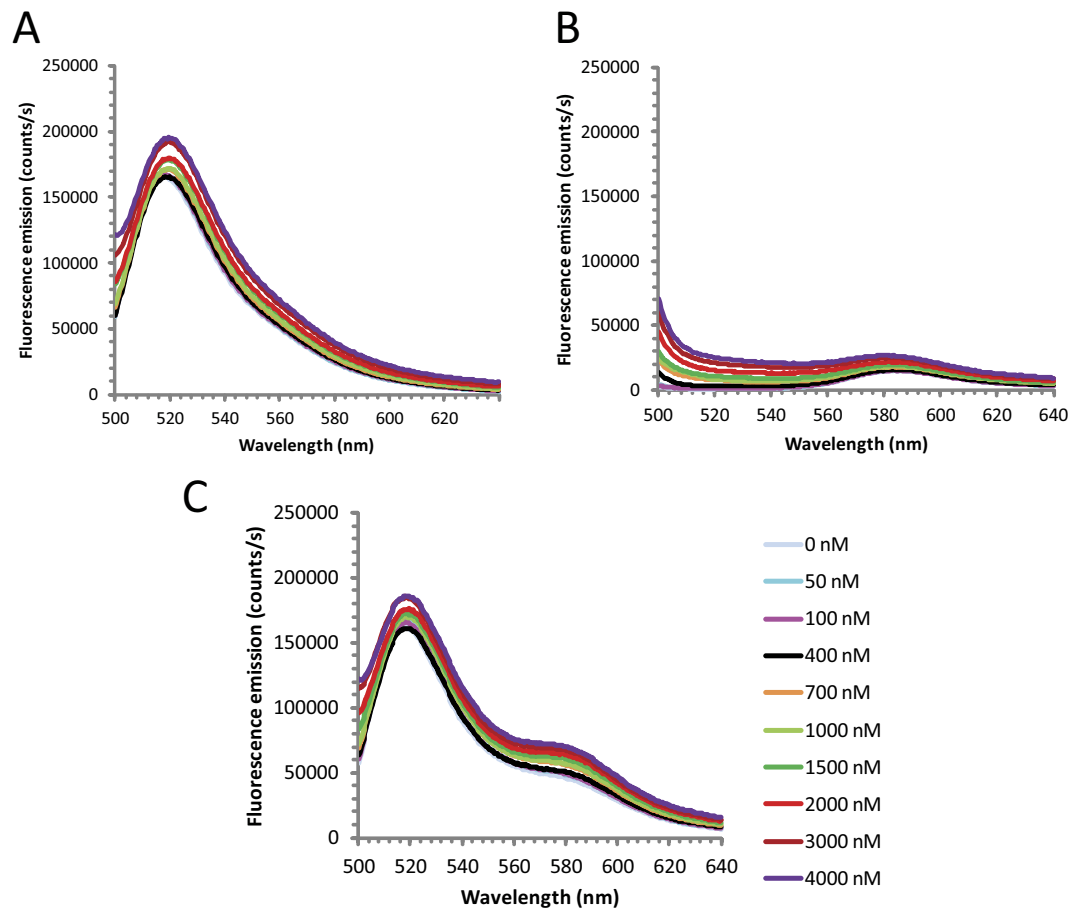
the rest migrated as monomers. In the presence of either  $K^+$  or  $Li^+$  ions under rapid cooling, however, this larger (trimer-like) band of  $G_{12}$  RNA diminished, likely due to transformation into more stable dimeric quadruplexes with concomitant quenching of fluorescein, similar to the non-thermally exposed RNA.

In short, KRIPP11 binds RNA with high affinity according to fluorescence polarization assays only if that RNA has a thermal history, specifically preparation under native conditions, that shows CD spectra and native-gel behavior consistent with a significant degree of G-quadruplex formation, suggesting that KRIPP11 binding is specific to quadruplex structures.

**KRIPP11 binding does not disrupt G-quadruplex structure.** Some proteins are known to remodel nucleic acid quadruplex structures upon binding<sup>61–63</sup>. To examine the state of KRIPP11-bound RNA, we applied fluorescence resonance energy transfer (FRET) analysis. We prepared a mixture of  $G_{12}$  RNA oligomers containing either a 5'-fluorescein (FLUO) or 3'-TAMRA fluorescent label. A strong FRET effect will be observed only if an intermolecular interaction such as G-quadruplex formation brings the fluorescein donor and TAMRA acceptor within the Förster radius (approximately 55 Å for this donor-acceptor pair), allowing efficient non-radiative donor-acceptor energy transfer. If KRIPP11 were to unfold the RNA quadruplexes, therefore, any FRET signal arising from the intermolecular quadruplex would reduce with KRIPP11 concentration. Here, we found that when annealed to 5'-FLUO- $G_{12}$ , the intensity of  $G_{12}$ -TAMRA-3' emission increases at least 3-fold relative to  $G_{12}$ -TAMRA-3' alone (Fig. 9A–C) following excitation at 490 nm. When the sum of the spectra of free 5'-FLUO- $G_{12}$  and free  $G_{12}$ -TAMRA-3' was deducted from the spectrum of annealed 5'-FLUO- $G_{12}$ / $G_{12}$ -TAMRA-3', a net positive TAMRA emission signal (peak intensity of about 8,000 counts/s) and no FLUO emission signal was observed, consistent with a FRET effect. Since the fluorescence donors and acceptors reside on different RNA chains, FRET arises unambiguously from the formation of intermolecular interactions, presumably in the form of G-quadruplex structures (cf. Figure 6). Specifically, the FRET signal arises solely from the subset of complexes containing both 5'-FLUO- $G_{12}$  and  $G_{12}$ -TAMRA-3'; complexes containing only multiple copies of a single species do not contribute to the observed transfer. The TAMRA emission intensity from FRET was only 5% of the emission intensity of free 5'-FLUO- $G_{12}$ , implying that there was low efficiency of resonance energy transfer. However, the low FRET signal was not unexpected as G-tracts without intervening heterogeneous nucleotides (A, U, and C) are very stable with melting points above 95 °C<sup>60,64</sup>. Thus, only a relatively small fraction of the RNA is unfolded and available to anneal to new partners.

The hypothesis that protein binding disrupts intermolecular quadruplex formation predicts that the magnitude of FRET would decrease upon protein-RNA complex formation. Upon titration with KRIPP11, however, the fluorescence emission of  $G_{12}$ -TAMRA-3' in the 5'-FLUO- $G_{12}$ / $G_{12}$ -TAMRA-3' FRET combination (i.e., the FRET signal) not only was not reduced but indeed increased (Fig. 9C). Besides structural remodeling, a change in fluorescence emission intensity of a ligand-conjugated fluorophore upon protein binding can also occur if the interaction alters the microenvironment around the fluorophore, which is very possible if the protein binds proximal to the fluorophore<sup>65,66</sup>. Given the small size of the oligos used in the KRIPP11 binding assay, the increase in  $G_{12}$ -TAMRA-3'





**Figure 9.** Intermolecular FRET in G-tract RNA. Fluorescence emission intensity of (A) 5'-FLUO-G<sub>12</sub>; (B) G<sub>12</sub>-TAMRA-3'; (C) An annealed mixture of FLUO-G<sub>12</sub> and G<sub>12</sub>-TAMRA as a function of KRIPP11 concentration. Excitation was at 490 nm in all cases. Traces presented are the average of three replicate experiments.

fluorescence emission intensity is not surprising. However, since the 5'-FLUO-G<sub>12</sub> is the source of fluorescence excitation for the G<sub>12</sub>-TAMRA-3' in the FRET experiment, the G<sub>12</sub>-TAMRA-3' emission would reduce or disappear in the event of quadruplex disruption regardless of the altered microenvironment. Thus, it appears that KRIPP11 does not in fact disrupt G-quadruplex structures upon binding. Taken together with the binding assays of RNAs with varying thermal history and therefore presumed quadruplex content (see above), this data suggests the possibility that G-tracts bind to KRIPP11 in quadruplex form. Given that all but one pan-edited transcript contain multiple long G-tracts (Table S3), these structures could plausibly arise in intramolecular fashion *in vivo*, in contrast to the intermolecular structures presumed to form in the *in vitro* experiments herein.

## Discussion

The family of pentatricopeptide proteins is significantly expanded in the evolutionary lineage leading to trypanosomatid parasites. Since these parasites are also characterized by numerous unusual features in RNA metabolism, including prolific U insertion/deletion editing in mitochondrial transcripts, it is tempting to speculate that the expansion of the PPR gene family in trypanosomatids represents a molecular-level adaptive radiation, in which ancestral PPR protein(s) have evolved a variety of RNA binding specificities and/or accessory domains to play distinct roles within the organelle. As one of the smallest and simplest PPR proteins in *T. brucei*, and one without any obvious accessory domains, KRIPP11 was our initial target for detailed characterization of RNA-binding specificity as a window into biological function. In this work, we found that KRIPP11 has a strong preferential affinity for non-duplex poly(G) sequences. KRIPP11 binds sequences of four to six or more Gs with high affinity, displays strong discrimination against other homooligomers and related sequences as well as double-stranded RNA, and shows modest specificity for RNA over identical sequences of single-stranded DNA. Bioinformatic analyses of *T. brucei* mitochondrial RNAs indicate a dramatic selective enrichment of G-tracts with at least four consecutive guanosines only in pre-edited transcripts of extensively edited mRNA genes (Fig. 6: Supplementary Tables S3–S7).

The simplest explanation of the above results would be that KRIPP11 interacts with transcripts destined for extensive editing and protects, sequesters, and/or targets the transcripts to the editing apparatus. Such functions are familiar roles for PPR proteins in other organisms<sup>21–24,34–36,67,68</sup>. Proteomic analyses of trypanosome mitochondrial ribosomes, however, indicate that KRIPP11 is associated with the small ribosomal subunit<sup>14,31</sup>. Indeed,

a total of 20 PPR proteins have been identified as components of either the large or small subunit of the mitochondrial ribosome. Although many of these PPRs are presumed to play a structural role, some have been shown to be involved in translational activation of the mitochondrial mRNAs.

The low affinity of KRIPP11 for U-rich and A-rich RNA suggests that direct recognition of rRNA or fully-edited, translation-competent mRNA by this protein is not likely. A variety of functions for a precursor/pre-edited RNA binding protein associated with the small ribosomal subunit<sup>31,32</sup> are possible. KRIPP11 binding could bridge mRNA maturation and translation, either by upregulating the substrates for the ribosome by stabilizing polycistronic precursors and pre-edited transcripts or by directly bringing the small ribosomal subunit into contact with mRNA. Alternatively, KRIPP11's differential affinity for G-tracts of different sizes could specify the residence time of some RNA processing enzymes on their sites of action, which in the case of RNA processing nucleases could prevent stray RNA damage. By masking potential sites of spurious translation initiation, KRIPP11 could prevent premature assembly of the complete translational machinery on unedited mRNA. Evidence from plants<sup>34–36,39,68–71</sup>, humans<sup>72–74</sup>, yeast<sup>75</sup>, and even trypanosomes<sup>26–28</sup>, among others, shows that PPR proteins are capable of executing any of these functions. A more speculative function for KRIPP11 may be suggested by the existence of an alternatively-edited product of cytochrome oxidase subunit III (COIII). In this transcript, a proposed alternative editing event brings an open reading frame (ORF) found in the pre-edited 5' end into frame with the ORF generated by editing of the 3' end of COIII. The resulting alternatively-edited mRNA codes for the AEP-1 protein, which is essential for mitochondrial DNA maintenance<sup>76,77</sup>. A substantial portion of the translation-competent AEP-1 mRNA therefore contains G-tracts that are potential binding sites for KRIPP11. Similar alternative editing and dual coding events are widespread in kinetoplast mitochondria<sup>78,79</sup>, and KRIPP11 may participate in stabilization or translational activation of such nonstandard mRNAs via their retained G-tracts. Further insight into the precise mechanistic role of KRIPP11 in organellar RNA metabolism will be obtained from knockdown and *in vivo* RNA-protein crosslinking studies, now in progress.

In the RNA binding assays, KRIPP11 preferentially interacted with G-tracts prepared under conditions favoring the quadruplex conformation over those in random coil states, raising the possibility that quadruplex forms may be the native target of the protein in its biological context. In recent years, it has become increasingly clear that G-quadruplex structures play important functional roles in both DNA and RNA, with specific protein targeting of RNA quadruplexes with concomitant stabilization or destabilization of the quadruplex often a key factor<sup>80–82</sup>. Quadruplex-containing structures have also been found in selected aptamers targeting biologically relevant proteins<sup>83</sup>. Thus, G-quadruplex recognition has the ability to modulate biological function in a variety of contexts. G-tracts in *T. brucei* pre-edited mRNAs have been shown to form G-quadruplex structures<sup>84</sup>, consistent with the analysis of KRIPP11 specificity shown here. If KRIPP11 does indeed bind G-tracts in their quadruplex form, one interesting unanswered question is how the structural mode of binding of PPR proteins, which in other contexts recognize non-quadruplex sequences in extended conformation<sup>45–47</sup>, is adapted to the very different overall shape of quadruplex RNA.

## Methods

**Materials.** RNA oligos labeled with FLUO (fluorescein) at the 5' end or TAMRA (tetramethylrhodamine) at the 3' end were synthesized by Thermo Scientific's Dharmacon RNA Technologies (Lafayette, CO) and deprotected according to the manufacturer's instructions. Dry RNA was dissolved in MQ-H<sub>2</sub>O. 5'-FLUO labeled dG<sub>12</sub> ssDNA was obtained from Integrated DNA Technologies (Coralville, IA). The pMalTEV-E30 plasmid vector was a kind gift of Dr. Alice Barkan (University of Oregon).

**Overexpression constructs and mutagenesis.** To achieve soluble expression, we fused KRIPP11 to the C-terminus of the maltose binding protein (MBP) in a pMalTEV-E30 plasmid vector<sup>85</sup> using the BamH1 and SalI restriction sites (Table S8). The pMalTEV-E30 vector encodes a TEV protease cleavage site just before the BamH1 restriction site. Initial expression of MBP-KRIPP11 resulted in significant leaky expression of MBP, which contaminated preparations purified on amylose alone. Thus, we inserted a His<sub>6</sub> tag at the C-terminus of KRIPP11 (Fig. 1B) to aid tandem metal ion- and amylose affinity chromatography. MBP-His<sub>6</sub> (used as a negative control) was expressed from a construct created by inserting His<sub>6</sub> at the C-terminus of MBP just before the TEV protease cleavage site (Fig. 1C) followed by conversion of the first residue of KRIPP11 in MBP-His<sub>6</sub>-KRIPP11 to a stop codon. Mutagenesis was performed by PCR following the QuikChange<sup>®</sup> procedure using the primers indicated in Table S8.

**Protein expression and purification.** Protein was expressed in BL21(DE3) *E. coli*. Recombinant colonies were inoculated into 10 ml LB broth and grown overnight at 37 °C and 250 rpm. 20 ml of this seed culture was then added to 980 ml of LB broth, grown at 37 °C, 250 rpm to an OD<sub>600</sub> of about 0.7, induced with 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG), and maintained at 22 °C for 6 hours. Cultures were harvested by centrifugation at 4 °C followed by protein purification under native conditions. Bacteria were resuspended in lysis buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 8.0, 300 mM NaCl, 10 mM imidazole, 5 mM β-ME), lysed by a microtip sonicator and clarified by centrifugation. The supernatant was loaded onto a Ni-NTA column and MB-KRIPP11-His<sub>6</sub> was eluted by 50 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 8.0, 300 mM NaCl, 200 mM imidazole, 5 mM β-ME. Ni-NTA eluates were immediately repurified using an amylose column. Amylose eluates were then dialyzed against at least 40 volumes of 10 mM Tris-HCl, pH 7.5, 50 mM NaCl at 4 °C for at least 12 hours with a 7 K MWCO SnakeSkin<sup>®</sup> tubing (Thermo Scientific, Rockford, IL). Dithiothreitol (DTT) was then added to the dialysates to 1 mM followed by concentration using 10 K MWCO centrifugal filters. Protein purity was assessed by 5–12% SDS-PAGE and nucleic acid contamination was monitored by the A<sub>260</sub>:A<sub>280</sub> ratio.

**KRIPP11 native gel electrophoresis.** After detachment of MBP by TEV protease, free KRIPP11 aggregated upon purification even in the presence of detergents and glycerol. To establish the oligomeric state of KRIPP11, we therefore performed blue native PAGE on the TEV protease cleavage mixture of MBP-KRIPP11-His<sub>6</sub>. Briefly, MBP-KRIPP11 was purified by tandem Ni-NTA and amylose as above. Dilute protein was then concentrated to 0.7 mg/ml followed by TEV protease cleavage in the presence of 50 mM Tris-HCl, 0.5 mM EDTA, 1.0 mM DTT, plus or minus 0.05% (w/v) DDM. The reaction was performed at 8 °C for 13 hours. 1.6 μl of 10x blue native Tris-glycine (BNTG) sample buffer (250 mM Tris, pH 7.0, 30% (w/v) sucrose, 5% (w/v) brilliant blue-G) was then added to 14.4 μl of TEV protease cleavage mixture and run on a 6–12% polyacrylamide gradient resolving gel with a 3.5% polyacrylamide stacking gel at 150 V constant voltage, 4 °C, for one hour. 25 mM Tris, 250 mM glycine was used as running buffer. Native TEV protease (27 kDa), thioredoxin-His<sub>6</sub> (17 kDa), MBP (45 kDa), MBP-KRIPP11 (69 kDa), and factor Xa protease (43 kDa) were the size markers.

**Electrophoretic mobility shift assays (EMSA).** 5'-FLUO labeled RNA and DNA oligonucleotides were used, namely, single stranded C<sub>12</sub>, U<sub>12</sub>, A<sub>12</sub>, G<sub>12</sub>, G<sub>9</sub>, G<sub>6</sub>, G<sub>6</sub>C<sub>2</sub>U<sub>2</sub>A<sub>2</sub>, G<sub>4</sub>, G<sub>4</sub>C<sub>2</sub>U<sub>3</sub>A<sub>3</sub> and (GGU)<sub>4</sub>, and double stranded (G.C)<sub>12</sub>. Double-stranded (G.C)<sub>12</sub> was prepared by heating a mixture of G<sub>12</sub> ssRNA, a three-fold excess of C<sub>12</sub> ssRNA, and a 5x annealing buffer (50 mM Tris-HCl, pH 7.5, 250 mM NaCl, 5 mM EDTA) on boiling water for two minutes followed by slow cooling at room temperature for one hour. Agarose rather than polyacrylamide gels were used for electrophoresis because the latter were associated with strong well retention of the RNA as the protein concentration rose. Due to the thickness of the agarose gels and the propensity of poly(G) to quench fluorescein, 150 nM RNA/DNA was required in the reaction for optimal signal in the gel images. The RNA binding reaction consisted of 2.5 μl of RNA/DNA, 7.5 μl of protein (MBP-KRIPP11-His<sub>6</sub> or MBP-His<sub>6</sub> control as labeled), and 2.5 μl of a 5x RNA binding buffer adapted from Stuart and coworkers (100 mM Tris-HCl, pH 7.5, 750 mM KCl, 25 mM MgCl<sub>2</sub>, 5 mM DTT, 0.5 μg/μl BSA, 2 U/μl RNasin)<sup>86</sup>. 25 mM Tris, 250 mM glycine was used as running buffer. Gels were scanned using a VersaDoc™ MP 4000 fluorescence imager (Bio-Rad Laboratories).

**Fluorescence polarization spectroscopy.** The anisotropy of 20 nM 5'-FLUO labeled ssRNA or ssDNA was assessed in the presence of increasing KRIPP11 concentration. A 450 μl reaction for each protein concentration was prepared from dialysis buffer, MBP-KRIPP11-His<sub>6</sub>, 45 μl of a 10x binding buffer (200 mM Tris-HCl, pH 7.5, 1.5 M KCl, 50 mM MgCl<sub>2</sub>, 10 mM DTT), and 50 μl of 180 nM RNA/DNA. After at least 15 minutes of equilibration at room temperature, the intensities of vertically- and horizontally polarized fluorescence emission (I<sub>vv</sub> and I<sub>vh</sub> respectively) following vertically polarized excitation were obtained at 25 °C using a 5 mm microsquare quartz sample cell in a QuantaMaster spectrofluorometer (Photon Technology International). Noise from Raman scattering was removed by subtracting the fluorescence emission of KRIPP11 dialysis buffer from data. For all assays, the instrument grating factor (G) was about 1.5. Excitation was at 450 nm with a 2.0 mm slit with emission measured at 520 nm with a 3.8 mm slit. Fluorescence anisotropy was computed from:

$$r = \frac{I_{vv} - GI_{vh}}{I_{vv} + 2GI_{vh}} \quad (1)$$

The mean anisotropies of at least three data sets were then normalized and used to determine the K<sub>d</sub> by non-linear fitting with IGOR Pro 6 according to

$$r = \frac{nc}{K_d + c} \quad (2)$$

where c is the protein concentration and n is an overall scaling factor.

Preliminary experiments indicated stronger and more reproducible interactions of KRIPP11 with slowly thawed G-tract RNA than with either rapidly thawed or thermally annealed RNA (cf. also Fig. 7A). Thus, reported anisotropy experiments used RNA that was thawed slowly on ice-cold water.

**In vitro selection of consensus RNA ligands for KRIPP11.** RNA aptamers for KRIPP11 were isolated from a pool of 2.4 × 10<sup>18</sup> random sequences (20 nucleotide variable region) by 11 rounds of *in vitro* selection using immobilized MBP-KRIPP11-His<sub>6</sub> on Ni-NTA magnetic beads using procedures based on published methods<sup>87</sup>. RNA was synthesized by T7 RNA polymerase from a DNA template containing a variable region of 20 nucleotides surrounded by primer binding sites followed by treatment with DNase and purification. Negative selection to remove sequences with affinity to the fusion partner was performed in every second round of selection using MBP-His<sub>6</sub> as a decoy. At each round, the RNA pool in binding buffer (50 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.5, 100 mM KCl, 1.5 mM MgCl<sub>2</sub>, 2.0 mM imidazole) was incubated with MBP-His<sub>6</sub> beads (if performed) at 4 °C for 20 minutes. The flow-through from the latter was then incubated with MBP-KRIPP11-His<sub>6</sub> for 20 minutes. Binding stringency was increased across the course of the selection by decreasing protein and (for the final round) RNA concentration (Table S10). After drainage of unbound RNA, the beads were washed with binding buffer followed by recovery of bound RNA by phenol:chloroform extraction and ethanol precipitation and reconstituted in MQ-H<sub>2</sub>O. The resulting RNA pool was converted to cDNA with M-MLV reverse transcriptase followed by PCR-amplification, transcription, and RNA purification in preparation for the next round of selection as described previously<sup>87</sup>. Enrichment of KRIPP11-bound RNA after each round of selection was quantitated with the Ribogreen assay<sup>88</sup> using a FLUOstar Omega microplate reader (BMG LABTECH). The cDNA from the 11<sup>th</sup> round RNA was cloned into a TOPO® TA plasmid and sequenced by dideoxynucleotide chain termination. Sequences were then aligned with MEME<sup>89</sup> and WebLogo<sup>90</sup> to identify conserved elements.

**Analysis of G-tract content in *T. brucei* mitochondrial transcripts.** Sequences of all the kinetoplast pre-mRNA and rRNA genes were decoded from the *T. brucei* maxicircle DNA sequence (GenBank accession # M94286.1). Edited mRNA sequences were obtained from Hong and Simpson<sup>91</sup>. A census of nonoverlapping dodecanucleotides containing at least one homopolymeric stretch of four or more Gs in each RNA was then made. Next, a count of each G<sub>n</sub> homopolymer (n = 4, 5, 6, 7, 8, or 11; no sequences of G<sub>9</sub> or G<sub>10</sub> were present) in each transcript category was made.

**Circular dichroism (CD) spectroscopy of guanosine tract RNA.** The native conformation of G-tract RNA ligands was investigated by CD spectroscopy. CD spectra were collected using a Chirascan CD spectrometer (Applied Photophysics, Leatherhead, UK) and a 0.1 mm path length cuvette. To assure native conformation, frozen RNA samples in MQ-H<sub>2</sub>O were thawed slowly on ice-cold water. For G<sub>12</sub>, the concentrations of RNA used were 4.1 μM in MQ-H<sub>2</sub>O and 3.7 μM in RNA binding buffer (20 mM Tris-HCl, pH 7.5, 150 mM KCl, 5.0 mM MgCl<sub>2</sub>, 1.0 mM DTT). For G<sub>4</sub>, the concentrations of RNA used were 7.5 μM in MQ-H<sub>2</sub>O and 6.8 μM in RNA binding buffer. An average of three spectra was obtained per sample. Before CD spectroscopy, the G<sub>12</sub> RNA was desalted by dialysis against two 1000-fold volumes of MQ-H<sub>2</sub>O at 12 hour intervals for 24 hours at 4 °C using a 3.5 K Slide-A-Lyzer<sup>®</sup> cassette (Thermo Scientific). G<sub>4</sub> was used in undialyzed form due to lack of retention by the molecular-weight filters used.

**RNA native gel electrophoresis.** To further understand the basis for better KRIPP11 binding to native G-tracts than thermally exposed ones, we investigated the gel mobility of four samples of 5'-FLUO-G<sub>12</sub> RNA with varying thermal history. One RNA sample was not exposed to elevated temperature, and represented the native RNA. The second RNA sample was heated at 95 °C for 10 minutes, then one of several different 10X RNA binding buffers (see caption to Fig. 7) was added after which tubes were transferred to 95 °C hot water and left to slowly cool to room temperature for 4 hours followed by storage at 4 °C for at least five days. The third RNA set was heated at 95 °C for 10 minutes, snap-cooled on ice for 2 minutes, transferred to room temperature, and RNA binding buffer added. The fourth RNA set was heated at 95 °C for 5 minutes, snap-cooled on ice for 2 minutes, transferred to room temperature, and RNA binding buffer added. For analysis, 12 μl of a mixture comprising 10.8 μl of 50 nM RNA and 1.2 μl of 10X sample buffer (10 mM Tris, 1.0 mM EDTA, pH 7.5, 50% (v/v) glycerol, 0.01% (w/v) bromophenol blue, 0.01% (w/v) xylene cyanol) were loaded on to a native 18% polyacrylamide (29:1 acrylamide: N,N'-methylene-bis-acrylamide) gel and run at 150 V at 4 °C using TBE (50 mM Tris, 50 mM boric acid, 1.0 mM EDTA). A 12 nt poly(A) ssRNA oligo (A<sub>12</sub>) was loaded as marker for monomeric G<sub>12</sub> RNA.

**Fluorescence resonance energy transfer (FRET).** FRET was used to determine whether KRIPP11 binding perturbs the native conformation of G-tract RNA. Separately labeled 5'-FLUO-G<sub>12</sub> and 3'-TAMRA labeled G<sub>12</sub> (G<sub>12</sub>-TAM-3') ssRNAs were desalted by dialysis against 200-fold volume of MQ-H<sub>2</sub>O at 4 °C overnight using a 2.0K MWCO Spectra/Por<sup>®</sup> dialysis membrane (Spectrum Laboratories Inc, Rancho Dominguez, CA). To anneal the 5'-FLUO-G<sub>12</sub> and G<sub>12</sub>-TAM-3', a mixture of 350 μl of each RNA at 3.0 μM monomeric concentration was heated at 95 °C for 10 minutes, then 245 μl of MQ-H<sub>2</sub>O and 105 μl of 10X RNA binding buffer were added, diluting the RNA to 2.0 μM monomeric concentration. Tubes containing the RNA mixture were then transferred to 95 °C water in a plastic dish and allowed to slowly cool to room temperature on the bench for 4 hours. Finally, the RNA was stored at 4 °C for at least five days to allow further renaturation before FRET assays. The intensities of fluorescence emission of 50 nM monomeric concentrations of 5'-FLUO-G<sub>12</sub> alone, G<sub>12</sub>-TAM-3' alone, and of the annealed FLUO-G<sub>12</sub>/G<sub>12</sub>-TAM mixture in the absence and presence of various KRIPP11 concentrations were then collected at 25 °C following excitation at 490 nm as described above. Three datasets were used to compute average spectra.

## Data Availability

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

## References

1. Lukes, J., Hashimi, H. & Zikova, A. Unexplained complexity of the mitochondrial genome and transcriptome in kinetoplastid flagellates. *Curr Genet* **48**, 277–299 (2005).
2. Lukes, J., Hashimi, H., Verner, Z. & Cicova, Z. The Remarkable Mitochondrion of Trypanosomes and Related Flagellates. *Structures and Organelles in Pathogenic Protists* **17**, 227–252 (2010).
3. Bringaud, F., Riviere, L. & Coustou, V. Energy metabolism of trypanosomatids: adaptation to available carbon sources. *Molecular and biochemical parasitology* **149**, 1–9 (2006).
4. Tielens, A. G. & van Hellemond, J. J. Surprising variety in energy metabolism within Trypanosomatidae. *Trends in parasitology* **25**, 482–490 (2009).
5. Michelotti, E. F., Harris, M. E., Adler, B., Torri, A. F. & Hajduk, S. L. Trypanosoma brucei mitochondrial ribosomal RNA synthesis, processing and developmentally regulated expression. *Molecular and biochemical parasitology* **54**, 31–41 (1992).
6. Vanhamme, L. & Pays, E. Control of gene expression in trypanosomes. *Microbiol Rev* **59**, 223–240 (1995).
7. Clayton, C. & Shapira, M. Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. *Molecular and biochemical parasitology* **156**, 93–101 (2007).
8. Palenchar, J. B. & Bellofatto, V. Gene transcription in trypanosomes. *Molecular and biochemical parasitology* **146**, 135–141 (2006).
9. Adhya, S. *et al.* Mitochondrial differentiation in kinetoplastid protozoa: a plethora of RNA controls. *Differentiation* **71**, 549–556 (2003).
10. Read, L. K., Lukes, J. & Hashimi, H. Trypanosome RNA editing: the complexity of getting U in and taking U out. *Wiley interdisciplinary reviews. RNA* **7**, 33–51 (2016).
11. Benne, R. RNA editing in trypanosomes. *European journal of biochemistry* **221**, 9–23 (1994).
12. Brennicke, A., Marchfelder, A. & Binder, S. RNA editing. *FEMS Microbiol Rev* **23**, 297–316 (1999).
13. Aphasizhev, R. & Aphasizheva, I. Mitochondrial RNA processing in trypanosomes. *Research in microbiology* **162**, 655–663 (2011).

14. Aphasizhev, R. & Aphasizheva, I. Emerging roles of PPR proteins in trypanosomes: switches, blocks, and triggers. *RNA biology* **10**, 1495–1500 (2013).
15. Aphasizheva, I. & Aphasizhev, R. U-Insertion/Deletion mRNA-Editing Holoenzyme: Definition in Sight. *Trends in parasitology* **32**, 144–156 (2016).
16. Zhang, L. *et al.* PPR polyadenylation factor defines mitochondrial mRNA identity and stability in trypanosomes. *The EMBO journal* **36**, 2435–2454 (2017).
17. Madina, B. R., Kumar, V., Mooers, B. H. & Cruz-Reyes, J. Native Variants of the MRB1 Complex Exhibit Specialized Functions in Kinetoplastid RNA Editing. *PLoS one* **10**, e0123441 (2015).
18. D'Andrea, L. D. & Regan, L. TPR proteins: the versatile helix. *Trends Biochem Sci* **28**, 655–662 (2003).
19. Small, I. D. & Peeters, N. The PPR motif - a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem Sci* **25**, 46–47 (2000).
20. Andres, C., Lurin, C. & Small, I. D. The multifarious roles of PPR proteins in plant mitochondrial gene expression. *Physiologia Plantarum* **129**, 14–22 (2007).
21. Delannoy, E., Stanley, W. A., Bond, C. S. & Small, I. D. Pentatricopeptide repeat (PPR) proteins as sequence-specific factors in post-transcriptional processes in organelles. *Biochemical Society Transactions*, 1643–1647 (2007).
22. Schmitz-Linneweber, C. & Small, I. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends Plant Sci* **13**, 663–670 (2008).
23. Sharma, M. & Pandey, G. K. Expansion and Function of Repeat Domain Proteins During Stress and Development in Plants. *Frontiers in plant science* **6**, 1218 (2015).
24. Barkan, A. & Small, I. Pentatricopeptide repeat proteins in plants. *Annual review of plant biology* **65**, 415–442 (2014).
25. Lurin, C. *et al.* Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* **16**, 2089–2103 (2004).
26. Mingler, M. K. *et al.* Identification of pentatricopeptide repeat proteins in *Trypanosoma brucei*. *Molecular and biochemical parasitology* **150**, 37–45 (2006).
27. Pusnik, M., Small, I., Read, L. K., Fabbro, T. & Schneider, A. Pentatricopeptide repeat proteins in *Trypanosoma brucei* function in mitochondrial ribosomes. *Molecular and cellular biology* **27**, 6876–6888 (2007).
28. Pusnik, M. & Schneider, A. A trypanosomal pentatricopeptide repeat protein stabilizes the mitochondrial mRNAs of cytochrome oxidase subunits 1 and 2. *Eukaryot Cell* **11**, 79–87 (2012).
29. Aphasizheva, I. *et al.* Ribosome-associated pentatricopeptide repeat proteins function as translational activators in mitochondria of trypanosomes. *Molecular microbiology* **99**, 1043–1058 (2016).
30. Etheridge, R. D., Aphasizheva, I., Gershon, P. D. & Aphasizhev, R. 3' Adenylation determines mRNA abundance and monitors completion of RNA editing in *T. brucei* mitochondria. *The EMBO journal* **27**, 1596–1608 (2008).
31. Zikova, A. *et al.* *Trypanosoma brucei* mitochondrial ribosomes: affinity purification and component identification by mass spectrometry. *Mol Cell Proteomics* **7**, 1286–1296 (2008).
32. Aphasizheva, I., Maslov, D., Wang, X., Huang, L. & Aphasizhev, R. Pentatricopeptide repeat proteins stimulate mRNA adenylation/uridylation to activate mitochondrial translation in trypanosomes. *Molecular cell* **42**, 106–117 (2011).
33. Uytewaal, M. *et al.* PPR336 is associated with polysomes in plant mitochondria. *Journal of molecular biology* **375**, 626–636 (2008).
34. Pfalz, J., Bayraktar, O. A., Prikryl, J. & Barkan, A. Site-specific binding of a PPR protein defines and stabilizes 5' and 3' mRNA termini in chloroplasts. *The EMBO journal* **28**, 2042–2052 (2009).
35. Prikryl, J., Rojas, M., Schuster, G. & Barkan, A. Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 415–420 (2011).
36. Okuda, K., Nakamura, T., Sugita, M., Shimizu, T. & Shikanai, T. A pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing. *Journal of Biological Chemistry* **281**, 37661–37667 (2006).
37. Hammami, K., Takenaka, M., Miranda, R. & Barkan, A. A PPR protein in the PLS subfamily stabilizes the 5'-end of processed rpl16 mRNAs in maize chloroplasts. *Nucleic acids research* **44**, 4278–4288 (2016).
38. Fujii, S. & Small, I. The evolution of RNA editing and pentatricopeptide repeat genes. *New Phytol* **191**, 37–47 (2011).
39. Fujii, S., Bond, C. S. & Small, I. D. Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 1723–1728 (2011).
40. Barkan, A. *et al.* A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet* **8**, e1002910 (2012).
41. Yagi, Y., Hayashi, S., Kobayashi, K., Hirayama, T. & Nakamura, T. Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS one* **8**, e57286 (2013).
42. Hall, T. M. De-coding and re-coding RNA recognition by PUF and PPR repeat proteins. *Current opinion in structural biology* **36**, 116–121 (2016).
43. Miranda, R. G., McDermott, J. J. & Barkan, A. RNA-binding specificity landscapes of designer pentatricopeptide repeat proteins elucidate principles of PPR-RNA interactions. *Nucleic acids research* **46**, 2613–2623 (2018).
44. Ban, T. *et al.* Structure of a PLS-class pentatricopeptide repeat protein provides insights into mechanism of RNA recognition. *The Journal of biological chemistry* **288**, 31540–31548 (2013).
45. Ke, J. *et al.* Structural basis for RNA recognition by a dimeric PPR-protein complex. *Nat Struct Mol Biol* **20**, 1377–1382 (2013).
46. Yin, P. *et al.* Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* **504**, 168–171 (2013).
47. Shen, C. *et al.* Structural basis for specific single-stranded RNA recognition by designer pentatricopeptide repeat proteins. *Nature communications* **7**, 11285 (2016).
48. Dean, S., Sunter, J. D. & Wheeler, R. J. TrypTag.org: A Trypanosome Genome-wide Protein Localisation Resource. *Trends Parasitol* **33**, 80–82 (2017).
49. Hill, J. J. & Royer, C. A. Fluorescence approaches to study of protein-nucleic acid complexation. *Methods in enzymology* **278**, 390–416 (1997).
50. Ellington, A. D. & Szostak, J. W. *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).
51. Gold, L., Polisky, B., Uhlenbeck, O. & Yarus, M. Diversity of oligonucleotide functions. *Annual review of biochemistry* **64**, 763–797 (1995).
52. Koslowsky, D., Sun, Y., Hindenach, J., Theisen, T. & Lucas, J. The insect-phase gRNA transcriptome in *Trypanosoma brucei*. *Nucleic acids research* **42**, 1873–1886 (2014).
53. Kirby, L. E., Sun, Y., Judah, D., Nowak, S. & Koslowsky, D. Analysis of the *Trypanosoma brucei* EATRO 164 Bloodstream Guide RNA Transcriptome. *PLoS Negl Trop Dis* **10**, e0004793 (2016).
54. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. & Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic acids research* **34**, 5402–5415 (2006).
55. Kyrp, J., Kejnovska, I., Rencuk, D. & Vorlickova, M. Circular dichroism and conformational polymorphism of DNA. *Nucleic acids research* **37**, 1713–1725 (2009).
56. Lane, A. N., Chaires, J. B., Gray, R. D. & Trent, J. O. Stability and kinetics of G-quadruplex structures. *Nucleic acids research* **36**, 5482–5515 (2008).

57. Masiero, S. *et al.* A non-empirical chromophoric interpretation of CD spectra of DNA G-quadruplex structures. *Org Biomol Chem* **8**, 2683–2692 (2010).
58. Viglasky, V., Bauer, L. & Tluczkova, K. Structural features of intra- and intermolecular G-quadruplexes derived from telomeric repeats. *Biochemistry* **49**, 2110–2120 (2010).
59. Juskowiak, B., Galezowska, E., Zawadzka, A., Gluszynska, A. & Takenaka, S. Fluorescence anisotropy and FRET studies of G-quadruplex formation in presence of different cations. *Spectrochim Acta A Mol Biomol Spectrosc* **64**, 835–843 (2006).
60. Rachwal, P. A. & Fox, K. R. Quadruplex melting. *Methods* **43**, 291–301 (2007).
61. Salas, T. R. *et al.* Human replication protein A unfolds telomeric G-quadruplexes. *Nucleic acids research* **34**, 4857–4865 (2006).
62. Dominguez, C., Fiset, J. F., Chabot, B. & Allain, F. H. Structural basis of G-tract recognition and engaging by hnRNP F quasi-RRMs. *Nat Struct Mol Biol* **17**, 853–861 (2010).
63. Samatanga, B., Dominguez, C., Jelesarov, I. & Allain, F. H. The high kinetic stability of a G-quadruplex limits hnRNP F qRRM3 binding to G-tract RNA. *Nucleic acids research* **41**, 2505–2516 (2013).
64. Grygoryev, D. & Zimbrick, J. D. Effect of quadruplex conformation on radiation-induced formation of 8-hydroxyguanine and unaltered base release in polyguanylic acid. *Radiat Res* **173**, 110–118 (2010).
65. Anderson, B. J., Larkin, C., Guja, K. & Schildbach, J. F. Using fluorophore-labeled oligonucleotides to measure affinities of protein-DNA interactions. *Methods in enzymology* **450**, 253–272 (2008).
66. Hovius, R., Vallotton, P., Wohland, T. & Vogel, H. Fluorescence techniques: shedding light on ligand-receptor interactions. *Trends Pharmacol Sci* **21**, 266–273 (2000).
67. Kotera, E., Tasaka, M. & Shikanai, T. A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature* **433**, 326–330 (2005).
68. Okuda, K. & Shikanai, T. A pentatricopeptide repeat protein acts as a site-specificity factor at multiple RNA editing sites with unrelated cis-acting elements in plastids. *Nucleic acids research* **40**, 5052–5064 (2012).
69. Hammami, K., Cook, W. B. & Barkan, A. RNA binding and RNA remodeling activities of the half-a-tetratricopeptide (HAT) protein HCF107 underlie its effects on gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 5651–5656 (2012).
70. Manavski, N., Guyon, V., Meurer, J., Wienand, U. & Brettschneider, R. An Essential Pentatricopeptide Repeat Protein Facilitates 5' Maturation and Translation Initiation of rps3 mRNA in Maize Mitochondria. *Plant Cell* (2012).
71. Okuda, K., Myouga, F., Motohashi, R., Shinozaki, K. & Shikanai, T. Conserved domain structure of pentatricopeptide repeat proteins involved in chloroplast RNA editing. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 8178–8183 (2007).
72. Davies, S. M. K. *et al.* Pentatricopeptide repeat domain protein 3 associates with the mitochondrial small ribosomal subunit and regulates translation. *FEBS letters* **583**, 1853–1858 (2009).
73. Rackham, O. *et al.* Pentatricopeptide repeat domain protein 1 lowers the levels of mitochondrial leucine tRNAs in cells. *Nucleic acids research* **37**, 5859–5867 (2009).
74. Sasarman, F., Brunel-Guitton, C., Antonicka, H., Wai, T. & Shoubridge, E. A. LRPPRC and SLIRP interact in a ribonucleoprotein complex that regulates posttranscriptional gene expression in mitochondria. *Mol Biol Cell* **21**, 1315–1323 (2010).
75. Maslov, A. M. & Agrawal, R. K. In *RNA Metabolism in Trypanosomes* Vol. 28 (ed. Bindereif, A.) Ch. 10, (Springer-Verlag, 2012).
76. Ochsenreiter, T. & Hajduk, S. L. Alternative editing of cytochrome c oxidase III mRNA in trypanosome mitochondria generates protein diversity. *EMBO reports* **7**, 1128–1133 (2006).
77. Ochsenreiter, T., Anderson, S., Wood, Z. A. & Hajduk, S. L. Alternative RNA editing produces a novel protein involved in mitochondrial DNA maintenance in trypanosomes. *Molecular and cellular biology* **28**, 5595–5604 (2008).
78. Ochsenreiter, T., Cipriano, M. & Hajduk, S. L. Alternative mRNA editing in trypanosomes is extensive and may contribute to mitochondrial protein diversity. *PloS one* **3**, e1566 (2008).
79. Kirby, L. E. & Koslowsky, D. Mitochondrial dual-coding genes in *Trypanosoma brucei*. *PLoS Negl Trop Dis* **11** (2017).
80. Fry, M. Tetraplex DNA and its interacting proteins. *Frontiers in bioscience: a journal and virtual library* **12**, 4336–4351 (2007).
81. Millevoi, S., Moine, H. & Vagner, S. G-quadruplexes in RNA biology. *Wiley interdisciplinary reviews. RNA* **3**, 495–507 (2012).
82. Fay, M. M., Lyons, S. M. & Ivanov, P. RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms. *Journal of molecular biology* **429**, 2127–2147 (2017).
83. Platella, C., Riccardi, C., Montesarchio, D., Roviello, G. N. & Musumeci, D. G-quadruplex-based aptamers against protein targets in therapy and diagnostics. *Biochimica et biophysica acta* **1861**, 1429–1447 (2017).
84. Leeder, W. M., Hummel, N. F. & Goring, H. U. Multiple G-quartet structures in pre-edited mRNAs suggest evolutionary driving force for RNA editing in trypanosomes. *Scientific reports* **6**, 29810 (2016).
85. Williams-Carrier, R., Kroeger, T. & Barkan, A. Sequence-specific binding of a chloroplast pentatricopeptide repeat protein to its native group II intron ligand. *Rna* **14**, 1930–1941 (2008).
86. Tarun, S. Z. *et al.* KREPA6 is an RNA-binding protein essential for editosome integrity and survival of *Trypanosoma brucei*. *Rna-a Publication of the Rna Society* **14**, 347–358 (2008).
87. Garrey, S. M., Voelker, R. & Berglund, J. A. An extended RNA binding site for the yeast branch point-binding protein and the role of its zinc knuckle domains in RNA binding. *Journal of Biological Chemistry* **281**, 27443–27453 (2006).
88. Wochner, A. & Glöckler, J. Nonradioactive fluorescence microtiter plate assay monitoring aptamer selections. *Biotechniques* **42**, 578, 580, 582 (2007).
89. Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic acids research* **37**, W202–208 (2009).
90. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188–1190 (2004).
91. Hong, M. & Simpson, L. Genomic organization of *Trypanosoma brucei* kinetoplast DNA minicircles. *Protist* **154**, 265–279 (2003).

## Acknowledgements

We extend our gratitude to Dr. Alice Barkan (University of Oregon) for the pMALTEV-E30 plasmid gift; Dr. John Wang and Dr. Honggao Yan (Michigan State University) for technical advice on the RNA binding assays; Dr. Lisa Lapidus (Michigan State University) for use of the fluorimeter and helpful discussions on quadruplex formation; Prof. Andrew Berglund (University at Albany) for protocols and helpful discussion on *in vitro* selection; Dr. Shelagh Ferguson-Miller (Michigan State University) for PCR and sonication facilities; the MSU Genomics Core facility for DNA sequencing; and the MSU DOE Plant Research Laboratory for the fluorescence imager. P.F.K. was supported by a Fulbright pre-doctoral scholarship from the U.S Department of State and MSU Cell and Molecular Biology program fellowships. This research was supported by the NSF (MCB-1413356 to C.G.H.), NIH (R03-AI0902 to D.J.K.), and a Makerere University-Carnegie Corporation Next Generation of African Academics grant (P.F.K.).

### Author Contributions

P.F.K. designed, performed, and analyzed experiments and drafted the manuscript. D.A.D. collected and analyzed anisotropy data. N.A.W. designed, performed, and analyzed the *in vitro* selection experiments. J.L.E. designed and supervised the cloning and expression work. D.J.K. designed research, interpreted data and drafted the manuscript. C.G.H. supervised and designed research, analyzed and interpreted data, and assembled the final manuscript. All authors reviewed and edited the manuscript.

### Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-34377-9>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018