This is to certify that the
thesis entitled

IDENTIFICATION AND CHARACTERIZATION OF TRYPANOSOMA BRUCEI PPR PROTEINS, PUTATIVE MITOCHONDRIAL RNA METABOLISM PROTEINS

presented by

Melissa Kay Mingler

has been accepted towards fulfillment
of the requirements for the

Master of Science degree in Biochemistry and Molecular Biology

Major Professor's Signature

12/6/05

Date

IDENTIFICATION AND CHARACTERIZATION OF TRYPANOSOMA BRUCEI PPR PROTEINS, PUTATIVE MITOCHONDRIAL RNA METABOLISM PROTEINS

By

Melissa Kay Mingler

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTERS OF SCIENCE

Department of Biochemistry and Molecular Biology

2005

ABSTRACT

IDENTIFICATION AND CHARACTERIZATION OF TRYPANOSOMA BRUCEI PPR PROTEINS, PUTATIVE MITOCHONDRIAL RNA METABOLISM PROTEINS

By

Melissa Kay Mingler

A new class of proteins, characterized by Pentatricopeptide Repeat (PPR) motifs, have been identified recently in plants. These proteins contain multiple 35-amino acid repeats that are proposed to form a superhelix capable of binding a strand of RNA. All PPR proteins characterized to date appear to be involved in RNA metabolism in organelles. Comparative genomic studies show that while there are 23 PPR proteins within *Trypanosoma brucei*, plants contain over 450 PPR proteins. In contrast, eukaryotes only contain 2-6 PPR proteins. Studies began with bioinformatics that characterized a *T. brucei*-specific PPR motif. One of the putative *T. brucei* mitochondrial PPR proteins, TbPPR1, was further characterized using RNA interference (RNAi). TbPPR1 is predicted to be mitochondrially targeted and contains 14 predicted PPR motifs, with the majority occurring in tandem. RNAi experiments designed to knockdown expression of TbPPR1 show a repeatable slow growth phenotype. The mitochondrial mRNA levels from the RNAi experiments were studied using Northern blot analysis and with a "poison primer" extension assay, and several messages showed an altered expression pattern. Through these studies we hope to identify relevant RNA processing factors of mitochondrial messages.

ACKNOWLEDGMENTS

Thank you to Donna Koslowsky, Ph.D. who supported and instructed me through my graduate studies. Also a thank you to Sandra L. Clement, Ph.D. for instructing me in many of the lab techniques used in this thesis as well as for all her helpful advice. I would like to acknowledge Kaillathe Padmanabhan, Ph.D. for his help in using the HMMER 2.2 package for my bioinformatics studies and all his other computer expertise. I would also like to acknowledge Annette Thelen, Ph.D. for all her helpful advise on the Real Time PCR studies and Shirley Owens for her help in the fluorescent microscopy studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

Introduction

Pentatricopeptide Repeat (PPR) proteins are an important family of α-helical repeat proteins involved in several aspects of organellar RNA post-transcriptional processing in many different eukaryotic organisms. A large number of these proteins have been identified in *Trypanosoma brucei.* We plan to investigate their role in mitochondrial mRNA processing, stability or expression. This chapter introduces the organism *T. brucei*, its significance for study, and its mitochondrial biogenesis, as well as the family of PPR proteins.

Trypanosomes: Lifecycle, Medical and Economical Significance.

Trypanosoma brucei are parasitic protozoa that cause African Sleeping Sickness in humans and Nagana in cattle. They threaten over 60 million people in 36 different countries in sub-Saharan Africa. In 1999, 45,000 cases of African Sleeping Sickness were diagnosed. However, the World Health Organization (WHO) estimates that the number of people affected is ten times greater due to the lack of screening. There are several villages in Angola, the Democratic Republic of Congo and Southern Sudan where Sleeping Sickness is the first or second greatest cause of mortality, even ahead of HIV/AIDS (www.who.int). *T. brucei* is not only detrimental to the health of the people of sub-Saharan Africa, but also to their well-being. Infection in animals causes decimation of cattle and abandonment of fertile land to avoid the disease.

The symptoms in the early stages of the disease are fever, headaches, pains in the joints, and itching. The second stage is the neurological phase where *T. brucei* crosses the blood-brain barrier into the central nervous system. The symptoms are confusion, sensory disturbances, poor coordination, and disturbances in the sleep cycle that all can culminate in fatality. Treatment in the initial stage is more effective and safer. The only second stage drug on the market today is Melarsoprol, which is an arsenic derivative with severe and deadly side effects, killing 4-12% of all patients who receive it [3]. Elfornithine also treats late stage infections, but only those caused by one of the subspecies. Elfornithine's maker, Adventis, has only guaranteed production of the injectable form until the year 2005 (www.who.int).

The *T. brucei* parasite is transmitted through the bite of the Tsetse fly, *Glossina*, when it takes a blood meal from a vertebrate host. It then proliferates in waves in the host's bloodstream where it evades the immune system by continually changing its antigenic coat of variant surface glycoproteins [4-6]. There are two life forms in the bloodstream. The long, slender form is able to undergo replication. The short, stumpy forms are nonreplicative and begin to change in metabolism in preparation to enter the next lifecycle stage in the tsetse fly. When the tsetse fly takes a blood meal from an infected individual, the short and stumpy bloodstream form will differentiate into the procyclic form in the mid-gut of the insect. This is the form that we study in the lab. The parasites then travel up to the salivary glands of the fly and differentiate again

into the epimastigotes and then metacyclics, which are then delivered to the mammals when the tsetse takes a bloodmeal (Fig. 1).

***Trypanosoma brucei* lifecycle.**

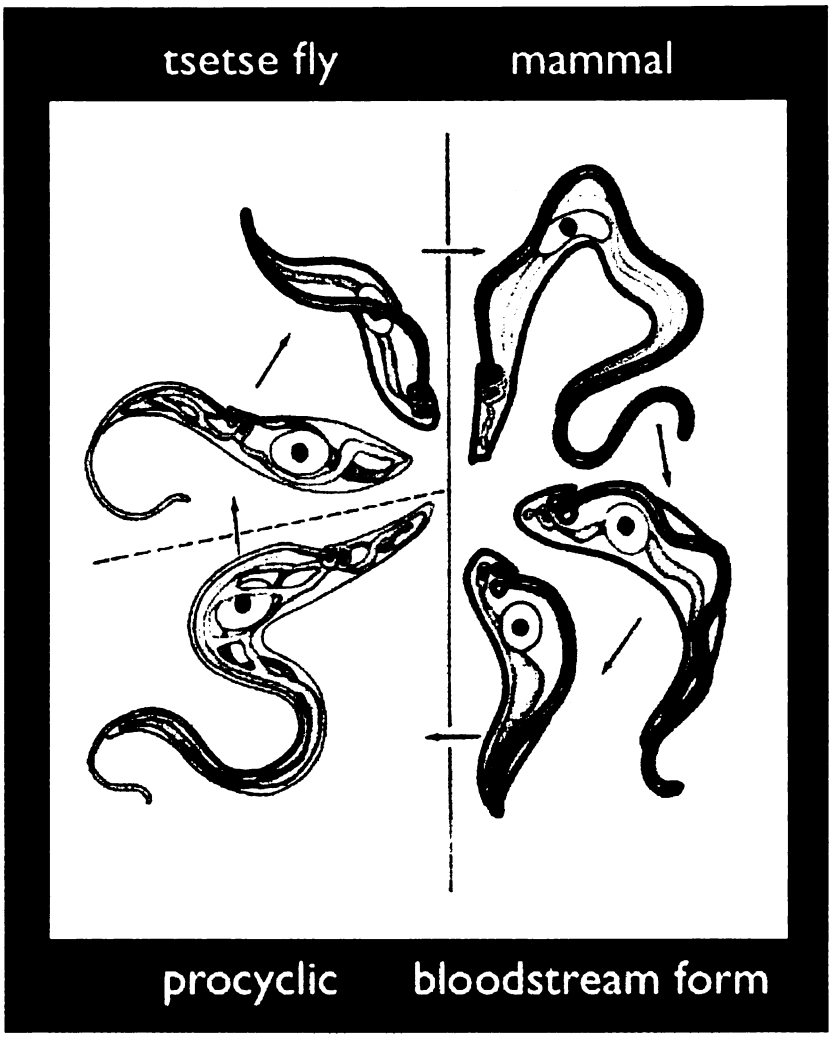


Figure 1. The lifecycle of Trypanosoma brucei. The light gray areas represent the mitochondria that have great morphological changes between lifecycle stages. The large circle with the dot is the nucleus and the smaller oval with the dot is the kinetoplast. Obtained from homepage.mac.com/mfield/lab/Images/lifecycle.gif.

FIGURE 1

Trypanosoma brucei Metabolism

Metabolism in *T. brucei* is quite complex due to the digenetic life cycle. The bloodstream form parasite, in the mammalian host, depends on glucose from the hosts' bloodstream for its energy production and secretes pyruvate [7, 8]. The procyclic or insect form parasite uses proline, glucose and threonine from the insects' gut for its energy production and secretes succinate, acetate, lactate, alanine, and CO_2 [9, 10].

The respiration rate in the bloodstream form of *Trypanosoma brucei* is quite high, 50-fold higher than any eukaryotic cell [11]. Metabolism in the bloodstream form occurs in a unique peroxisome-like organelle called the glycosome. Glycosomes have a single phospholipid bilayer with an electron dense proteinaceous matrix and no DNA. The glycosome contains glycolytic enzymes and enzymes of peroxide metabolism, fatty acid oxidation, and ether lipid biosynthesis. Metabolism occurs here through substrate level phosphorylation of glucose that produces 2 moles of ATP for every mole of glucose consumed. No net ATP or NADH is produced in the glycosome, as the ATP is actually produced in the cytoplasm by pyruvate kinase [10]. In the mitochondria of the bloodstream forms, there is no TCA cycle and most of the cytochromes are not expressed [12]. The intermembrane space of the mitochondria does contain a glycerol-3-phosphate dehydrogenase and a terminal alternative oxidase that are principal in the glycerol-3-phosphate and

dihydroxyacetone phosphate shunt between the glycosome and the mitochondria. This shunt maintains the NAD+/NADH balance within the glycosome. There is also evidence of an active Complex I in the mitochondria that may transfer electrons via the alternative oxidase [13]. The F_o/F_1-ATP synthase of the mitochondria is responsible for maintaining a proton gradient across the mitochondrial membrane in the bloodstream form trypanosomes [14].

The metabolism in the procyclic form is quite different. At low levels of glucose, oxidative phosphorylation within the mitochondria is essential. The mitochondria contain and use an incomplete TCA cycle. The succinyl-CoA synthetase of the TCA has been found to be essential as the last step in both the glucose and proline degradation pathways [15-17]. The TCA cycle is thought to feed into the electron transport chain when Complex II transfers the electrons from succinate of the TCA cycle to ubiquinone of the electron transport chain, thereby skipping Complex I in the procyclic forms [10, 18]. The procyclic form contains two terminal oxidases in its electron transport chain, the cytochrome c oxidase and the plant-like alternative oxidase. The cytochrome c oxidase is active in the mitochondrial oxidative phosphorylation, but the role of the alternative oxidase within the procyclic form is unknown. The protein levels of the alternative oxidase are lower in the procyclic then they are in the bloodstream form [19]. There is some evidence that the alternative oxidase is responsible for decreasing the reactive oxygen species (ROS) in procyclics [20]. The F_o/F_1-ATP synthase is the principal site of ATP generation.

Metabolism Summary. The glycosome is highly active in the bloodstream form of the parasite producing energy through substrate level phosphorylation. The role of the mitochondria in the bloodstream form is minor in that it contains a glycerol-3-phosphate dehydrogenase and the alternative oxidase that are important in maintaining the NAD+/NADH ratio in the glycosome, has some Complex I activity and uses the F_o/F_1-ATP synthase to maintain the proton gradient. Mitochondrial function in the procyclic is increased with its use of the oxidative phosphorylation, substrate level phosphorylation, and TCA cycle.

Mitochondrial Genome and mRNA Editing

The mitochondrial genome of *T. brucei* consists of ~50 copies of a ~23 kb maxicircle (Fig. 2) and 5,000-10,000 1.0 kb minicircles (~300 sequence families)(Fig. 3). These circles are topologically interlocked forming a network in the form of a disk called the kinetoplast DNA (kDNA). The maxicircles are analogous to mtDNA of other organisms. They encode 18 messenger RNAs (mRNAs), two guide RNAs (gRNA), a large and small ribosomal RNA (rRNA), but surprisingly, there are no transfer RNAs (tRNA) encoded in the maxicircles. Though the maxicircles are ~23 kb in size, only about 15 kb of this is actually coding region (Fig. 2). The remaining 8 kb of the maxicircle is called the variable region and it is thought that transcription of the polycistronic message begins somewhere in this region [21-23] (Fig. 2). Michelotti et al. have shown the existence of a transient precursor element, which would place the

transcription start site about 1,200 nt upstream of the 12s rRNA mature 5′ end [22], but the exact location is still unknown. The only mapped transcription start site on the maxicircle is for gMURF2-II, one of the two gRNAs that are encoded on the *T. brucei* maxicircle. The gMURF2-II gene is found completely within the 5′ end of the gene ND4, introducing more complexity to the transcription and processing of the maxicircle [24] (Fig. 4).

Minicircle transcription is somewhat different; each of the three gRNAs on a minicircle are primary transcription products located within cassettes of imperfect inverted 18-bp repeats [25, 26]. The 5′ ends of many gRNAs have been mapped 29-33 bp from the upstream repeat [26, 27]. The 18-bp inverted repeats have been implicated in playing a role in transcription, but the few gRNA genes located outside of these 18-bp repeats have also been found to be primary transcripts [24, 28, 29]. Even though each gRNA gene has the ability to initiate transcription, it appears that minicircle gRNA genes can be transcribed polycistronically [30], but the majority of the gRNAs found in the cell have 5′ di- or tri-phosphates, indicating they are not processed at their 5′ ends [24, 26].

Though the transcription start site of the maxicircle is unknown, it is known that transcription of the maxicircle results in polycistronic transcripts from which the individual RNAs are then processed out. The 5′ and 3′ ends of many maxicircle genes have been mapped. The coding region is very compact, with the majority of the genes overlapping [31](Table 1) at the 5′ and 3′ ends. Thus processing of one message will often result in the destruction of its neighbor(s)

[31, 32]. The control of which of the neighboring messages is processed out and translated remains unknown.

Extensive RNA processing of maxicircle transcripts must occur before a translatable message is obtained, including endonucleolytic cleavage, polyadenylation and RNA editing [22, 24, 31]. Kinetoplastid RNA editing inserts or deletes uridylates into the mRNA transcripts, forming the correct open reading frames, start and stop codons for translation. Once transcribed, many of the mRNAs are edited by gRNAs, primary transcripts encoded mainly by the minicircles [27] (Fig. 3). These gRNAs are present in both life cycle stages, though almost nothing is known about gRNA transcription, processing or transcript level regulation.

Linearized Maxicircle Map.

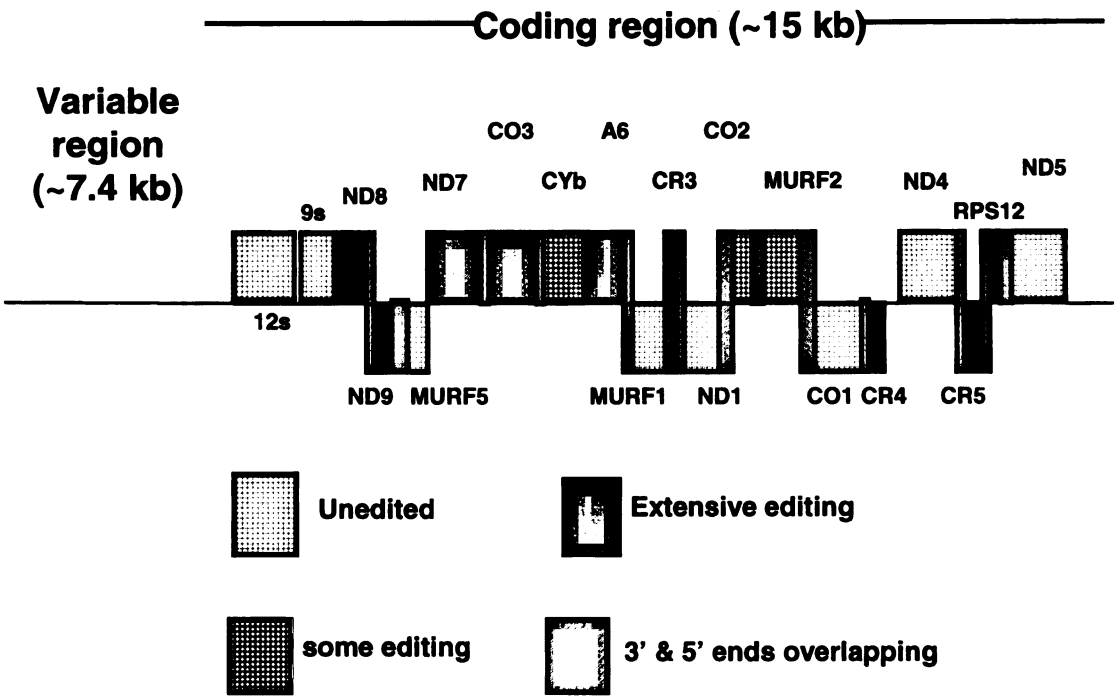


Figure 2. Linearized maxicircle map (coding region). 12s and 9s = rRNA subunits; ND=NADH dehydrogenase; CO=cytochrome oxidase; CYb=Cytochrome b; MURF=maxicircle unidentified reading frame; CR=C-rich region; RSP12=ribosomal protein 12. Overlapping regions are shown in gray, unedited genes are striped, extensively edited genes are in black, and the dimpled genes are genes that go through some mRNA editing. Courtesy of Sandra Clement.

FIGURE 2

***T. brucei* Minicircle**

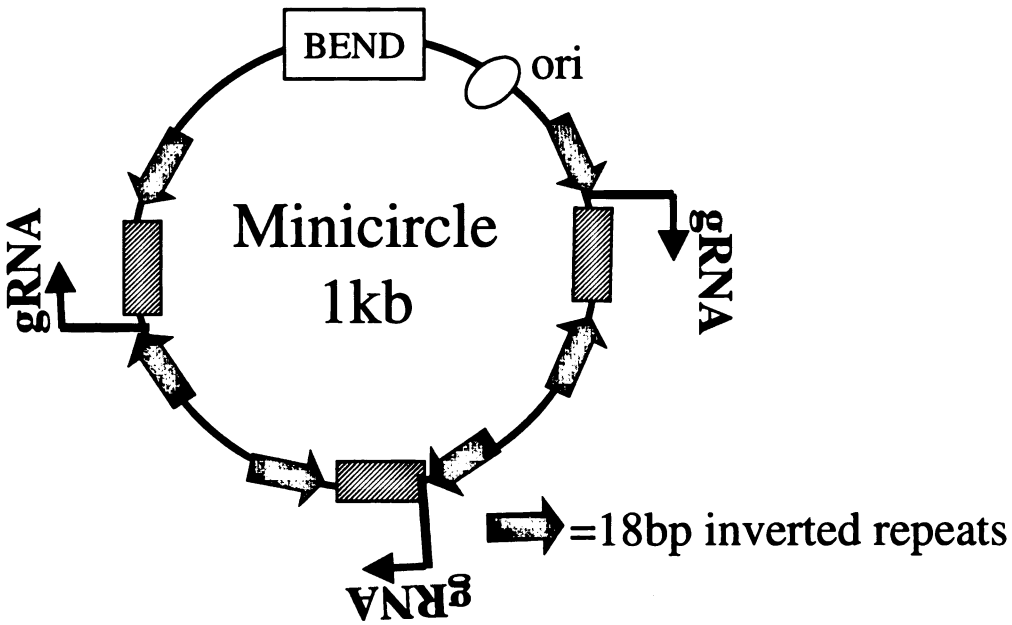


Figure 3. *T. brucei* minicircle. The gray arrows represent the 18bp inverted repeats that flank most gRNA genes. The striped regions are the gRNAs, typically 3 per minicircle. The oval, ori, represents the origin of replication and the box indicates the bend present in the minicircle DNA structure. The black arrows indicate the direction of the gRNA transcription.

FIGURE 3

Position of the maxicircle encoded gMURFII-2 gene in *T. brucei*

Figure 4. Position of the maxicircle encoded gMurfII-2 gene in *T. brucei*. The mapped 5′ ends of both ND4(→) and gMurfII-2 (⇒) are indicated with arrows pointing in the direction of transcription. The 5′ end of ND4 is processed, whereas the 5′ end of gMurfII-2 is a primary transcript. The arrow pointing down indicates the position of gRNA mapped 3′ end. The ATG start codon for ND4 is shown in bold. Courtesy of Donna Koslowsky.

FIGURE 4

Gene overlap of *T. brucei* Maxicircle.

Opposite Strands	Orientation	Overlap	Same Strand	Orientation	Overlap
ND8/ND9	3'/3'	49nts	ND7/COIII	3'/5'	32nts
A6/MURFI	3'/3'	57nts	COIII/Cyb	3'/5'	3nts
MURFI/CR3	5'/5'	61nts	Cyb/A6	3'/5'	0nts
CR3/ND1	3'/3'	41nts	COII/MURFII	3'/5'	31nts
ND1/COII	5'/5'	129nts	COI/CR4	5'/3'	56nts
MURFII/COI	3'/3'	76nts	RSP12/ND5	3'/5'	39nts
CR4/ND4	5'/5'	8nt intergen			
ND4/CR5	3'/3'	44nts			
CR5/RSP12	5'/5'	50nts			

Table 1. Table representing gene overlap on the maxicircle for border regions of different genes. The left half of the table lists neighboring genes on opposite strands with their orientation and overlap region. The right half of the table lists the neighboring genes on the same strand with their orientation and overlap region. Courtesy of Donna Koslowsky.

TABLE 1

Mitochondrial Regulation

During the life cycle, the changes in energy metabolism, discussed previously, are accompanied by distinct changes in mitochondrial gene expression. Steady state levels of mitochondrial mRNAs and rRNAs, polyadenylation of mitochondrial mRNAs, and editing of these same RNAs are also developmentally regulated in a transcript specific manner[33-36].

mRNA Editing

T. brucei mitochondrial mRNA editing is regulated in a transcript specific manner between the two life cycle stages. This regulation is not controlled by the gRNA since they are present in both life cycle stages [27]. The Complex I edited mRNAs NADH Dehydrogenase 8 (ND8) and the 3' editing domain of ND7 are only fully edited in the bloodstream forms, whereas ND9 edited mRNA is similar in abundance in both life cycle stages [32, 37-39]. Ribosomal protein subunit 12, RPS12, has a higher level of edited mRNA in the bloodstream form [32] and C-rich region 4 (CR4) is only fully edited in the bloodstream form, just like ND7 and ND8 [40]. In contrast, the edited form of Cytochrome Oxidase III (COIII), Cytochrome b (CYb), and COII occurs to a higher extent in procyclic than bloodstream forms [36, 39, 41-44]. ATP Synthase 6 (A6) and Maxicircle Unidentified Reading Frame II (MURFII) have similar levels of the edited form

between life cycle stages [43, 45]. MURFI, COI, ND4, and ND5 are never edited (Table 2).

Steady State mRNA Levels

Though the transcription rates of the mitochondrial RNAs are unchanged between the two life cycle forms, the steady state levels of many of the mitochondrial transcripts differ in a transcript specific manner [22]. Steady state levels of the mitochondrial 12s and 9s rRNAs are 30 fold higher in the procyclic form compared to the long, slender bloodstream form [21]. Similarly, transcripts for members of both Complex III (CYb) and Complex IV (COII and COI) are also upregulated in procyclic forms [29, 46]. In contrast, Complex I subunits ND3-5 and ND7-9, MURFI, and RPS12 are elevated in the bloodstream forms of *T. brucei* [32, 35, 37, 38, 47, 48]. This supports the suggestion that the procyclics bypass Complex I in the electron transport chain [41]. Other transcripts, A6, ND1, and MURFII are constitutively expressed and show no difference in transcript levels between the two forms [35, 45, 46] (Table 2).

Polyadenylation

All protein coding mRNAs in the mitochondria of *T. brucei*, with the exception of ND5, occur in 2 size classes that differ by 120-200 nucleotides in length, a difference which cannot be accounted for by RNA editing [33, 34, 36, 49]. RNase H digestion using oligo(dT) showed that this can be accounted for

by the addition of two separate poly(A) tail lengths, short (20 nucleotides) and long (120-200 nucleotides)[46]. Evidence is beginning to emerge that this poly(A) tail is involved in mitochondrial mRNA decay in *T. brucei*. Unedited RPS12 is targeted for degradation with the addition of a poly(A) tail, but has a much longer half life if a poly(A) tail is not added. In contrast, an edited RPS12 mRNA or even a 10% partially 3' edited RPS12 mRNA is quickly degraded if it does not possess a poly(A) tail. The longer (120-200 nt) poly(A) tails on edited RPS12 mRNAs actually are degraded faster then the shorter (20 nt) tails, but slower then a non-poly(A) tail edited mRNA. This suggests a role of both the *cis* edited portion of the mRNA and the poly(A) tail in stabilization [50]. This pathway of mRNA degradation is also dependent on the addition of UTP [51]. This brings up the question of why the larger size transcript (larger poly(A) tail) class consists of mainly edited mRNAs [32, 37, 46, 47, 52, 53] if a longer poly(A) tail is more destabilizing? Complex I mRNA ND8 has a higher percentage of the longer poly(A) tails in the bloodstream form compared to the procyclic form [46]. COIII, CYb, COII, and COI on the other hand have a higher ratio of long poly(A) tails in the procyclic form [42, 46]. ND4 and MURFI have a similar size distribution between the two forms [46] (Table 2).

Poly(A) tail addition in prokarya, such as *Escherichia coli,* acts as a targeting signal for mRNA decay. In the cytosol of eukarya, including humans and *Saccharomyces cerevisiae*, poly(A) tail addition protects them from mRNA degradation. In the mitochondria of *S. cerevisiae*, the poly(A) tail addition is

dispensable without effects on the mRNA metabolism. Plant mitochondrial and chloroplast mRNAs are also not constitutively polyadenylated, although there is evidence that a small proportion of each mRNA can be polyadenylated. Recent data have shown that the polyadenylation in plant mitochondria and chloroplasts can actually trigger degradation of the mRNA [54-57]. Human H-strand mitochondrial mRNAs are all polyadenylated with 50-60 residues, generating functional stop codons and conferring mt-mRNA stability [58-60]. It seems that *T. brucei* mitochondria is another decay system regulated partially by polyadenylation, but unlike other systems, the edited state also seems to regulate the mRNA decay.

Regulation Summary. *T. brucei* mitochondria show numerous levels of post transcriptional regulation. Alhough very little is known about mRNA processing, it is clear that this processing is unique and important in the overall mitochondrial gene expression. Therefore, identifying and learning about the machinery that performs these tasks is important. We believe the PPR proteins will play a important role, as described below.

Regulation of *T. brucei* Mitochondrial Expression

mRNA	Higher Steady State mRNA level	Higher 200:20 nt poly(A) tail ratio	Fully Edited/Higher Levels	Involved in Mitochondrial Respiration	Mitochondrial Complex
ND1	Constitutive		Never Edited	Bloodstream	I
ND3	Bloodstream	Bloodstream	Bloodstream	Bloodstream	I
ND4	Bloodstream	Similar	Never Edited	Bloodstream	I
ND5	Bloodstream	Single size	Never Edited	Bloodstream	I
ND7	Bloodstream	Bloodstream	Bloodstream	Bloodstream	I
ND8	Bloodstream	Bloodstream	Bloodstream	Bloodstream	I
ND9	Bloodstream		Constitutive	Bloodstream	I
CYb	Procyclic	Procyclic	Procyclic	Procyclic	III
COI	Procyclic	Procyclic	Never Edited	Procyclic	IV
COII	Procyclic	Procyclic	Procyclic	Procyclic	IV
COIII	Procyclic			Procyclic	IV
A6	Constitutive		Constitutive	Both	V
MURFI	Bloodstream	Similar	Never Edited		
MURFII	Constitutive				
CR3					
CR4			Bloodstream		
RSP12	Bloodstream	Bloodstream			Ribosomal

Table 2. Regulation of *T. brucei* mitochondrial gene expression. For each mitochondrial mRNA it is listed which lifecycle form has the higher steady state mRNA level, the higher 200nt:20nt poly(A) tail ratio, the higher level of fully edited mRNA, the lifecycle form in which it is thought to function, and in the mitochondrial complex in which it acts.

TABLE 2

PPR Proteins

Pentatricopeptide Repeat (PPR) proteins contain multiple, tandem PPR motifs that are degenerate 35-amino acid sequences that form two antiparallel α-helices with characteristic distributions of hydrophobic and hydrophilic amino acids. Multiple PPR domains are thought to form a superhelix with a central groove proposed to serve as the ligand-binding surface. The width of the groove is sufficient to hold a single strand of RNA. The sidechains lining the central groove are almost all hydrophobic, with positive residues at the bottom to bind the phosphate backbone. This suggests that the PPR motifs may be RNA-binding instead of protein-binding. The PPR proteins have been implicated in stability, translation, mRNA editing and mRNA processing of mitochondrial and chloroplast encoded messages. The largest groups of PPR proteins are found in the higher plants like Arabidopsis, rice, and maize probably due to the complex organelle gene expression in these systems. In all studies with PPR proteins thus far, mutants were not rescued by the other PPR proteins present in the cell, indicating that although there are many PPR proteins, they are not completely redundant.

Pet309 and cya5 are PPR proteins implicated in stability and translation of mitochondrial cytochrome c oxidase subunit 1 (COXI) transcripts in yeast and neurospora respectively [61-64]. Maize CRP1 is required for translation of plastid petA and petD transcripts and in processing of petD from the polycistronic

precursor in the chloroplast [65, 66]. There are cytoplasmic male sterility (CMS) restorer genes that encode PPR proteins in petunia (Rf1)[67], radish (Rfk and Rfo)[68-70], as well as rice (Rf-1). Rf-1 is responsible for processing atp6 mRNA from polycistronic precursor [71, 72]. Most of the other PPR CMS restorers are involved through mRNA stability or processing as well [1]. *Arabidopsis* HCF152 null mutant shows impaired 5'-end processing and splicing of petB transcripts and the HCF152 protein has been found to bind the exon-intron junction of this RNA with high affinity [73-75]. In higher eukaryotes, *Drosophila melanogaster* BSF PPR protein binds a region of the *bicoid* mRNA 3' untranslated region that supports normal mRNA maternal deposition and localization in the embryo during oogenesis through its role in mRNA stabilization. *Drosophila* BSF PPR protein is the only one studied to date that does not have an organellar localization [76]. A single missense mutation in the human LRP130/LRPPC gene, encoding the only human PPR protein studied, is the cause of the genetic LSFC disease, Leigh syndrome French Canadian, characterized by COX1 deficiency [77]. Levels of COXI and COXIII mRNAs were measurably reduced in LSFC patients. Translation of COXI is also reduced, indicating a role in translation or stability of mRNA for COXI [78]. The LRPPC protein has also been found to associate with mRNA/mRNP complexes [79-81] (Table 3). These studies show the involvement of PPR proteins in different aspects of organellar biogenesis. Several of the processes affected in plants are similar in *T. brucei*, such as processing of polycistronic messages.

There have been a few PPR proteins studied that do not fit in to the umbrella description of being organellar targeted and involved in the processes of mitochondrial mRNA stability, processing and translation. There has only been one PPR protein studied to date that is involved in mRNA editing. Mutations in the *crr4* gene of *Arabidopsis thaliana* have been found to be responsible for a decrease in the level of the NA(P)DH dehydrogenase (NDH) complex. There is no alteration in the size or level of the NDH transcripts, but by restriction analysis the mutants were found to be defective in one of their C to U editing sites in the initiation codon of the *ndhD* message [82]. The authors concluded that the PPR mutants are unable to edit this necessary NDH complex message [82] (Table 3). Researchers studying other PPR proteins in plants have looked for effects on RNA editing, but have not found any others as of yet. This may be due to editing mutations not presenting a obvious phenotype. This also may be due to the enormity of possible editing sites within the plant organellar genomes, making it hard to scan for these mutations. In the future, many other PPR proteins will probably be found to be involved in plant organellar RNA editing.

Eukaryotic PPR Proteins

Organism	PPR protein	Gene Affected	Affect
Yeast	Pet309	COXI	Stability & Translation
Neurospora	Cya5	COXI	Stability & Translation
Human	LRP130	COXI	Stability & Translation ?
Drosophila	BSF	bicoid	Stability
Maize	Crp1	petA & psaC & petD	Translation
		petB/petD	Processing
Rice	Rf-1	atp6	Processing
Arabidopsis	HCF152	psbB-psbT-psbH-petB-petD	Processing
Wheat	p63	COXII	Transcription Initiation
Arabidopsis	CRR4	ndhD	mRNA editing

Table 3. Eukaryotic PPR proteins. The table lists representative PPR proteins that have been studied as well as their parent organism, what genes/mRNAs they affect and which mRNA process they affect.

TABLE 3

There are many PPR proteins that may bind mRNA, but there is little direct evidence of this activity. Radish p67 and Drosophila BSF were both purified as sequence-specific RNA binding proteins in vitro [66, 76]. The C-terminal half of LRP130 has been shown to bind single-stranded polyadenylated RNA in vivo [83, 84]. *Arabidopsis thaliana* Hcf152 is a chloroplast protein that binds as a homodimer to the petB exon-intron junctions with high affinity [73-75]. The strongest evidence comes from recent immunoprecipitations and microarray analysis showing that CRP1 binds specifically to petA and psaC mitochondrial mRNA messages in maize [85].

The same group that identified over 400 PPR proteins in plants have identified 18 PPR proteins in the *T. brucei* database [86](Table 4). Next to plants, Trypanosomes have the most PPR proteins of all the organisms investigated. This may be relevant because during evolution Trypanosomes may have had a secondary loss of chloroplasts and they still contain some chloroplast-specific metabolic proteins today [87, 88]. They also both have complex organellar mRNA processing, including polycistronic precursor processing and mRNA editing. We have initiated a project to identify and characterize PPR proteins within *T. brucei*. We then proceeded to investigate the role of one of these proteins, TbPPR1, in *T. brucei* mitochondrial biology. TbPPR1 was found to influence the growth rate of the *T. brucei* cells and affect the expression of several mitochondrial messages.

PPR proteins are specific to Eukaryotes

Organism	Genes	PPR Hits
Homo sapiens	37,490	6
Drosophila melanogaster	17,087	2
Caenorhabditis elegans	20,673	2
Schizosaccharomyces pombe	5,010	6
Saccharomyces cerevisiae	6,304	5
Trypanosoma brucei	16,757	19
Cyanidioschyzon merolae	4,772	10
Arabidopsis thaliana	28,581	470
Oryza sativa	74,385	655
Ralstonia solanacearum	5,118	1
Rickettsia prowazekii	834	0
Synechocystis sp.	3,169	0

Table 4. PPR Proteins are Specific to Eukaryotes. The number of PPR genes (PPR Hits) found through bioinformatics in the eukaryotic organisms that have been sequenced is shown. Genes=number of identified genes for each organism. Modified from: Lurin, C., et al., *Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis.* Plant Cell, 2004. **16**(8): p. 2089-103.

TABLE 4

Overview

A new family of over 400 proteins with PPR domains have been identified, but most of which are in plants. All of the investigated proteins in this family have been implicated in RNA metabolism in the mitochondria and chloroplasts, with the exception of *Drosophila melanogaster* BSF. The massive expansion of this protein family in the plant kingdom is most likely due to the high complexity of organellar expression in plants, including polycistronic processing and RNA editing. *T. brucei* also undergoes complex mitochondrial RNA processing to obtain a complete message. The mRNAs must be excised out from the polycistronic precursor, some are edited, and all of them must be polyadenylated before translation can occur. There is a great deal of information on *T. brucei* mRNA editing and research on polyadenylation is progressing, but nothing is known about the processing of these messages from the polycistronic precursors. I have performed bioinformatics to characterize this family of proteins in *T. brucei*, which, next to plants, has the highest number of PPR proteins of the sequenced eukaryotes. The majority of the 23 PPR proteins identified in *T. brucei* have the traditional mitochondrial targeting signals. Knockdown of one of these proteins produces a slow growth phenotype and alterations in mitochondrial mRNA expression. Based on the current PPR literature and our own research, this family of proteins is highly likely to be involved in *T. brucei* mitochondrial mRNA expression.

My hypothesis is that PPR proteins in *T. brucei* are involved in mitochondrial mRNA metabolism through processing of the polycistronic maxicircle transcript and/or mRNA stabilization of the processed products or mRNA translation. With this study, we have started to characterize the role these important proteins play in *T. brucei* mitochondrial biogenesis.

CHAPTER 2
BIOINFORMATICS

Introduction

Trypanosoma brucei is a parasitic protozoan with a very complex lifecycle. Throughout the lifecycle the metabolism and the mitochondria of the protozoan go through many changes. The factors that control this complex regulation of metabolism and mitochondrial expression between the two lifecycle stages is unknown, but may be related to a family of Pentatricopeptide Repeat (PPR) proteins present in other organisms and *T. brucei* itself. Through studies in plants[1, 2, 65-75, 82, 85, 86, 89-98], humans [76-81, 83, 84, 99], *Drosophila* [76], *Chlamydomonas reinhardtii* [100], yeast [61-63], and *Neurosopora crassa* [64, 101], the PPR proteins are emerging as important factors in organellar mRNA metabolism.

It was not until the *Arabidopsis* sequencing project, that the enormity of this family, over 400 putative PPR proteins, within *Arabidopsis* was realized. Almost all of the PPR proteins studied are implicated in organellar RNA metabolism. Along with all the PPR proteins identified in *Arabidopsis*, Lurin et. al. [86] identified 18 PPR proteins within the *T. brucei* genome. This is the highest number of PPR proteins identified in non-plant eukaryotes. We began looking at the 18 putative *T. brucei* PPR proteins through bioinformatics. Of these 18 proteins, 13 have conventional mitochondrial targeting signals predicted by both Predotar (http://genoplante-info.infobiogen.fr/predotar/) and MitoProt [102], suggesting that this class of proteins plays an important role in Trypanosome mitochondrial biology. We identified the PPR domains in the 18

putative proteins through Pfam and created a probability matrix of the motif in HMMER and MEME. Using this *T. brucei* specific probability matrix, we identified 6 other putative PPR proteins within *T. brucei*, and additional PPR domains were found in almost all of the PPR proteins previously identified in Lurin et al. [86].

Materials and Methods

Protein and Motif Identification.

The PPR domains from the 18 identified proteins [86] were found with Pfam [103] and aligned in ClustalX version 1.81 [104]. The ClustalX alignment was imported into both the HMMER 2.2 package (http://hmmer.wustl.edu/) [105, 106] and MEME [107]. Using hmmbuild and hmmcalibrate, HMMER constructed a probability matrix or position-specific scoring system of the *T. brucei* PPR motif using profile hidden Markov models which can put in additions and deletions anywhere within the alignment. This was then used in hmmsearch for a more extensive search of the *T. brucei*, *Leishmania major*, and *Trypanosoma cruzi* databases for other PPR proteins. MEME [107] used a slightly different algorithm, two-component finite mixture model, which uses a "motif" rather then "profile" method to produce alignments of ungapped blocks. MEME also gave us a slightly more *T. brucei* specific motif to search the databases with (Fig. 5). All amino acids in the MEME have a probability of 0.2 or higher, with the highest probability amino acids listed on the first line. For the consensus sequence from the HMMER the capital letters have probability >0.5 of occurrence.

Mitochondrial Targeting.

Mitochondrial targeting was determined through TargetP [108], Predotar (http://genoplante-info.infobiogen.fr/predotar/), and MitoProt [102].

Databases.

We also used the following databases in our search for kinetoplast and other eukaryotic Orthologs:

Sanger www.sanger.ac.uk

TIGR www.tigr.org

Trypanosoma cruzi Database tcruzidb.org

Results

PPR Consensus Motifs identified with HMMER and MEME

We began by formulating a *T. brucei* specific consensus PPR motif and comparing it to the Arabidopsis to see if we could find species specific differences. A Clustal X alignment of all the Pfam identified PPR domains within the 18 PPR proteins identified by Lurin et al. [86] was imported into both HMMER and MEME. HMMER then calculated a probability matrix based on profile hidden Markov models for the consensus sequence. Capital letters within the consensus indicate a probability of occurrence >0.5. For 16 out of the 35 positions, the *T. brucei* PPR HMMER consensus sequence was identical to the Arabidopsis PPR consensus sequence (Fig. 5). The only differences between the two sequences were conservative replacements such as leucine in position 6 in place of

isoleucine found in Arabidopsis, arginine at position 8 in place of asparagine, cysteine at position 10 in place of tyrosine, and arginine at position 12 in place of lysine. The *T. brucei* HMMER consensus included additional amino acids where the Arabidopsis consensus did not list any, including an alanine at position 13, an aspartic acid at position 15, tryptophan at position 16, glutamic acids at 20, 21, 24, 25, 28 and 36, arginines at positions 27 and 29, a valine at position 31, and proline at 32 and 35. These additions were probably based on the smaller number of PPR domains aligned in *T. brucei* versus the enormous number of Arabidopsis PPR domains that were used for the same type of consensus sequence. All of these additions do provide a consensus sequence that is more specific for *T. brucei*, but with only 5 differences between the two sequences being conservative replacements and 16 out of the 35 amino acids being identical, this gives us a highly conserved PPR motif (Fig. 5).

The Clustal X alignment of PPR domains was also imported into MEME [107] for a consensus sequence based on the two-component finite mixture model. The *T. brucei* MEME consensus included amino acids above a probability of occurrence at 0.2. The only differences the MEME and HMMER programs found between the *T. brucei* consensus sequences were at positions 21 and 23, with the tyrosine to leucine change at position 23 being a conservative replacement (Fig. 5).

T. brucei Specific PPR Motif

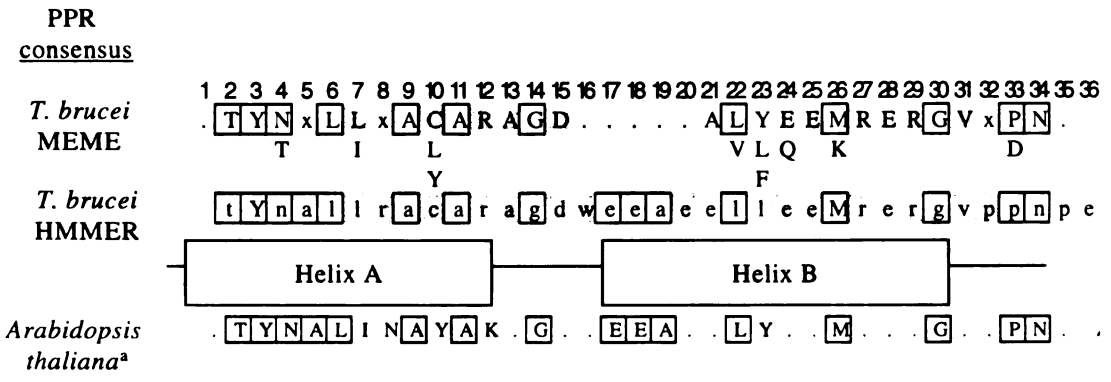


Figure 5. *T. brucei* specific PPR domains. Helix A and Helix B denote the area where the two α-helices of the PPR motif are predicted to be. The MEME consensus sequence is listed first with alternative amino acids on the second and third line. All amino acids in the MEME have a probability of 0.2 or higher, with the highest probability amino acids listed on the first line. The consensus sequence generated from the HMMER program is aligned underneath the MEME with the capital letters having a probability of >0.5. The *Arabidopsis thaliana* PPR consensus sequence published by Small et al. [1] is shown last. Amino acids conserved between both organisms are boxed and the *T. brucei* specific amino acids found are

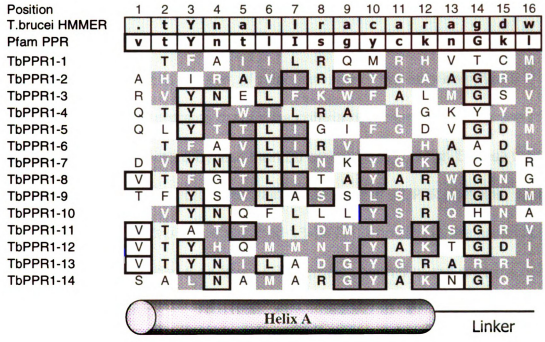
FIGURE 5

Protein Characterization

With the HMMER program [105, 106] we identified 1-4 additional PPR motifs on most (10 of 17) of the *T. brucei* PPR proteins already identified by Lurin et al. [86]. (One of the PPR proteins found by Lurin et al. (Tb11.01.5980) was also dropped because the 2 PPR domains identified were not in tandem.) A representation of the conservation in the *T. brucei* domains aligned with the HMMER *T. brucei* and the Pfam PPR consensus sequences is shown in Figure 6, using the Tb927.2.3180 PPR domains as representative domains.

Figure 6. TbPPR1 PPR motif sequences. Depicted here is an alignment of all the PPR motifs found in TbPPR1 as well as the Pfam consensus motif and the HMMER consensus motif. The light gray amino acids are the amino acids in the TbPPR1 motifs that match the HMMER consensus motif. The medium gray amino acids in the TbPPR1 motifs are what HMMER considers a conservative replacement to the HMMER consensus motif. The outlines boxes of the TbPPR1 motif sequences are amino acids that match the Pfam consensus motif sequence. The portions of the sequence corresponding to helix A or B are above the labeled cylinder for that helix.

TbPPR1 PPR motif sequences



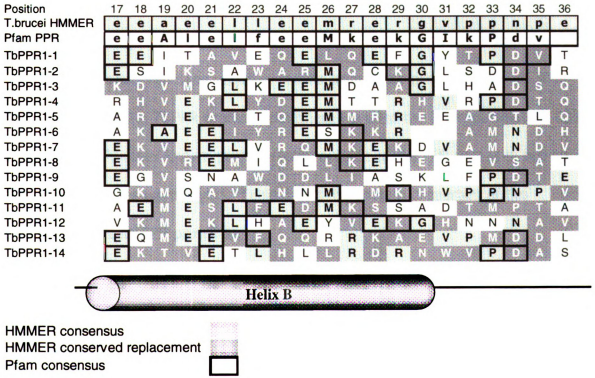


FIGURE 6

Not surprisingly, the new domains were all in tandem with the domains previously identified. PPR domains are usually found in tandem and the new domains that we identified filled in some "holes" in our tandem arrays, making almost every PPR repeat found part of a tandem array. We also identified 6 additional *T. brucei* PPR proteins that fit the criteria of having more then one PPR domain in tandem (separated by less then 60 amino acids) and an E-value less then 10. *Arabidopsis* PPR proteins contain an average of 12 PPR domains. Only 3 out of the total 23 *T. brucei* proteins had 13-14 PPR domains in the proteins. The rest of the proteins had 6 or fewer PPR domains, with 9 of the proteins only containing 2 tandem repeats. Nineteen out of the 23 *T. brucei* PPR proteins were predicted to be mitochondrially targeted by at least two of the following prediction programs: TargetP [108], Predotar (http://genoplante-info.infobiogen.fr/predotar/), and MitoProt [102]. It should be noted that the percentage of observed false negative results from these programs is ~20-30% [76]. With all of the TbPPR proteins identified, we used the Pfam database [103] to try to identify other known protein motifs outside of the PPR domains, but no significant matches were found. Table 5 lists the protein properties of the identified *T. brucei* PPR proteins.

Table 5. *T. brucei* PPR proteins. The table lists each putative *T. brucei* PPR protein found through bioinformatics. Lurin, Pfam and HMMER indicate the number of PPR motifs identified by Lurin et al. [86], by Pfam as listed in the *T. brucei* database, and by our HMMER search. The predicted protein length, molecular weight in Daltons, and the presence of a mitochondrial targeting signal are listed for each protein. The proteins below the bold line are the additional PPR proteins found with using HMMER with the *T. brucei* specific probability matrix.

T. brucei PPR proteins

	Number of PPR Motifs						
ID number	Lurin et al.	Pfam	HMMER	HMMER E-value	Protein Length	MW	Predicted MTS
Tb927.2.3180	11	12	14	3.9E-95	1004	114384.7	yes
Tb11.01.6040	11	10	14	8.1E-83	823	94413.33	no
Tb10.6k15.0120	5	6	6	3.2E-55	459	50695.05	no
Tb09.211.3720	9	10	13	2.6E-81	1023	116545.4	yes
Tb927.1.2990	3	4	5	2.1E-26	1025	115951.1	yes
Tb927.7.1350	2	3	2	5.1E-11	351	40688.09	yes
Tb927.4.4720	3	2	2	5.2E-11	528	59081.97	yes
Tb10.70.7360	2	3	2	4E-11	604	67488.74	no
Tb927.1.1160	3	3	4	3.4E-19	532	58512.57	yes
Tb11.02.5120	2	4	4	9E-16	947	104092.3	yes
Tb10.70.5780	3	4	5	9.1E-13	614	69058.75	yes
Tb10.389.0260	3	3	4	1.8E-17	929	102152.8	yes
Tb10.61.2890	2	1	2	0.000017	362	39633.05	yes
Tb927.3.4450	2	3	3	6.1E-19	676	74840.09	yes
Tb927.3.5240	2	3	3	9.2E-08	582	66655.51	yes
Tb927.8.6040	3	3	4	7E-19	240	26898.62	no
Tb927.6.4190	3	3	6	4.2E-22	552	60803.04	yes
Tb927.8.4860		2	3	0.00023	679	76539.96	yes
Tb927.3.1330		1	2	0.005	1425	156051.6	yes
Tb927.6.4400		0	3	0.0088	603	67159.65	yes
Tb927.5.1790		2	2	0.18	631	71956.81	yes
Tb11.01.7930		0	2	9.3	607	67981.12	yes
Tb09.160.4750		0	2	9.6	667	75135.48	yes

TABLE 5

We used the *Leishmania major* and *T. cruzi* genomic databases (www.sanger.ac.uk, www.tigr.org, and tcruzidb.org) to look for homologues. Homologues were found in both *L. major* and *T. cruzi* for every putative *T. brucei* PPR protein (Table 6) indicating that this is a highly conserved family of proteins across the Kinetoplastida family. Future work should be done to further characterize the *L. major* and *T. cruzi* homologues to see if they also have predicted mitochondrial targeting signals and if their PPR motif distribution throughout the protein is similar to their homologue in *T. brucei*.

Orthologs of *T. brucei* PPR Proteins

Name	T.cruzi Ortho. Tc00.10470535...	L. major Ortho.	Other Ortho	E-value
Tb927.2.3180	..09123.30	LmjF18.0010	Zea mays Crp1	2.60E-18
Tb11.01.6040	..06871.10	LmjF32.1210		
Tb10.6k15.0120	..08307.50	LmjF36.4810		
Tb09.211.3720	..08463.20	LmjF35.2950		
Tb927.1.2990	..06363.50	LmjF20.1580		
Tb927.7.1350	..09505.80	LmjF26.0610		
Tb927.4.4720	..10087.50	LmjF31.1700		
Tb10.70.7360	..03479.10	LmjF21.1270		
Tb927.1.1160	..07949.80	LmjF20.0480		
Tb11.02.5120	..08239.20	LmjF28.0040		
Tb10.70.5780	..08951.10	LmjF21.0260		
Tb10.389.0260	..11275.40	LmjF18.0910		
Tb10.61.2890	..11277.310	LmjF18.0500		
Tb927.3.4450	..07257.10	LmjF29.2010		
Tb927.3.5240	..10143.129	LmjF29.0430		
Tb927.8.6040	..04037.40	LmjF24.2200		
Tb927.6.4190	..11733.30	LmjF30.2880		
Tb927.8.4860	..08837.160	LmjF16.156		
Tb927.3.1330	..09451.24	LmjF25.1230		
Tb927.6.4400	..04257.20	LmjF30.3100		
Tb927.5.1790	..10825.10	LmjF15.041		
Tb11.01.7930	..08525.40	LmjF32.3940		
Tb09.160.4750	..04881.40	LmjF15-1.76		

Table 6. Orthologs of *T. brucei* PPR proteins. This table lists the orthologs found in *T. cruzi* and *Leishmania major* through www.sanger.ac.uk, www.tigr.org, and tcruzidb.org for each *T. brucei* PPR protein.

TABLE 6

Discussion

Through our bioinformatics studies we found a PPR consensus sequence that was slightly more *T. brucei* specific, but aside from 5 conservative replacements, the sequence was identical to the Arabidopsis derived PPR consensus sequence. The bioinformatics studies were able to expand the list of previously identified PPR genes from 18 [86] to 23 and identify homologs within *Leishmania major* and *Trypanosoma cruzi*. We also identified more PPR domains in almost every putative PPR protein previously identified.

Based on the high conservation of the PPR consensus motifs that were found between Arabidopsis and *T. brucei*, the RNA binding properties suspected in the plant PPR proteins were compared to the PPR domains of *T. brucei*. Williams et al. predicted that based on the TPR structure, positions 4 and 12 in helix A of PPR motifs will be important in RNA binding. In a comparison of several plant PPR motifs he found the pattern that position 4 is mostly uncharged polar amino acids, asparagine being the most prevalent. Residues at position 12 are highly charged amino acids where lysine and arginine are most likely [2]. If we look at a representative alignment of *T. brucei* PPR motifs (Fig. 6 & 7), 9 out of 14 of the motifs have an asparagine, serine or threonine at position 4 and 8 out of the 14 motifs have a lysine or arginine at position 12. Five out of the 14 of the representative motifs have both of these elements. Extending this analysis to the total of the *T. brucei* PPR domains, about half of them follow the

trend at position 4, one-third of them at position 12, and about a quarter of them at both positions. Arginine, asparagine, serine, and lysine are the most preferred amino acids at the RNA-protein interfaces, based on a statistical study of 45 crystal structures between RNA and protein complexes by Treger et al. [109]. This corresponds to the trends seen at positions 4 and 12 of the *T. brucei* PPR motifs and the plant PPR motifs investigated by Williams et al. [2]. Another interesting statistic from Treger et al. was that amino acids proline and asparagine prefer contacts with RNA bases, and arginine and lysine prefer phosphate contacts [109]. Asparagine is highly represented at position 4 in our *T. brucei* motif alignments, implying possible contact with the bases of RNA through hydrogen bonding. Arginine and lysine are highly represented at position 12 inferring RNA phosphate contacts at this position of the *T. brucei* motifs.

PPR Putative RNA Contacting Residues

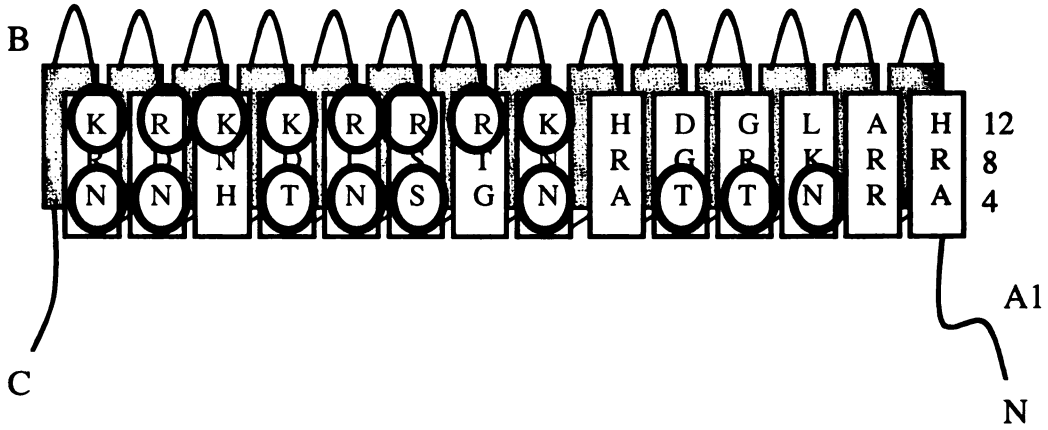


Figure 7. PPR Putative RNA contacting residues. Depiction of the amino acids from positions 4 and 12 of each PPR motif in Tb927.2.3180 , that are hypothesized to be involved in the substrate contacts. N is the N-terminus and C, the C-terminus, of the protein, with A1 representing the first A helix of the protein and progressing towards the C-terminus. 4, 8 and 12 are the positions within the PPR motif sequence. The gray boxes in the back represent the B helices. The amino acids circled in bold and clustered towards the C-terminus are the ones that are predicted by Williams et al. [2] to determine RNA binding of the PPR domains.

FIGURE 7

PPR proteins belong to the a-helical tandem repeat family of proteins that include Tetracopeptide Repeat (TPR) proteins, HEAT repeat proteins, the Puf family of proteins and armadillo (Arm) repeat proteins. Of these family members, the Puf family of proteins has been shown to be RNA binding proteins (Fig. 8). Puf family proteins are regulators of translation and mRNA stability that are characterized by eight imperfect repeats of 36 amino acids. Each Puf repeat forms a trihelical bundle that combines with other tandem Puf repeats to form a right-handed superhelical arc type structure (Fig. 8). The Puf family of proteins have been found in flies, worms, slime mold, frogs, mouse, humans and yeast. The crystal structure of the *Drosophila* Pumilio protein, a protein involved in the control of *hunchback* mRNA in early *Drosophila* embryongenesis, showed that the repeated α-helical structure of tandem Puf domains actually presents an RNA binding surface for the 3' UTR of the *hunchback* mRNA on its inner, concave surface. It is thought to bind other repressional proteins on its outer, convex surface [110]. The crystal structure of human Pumilio1 showed that each of its eight tandem Puf repeats makes contacts with a different RNA base on its concave, inner surface and it is sequence specific. Based on the structural similarities of Puf repeat proteins and that of the PPR proteins, it is conceivable that PPR proteins may bind RNA and other components of protein complexes as well.

Figure 8. Family of α-helical repeat proteins. (A.) Pumillo protein made up of 8 Puf domains, a triple helix repeat protein found in *Drosophila* and humans. (B.) Cartoon of how the Pumillo protein is hypothesized to contact both the mRNA it stabilizes as well as the other proteins in the complex. (C.) Three of the α-helical repeat family members. The first is β-catenin, a protein containing armadillo repeats. The second is a Pumillo protein containing 8 Puf repeats. The final one is a pp2A protein containing HEAT motifs. Obtained from Edwards, T.A., et al., *Structure of Pumilio reveals similarity between RNA and peptide binding motifs.* Cell, 2001. **105**(2): p. 281-9.

Family of α-helical Repeat Proteins

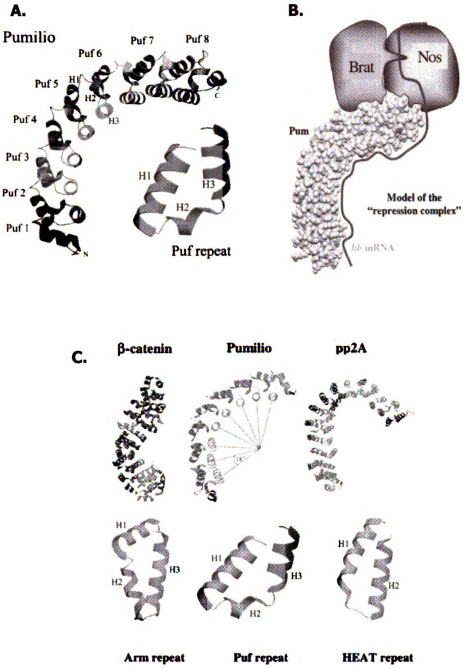


FIGURE 8

In general most of the studies in which PPR protein expression has been altered,have specific effects on certain organellar RNAs [2, 62, 64, 65, 69, 71, 73-75, 78, 82, 85, 91, 94, 100]. However, direct evidence of PPR proteins binding to RNA is limited. Both Radish p67 and Drosophila BSF were purified as sequence-specific RNA binding proteins *in vitro* [66, 76]. The C-terminal half of LRP130 has been shown to bind single-stranded polyadenylated RNA *in vivo* [83, 84]. *Arabidopsis thaliana* Hcf152 is a chloroplast protein that binds as a homodimer to the petB exon-intron junctions with high affinity [73-75]. The strongest evidence comes from recent immunoprecipitations and microarray analysis showing that CRP1 binds specifically to petA and psaC mitochondrial mRNA messages in maize [85]. The statistical evidence presented by Williams et al. [2] and Treger et al. [109] also support PPR proteins binding to RNA.

Accumulated evidence suggests that PPR proteins are organellar proteins involved in RNA metabolism. While the number of definitive studies is limited, the mutant phenotypes and the characteristics of the PPR motif strongly suggest that they are RNA binding proteins. In plants, the explosive expansion of this family is thought to be due to the unique nature of the RNA processing that is required. In trypanosomes, the mitochondrial genome is also transcribed in polycistronic units and requires extensive processing to generated translatable messages. The unique organization of the trypanosome mitochondrial genome suggests that specific RNA binding proteins will be required for many of the

maturation steps. The identification of a large number of *T. brucei* PPR proteins, most of which are mitochondrially targeted, suggests that PPR proteins will play important roles in *T. brucei* mitochondrial biogenesis.

CHAPTER 3
RNAi DATA

Introduction

We have identified a large class of PPR proteins in *Trypanosoma brucei*, most of which have predicted mitochondrial targeting signals. This class of proteins appears to play important roles in organellar RNA metabolism in plants. In order to determine if one is important in organellar gene expression in *T. brucei* as well, we have studied the effects of TbPPR1 (Tb927.2.3180) down regulation on mitochondrial RNA expression. TbPPR1 was chosen because it had the highest e-value of its PPR domains. TbPPR1 has the highest number of PPR domains, when compared to the other identified *T. brucei* PPR proteins, at 14 (Fig. 9). TbPPR1 is predicted to be 1004 amino acids in length with a molecular weight of about 114 kD. Both MitoProt [102] and Predotar (http://genoplante-info.infobiogen.fr/predotar/) predict TbPPR1 to contain a traditional mitochondrial targeting signal.

We have targeted TbPPR1 for RNAi knockdown regulation in *T. brucei* and observed a reproducible slow growth phenotype when compared to non-induced procyclic cells. The decrease in TbPPR1 mRNA was associated with specific changes in the levels of several mitochondrial mRNAs and in RNA editing. For the CYb transcript, we observed a decrease in the longer poly(A) tail and increase in the shorter poly(A) tail size class. The 12s rRNA, 9s rRNA, and the ND4 transcript showed a steady decrease throughout TbPPR1 RNAi induction. We then looked further to investigate mRNA editing in our TbPPR1 deficient cell

line through using a "poison primer extension" assay and showed that the edited CYb message is greatly decreased when TbPPR1 is knocked down.

Materials and Methods

RNAi construct

A double-stranded RNA target homologous to PPR1 was chosen with the help of the TrypanoFAN: RNAit program (Whitehead Institute for Biomedical Research). The program uses both MIT primer 3 [111] and NCBI Blast [112] to both identify good fragments for PCR amplification and avoid crosstalk between related gene products. A fragment from TbPPR1 was amplified through PCR of 29-13 *T. brucei* genomic DNA with primers synthesized by IDT:

PPR1F1XhoI 5' TAGCTCGAGTGATTGTGCTGCAGGAGTTC 3'

PPR1R1HindIII 5' TAGAAGCTTCTAAATCATCCGGCTCCAAA 3'

This amplified TbPPR1 from nucleotides 224-721 (Fig. 9).

TbPPR1 PPR domains

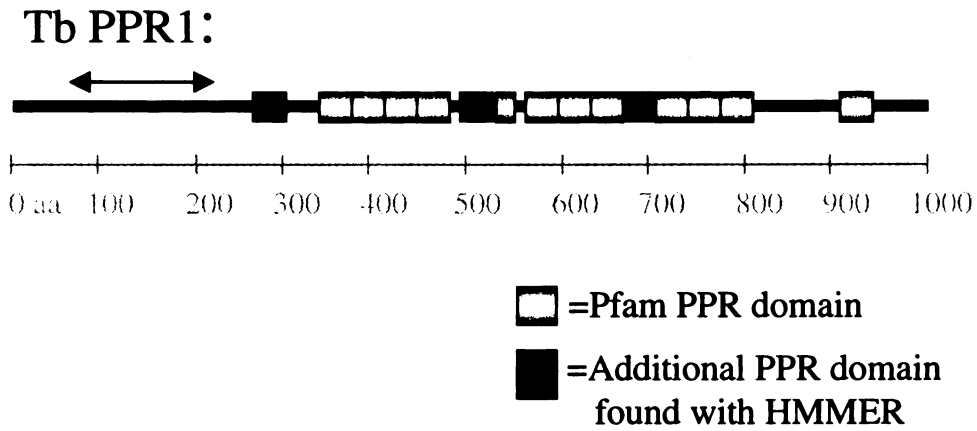


Figure 9. TbPPR1 PPR domains. This figure is a graphical depiction of the location of PPR domains on the TbPPR1 protein and the ones that are located in tandem. They are concentrated between the middle to the C-terminal end of the protein, with the majority of the motifs being in tandem. The light gray boxes are PPR motifs found through Pfam and the dark gray boxes are additional PPR motifs found with HMMER. The double headed arrow indicates the region of the protein that was targeted for RNAi in the pZJM vector.

FIGURE 9

Both the TbPPR1 fragment and pZJM vector (Fig. 10) were digested with XhoI and HindIII (New England Biolabs), gel purified and ligated with T4 DNA Ligase (Invitrogen) according to manufacturers instructions.

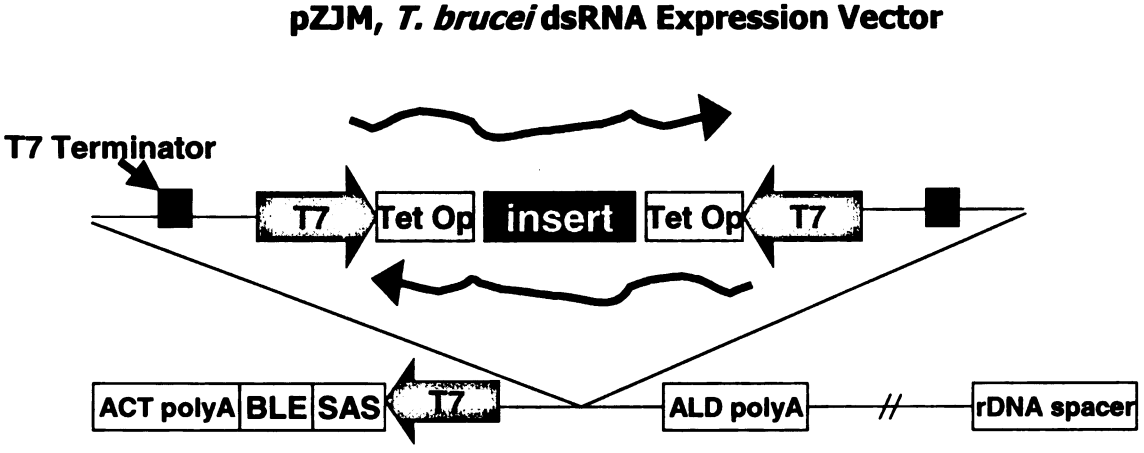


Figure 10. pZJM T. brucei dsRNA expression vector. The pZJM construct has two opposable T7 promoters (T7) that drive the expression of the insert under the control of a tetracycline operator (Tet Op). Integration is selected for by bleomycin (BLE) resistance, driven by a separate, unregulated T7 promoter. The construct contains rDNA sequence for plasmid integration. The SAS is the splice acceptor site. ACT poly A and ALD poly A are the actin and aldolase poly adenylation sites respectively. The construct is specific for the *T. brucei* cell line, 29-13, that stably expresses the T7 polymerase and Tetracycline repressors.

FIGURE 10

Transfection, RNAi induction, and Growth Curve

Procyclic *T.brucei* strain 29-13 (generated by the George Cross lab) cells are maintained in SDM-79 medium with 10% fetal bovine serum (FBS) (Sigma), 15 ug/ml G418 and 50 ug/ml hygromycin B. While on ice, 20 ug of linearized plasmid in 100ul Elution Buffer (EB)(10 mM Tris-Cl, pH 8.5) was added to 0.5 ml of 1 x 10^8 cells that had been washed and resuspended in three-quarters cytomix (120mM KCl, 150μM $CaCl_2$, 10mM K_2HPO_4, 25mM HEPES, 2mM EDTA, and 5mM $MgCl_2$, pH 7.6) and one-quarter PB (277mM sucrose, 7mM KH_2PO_4, 1mM $MgCl_2$, pH 7.4). The cells were electroporated with a BioRad Gene Pulser at 1.5 kV, 25 uF, ∞ resistance for 2 pulses with 10 seconds between each pulse. Electroporated cells were immediately transferred to 9.5 ml of SDM-79 medium with 10% FBS and allowed to recover in a non-shaking 27°C incubator for ~16 hours. The cells were then centrifuged at 1500 x g and resuspended in SDM-79 with 10% FBS and 2.5 μg/ml phleomycin and put into a 27°C non-shaking incubator for selection of stable transfectants. Cells were then induced for RNAi production in the same medium with 1 ug/ml tetracycline and harvested in log phase at the indicated times by centrifugation. Growth curves were plotted with the cell counts multiplied by the dilution factors used to keep the cells in a constant log phase of growth. Excel (Microsoft) was used to fit an exponential line to the growth curves and the slope of the plus Tet divided by the slope of the no Tet controls was used to calculate the growth rate.

Northern Anaylses

Total cellular RNA was isolated using a guanidinium-phenol-choloroform procedure [113]. Ten micrograms of total RNA were run for each time point in the RNAi growth curve on 1.5% formaldehyde denaturing agarose gels. RNA was then transferred by downward capillary transfer overnight to a Nytran membrane (Schliecher & Schuell Bioscience Inc.) and crosslinked. Probes for β-tubulin, ND8 unedited, ND7 unedited, CYb, ND1 and ND4 were DNA probes made with a random priming system (Invitrogen, Cat. No 18187-013) using a PCR template with α^{32}P-dATP. The CYb probe was designed to pick up both edited and unedited messages. Probes specific for 12s and 9s ribosomal RNAs were oligodeoxyribonucleotides (IDT), ^{32}P-ATP end-labeled (Invitrogen T4 Kinase) according to the manufacturer's instructions. Riboprobes made with the T7 Maxiscript kit (Ambion, Cat. No 1312) and α^{32}P-ATP were used for PPR1 and dsRNA-PPR1. Bands were quantitated using ImageQuant software (Amersham Biosciences).

Poison Primer Extensions

Total cellular RNA was isolated using a guanidinium-phenol-choloroform procedure and DNase treated to remove contaminating genomic DNA [113]. The following oligo was 5' end-labeled with T4 kinase (Invitrogen) and γ^{32}-ATP according to the manufacturer's instructions:

CYb poison RT 5' CTATATAAACAACCTGACATTAAAAGAC 3'

End-labeled primers were mixed in 1x RT Buffer (50 mM KCl, 20 mM Tris pH 8.3, 0.5 mM EDTA, and 8 mM $MgCl_2$) with the indicated amounts of RNA in a volume of 20 μl. The mixture was heated to 70°C for 2 minutes and then slowly cooled 2°C per minute to 50°C. Five μl of extension cocktail (1x RT Buffer, 10 mM dATP, 5 mM dTTP, 5 mM dCTP, 2 mM ddGTP, 20 units RNasin, 15 units of Seikagaku AMV) pre-warmed to 50°C was added to the RNA-oligo mixture, and the reaction incubated at 50°C for 45 minutes. Reactions were quenched on ice, phenol/chloroform extracted and extension products were ethanol precipitated. Primer extension products were analyzed by electrophoresis on 8% polyacrylamide gels containing 8M urea.

Results

RNAi

Induction of double-stranded RNA transcription causes a slow growth phenotype when compared to uninduced cells (Fig. 11). Stable transfectants of an RNAi strain against TbPPR1 were induced for RNAi production with 1 μg/ml tetracycline. The growth rate is decreased on average to 64% of uninduced for trials A through D. No morphological phenotype was seen by light microscopy in these cells. Cell growth recovery was observed in every case after about 2 weeks, which is normal for RNAi in this system [114]. Cells were harvested at time points throughout the growth curves from induced and non-induced cells

and the total RNA from these cells was isolated (Fig. 11). A control growth curve is shown in Figure 12 to show that tetracycline has no significant effect on untransfected 29-13 cells, the parent cell line.

Figure 11. TbPPR1 RNAi Growth Curves. The top graph depicts the average growth curve for A and B trials for the TbPPR1 RNAi and the bottom graph is for C and D trials. Total cells is calculated by the cell count multiplied by the dilution factor used to keep the cells in a log phase of growth. The black/diamond lines are cells not exposed to tetracycline, therefore not induced for TbPPR1 RNAi. The gray/square lines are cells exposed to tetracycline and are induced for TbPPR1 RNAi. A and B TbPPR1 RNAi cells decrease growth to 60% of uninduced and C and D to 67.7% of uninduced.

TbPPR1 RNAi Growth Curves

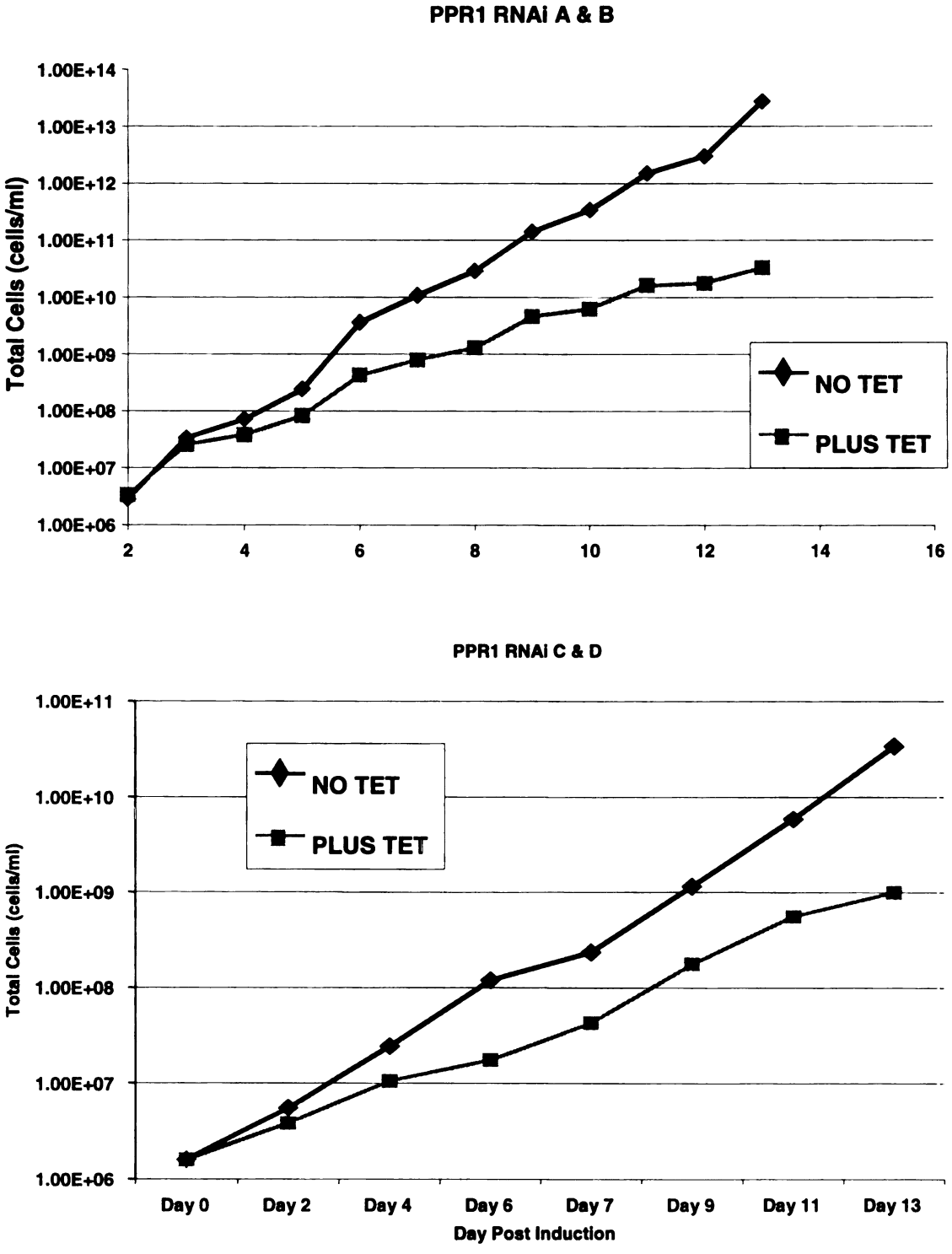


FIGURE 11

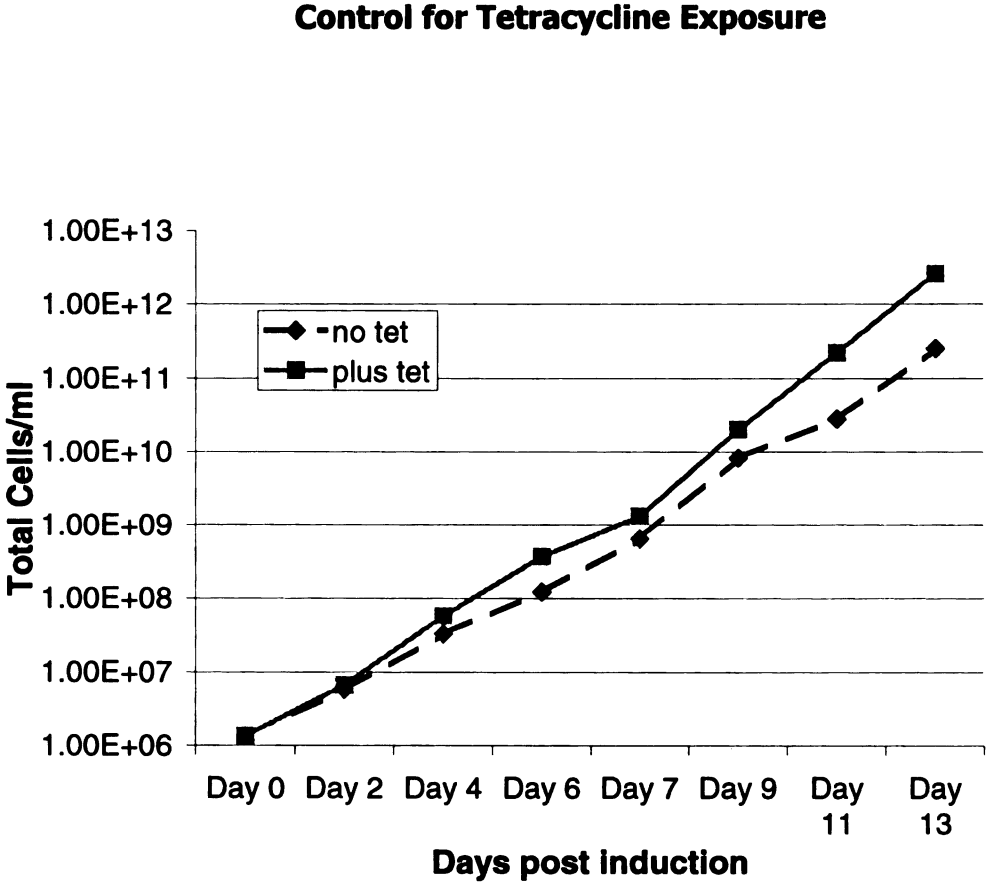


Figure 12. Control for Tetracycline Exposure. The graph depicts the growth curve for untransfected 29-13 cells exposed to tetracycline. Total cells is calculated by the cell count multiplied by the dilution factor used to keep the cells in a log phase of growth. The black/diamond, dotted lines are cells not exposed to tetracycline. The gray/square lines are cells exposed to tetracycline. The growth rate of the cells show no significant difference upon exposure to tetracycline.

FIGURE 12

The RNA was then used in Northern blot analysis and hybridized with end-labeled oligos, random primed labeled oligos, or riboprobes to recognize TbPPR1 mRNA, the dsRNA-TbPPR1 that was used in the vector construct, and other mitochondrial messages.

The TbPPR1 message, from growth curve A, was probed with a riboprobe that hybridized to the same portion of TbPPR1 message that is in the RNAi construct. This will give bands for the TbPPR1 message and for the smaller dsRNA that is produced from our pZJM-TbPPR1 construct. The TbPPR1 message was visibly decreased in the cells that were induced for RNAi with tetracycline. There is some noticeable recovery of the cells by Day 15, which is a common occurrence for RNAi in trypanosomes [114-116]. There was a dramatic increase in production of the dsRNA upon tetracycline addition to the media, with the highest production on Day 3. There is also some leaky expression of dsRNA that is apparent in the no tetracycline lanes. This has also been reported in Trypanosomes [114, 117]. However, TbPPR1 levels were not visibly reduced by the leaky dsRNA expression when compared to the parent cell line (data not shown) (Fig. 13).

Northern Blot of TbPPR1 RNAi Trial A: TbPPR1 and dsRNA Expression

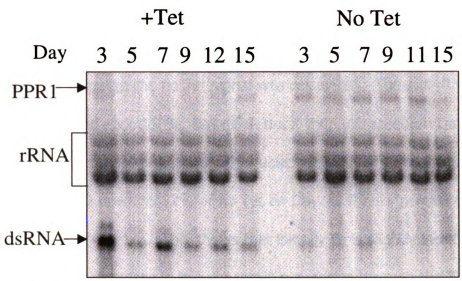


Figure 13. Northern blot of TbPPR1 RNAi Trial A: TbPPR1 mRNA and dsRNA expression. The left side of the blot shows samples from cells induced for TbPPR1 RNAi expression (+Tet) and the right side shows samples from uninduced cells (No Tet) RNA was isolated on Days 3, 5, 7, 9, 11/12, 15 post induction. The top band is the TbPPR1 full-length message, which is visibly decreased in the + tetracycline lanes. The middle three bands represent the rRNAs of the cell from nonspecific binding of the probe. The bottom band is the dsRNA TbPPR1 message produced from the pZJM-TbPPR1 vector when cells are induced with tetracycline. There is evidence of the dsRNA production without the presence of tetracycline.

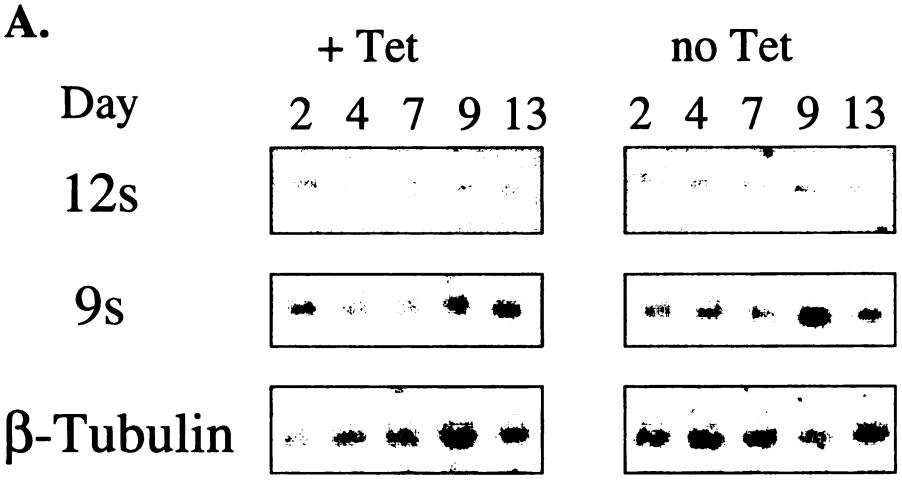
FIGURE 13

Evaluation of TbPPR1 Depletion Effects on Mitochondrial RNA Transcripts

The Northerns from Growth Curves C/D were probed and the signals were normalized against the β-Tubulin signal for that lane (Fig. 14-16). Then the ratio of the plus Tet signal to the no Tet signal was calculated. The 12s rRNA signal at ~2 kb decreased about 50%, Days 2-9, but it then increased on Day 13 (Fig. 14). The 9s rRNA signal is at about 606 nucleotides (nt) and went from an slight induction on Day 2 down to 52% of no Tet on Day 9. It also increased above 100% on Day 13 (Fig. 14). This shows that initially 9s is slightly induced, but there is an overall decrease in the ribosomal expression through the TbPPR1 RNAi induction Days 2-9. There then is an increase in steady state levels on Day 13 indicating that the cells may be recovering from the RNAi effect on this day. Days 12 and 15 on the Northern of the TbPPR1 expression also show that the TbPPR1 expression is increasing at this time as well (Fig. 13).

Northern Blot of TbPPR1 RNAi Trial C/D: 12s and 9s rRNA Expression

A.



B.

	Day 2	Day 4	Day 7	Day 9	Day 13
12s	0.98	0.64		0.52	2.01
9s	1.20	0.74	0.90	0.52	1.48

Figure 14. Northern blot of TbPPR1 RNAi Trial C/D: 12s and 9s rRNA expression. A. + Tet=induced (+ tetracycline) and no Tet=uninduced (no tetracycline) for TbPPR1 RNAi expression on Days 2-13 post induction. 12s=12s ribosomal RNA, 9s=9s ribosomal RNA, and β-Tubulin message used as a loading control. B. Quantitation of the signal normalized against β-Tubulin indicates decrease in expression of both rRNAs with a recovery observed by Day 13. 12s Day 7 not quantitated due to defects in gel that could not be accounted for in the quantitation.

FIGURE 14

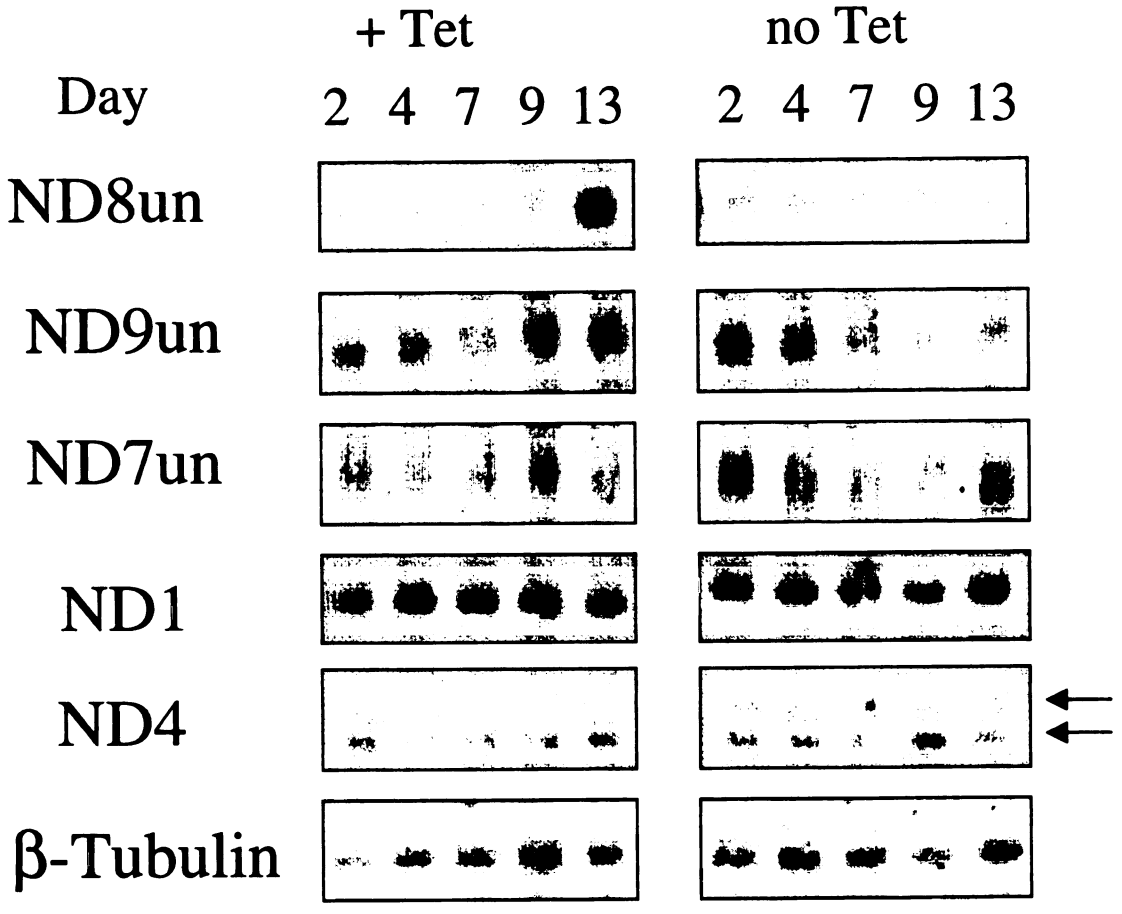
The ND7, 8, and 9 from mitochondrial Complex I, all had no significant change in the expression for the unedited forms when induced for TbPPR1 RNAi (Fig. 15) Days 2-9 post induction. The never edited ND1 also remains unchanged during the TbPPR1 RNAi induction. The single band at 1300 nt ND1 is between 102% and 115% Days 2-13, remaining unchanged (Fig. 15). ND7 unedited signal, at ~850 nt, remained between 89% and 99% of no Tet. The ND8 unedited signal at about 520-550 nt remained between 94% and 84% of no Tet Days 2-9. On Day 13 ND8 unedited signal increased. The ND9 unedited at 340 nt was between 91% and 97% Days 2-7 and increased by Day 13 (Table 7). This shows little overall change in most Complex I mitochondrial mRNA expression throughout Days 2-9 of our TbPPR1 RNAi induction. Some of the mRNA messages do show an increase on Day 13, similar to that seen in the 12s and 9s rRNAs. This may also indicate recovery of the cells from the tetracycline induced RNAi effect.

While most Complex I members show no change in expression Days 2-9 of the TbPPR1 RNAi induction, there is a specific decrease in both size transcripts (1600 and 1400 nt) of ND4 for the same time points. The larger transcripts are the long size class of poly(A) tail (~200 nt) and the smaller transcripts are the short poly(A) tail size classes (~20 nt). ND4 decreases from 135% of no Tet on Day 2 to 46% on Day 9. Similar to 12s, 9s and some other Complex I members, the ND4 expression increases back up to 112% of no Tet on Day 13 (Fig. 15).

Figure 15. Northern blots of TbPPR1 RNAi Trial C/D: Complex I expression. A. + Tet= induced (+tetracycline) and no Tet=uninduced (no tetracycline) for TbPPR1 RNAi expression Days 2 –13 post induction. The top panels are the Complex I Northern blots: ND8un= NADH dehydrogenase subunit 8 unedited message, ND9un= ND9 unedited message, ND7un=ND7 unedited message, ND1, and ND4 message. The bottom panel is the β-Tubulin message used as a loading control. B. Quantitation of the signals normalized to β-Tubulin indicate no overall change in expression of Complex I mRNAs during the TbPPR1 induction Days 2-9, except for a decrease in both bands of the ND4 message. Many messages also show an increase in their steady state levels on Day 13. The arrows point to the two size class transcripts of ND4.

Northern Blots of TbPPR1 Trial C/D: Complex I Expression

A.



B.

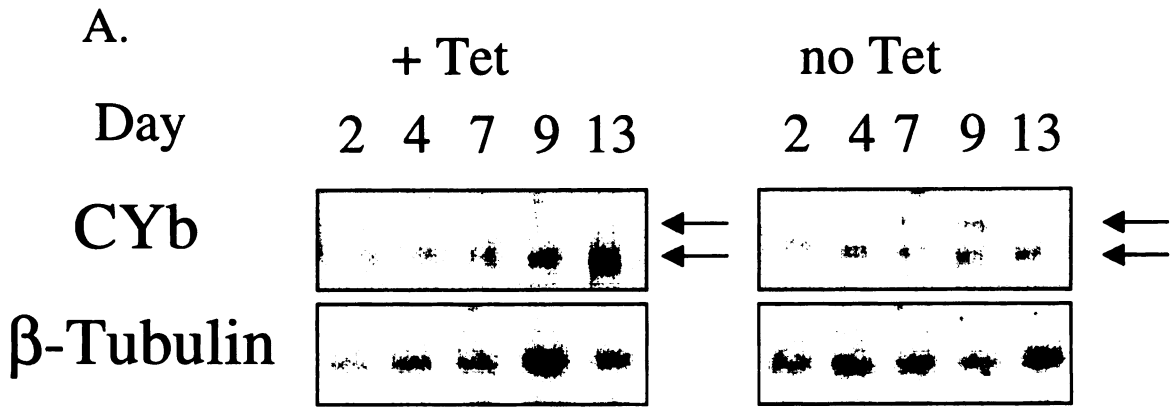
	Day 2	Day 4	Day 7	Day 9	Day 13
ND8	0.94	0.84	0.92	0.93	4.02
ND9	0.95	0.91	0.97	1.31	1.69
ND7	0.99	0.89	0.88	0.99	0.90
ND1	1.14	1.15	1.04	1.02	1.14
ND4	1.37	0.80	0.78	0.46	1.12

FIGURE 15

T. brucei mitochondrial Complex III's CYb transcripts showed some interesting results on the Northern. The probe used recognized both the edited and unedited CYb message. The upper transcript (1350 nt), long poly(A) tail size class, decreases along with the induction of TbPPR1 RNAi in the plus Tet induced lanes. It is also evident that the lower transcript (1200 nt), short poly(A) tail size class, increases when induced for TbPPR1 RNAi (Fig. 16). The steady state levels of the total CYb transcripts increase during all 13 Days of TbPPR1 RNAi induction. It begins at 165% of no Tet on Day 2, decreases to 103% on Day 4, and then progressively increases to 459% on Day 13 (Fig. 16). This shows the overall steady state levels of CYb message increases during TbPPR1 RNAi induction, but the size classes of poly(A) tail steady state levels change with the longer class decreasing and the shorter size class increasing. The longer poly(A) size class of CYb is primarily edited CYb messages while the unedited CYb messages are only in the shorter poly(A) size class [43].

Northern Blot of TbPPR1 RNAi Trial C/D: CYb Expression

A.



B.

	Day 2	Day 4	Day 7	Day 9	Day 13
CYb	1.65	1.03	1.41	2.28	4.59

Figure 16. Northern blot of TbPPR1 RNAi Trial C/D: CYb expression. A. + Tet=induced (+ tetracycline) and no Tet=uninduced (no tetracycline) for TbPPR1 RNAi expression on Days 2-13 post induction. CYb=Cytochrome b mRNA and β-Tubulin message is used as a loading control. There is an obvious decrease in the upper band or longer poly(A) tail size class when induced for TbPPR1 RNAi. The arrows point to both poly(A) size classes of CYb. B. Quantitation of the signal normalized against β-Tubulin indicates an increase in expression of both size class of transcripts combined through Day 13.

FIGURE 16

Northern Summary. The Northern results show a decrease in expression of the ribosomal RNAs, though the 9s rRNA may be slightly induced initially. Complex I messages, ND1, 7, 8, and 9 also show no significant change in steady state expression levels throughout the TbPPR1 RNAi induction Days 2-9. CYb of Complex III, had a great decrease in the intensity of the long poly(A) size class transcripts during TbPPR1 RNAi, but it had an overall increase in its steady state levels of the shorter poly(A) size class transcripts. ND4 also looked like it followed the same pattern as CYb, loss of longer poly(A) size class transcripts, but when it was further quantitated there was a decrease in both of its bands, not just the longer poly(A) size class. ND4 is the only Complex I member to decrease in expression that we have identified.

Poison Primer Extension

The longer poly(A) tail size class of CYb is primarily edited CYb and the unedited CYb messages are only in the short poly(A) size classes [43]. Based on the loss of the longer poly(A) size class transcripts of CYb during TbPPR1 RNAi, we investigated whether the editing of CYb was truly affected through "poison" primer extension analysis. CYb is edited by the insertion of 34 U-residues at 13 different sites near the 5′ end of the transcript [36]. Placing a reverse transcription (RT) primer directly 3′ of the first CYb editing site and replacing dGTP in the Poison Primer reaction cocktail with ddGTP will stop synthesis at the first cytosine residue encountered, producing a 74 nt band for the edited CYb

messages and a 45 nt band for the unedited CYb messages (Fig. 17). Both bands are then quantitated and the ratio of edited to unedited CYb message is calculated. Our analyses indicate that there is a distinct loss of editing activity in the samples induced for TbPPR1 RNAi induction (+Tet) compared with those that were not (- Tet). The samples from the cells induced for TbPPR1 RNAi induction showed an edited:unedited ratio of ~1% to 9%, indicating that most of the messages are unedited. The cells that were not induced for TbPPR1 RNAi showed ratios between 47% to 114%. In the no Tet lanes there is a lower ratio for the 10ug RNA samples versus the 5 ug RNA samples. This may be due to a limiting factor in our Poison Primer Reaction Buffer, most likely dATP due to all of the editing that is present. This experiment shows a clear loss of CYb editing activity that accompanies the loss of the long poly(A) tail size class of transcripts when the cells are induced for TbPPR1 RNAi.

Figure 17. Poison Primer Extension of CYb on TbPPR1 RNAI Total RNA. A. The RNA sequences of both unedited (top) and edited (bottom) where the capital letters are nucleotides encoded by the maxicircle genome and the lowercase, bold **u**'s are the u's added during mRNA editing as directed by the gRNA. The outlined "ℂ" is where the reverse transcription will stop when ddGTP is incorporated. Position of the primer is indicated with black boxes. The arrow shows the extension that we expect with the size of each product, including the primer length, listed. B. Primer extension analysis of CYb editing during PPR knockdowns when both 5 and 10 micrograms of + tetracycline and no tetracycline total RNA from the corresponding time points were analyzed. The ratio of the edited to unedited band was then quantified and shown below its corresponding lane.

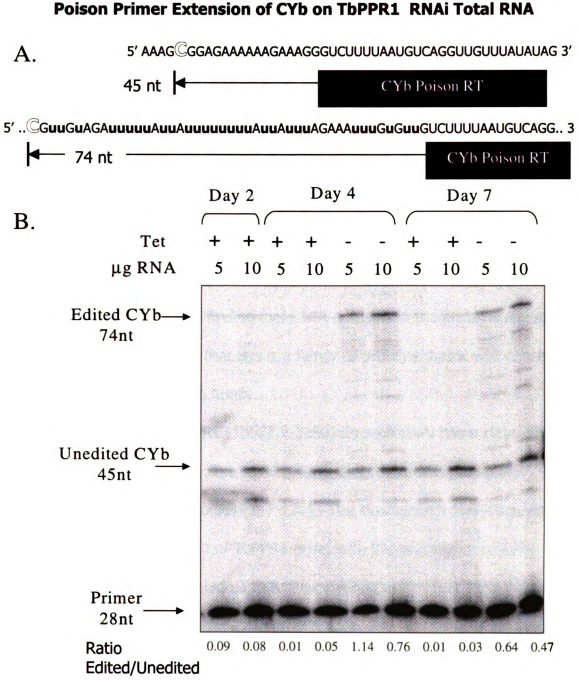

Poison Primer Extension of CYb on TbPPR1 RNAi Total RNA
A.
5′ AAAGCGGAGAAAAAAGAAAGGGUCUUUUAAUGUCAGGUUGUUUAUAUAG 3′
45 nt
CYb Poison RT
5′ ..CGuuGuAGAuuuuuAuuAuuuuuuuuAuuAuuuAGAAAuuuGuGuuGUCUUUUAAUGUCAGG.. 3
74 nt
CYb Poison RT
B.
Day 2
Day 4
Day 7
Tet + + + + - - + + - -
µg RNA 5 10 5 10 5 10 5 10 5 10
Edited CYb 74nt
Unedited CYb 45nt
Primer 28nt
Ratio Edited/Unedited
0.09 0.08 0.01 0.05 1.14 0.76 0.01 0.03 0.64 0.47

FIGURE 17

Discussion and Conclusion

The family of PPR proteins is an interesting and greatly expanding family in plants. Therefore their functions in *T. brucei* should also prove to be of great interest. Based on analysis of the other RNA binding proteins, PPR proteins show a high probability of being RNA binding proteins. Studies of plant PPR proteins and those in other eukaryotes, show that they are involved in different processes of organellar mRNA expression. Our bioinformatics studies built on the 18 *T. brucei* PPR proteins found by Lurin et al. [86], expanding the family to 23 proteins in *T. brucei* and finding more PPR domains in the proteins already discovered. It also showed that this is a family of proteins that is well conserved throughout the kinetoplastid family.

Knocking down TbPPR1 (Tb927.2.3180) through RNAi has a clear effect on the growth of *T. brucei*. This same growth defect was seen using this plasmid for RNAi in other studies [117-121]. The slow growth phenotype that accompanied the knockdown of TbPPR1 along with the importance of PPR proteins in other organisms led us to further investigate the function of TbPPR1 in the mitochondrial RNA metabolism in *T. brucei*. The ribosomal RNAs decrease in expression and ND1, ND7, ND8 and ND9 mRNAs all remain relatively unchanged during TbPPR1 RNAi induction. The ribosomal decrease may be a direct or indirect effect from the loss of TbPPR1, but not enough is understood about *T. brucei* mitochondrial biogenesis to know for sure. A similar ribosomal

RNA decrease also occurs in the cytoplasm of bacteria when undergoing stress induced by pH changes [122].

For CYb mRNA, the amount of long poly(A) tail size class present during TbPPR1 RNAi induction, seems to decrease whereas the short poly(A) tail size class is stabilized. ND4 also appeared to lose the long poly(A) tail size class, but without any stabilization of the short poly(A) tail size class; however quantification showed that there is a decrease in both poly(A) tail size classes. There are several possibilities to explain why the long poly(A) tail size class of CYb decreased. One theory is that there may be a defect in polyadenylation. Loss of TbPPR1 could be affecting the poly(A) machinery itself or a simple loss of ATP to incorporate into the poly(A) tails due to cellular stress as the TbPPR1 RNAi progresses. It could also be due to loss of CYb editing as shown in the poison primer extension analysis. It has been shown that the majority of the long poly(A) tail size class is composed of edited CYb messages, where the unedited messages are in the short poly(A) tail size class [43]. We do know that for the RPS12 mRNA, the edited form is targeted for destruction when it does not posses a poly(A) tail. In contrast the unedited form is targeted for destruction when it does contain a poly(A) tail [50, 123]. So for this mRNA, there is a connection between the editing state, poly(A) tail addition and mRNA degradation which may also be occurring for CYb. Both mRNA editing and processing/polyadenylation have been shown to be two independent events [31]. Based on this information, the composition of the CYb poly(A) tails in the TbPPR1

knockdowns should be studied next. The loss of ND4 would not fit this hypothesis since ND4 is never edited. Unfortunately, there is not yet enough information available on *T. brucei* mitochondria RNA metabolism to speculate any further.

Further studies should then tell us if transcripts within the bloodstream form are affected by TbPPR1 RNAi and whether the protein expression is affected confirming a role in mitochondrial mRNA biogenesis. Further structural studies should also confirm that these PPR proteins are in fact RNA binding proteins.

Future Work

We have been attempting to confirm the Northern results using Real Time PCR, but we are still optimizing this procedure. We have developed good primer pairs that produce high quality melt curves for almost every mitochondrial transcript. Primer concentrations and cDNA concentrations are still being optimized for each primer set. We have also begun to characterize the poly(A) tails in the TbPPR1 RNAi cells through 5′ and 3′ RACE, but more primer pair optimization must also be worked out here.

Regulation of both the TbPPR1 protein and RNA from both bloodstream and procyclic Trypanosomes can be further studied with Westerns and Northerns respectively. Many characterization studies can also be done when we *in vivo* epitope label the PPR1 protein or obtain antibodies against the TbPPR1 protein. Western blot protein characterization, immunolocalization, subcellular

fractionation, and protein complex studies can all be done with these epitope tags and/or antibodies.

We are also working on a collaboration to express the TbPPR proteins and perform crystallography and NMR studies with the proteins and their RNA substrates (determined through mRNA expression alterations via Northern blots and ribonuclease protection assays) to look at the binding surface of the PPR motif. These proteins and RNA substrates will also be used in gel mobility shift assays to determine the kinetics of binding. Eventually we hope to use several different approaches to find some drugs that target these PPR proteins.

BIBLIOGRAPHY

1. Small, I.D. and N. Peeters, *The PPR motif - a TPR-related motif prevalent in plant organellar proteins.* Trends Biochem Sci, 2000. **25**(2): p. 46-7.

2. Williams, P.M. and A. Barkan, *A chloroplast-localized PPR protein required for plastid ribosome accumulation.* Plant J, 2003. **36**(5): p. 675-86.

3. Pepin, J., et al., *Trial of prednisolone for prevention of melarsoprol-induced encephalopathy in gambiense sleeping sickness.* Lancet, 1989. **1**(8649): p. 1246-50.

4. Vickerman, K. and A.G. Luckins, *Localization of variable antigens in the surface coat of Trypanosoma brucei using ferritin conjugated antibody.* Nature, 1969. **224**(224): p. 1125-6.

5. Le Ray, D., J.D. Barry, and K. Vickerman, *Antigenic heterogeneity of metacyclic forms of Trypanosoma brucei.* Nature, 1978. **273**(5660): p. 300-2.

6. Jenni, L., *Comparisons of antigenic types of Trypanosoma (T)brucei strains transmitted by Glossina m. morsitans.* Acta Trop, 1977. **34**(1): p. 35-41.

7. Ryley, J.F., *Studies on the metabolism of the protozoa. 9. Comparative metabolism of blood-stream and culture forms of Trypanosoma rhodesiense.* Biochem J, 1962. **85**: p. 211-23.

8. Flynn, I.W. and I.B. Bowman, *The metabolism of carbohydrate by pleomorphic African trypanosomes.* Comp Biochem Physiol B, 1973. **45**(1): p. 25-42.

9. Cazzulo, J.J., *Aerobic fermentation of glucose by trypanosomatids.* Faseb J, 1992. **6**(13): p. 3153-61.

10. Lamour, N., et al., *Proline metabolism in procyclic Trypanosoma brucei is down-regulated in the presence of glucose.* J Biol Chem, 2005. **280**(12): p. 11902-10.

11. Grant, P.T. and J.R. Sargent, *L-alpha-Glycerophosphate dehydrogenase, a component of an oxidase system in Trypanosoma rhodesiense.* Biochem J, 1961. **81**: p. 206-14.

12. Bienen, E.J., E. Hammadi, and G.C. Hill, *Trypanosoma brucei: biochemical and morphological changes during in vitro transformation of bloodstream- to procyclic-trypomastigotes.* Exp Parasitol, 1981. **51**(3): p. 408-17.

13. Beattie, D.S. and M.M. Howton, *The presence of rotenone-sensitive NADH dehydrogenase in the long slender bloodstream and the procyclic forms of Trypanosoma brucei brucei.* Eur J Biochem, 1996. **241**(3): p. 888-94.

14. Nolan, D.P. and H.P. Voorheis, *The mitochondrion in bloodstream forms of Trypanosoma brucei is energized by the electrogenic pumping of protons catalysed by the F1F0-ATPase.* Eur J Biochem, 1992. **209**(1): p. 207-16.

15. Bochud-Allemann, N. and A. Schneider, *Mitochondrial substrate level phosphorylation is essential for growth of procyclic Trypanosoma brucei.* J Biol Chem, 2002. **277**(36): p. 32849-54.

16. Coustou, V., et al., *ATP generation in the Trypanosoma brucei procyclic form: cytosolic substrate level is essential, but not oxidative phosphorylation.* J Biol Chem, 2003. **278**(49): p. 49625-35.

17. van Weelden, S.W., et al., *Procyclic Trypanosoma brucei do not use Krebs cycle activity for energy generation.* J Biol Chem, 2003. **278**(15): p. 12854-63.

18. Schultz, B.E. and S.I. Chan, *Structures and proton-pumping strategies of mitochondrial respiratory enzymes.* Annu Rev Biophys Biomol Struct, 2001. **30**: p. 23-65.

19. Walker, R., Jr., et al., *The effect of over-expression of the alternative oxidase in the procyclic forms of Trypanosoma brucei.* Mol Biochem Parasitol, 2005. **139**(2): p. 153-62.

20. Fang, J. and D.S. Beattie, *Alternative oxidase present in procyclic Trypanosoma brucei may act to lower the mitochondrial production of superoxide.* Arch Biochem Biophys, 2003. **414**(2): p. 294-302.

21. Michelotti, E.F. and S.L. Hajduk, *Developmental regulation of trypanosome mitochondrial gene expression.* J Biol Chem, 1987. **262**(2): p. 927-32.

22. Michelotti, E.F., et al., *Trypanosoma brucei mitochondrial ribosomal RNA synthesis, processing and developmentally regulated expression.* Mol Biochem Parasitol, 1992. **54**(1): p. 31-41.

23. Myler, P.J., et al., *Structural organization of the maxicircle variable region of Trypanosoma brucei: identification of potential replication origins and topoisomerase II binding sites.* Nucleic Acids Res, 1993. **21**(3): p. 687-94.

24. Clement, S.L., M.K. Mingler, and D.J. Koslowsky, *An Intragenic Guide RNA Location Suggests a Complex Mechanism for Mitochondrial Gene Expression in Trypanosoma brucei.* Eukaryot Cell, 2004. **3**(4): p. 862-9.

25. Jasmer, D.P. and K. Stuart, *Conservation of kinetoplastid minicircle characteristics without nucleotide sequence conservation.* Mol Biochem Parasitol, 1986. **18**(3): p. 257-69.

26. Pollard, V.W. and S.L. Hajduk, *Trypanosoma equiperdum Minicircles Encode Three Distinct Primary Transcripts Which Exhibit Guide RNA Characteristics.* molecular and Cellular Biology, 1991: p. 1668-1675.

27. Koslowsky, D.J., et al., *Guide RNAs for transcripts with developmentally regulated RNA editing are present in both life cycle stages of Trypanosoma brucei.* Mol Cell Biol, 1992. **12**(5): p. 2043-9.

28. Hong, M. and L. Simpson, *Genomic organization of Trypanosoma brucei kinetoplast DNA minicircles.* Protist, 2003. **154**(2): p. 265-79.

29. Riley, G.R., R.A. Corell, and K. Stuart, *Multiple guide RNAs for identical editing of Trypanosoma brucei apocytochrome b mRNA have an unusual minicircle location and are developmentally regulated.* J Biol Chem, 1994. **269**(8): p. 6101-8.

30. Grams, J., M.T. McManus, and S.L. Hajduk, *Processing of polycistronic guide RNAs is associated with RNA editing complexes in Trypanosoma brucei.* The EMBO Journal, 2000. **19**(20): p. 5525-5532.

31. Koslowsky, D.J. and G. Yahampath, *Mitochondrial mRNA 3' cleavage/polyadenylation and RNA editing in Trypanosoma brucei are independent events.* Mol Biochem Parasitol, 1997. **90**(1): p. 81-94.

32. Read, L.K., P.J. Myler, and K. Stuart, *Extensive editing of both processed and preprocessed maxicircle CR6 transcripts in Trypanosoma brucei.* J Biol Chem, 1992. **267**(2): p. 1123-8.

33. Feagin, J.E., D.P. Jasmer, and K. Stuart, *Apocytochrome b and other mitochondrial DNA sequences are differentially expressed during the life cycle of Trypanosoma brucei.* Nucleic Acids Res, 1985. **13**(12): p. 4577-96.

34. Jasmer, D.P., J.E. Feagin, and K. Stuart, *Diverse patterns of expression of the cytochrome c oxidase subunit I gene and unassigned reading frames 4 and 5 during the life cycle of Trypanosoma brucei.* Mol Cell Biol, 1985. **5**(11): p. 3041-7.

35. Feagin, J.E. and K. Stuart, *Differential expression of mitochondrial genes between life cycle stages of Trypanosoma brucei.* Proc Natl Acad Sci U S A, 1985. **82**(10): p. 3380-4.

36. Feagin, J.E., D.P. Jasmer, and K. Stuart, *Developmentally regulated addition of nucleotides within apocytochrome b transcripts in Trypanosoma brucei.* Cell, 1987. **49**(3): p. 337-45.

37. Souza, A.E., P.J. Myler, and K. Stuart, *Maxicircle CR1 transcripts of Trypanosoma brucei are edited and developmentally regulated and encode a putative iron-sulfur protein homologous to an NADH dehydrogenase subunit.* Mol Cell Biol, 1992. **12**(5): p. 2100-7.

38. Souza, A.E., et al., *Extensive editing of CR2 maxicircle transcripts of Trypanosoma brucei predicts a protein with homology to a subunit of NADH dehydrogenase.* Mol Cell Biol, 1993. **13**(11): p. 6832-40.

39. Koslowsky, D.J., et al., *The MURF3 gene of T. brucei contains multiple domains of extensive editing and is homologous to a subunit of NADH dehydrogenase.* Cell, 1990. **62**(5): p. 901-11.

40. Corell, R.A., P. Myler, and K. Stuart, *Trypanosoma brucei mitochondrial CR4 gene encodes an extensively edited mRNA with completely edited sequence only in bloodstream forms.* Mol Biochem Parasitol, 1994. **64**(1): p. 65-74.

41. Priest, J.W. and S.L. Hajduk, *Developmental regulation of Mitochondrial Biogenesis in Trypanosoma brucei.* Journal of Bioenergetics and Biomemebranes, 1994. **26**(2): p. 179-191.

42. Feagin, J.E., J.M. Abraham, and K. Stuart, *Extensive editing of the cytochrome c oxidase III transcript in Trypanosoma brucei.* Cell, 1988. **53**(3): p. 413-22.

43. Feagin, J.E. and K. Stuart, *Developmental aspects of uridine addition within mitochondrial transcripts of Trypanosoma brucei.* Mol Cell Biol, 1988. **8**(3): p. 1259-65.

44. Blum, B., N. Bakalara, and L. Simpson, *A model for RNA editing in kinetoplastid mitochondria: "guide" RNA molecules transcribed from maxicircle DNA provide the edited information.* Cell, 1990. **60**(2): p. 189-98.

45. Bhat, G.J., et al., *An extensively edited mitochondrial transcript in kinetoplastids encodes a protein homologous to ATPase subunit 6.* Cell, 1990. **61**(5): p. 885-94.

46. Bhat, G.J., et al., *Transcript-specific developmental regulation of polyadenylation in Trypanosoma brucei mitochondria.* Mol Biochem Parasitol, 1992. **52**(2): p. 231-40.

47. Militello, K.T. and L.K. Read, *Coordination of kRNA editing and polyadenylation in Trypanosoma brucei mitochondria: complete editing is not required for long poly(A) tract addition.* Nucleic Acids Res, 1999. **27**(5): p. 1377-85.

48. Read, L.K., et al., *Editing of Trypanosoma brucei maxicircle CR5 mRNA generates variable carboxy terminal predicted protein sequences.* Nucleic Acids Res, 1994. **22**(8): p. 1489-95.

49. Feagin, J.E., D.P. Jasmer, and K. Stuart, *Differential mitochondrial gene expression between slender and stumpy bloodforms of Trypanosoma brucei.* Mol Biochem Parasitol, 1986. **20**(3): p. 207-14.

50. Kao, C.Y. and L.K. Read, *Opposing effects of polyadenylation on the stability of edited and unedited mitochondrial RNAs in Trypanosoma brucei.* Mol Cell Biol, 2005. **25**(5): p. 1634-44.

51. Militello, K.T. and L.K. Read, *UTP-dependent and -independent pathways of mRNA turnover in Trypanosoma brucei mitochondria.* Mol Cell Biol, 2000. **20**(7): p. 2308-16.

52. Decker, C.J. and B. Sollner-Webb, *RNA editing involves indiscriminate U changes throughout precisely defined editing domains.* Cell, 1990. **61**(6): p. 1001-11.

53. Read, L.K., et al., *Developmental regulation of RNA editing and polyadenylation in four life cycle stages of Trypanosoma congolense.* Mol Biochem Parasitol, 1994. **68**(2): p. 297-306.

54. Gagliardi, D. and C.J. Leaver, *Polyadenylation accelerates the degradation of the mitochondrial mRNA associated with cytoplasmic male sterility in sunflower.* Embo J, 1999. **18**(13): p. 3757-66.

55. Gagliardi, D., et al., *Plant mitochondrial polyadenylated mRNAs are degraded by a 3'- to 5'-exoribonuclease activity, which proceeds unimpeded by stable secondary structures.* J Biol Chem, 2001. **276**(47): p. 43541-7.

56. Lupold, D.S., A.G. Caoile, and D.B. Stern, *Polyadenylation occurs at multiple sites in maize mitochondrial cox2 mRNA and is independent of editing status.* Plant Cell, 1999. **11**(8): p. 1565-78.

57. Kuhn, J., U. Tengler, and S. Binder, *Transcript lifetime is balanced between stabilizing stem-loop structures and degradation-promoting polyadenylation in plant mitochondria.* Mol Cell Biol, 2001. **21**(3): p. 731-42.

58. Ojala, D., J. Montoya, and G. Attardi, *tRNA punctuation model of RNA processing in human mitochondria.* Nature, 1981. **290**(5806): p. 470-4.

59. Temperley, R.J., et al., *Investigation of a pathogenic mtDNA microdeletion reveals a translation-dependent deadenylation decay pathway in human mitochondria.* Hum Mol Genet, 2003. **12**(18): p. 2341-8.

60. Gagliardi, D., et al., *Messenger RNA stability in mitochondria: different means to an end.* Trends Genet, 2004. **20**(6): p. 260-7.

61. Manthey, G.M., B.D. Przybyla-Zawislak, and J.E. McEwen, *The Saccharomyces cerevisiae Pet309 protein is embedded in the mitochondrial inner membrane.* Eur J Biochem, 1998. **255**(1): p. 156-61.

62. Krause, K., et al., *The mitochondrial message-specific mRNA protectors Cbp1 and Pet309 are associated in a high-molecular weight complex.* Mol Biol Cell, 2004. **15**(6): p. 2674-83.

63. Naithani, S., et al., *Interactions among COX1, COX2, and COX3 mRNA-specific translational activator proteins on the inner surface of the mitochondrial inner membrane of Saccharomyces cerevisiae.* Mol Biol Cell, 2003. **14**(1): p. 324-33.

64. Coffin, J.W., et al., *The Neurospora crassa cya-5 nuclear gene encodes a protein with a region of homology to the Saccharomyces cerevisiae PET309 protein and is required in a post-transcriptional step for the expression of the mitochondrially encoded COXI protein.* Curr Genet, 1997. **32**(4): p. 273-80.

65. Fisk, D.G., M.B. Walker, and A. Barkan, *Molecular cloning of the maize gene crp1 reveals similarity between regulators of mitochondrial and chloroplast gene expression.* Embo J, 1999. **18**(9): p. 2621-30.

66. Lahmy, S., et al., *A chloroplastic RNA-binding protein is a new member of the PPR family.* FEBS Lett, 2000. **480**(2-3): p. 255-60.

67. Bentolila, S., A.A. Alfonso, and M.R. Hanson, *A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants.* Proc Natl Acad Sci U S A, 2002. **99**(16): p. 10887-92.

68. Brown, G.G., et al., *The radish Rfo restorer gene of Ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats.* Plant J, 2003. **35**(2): p. 262-72.

69. Desloire, S., et al., *Identification of the fertility restoration locus, Rfo, in radish, as a member of the pentatricopeptide-repeat protein family.* EMBO Rep, 2003. **4**(6): p. 588-94.

70. Koizuka, N., et al., *Genetic characterization of a pentatricopeptide repeat protein gene, orf687, that restores fertility in the cytoplasmic male-sterile Kosena radish.* Plant J, 2003. **34**(4): p. 407-15.

71. Kazama, T. and K. Toriyama, *A pentatricopeptide repeat-containing gene that promotes the processing of aberrant atp6 RNA of cytoplasmic male-sterile rice.* FEBS Lett, 2003. **544**(1-3): p. 99-102.

72. Komori, T., et al., *Map-based cloning of a fertility restorer gene, Rf-1, in rice (Oryza sativa L.).* Plant J, 2004. **37**(3): p. 315-25.

73. Nakamura, T., et al., *Chloroplast RNA-binding and pentatricopeptide repeat proteins.* Biochem Soc Trans, 2004. **32**(Pt 4): p. 571-4.

74. Nakamura, T., et al., *RNA-binding properties of HCF152, an Arabidopsis PPR protein involved in the processing of chloroplast RNA.* Eur J Biochem, 2003. **270**(20): p. 4070-81.

75. Meierhoff, K., et al., *HCF152, an Arabidopsis RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast psbB-psbT-psbH-petB-petD RNAs.* Plant Cell, 2003. **15**(6): p. 1480-95.

76. Mancebo, R., et al., *BSF binds specifically to the bicoid mRNA 3' untranslated region and contributes to stabilization of bicoid mRNA.* Mol Cell Biol, 2001. **21**(10): p. 3462-71.

77. Mootha, V.K., et al., *Identification of a gene causing human cytochrome c oxidase deficiency by integrative genomics.* Proc Natl Acad Sci U S A, 2003. **100**(2): p. 605-10.

78. Xu, F., et al., *The role of the LRPPRC (leucine-rich pentatricopeptide repeat cassette) gene in cytochrome oxidase assembly: mutation causes lowered levels of COX (cytochrome c oxidase) I and COX III mRNA.* Biochem J, 2004. **382**(Pt 1): p. 331-6.

79. Ostrowski, J., et al., *Heterogeneous nuclear ribonucleoprotein K protein associates with multiple mitochondrial transcripts within the organelle.* J Biol Chem, 2002. **277**(8): p. 6303-10.

80. Liu, L. and W.L. McKeehan, *Sequence analysis of LRPPRC and its SEC1 domain interaction partners suggests roles in cytoskeletal organization, vesicular trafficking, nucleocytosolic shuttling, and chromosome activity.* Genomics, 2002. **79**(1): p. 124-36.

81. Mili, S., et al., *Distinct RNP complexes of shuttling hnRNP proteins with pre-mRNA and mRNA: candidate intermediates in formation and export of mRNA.* Mol Cell Biol, 2001. **21**(21): p. 7307-19.

82. Kotera, E., M. Tasaka, and T. Shikanai, *A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts.* Nature, 2005. **433**(7023): p. 326-30.

83. Mili, S. and S. Pinol-Roma, *LRP130, a pentatricopeptide motif protein with a noncanonical RNA-binding domain, is bound in vivo to mitochondrial and nuclear RNAs.* Mol Cell Biol, 2003. **23**(14): p. 4972-82.

84. Tsuchiya, N., et al., *LRP130, a single-stranded DNA/RNA-binding protein, localizes at the outer nuclear and endoplasmic reticulum membrane, and interacts with mRNA in vivo.* Biochem Biophys Res Commun, 2004. **317**(3): p. 736-43.

85. Schmitz-Linneweber, C., R. Williams-Carrier, and A. Barkan, *RNA Immunoprecipitation and Microarray Analysis Show a Chloroplast*

Pentatricopeptide Repeat Protein to Be Associated with the 5' Region of mRNAs Whose Translation It Activates. Plant Cell, 2005.

86. Lurin, C., et al., *Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis.* Plant Cell, 2004. **16**(8): p. 2089-103.

87. Martin, W. and P. Borst, *Secondary loss of chloroplasts in trypanosomes.* Proc Natl Acad Sci U S A, 2003. **100**(3): p. 765-7.

88. Hannaert, V., et al., *Plant-like traits associated with metabolism of Trypanosoma parasites.* Proc Natl Acad Sci U S A, 2003. **100**(3): p. 1067-71.

89. Ikeda, T.M. and M.W. Gray, *Characterization of a DNA-binding protein implicated in transcription in wheat mitochondria.* Mol Cell Biol, 1999. **19**(12): p. 8113-22.

90. Oguchi, T., et al., *Genomic structure of a novel Arabidopsis clock-controlled gene, AtC401, which encodes a pentatricopeptide repeat protein.* Gene, 2004. **330**: p. 29-37.

91. Hashimoto, M., et al., *A nucleus-encoded factor, CRR2, is essential for the expression of chloroplast ndhB in Arabidopsis.* Plant J, 2003. **36**(4): p. 541-9.

92. Yamazaki, H., M. Tasaka, and T. Shikanai, *PPR motifs of the nucleus-encoded factor, PGR3, function in the selective and distinct steps of chloroplast gene expression in Arabidopsis.* Plant J, 2004. **38**(1): p. 152-63.

93. Cushing, D.A., et al., *Arabidopsis emb175 and other ppr knockout mutants reveal essential roles for pentatricopeptide repeat (PPR) proteins in plant embryogenesis.* Planta, 2005.

94. Barkan, A., et al., *A nuclear mutation in maize blocks the processing and translation of several chloroplast mRNAs and provides evidence for the differential translation of alternative mRNA forms.* Embo J, 1994. **13**(13): p. 3170-81.

95. Gothandam, K.M., et al., *OsPPR1, a pentatricopeptide repeat protein of rice is essential for the chloroplast biogenesis.* Plant Mol Biol, 2005. **58**(3): p. 421-33.

96. Akagi, H., et al., *Positional cloning of the rice Rf-1 gene, a restorer of BT-type cytoplasmic male sterility that encodes a mitochondria-targeting PPR protein.* Theor Appl Genet, 2004. **108**(8): p. 1449-57.

97. Klein, R.R., et al., *Fertility restorer locus Rf1 of sorghum (Sorghum bicolor L.) encodes a pentatricopeptide repeat protein not present in the colinear region of rice chromosome 12.* Theor Appl Genet, 2005.

98. Hattori, M., M. Hasebe, and M. Sugita, *Identification and characterization of cDNAs encoding pentatricopeptide repeat proteins in the basal land plant, the moss Physcomitrella patens.* Gene, 2004. **343**(2): p. 305-11.

99. Tsuchiya, N., et al., *LRP130, a protein containing nine pentatricopeptide repeat motifs, interacts with a single-stranded cytosine-rich sequence of mouse hypervariable minisatellite Pc-1.* Eur J Biochem, 2002. **269**(12): p. 2927-33.

100. Auchincloss, A.H., et al., *Characterization of Tbc2, a nucleus-encoded factor specifically required for translation of the chloroplast psbC mRNA in Chlamydomonas reinhardtii.* J Cell Biol, 2002. **157**(6): p. 953-62.

101. Lown, F.J., A.T. Watson, and S. Purton, *Chlamydomonas nuclear mutants that fail to assemble respiratory or photosynthetic electron transfer complexes.* Biochem Soc Trans, 2001. **29**(Pt 4): p. 452-5.

102. Claros, M.G. and P. Vincens, *Computational method to predict mitochondrially imported proteins and their targeting sequences.* Eur J Biochem, 1996. **241**(3): p. 779-86.

103. Bateman, A., et al., *The Pfam protein families database.* Nucleic Acids Res, 2004. **32**(Database issue): p. D138-41.

104. Thompson, J.D., et al., *The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.* Nucleic Acids Res, 1997. **25**(24): p. 4876-82.

105. Eddy, S.R., *Profile hidden Markov models.* Bioinformatics, 1998. **14**(9): p. 755-63.

106. Eddy, S.R., *Hidden Markov models.* Curr Opin Struct Biol, 1996. **6**(3): p. 361-5.

107. Bailey, T.L. and C. Elkan, *Fitting a mixture model by expectation maximization to discover motifs in biopolymers.* Proc Int Conf Intell Syst Mol Biol, 1994. **2**: p. 28-36.

108. Emanuelsson, O., et al., *Predicting subcellular localization of proteins based on their N-terminal amino acid sequence.* J Mol Biol, 2000. **300**(4): p. 1005-16.

109. Treger, M. and E. Westhof, *Statistical analysis of atomic contacts at RNA-protein interfaces.* J Mol Recognit, 2001. **14**(4): p. 199-214.

110. Edwards, T.A., et al., *Structure of Pumilio reveals similarity between RNA and peptide binding motifs.* Cell, 2001. **105**(2): p. 281-9.

111. Rozen, S. and H. Skaletsky, *Primer3 on the WWW for general users and for biologist programmers.* Methods Mol Biol, 2000. **132**: p. 365-86.

112. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* Nucleic Acids Res, 1997. **25**(17): p. 3389-402.

113. Chomczynski, P. and N. Sacchi, *Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction.* Anal Biochem, 1987. **162**(1): p. 156-9.

114. Motyka, S.A. and P.T. Englund, *RNA interference for analysis of gene function in trypanosomatids.* Curr Opin Microbiol, 2004. **7**(4): p. 362-8.

115. Shi, H., et al., *Genetic interference in Trypanosoma brucei by heritable and inducible double-stranded RNA.* Rna, 2000. **6**(7): p. 1069-76.

116. Djikeng, A., et al., *RNA interference in Trypanosoma brucei: cloning of small interfering RNAs provides evidence for retroposon-derived 24-26-nucleotide RNAs.* Rna, 2001. **7**(11): p. 1522-30.

117. Wang, Z., et al., *Inhibition of Trypanosoma brucei gene expression by RNA interference using an integratable vector with opposing T7 promoters.* J Biol Chem, 2000. **275**(51): p. 40174-9.

118. Liu, Q., et al., *Identification and functional characterization of lsm proteins in Trypanosoma brucei.* J Biol Chem, 2004. **279**(18): p. 18210-9.

119. Tu, X. and C.C. Wang, *Pairwise knockdowns of cdc2-related kinases (CRKs) in Trypanosoma brucei identified the CRKs for G1/S and G2/M transitions and demonstrated distinctive cytokinetic regulations between two developmental stages of the organism.* Eukaryot Cell, 2005. **4**(4): p. 755-64.

120. Vondruskova, E., et al., *RNA interference analyses suggest a transcript-specific regulatory role for MRP1 and MRP2 in RNA editing and other RNA processing in trypanosoma brucei.* J Biol Chem, 2004.

121. Downey, N., et al., *Mitochondrial DNA Ligases of Trypanosoma brucei.* Eukaryot Cell, 2005. **4**(4): p. 765-74.

122. El-Sharoud, W.M. and G.W. Niven, *The activity of ribosome modulation factor during growth of Escherichia coli under acidic conditions.* Arch Microbiol, 2005. **184**(1): p. 18-24.

123. Ryan, C.M., K.T. Militello, and L.K. Read, *Polyadenylation Regulates the Stability of Trypanosoma brucei Mitochondrial RNAs.* The Journal of Biological Chemistry, 2003. **278**(35): p. 32753-32762.