This is to certify that the
dissertation entitled

MITOCHONDRIAL TRANSCRIPTION IN *TRYPANOSOMA BRUCEI*: CONSERVED PROTEINS AND UNUSUAL INITIATIONS

presented by

SANDRA L. CLEMENT

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Microbiology and Molecular Genetics

Major Professor's Signature

12/16/04

Date

MITOCHONDRIAL TRANSCRIPTION IN *TRYPANOSOMA BRUCEI*: CONSERVED PROTEINS AND UNUSUAL INITIATIONS

by

Sandra L. Clement

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Microbiology and Molecular Genetics

2004

ABSTRACT

MITOCHONDRIAL TRANSCRIPTION IN TRYPANOSOMA BRUCEI: CONSERVED PROTEINS AND UNUSUAL INITIATIONS

By

Sandra L. Clement

Trypanosoma brucei has an extraordinary mitochondrial genome and is one of the earliest organisms to possess mitochondria. Mitochondrial gene expression in this parasitic protist is complex and developmentally regulated, but little is known about the transcription of this unusual kinetoplast DNA. In order to pinpoint promoter regions within the *T. brucei* mitochondrial genome, the 5' ends of the guide RNAs gMURF2-II and gCYb(560) were characterized. Each of these genes have uncommon kinetoplast DNA locations not typically associated with transcription initiation events. The gMURF2-II gene is located entirely within the 5' end of the ND4 gene. Interestingly, the gMURF2-II and ND4 transcripts are generated by distinctly different events; the ND4 mRNA is processed from a polycistronic precursor, while transcription of the gRNA initiates downstream of the 5' end of the ND4 gene. The data presented in this dissertation likewise show that the mature gCYb(560) gRNA is also a primary transcript and that the 5' end heterogeneity previously observed for this gRNA is a result of multiple transcription initiation sites and not imprecise 5' end processing. Together, these data indicate that gRNA genes represent individual transcription units regardless of

genomic context and suggest a complex mechanism for mitochondrial gene expression in *T. brucei.*

In spite of its phylogenetic position near the acquisition of mitochondria by eukaryotes and its bizarre mitochondrial genome structure, the core components of mitochondrial transcription found in other eukaryotes are conserved in trypanosomes and their kin. The gene encoding the mitochondrial RNA polymerase, *TbmtRNAP* was cloned and shown to be developmentally regulated via differential RNA processing and stability. In addition, *TbmtTFB*, a homologue to the universally conserved mitochondrial transcription factor, mtTFB, was identified in trypanosomes although no homologue to the vertebrate factor TFAM was found. Over-expression of *TbmtTFB* caused a slight increase in growth rate, while RNAi against the gene resulted in slower growth. This together with the inability to generate double knockouts of this gene suggests an important role for this TbmtTFB in *T. brucei*, although the precise function of the protein is still unresolved.

ACKNOWLEDGEMENTS

Many people deserve credit for helping me to complete this long journey, and I would like to thank a few of them specifically. My dissertation advisor, Dr. Donna J. Koslowsky, gave me the opportunity to develop a sense of independence and confidence by allowing me to work on my own project. Yet, she also gave me the tools to get my project off the ground, without which I never would have succeeded. In addition, she has always been respectful of my career goals and has allowed me to pursue opportunities to obtain additional teaching experience. I would also like to thank some other very important faculty, specifically Dr. Ron Patterson, for frequent advice about graduate school and beyond and for pointing me in the right direction (towards RNA). Both Dr. John Merrill and Dr. Martha Mulks have been excellent mentors in teaching college science to a wide variety of students.

I am also very lucky to have been surrounded by extremely intelligent and wise peers. In particular, I would like to thank Dr. Sheldon Leung, who took me under his wing and gave me constant advice and support during the initial and toughest years in the lab. He taught me the most valuable lesson I have learned in graduate school, and perhaps in life, that of perpetual optimism. The current members of the Koslowsky lab, Laura Yu, Larissa Reifur, and Mel Mingler, are each among my best friends. I hope they have learned at least half as much from me as I have learned from them.

Of course, I would have given up on everything long ago if not for the support of my Matthew, whose importance to me would take more than all the pages of all the dissertations in the world to describe.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

CHAPTER II

CHAPTER III

CHAPTER IV

LIST OF ABBREVIATIONS

T. brucei	*Trypanosoma brucei*
mtDNA	mitochondrial DNA
kDNA	kinetoplast DNA
mtTFA, TFAM	mitochondrial transcription factor A
mtTFB	mitochondrial transcription factor B
mtRNAP	mitochondrial RNA polymerase
gRNA	guide RNA
mRNA	messenger RNA
tRNA	transfer RNA
RNAi	RNA interference
dsRNA	double stranded RNA
PCR	polymerase chain reaction
DDRLACE	differential display of RNA ligase-mediated cDNA amplification
RACE	rapid amplification of cDNA ends
cDNA	copy DNA
ND4	NADH dehydrogenase subunit 4
A6	ATPase subunit 6
MURF	maxicircle unidentified reading frame

CHAPTER I

INTRODUCTION

Mitochondria

Origin and Evolution

The acquisition of mitochondria was a defining moment in the early evolution (and perhaps even the origin) of eukaryotic organisms. Approximately 1.5 to 2 billion years ago, an α-proteobacterium took up residence in a proto-eukaryotic cell or possibly an Archea (3, 50). There are several theories as to the nature of this initial symbiotic relationship. Because the acquisition of mitochondria occurs around the time of increasing oxygen levels in the atmosphere, the "Ox-Tox" model suggests that the role of the symbiont was to protect the host from rising oxygen levels by using O_2 as an electron acceptor (65). In two other common theories, the ability to utilize molecular oxygen is secondary. The metabolic syntrophy theory suggests that the symbiont fed on methane while the hydrogen hypothesis proposes the role was to produce hydrogen (4, 65). Regardless of the impetus of this initial relationship, since bacteria do not have a mechanism for exporting ATP, it is unlikely that the common explanation that the relationship involved an exchange of metabolites for ATP is accurate. However, the host cell rapidly invented many proteins to take advantage of this opportunity early in the association. As a result, modern mitochondria have evolved to play important roles in energy production, fatty acid metabolism, ion homeostasis, and apoptosis (52). In addition, mitochondrial dysfunction is associated with aging and disease (66, 103, 113). Consequently, learning more about mitochondrial function will improve our knowledge of how this distinct and often enigmatic organelle contributes to the overall function and health of the cell.

Aerobic respiration is more efficient than anaerobic respiration, and nearly all eukaryotes have mitochondria. The only eukaryotes to lack them are unicellular anaerobic organisms containing specialized organelles called hydrogenosomes which are thought to be related to mitochondria (4). There are conflicting data and opinions as to whether the latter group of organisms are primitively or secondarily amitochondriate (3). In any case, it is clear that mitochondrial function has played a pivotal role in eukaryotic biology and evolution, and likely fomented the remarkable diversification and boom in complexity (40). One can imagine that this increase in energy was at least part of what promoted the development of the much larger and more complex eukaryotic cell.

Mitochondrial DNA structure and gene content

All mitochondria retain small genomes that are relics of their eubacterial ancestors. These genomes vary in size, gene content, and structure, but have in common a dramatic reduction of genes relative to the original endosymbiont. During the course of evolution, many of the genes originally encoded within the ancestor were lost; a phenomenon also observed for many other endosymbiotic and intracellular pathogens. Like bacterial genomes, mitochondrial DNA (mtDNA) maps as a circular molecule in most organisms, but recent evidence suggests that many mtDNAs are in fact linear (9, 61). While mtDNA varies dramatically in size from 6 kb in the malaria parasite, *Plasmodium falciparum* (43), to 2000 kb in cucumbers and melons (115), there is no correlation between genome size and coding content. In general, the larger mitochondrial genomes are often the result of the presence of group I and II introns and associated proteins as well as a larger proportion of intergenic space.

Although there is some variation in the precise gene content among mitochondria, a core set of proteins remain encoded within the mitochondrial genome. These genes include rRNAs, usually tRNAs, and mRNAs for proteins involved in oxidative phosphorylation and electron transport (22). This common set of retained genes indicates that the modern mtDNA gene content was established early in evolution, prior to the divergence of the major groups of eukaryotes. However, the specific genes retained can vary from group to group by as much as 20-fold (1) indicating that the frequent and ongoing process of differential loss and migration have also helped to shape mitochondrial gene content.

Mitochondrial Genomes and Promoters

Animals

Animal mitochondrial genomes are typically small, compact molecules. The structure and content is highly conserved among animals, especially in mammals, but some variation does exist. Nearly all multi-cellular animals have a closed circular genome while phylogenetically early animals such as cnidarians have linear genomes (17). In general, animal mtDNA encodes 37 or fewer genes including 22 tRNAs, large and small subunit rRNA, and 13 genes encoding subunits for proteins involved in oxidative phosphorylation. (For reviews see (30, 44, 106). Coding content, as well as G+T content, is asymmetric; 14 tRNAs, 2 rRNAs, and 12 protein genes are encoded on the heavy strand and 8 tRNAs and one protein gene are found on the light strand.

These genomes are compact and in some cases adjacent genes overlap. The only region of intergenic space is known as the regulatory region or D loop, which contains

promoter elements for each strand. These two transcription units are each transcribed a polycistronic unit from which individual transcripts are processed. Mitochondrially encoded genes in animals lack introns and often have no or small (3nt) 5' untranslated regions and stop codons are often generated by polyadenylation.

The two promoters within the vertebrate D-loop region have been defined using *in vitro* assays with partially purified mitochondrial proteins (15, 16, 99). These promoters are located about 150 bp from each other and are known as the heavy strand promoter (HSP) and light strand promoter (LSP). The promoter elements consist of a 50 bp bipartite sequence with distinct protein binding sites for factors involved in transcription initiation. Both the HSP and LSP consist of a 15 bp consensus motif that surrounds the initiation site and another element located upstream (-12 to -39) which includes the binding sites for the transcription factor, TFAM (discussed in more detail below). The LSP also plays a role in the replication of the mitochondrial genome. Processing of transcripts from this promoter by RNase MRP generates primers for replication of the heavy strand.

Fungi

Fungal mitochondrial genomes range in size from 19 kb in *Schizosaccharomyces pombe* to ~100 kb in *Podosopora anserina.* They contain several regions of intergenic space, but gene content is fairly consistent (21). The most well characterized mitochondrial genome is that of the baker's yeast, *Saccharomyces cerevisiae.* The yeast mtDNA is ~86 kb and encodes both large and small subunit ribosomal RNA, 24 tRNAs and 30 proteins (36). An unusual feature relative to animal mitochondrial DNA is that

only 1/3 of these protein genes encode vital mitochondrial OXPHOS genes. As many as 10 of the protein-coding genes in the yeast mitochondria code for endonucleases, reverse transcriptases and mRNA maturases, and yet another 10 encode genes of unknown function (3).

In contrast to the mammalian mtDNA, the larger yeast mitochondrial genome contains ~20 promoters and there have been at least 12 identified polycistronic transcription units (36). All but one gene, $tRNA_2^{thr}$, are transcribed from the same strand of DNA. In contrast to the bipartite promoters found in animal mtDNA, yeast mitochondrial promoters share a nonanucleotide sequence near the start site (83). The majority of these promoters share the consensus sequence 5' ATATAAGTA 3' and transcription initiates from the 3' terminal A residue. Interestingly, yeast mitochondrial promoters are classified as weak or strong depending on the type of nucleotide at position +2. Weak promoters have a pyrimidine, while strong promoters have a purine (14).

Plants

Mitochondrial genomes of higher plants display a number of unique features relative to their animal and fungal counterparts. They are the largest of the mitochondrial genomes. However, with the exception of some ribosomal proteins not found in other mitochondrial genomes, they encode basically the same core set of proteins found in mtDNA of other organisms. There is some variability in gene content among the flowering plants, including several independent gene losses, particularly of ribosomal proteins (1). The overall structure of plant mtDNA is peculiar, consisting of a large master circle encoding a full complement of genes, as well as several sub-genomic

minicircles that are generated by recombination. Several plant mitochondrial genomes have been sequenced and range in size from ~184 kb in the non-vascular plant, *Marchantia*(82) to > 2000 kb in curcurbits (115). The mitochondrial genome of the model organism *Arabidopsis thaliana* is >360 kb and encodes 57 genes including genes for rRNA, tRNA and OXPHOS components. These genes make up less than 10% of the genome (111), as plant mtDNA is rich in intergenic spaces (2-60 kb) and pseudogenes. Gene order is very dynamic due to intra- and inter- chromosomal rearrangements. As in fungi, plant mitochondrial genomes also encode maturases and introns. In addition, plant mtDNA appears to be rather promiscuous, with traces of nuclear and chloroplast DNA found in some plant mitochondrial genomes.

There are multiple transcription initiation sites in plant mitochondria (12, 19). Mapping of these transcription start sites and comparison of the surrounding sequences revealed conserved sequence motifs that are distinct between mono- and dicotyledoneous plants. (For reviews see (11, 18). Further *in vitro* characterization with partially purified mitochondrial extracts has determined in greater detail the nucleotide elements required for transcription initiation. Both classes share a 5′-CRTA-3′ tetranucleotide motif (45). In dicots, this sequence is extended to a conserved nonanucleotide motif (58).

Protists

Although protists make up the majority of eukaryotic evolutionary diversity, they are poorly sampled with respect to their mitochondrial genomes. Current genome sequencing projects such as the Organelle Genome Megasequencing Program (OGMP; http://megasun.bch.umontreal.ca/ogmp) have revealed that protist mtDNA varies

dramatically in structure, gene organization and content (21, 22, 51). From an evolutionary perspective, one of the most exciting mitochondrial genomes belongs to the early jakobid protist, *Reclinomonas americana*. This genome not only contains the largest gene complement of any sequenced mitochondrial genome, it is also thought to be the most ancestral . At 69 kbp, it is dwarfed in size by many plant mitochondrial genomes, but contains close to 100 genes, and retains similar gene order and even operon organization with that of eubacteria (5). Unlike other mitochondrial genomes, which lack genes required for their own expression, *Reclinomonas* mtDNA encodes genes for the eubacterial RNA polymerase subunits α, β, β' and σ and contains identifiable eubacterial promoter regions. There have been very few studies of mitochondrial transcription in protists, although some transcript mapping in *Plasmodium* (89)and *Dictyostelium* has been described (6, 7).

Mitochondrial Transcription Machinery

RNA polymerase (mtRNAP)

Mitochondrial transcription is the initial step in a cascade of regulatory processes responsible for the expression of the mitochondrial genome. The origin and evolution of the components involved in this step is still a bit of a mystery. Interestingly, with the exception of *Reclinomonas americana*, mitochondria are not known to encode genes for their own expression. Instead, most organisms appear to use a highly conserved, nuclear encoded RNA polymerase homologous to that of the single subunit bacteriophage T7 RNAP. The first such protein to be identified is the product of the *RPO41* gene in *S. cerevisiae* (73). Homologues have since been described in vetebrates as well. (93, 109).

A PCR survey performed in 1996 identified homologues in a phylogenetically diverse variety of organisms, but did not find homologues in the kinetoplastid, *Crithidia*, or in the jakobid *Reclinomonas*. (26) Further phylogenetic studies on the evolution of single subunit RNAPs did little to shed light on the origin of the mitochondrial RNA polymerase, other than to indicate that it was acquired shortly after the initial endosymbiotic event (27). Single subunit RNAPs have been identified in plants as well (56, 60, 118), although no functional studies have been performed to demonstrate their role in mitochondrial transcription. Interestingly, *Arabidiopsis thaliana* has three ssRNAPs; one targeted to the chloroplast where it transcribes a subset of genes, one targeted to the mitochondria, and one targeted to both organelles (57).

Multiple sequence alignments of mitochondrial RNAPs with the phage RNAPs show that nearly all the active domains are highly conserved, but that the mtRNAPs possess an amino terminal domain (ATD) extension relative to the bacteriophage proteins . Interestingly, there is little conservation among various species in the ATD sequence, although closely related groups do share some homology (114). These ATD sequences are thought to play a role in species specific interactions with other proteins. For example, fungal mtRNAPs interact with proteins that are important for RNA processing and translation (20, 91, 114), while vertebrate mtRNAPs have a weakly conserved PPR motif that is speculated to play a role in RNA processing (91).

Accessory factors

The major feature that distinguishes the mitochondrial RNA polymerases from their bacteriophage cousins is the requirement for one or more accessory factors. The

bacteriophage enzymes can recognize promoter regions, initiate, elongate and terminate transcription alone. In contrast, although recent data do suggest that the promoter recognition domain resides within the mtRNAP subunit (48, 74), all known mtRNAPs require one or more nuclear encoded accessory factors for accurate transcription initiation. As is the case for many other host cell inventions, these differences appear to reflect the different evolutionary pressures experienced by the different groups.

mtTFB

Although the mitochondrial RNA polymerases of various organisms are highly conserved, there is less conservation among mitochondrial transcription factors (24). The first transcription factor to be characterized was the product of the *MTF1* gene in *Saccharomyces cerevisiae*. Together with mtRNAP catalytic subunit, Rpo41, Mtf1 was shown to be necessary and sufficient for transcription initiation using in vitro studies (62). Early sequence analysis revealed limited homology to eubacterial sigma factors, particularly in domains 2 and 3 (31, 32, 62), though mutagenesis studies did not show strong support for a function of these putative conserved domains in promoter recognition or interaction with Rpo41 (100). However, Mtf1 was shown to behave in a manner similar to that of sigma in that it associates with Rpo41 at the promoter during transcription initiation and during the early abortive transcription stages but is released upon conversion of the polymerase into an elongation competent form (71). Interestingly, the crystal structure of Mtf1 demonstrated that it belongs to a large and highly conserved family of rRNA methyltransferases (96). This information helped two groups working independently to identify the vertebrate homologue, TFBM1, which

binds SAM, and is able to methylate *E. coli* small subunit (SSU) rRNA, but does not require this function for transcription initiation (42, 76, 98). Intriguingly, a paralogue of TFB1M, TFB2M is 1-2 orders of magnitude more effective at activating transcription initiation from promoters *in vitro* and is found only in higher animals (42, 88) including *Drosophila* (75).

mtTFA (aka TFAM)

The first mitochondrial transcription factor identified in animals was not homologous to the yeast Mtf1 protein. This protein, mtTFA or Tfam, belongs to a family of proteins bearing two high mobility group or HMG domains (46). HMG proteins are eukaryotic inventions not found in eubacteria or archea and consequently were recruited to play a role in mitochondrial gene expression by the host. In general, HMG proteins bind the minor groove of DNA in a non-sequence specific fashion and include non-histone chromosomal proteins and nucleolar transcription factors (108). These proteins bind and distort or bend DNA, thereby facilitating access to template DNA by the RNA polymerase. Mitochondrially targeted proteins containing HMG box domains are found in a variety of organisms including the Glom protein in the slime mold *Physarum* (94), mthmg1 in the filamentous fungus *Podospora anserina* (35)and Abf2 in yeast (92, 117). These proteins all play important roles in mtDNA maintenance, but do not appear to be transcription factors. TFAM likewise plays a role in mtDNA maintenance, apparently through two distinct mechanisms. One way in which TFAM helps to maintain mtDNA is by coating it and condensing it, preventing nucleoid loss (2, 39). Another way is by playing a role in transcription of the primer at the LSP for heavy strand replication. The

other DNA binding proteins are not required for transcription and lack both the hinge region found in the metazoan TFAMS as well as 29 amino acid basic CTD also unique to the metazoan TFAMs. This CTD is required for transcription activation, and can convert the Abf2 protein into a transcriptional activator on human LSP promoters in vitro (33). This C-tail of TFAM has also been shown recently to interact with mtTFB in humans (77). Like other HMG proteins, TFAM binds mtDNA non-specifically, but does show a higher affinity for promoter regions.

Putative mitochondrial transcription factors

Although *in vitro* transcription systems exist for plants, *in vitro* promoter recognition is very fragile and rapidly loses activity in most manipulations (11). Consequently, attempts at fractionation abolish all activity. This suggests that several loosely connected proteins may be involved in promoter recognition. One such protein, p63, has been characterized for wheat mitochondria (59). This protein contains a DNA binding domain and stimulates transcription from the cox2 promoter *in vitro*. Interestingly, this protein also belongs to a large family of RNA processing proteins in plants. This family of PPR or pentatricopeptide repeat containing proteins contain putative RNA binding domains and are thought to play a role in RNA processing (69). Another protein with putative promoter-binding activity was isolated from pea mitochondria and is homologous to an isovaleryl CoA dehydrogenase enzyme (13, 55). The authors speculate that it could be dual function protein, as is the style in mitochondria. However, further studies demonstrating a role in transcription have not been described.

Mechanism of promoter recognition and likelihood for diversity in accessory factors

Exciting new data have become available recently about the mechanism of promoter recognition by the mitochondrial RNA polymerase. The Jaehning lab has shown that the mtRNAP itself is capable of promoter recognition *in vitro* and can even initiate transcription on its own from an open "bubble" template (74). This result indicates that the mtTFB does not confer promoter specificity, but rather contributes to promoter opening. This innate ability to recognize promoter sequences is not unique to the yeast mitochondrial RNAP. Studies of the human mtRNAP indicate that it too has promoter recognition ability (48), but that this capacity is absolutely dependent on both TFAM and the human mtTFB2. Consequently, although yeast and mammals use similar machinery, the precise mechanism of mitochondrial transcription initiation must be substantially different in these two organisms. Fungi and animals are more closely related to each other than either of them are to other groups of organisms. However, they show distinct differences in the composition of the mtRNAP holoenzyme, making it likely that other variations will become evident as more systems are studied. In general, we know very little about the scope of mitochondrial transcription in eukaryotes because this has only been elaborated for two relatively close clades of crown eukaryotes. Studying protists, especially early branching organisms such as the parasitic protist *Trypanosoma brucei,* can reveal valuable information about the evolution of the mitochondrial transcription machinery. *T. brucei* is not only an ancient eukaryote but also has a remarkable mitochondrial genome and a bizarre mode of mitochondrial gene expression.

Trypanosoma brucei

Evolutionary, Medical and Economic Significance

Trypanosoma brucei is a unicellular flagellated protist that belongs to the order Kinetoplastida. This evolutionarily ancient group is so named for their unusual mitochondrial genome, the kinetoplast. Together, *T. brucei*, and other kinetoplastid species such as *T. cruzi* and *Leishmania major*, cause a wide variety of diseases affecting hundreds of millions of people in the developing countries of Asia, Africa and South America (http://www.who.int). Other members parasitize a phylogenetically diverse array of organisms including reptiles, fish, amphibians and even plants; and free living forms exist as well (23, 53, 54). *T. brucei* is perhaps one of the best studied organisms of the group. This parasite has had a devastating impact in sub-Saharan Africa by causing African Sleeping Sickness in humans and nagana (a wasting disease) in African livestock. Dramatic inroads had been made in reducing the spread of the disease, mainly by limiting the range of its insect vector, the tsetse fly. However, recent political and socioeconomic instability in the region has led to an increase in the incidence of Sleeping Sickness, and has placed millions more people at risk. There is no vaccine against the disease, the treatments are harsh, and left untreated, the disease is inevitably fatal (37, 41, 81). Fortunately, members of the Kinetoplastida share many unusual biochemical features relative to their mammalian hosts. Learning more about these will hopefully provide more drug targets to combat the parasite.

Life cycle

T. brucei is transmitted from the bloodstream of one host to that of another by the bite of the tsetse fly. This rapid environmental fluctuation is accompanied by dramatic morphological and biochemical changes in the parasite (87, 107, 112). Each trypanosome possesses a single mitochondrion that is fully functional during the insect or procyclic phase of the life cycle but greatly reduced in size and function during the bloodstream phase. While living in the midgut of the tsetse fly, procyclic form parasites generate ATP primarily via oxidative phosphorylation. Unlike other organisms which predominantly use carbohydrates as an energy source, procyclic trypanosomes use amino acids such as proline which is fed directly into the Krebs' cycle as a source of energy. Glycolysis in *T. brucei* is also unique in that the first seven steps are contained within a specialized organelle called the glycosome (84). In the mammalian bloodstream, trypanosomes rely entirely on glycolysis and substrate level phosphorylation for ATP production, and the mitochondrion lacks complexes III and IV (95). However, the mitochondrion does retain important functions in redox balance for the glycosome during the bloodstream infection (54). A plant-like alternative oxidase exists within the bloodstream form mitochondrion to oxidize the excess reducing power generated by glycolysis (28, 54). As with other eukaryotes, the mitochondrion of *T. brucei* may also play a role in Ca^{2+} homeostasis , fatty acid metabolism, and strangely enough for a unicellular organism, apoptosis (25, 34, 90).

The Kinetoplast

kDNA structure and gene content

Aside from its significance as a disease-causing organism, *T. brucei* is perhaps best known for its unusual mitochondrial genome and its bizarre mode of expression. The mitochondrial genome, or kinetoplast, is a disk of highly condensed DNA that is physically associated with the basal body of the flagellum (102, 104). Consequently, early observations led people to believe this specialized organelle played a direct role in flagellar movement. This unusual structure is historically significant as it was one of the first demonstrations that mitochondria contained DNA (95).

Kinetoplast DNA (kDNA) in *T. brucei* consists of a network of thousands of interlocked molecules (68). Although kDNA is one of the most remarkable mitochondrial genomes studied, the maxicircle encodes similar genes to those found in other mitochondria including: apocytochrome B (Cyb), cytochrome oxidase (CO) subunits I, II, and III; ATPase subunit 6 (A6), one ribosomal protein (RPS12), both large and small rRNA genes, and five other open reading frames of unknown homology. Many of these genes are in fact "cryptogenes" whose mRNAs must be edited by the post-transcriptional insertion and/or deletion of U residues to create mature transcripts (105). The maxicircle genes are tightly packed within ~15 kb of the 22 kb maxicircle sequence, and there is a notable lack of tRNA genes. As with animal mtDNA, these protein-coding genes lack introns and in many cases the 5' and 3' ends of adjacent genes overlap. However, unlike animal mitochondrial mRNAs, maxicircle encoded mRNAs have both 5' and 3' UTRs. The remaining 7 kb of the maxicircle is known as the variable region due to species- and strain- specific variability (80). This region lacks open reading

frames and is thought to contain regulatory elements for transcription and replication, although none have been precisely mapped.

The majority of the kinetoplast genome is composed of a distinct minicircle component consisting of thousands of 1 kb interlocked circular molecules. Minicircles encode the small guide RNA (gRNA) transcripts that play a role in editing the cryptogenes encoded on the maxicircle genome. In general, *T. brucei* minicircles each encode three gRNA genes (101) . There are estimated to be about 300 sequence classes of minicircles (although only 27 have been sequenced) providing the hundreds of gRNAs necessary for the extensive editing required for some of the maxicircle mRNAs . Nearly all gRNA genes are flanked by 18 bp imperfect inverted repeats that have been postulated to play a role in transcription and/or processing of gRNA transcripts and/or gRNA mobility, but none of these functions have been demonstrated (85, 86).

kDNA Transcription

Although much is known about the post-transcriptional process of RNA editing in *T. brucei*, very little is known about transcription of kDNA. Similar to animal mitochondrial transcription, transcription of the maxicircle appears to generate a polycistronic precursor from which mature transcripts must be processed by cleavage and polyadenylation (78, 79). In contrast, minicircle encoded gRNA genes are monocistronic, although read-through transcription produces oligocistronic precursors from which only the initiating transcript survives downstream processing (49, 85, 86). No promoter sequences have been identified for either maxicircle or minicircle genes. However, transcription of the maxicircle has been shown to initiate at least 1200 nt

upstream of the mature 5' end of the 12S rRNA (79). Due to the instability of this precursor, the precise transcription initiation site has not been mapped.

Developmental Regulation of Gene Expression

T. brucei and many other members of this family have alternating forms during which part of the life cycle is spent in an invertebrate vector before being transmitted to a vertebrate (and sometimes plant) host. These parasites must respond quickly to extreme changes in their environment with dramatic changes in morphology, surface structure, and metabolism. Consequently, precise and rapid gene regulation is essential. In *T. brucei*, this regulation occurs via the coordinated expression of both the nuclear and mitochondrial genomes and is indispensable for completion of the trypanosome life cycle. The dual nature of the kinetoplast genome adds another layer of complexity as both maxicircle and minicircle genes are required to produce functional mRNAs.

Regulation of kDNA expression

Interestingly, the steady state levels of various maxicircle transcripts are developmentally regulated in *T. brucei*. The 9S and 12S ribosomal RNAs are 30 fold more abundant in the procyclic stage than in the bloodstream stage, reflecting the relative importance of translation of the mitochondrial mRNAs in the different life stages (78, 79). In spite of this dramatic difference, transcription rates measured *in organello* and *in vivo* in bloodstream and procyclic trypanosomes are identical, suggesting that the regulation of transcript abundance, at least for the rRNAs, occurs at the level of processing or stability (79). Several mitochondrial mRNAs also show stage specific

differences in abundance that reflect the relative importance of the specific functions of the mitochondrion for each stage. For example, cytochrome subunits are very low or undetectable in the relatively inactive bloodstream form mitochondrion which are inactive for electron transport (87). In contrast, NADH dehydrogenase subunit 7 and 8 mRNAs are upregulated in the bloodstream, possibly reflecting their importance in maintaining redox balance in the glycosome by transferring electrons via ubiquinone to the alternative oxidase (8, 54, 87). Editing of various transcripts is also developmentally regulated, often in a similar pattern (10, 64). Complex I transcripts are predominantly edited in the bloodstream form stage, while complexes III and IV are edited preferentially in the procyclic stage.

Nuclear gene regulation

As with other eukaryotes, the majority of the genes required for mitochondrial biogenesis and function are encoded in *T. brucei* nuclear genome. Many of these genes are also developmentally regulated according to the life cycle stage of the parasite. Nuclear gene expression in the kinetoplastida is remarkably different than that studied in model, crown eukaryotes in that there is very little regulation at the level of transcription initiation (29, 112). This is likely a consequence of the unique genomic organization of the nuclear genes. In the chromosomes assembled from sequence data for both *T. brucei* and *Leishmania major*, intron-less open reading frames are clustered in groups of 30 -50 or more (72, 116). Interestingly, these genes are not organized as operons since genes for similar functions are typically not coded near each other. These genes are co-transcribed by a highly proccessive α-amanitin sensitive polymerase, presumably pol II, although pol

I has been shown to be responsible for the transcription of certain classes of cell surface proteins (112). Individual genes are processed from this polycistronic precursor by the addition of a capped ~40 nt trans-spliced leader RNA to the 5' end and cleavage/polyadenylation at the 3' end. No developmental regulation at the level of transcription, trans-splicing or polyadenylation has been observed (29). It appears that the predominant element for regulation is the 3' UTR of mature mRNAs, which affects both the stability and in some cases translatability of the mRNA (38, 47, 67, 97). 3' UTR elements include short 16 mer sequences that form secondary structure and AU rich elements (AREs). So far, the only role for the 5' UTR in regulating transcript abundance has been seen in Crithidia, in which the 5' UTR in several nuclear-encoded mitochondrial genes have been shown to play a role in cell-cycle regulation (70). In addition, regulation at the level of protein stability has been observed (110).

Overview

The modern mitochondrial proteome consists of a limited set of genes encoded within the mitochondria, and a larger set that have been transferred to the nucleus. In addition, the majority of the mitochondrial proteome consists of eukaryotic, nuclear-encoded inventions (63). Whole genome sequence comparisons suggest that new components of the mitochondrial proteome have continued to evolve after the divergence of the major taxa. (3). This simple fact is often overlooked in papers describing how particular mitochondrial properties are conserved throughout evolution from "yeast to fly to man". It is clear from sampling even these relatively few taxa that group specific evolution has resulted in differing requirements for various mitochondrial processes in

animals, fungi, plants, and protists. Indeed, distinct differences exist in the mechanisms for transcription of the mitochondrial genome, both in *cis* elements and *trans*-acting factors.

The origins and evolution of the mitochondrial transcription apparatus remain an evolutionary enigma. Since *T. brucei* belongs to one of the earliest branches of eukaryotes that possess mitochondria, learning more about the differences and similarities between trypanosomes and model organisms will provide evolutionary insight into the development of this unusual transcription machinery. Additionally, the potential to discover novel components and/or mechanisms of transcription in this ancient organism opens up the possibility of finding much needed chemotherapeutic agents against this pathogen.

In order to learn more about mitochondrial transcription in *T. brucei,* my thesis work has approached some basic questions about the nature of the mitochondrial transcription apparatus using several different strategies. Very little was known about mitochondrial transcription in this group of organisms when I began this project. The mitochondrial transcription machinery was unknown, as were maxicircle transcription start sites. In the work described in this dissertation, I have sought to identify important components of mitochondrial transcription in *T. brucei* including the proteins and *cis*-elements required for transcription initiation. Chapter II describes the identification and mapping of a transcription start site for a gRNA gene with an unusual intergenic location, while chapters III and IV describe the identification of homologues to the mitochondrial RNAP and accessory factor mtTFB, respectively.

LITERATURE CITED

1. **Adams, K. L., and J. D. Palmer.** 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Mol Phylogenet Evol **29:**380-95.

2. **Alam, T. I., T. Kanki, T. Muta, K. Ukaji, Y. Abe, H. Nakayama, K. Takio, N. Hamasaki, and D. Kang.** 2003. Human mitochondrial DNA is packaged with TFAM. Nucleic Acids Res **31:**1640-5.

3. **Andersson, S. G., O. Karlberg, B. Canback, and C. G. Kurland.** 2003. On the origin of mitochondria: a genomics perspective. Philos Trans R Soc Lond B Biol Sci **358:**165-77; discussion 177-9.

4. **Andersson, S. G., and C. G. Kurland.** 1999. Origins of mitochondria and hydrogenosomes. Curr Opin Microbiol **2:**535-41.

5. **Andersson, S. G., A. Zomorodipour, J. O. Andersson, T. Sicheritz-Ponten, U. C. Alsmark, R. M. Podowski, A. K. Naslund, A. S. Eriksson, H. H. Winkler, and C. G. Kurland.** 1998. The genome sequence of Rickettsia prowazekii and the origin of mitochondria. Nature **396:**133-40.

6. **Barth, C., U. Greferath, M. Kotsifas, and P. R. Fisher.** 1999. Polycistronic transcription and editing of the mitochondrial small subunit (SSU) ribosomal RNA in Dictyostelium discoideum. Curr Genet **36:**55-61.

7. **Barth, C., U. Greferath, M. Kotsifas, Y. Tanaka, S. Alexander, H. Alexander, and P. R. Fisher.** 2001. Transcript mapping and processing of mitochondrial RNA in Dictyostelium discoideum. Curr Genet **39:**355-64.

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

9. **Bendich, A. J.** 1993. Reaching for the ring: the study of mitochondrial genome structure. Curr Genet **24:**279-90.

10. **Bhat, G. J., A. E. Souza, J. E. Feagin, and K. Stuart.** 1992. Transcript-specific developmental regulation of polyadenylation in Trypanosoma brucei mitochondria. Mol Biochem Parasitol **52:**231-40.

11. **Binder, S., and A. Brennicke.** 2003. Gene expression in plant mitochondria: transcriptional and post-transcriptional control. Philos Trans R Soc Lond B Biol Sci **358:**181-8; discussion 188-9.

12. **Binder, S., and A. Brennicke.** 1993. Transcription initiation sites in mitochondria of Oenothera berteriana. J Biol Chem **268:**7849-55.

13. **Binder, S., F. Hatzack, and A. Brennicke.** 1995. A novel pea mitochondrial in vitro transcription system recognizes homologous and heterologous mRNA and tRNA promoters. J Biol Chem **270:**22182-9.

14. **Biswas, T. K.** 1990. Control of mitochondrial gene expression in the yeast Saccharomyces cerevisiae. Proc Natl Acad Sci U S A **87:**9338-42.

15. **Bogenhagen, D. F., and M. F. Romanelli.** 1988. Template sequences required for transcription of Xenopus laevis mitochondrial DNA from two bidirectional promoters. Mol Cell Biol **8:**2917-24.

16. **Bogenhagen, D. F., and B. K. Yoza.** 1986. Accurate in vitro transcription of Xenopus laevis mitochondrial DNA from two bidirectional promoters. Mol Cell Biol **6:**2543-50.

17. **Boore, J. L.** 1999. Animal mitochondrial genomes. Nucleic Acids Res **27:**1767-80.

18. **Brennicke, A., E. Zabaleta, S. Dombrowski, M. Hoffmann, and S. Binder.** 1999. Transcription signals of mitochondrial and nuclear genes for mitochondrial proteins in dicot plants. J Hered **90:**345-50.

19. **Brown, G. G., A. H. Auchincloss, P. S. Covello, M. W. Gray, R. Menassa, and M. Singh.** 1991. Characterization of transcription initiation sites on the soybean mitochondrial genome allows identification of a transcription-associated sequence motif. Mol Gen Genet **228:**345-55.

20. **Bryan, A. C., M. S. Rodeheffer, C. M. Wearn, and G. S. Shadel.** 2002. Sls1p is a membrane-bound regulator of transcription-coupled processes involved in Saccharomyces cerevisiae mitochondrial gene expression. Genetics **160:**75-82.

21. **Bullerwell, C. E., and M. W. Gray.** 2004. Evolution of the mitochondrial genome: protist connections to animals, fungi and plants. Curr Opin Microbiol **7:**528-34.

22. **Burger, G., M. W. Gray, and B. F. Lang.** 2003. Mitochondrial genomes: anything goes. Trends Genet **19:**709-16.

23. **Camargo, E. P.** 1999. Phytomonas and other trypanosomatid parasites of plants and fruit. Adv Parasitol **42:**29-112.

24. **Carrodeguas, J. A., S. Yun, G. S. Shadel, D. A. Clayton, and D. F. Bogenhagen.** 1996. Functional conservation of yeast mtTFB despite extensive sequence divergence. Gene Expr **6:**219-30.

25. **Catisti, R., S. A. Uyemura, R. Docampo, and A. E. Vercesi.** 2000. Calcium mobilization by arachidonic acid in trypanosomatids. Mol Biochem Parasitol **105:**261-71.

26. **Cermakian, N., T. M. Ikeda, R. Cedergren, and M. W. Gray.** 1996. Sequences homologous to yeast mitochondrial and bacteriophage T3 and T7 RNA polymerases are widespread throughout the eukaryotic lineage. Nucleic Acids Res **24:**648-54.

27. **Cermakian, N., T. M. Ikeda, P. Miramontes, B. F. Lang, M. W. Gray, and R. Cedergren.** 1997. On the evolution of the single-subunit RNA polymerases. J Mol Evol **45:**671-81.

28. **Chaudhuri, M., W. Ajayi, and G. C. Hill.** 1998. Biochemical and molecular properties of the Trypanosoma brucei alternative oxidase. Mol Biochem Parasitol **95:**53-68.

29. **Clayton, C. E.** 2002. Life without transcriptional control? From fly to man and back again. Embo J **21:**1881-8.

30. **Clayton, D. A.** 2000. Transcription and replication of mitochondrial DNA. Hum Reprod **15 Suppl 2:**11-7.

31. **Cliften, P. F., S. H. Jang, and J. A. Jaehning.** 2000. Identifying a Core RNA Polymerase Surface Critical for Interactions with a Sigma-Like Specificity Factor. Mol Cell Biol **20:**7013-7023.

32. **Cliften, P. F., J. Y. Park, B. P. Davis, S. H. Jang, and J. A. Jaehning.** 1997. Identification of three regions essential for interaction between a sigma-like factor and core RNA polymerase. Genes Dev **11:**2897-909.

33. **Dairaghi, D. J., G. S. Shadel, and D. A. Clayton.** 1995. Addition of a 29 residue carboxyl-terminal tail converts a simple HMG box-containing protein into a transcriptional activator. J Mol Biol **249:**11-28.

34. **Debrabant, A., and H. Nakhasi.** 2003. Programmed cell death in trypanosomatids: is it an altruistic mechanism for survival of the fittest? Kinetoplastid Biol Dis **2:**7.

35. **Dequard-Chablat, M., and C. Allandt.** 2002. Two copies of mthmg1, encoding a novel mitochondrial HMG-like protein, delay accumulation of mitochondrial DNA deletions in Podospora anserina. Eukaryot Cell **1:**503-13.

36. **Dieckmann, C. L., and R. R. Staples.** 1994. Regulation of mitochondrial gene expression in Saccharomyces cerevisiae. Int Rev Cytol **152:**145-81.

37. **Docampo, R., and S. N. Moreno.** 2003. Current chemotherapy of human African trypanosomiasis. Parasitol Res **90 Supp 1:**S10-3.

38. **Drozdz, M., and C. Clayton.** 1999. Structure of a regulatory 3' untranslated region from Trypanosoma brucei. Rna **5:**1632-44.

39. **Ekstrand, M. I., M. Falkenberg, A. Rantanen, C. B. Park, M. Gaspari, K. Hultenby, P. Rustin, C. M. Gustafsson, and N. G. Larsson.** 2004. Mitochondrial transcription factor A regulates mtDNA copy number in mammals. Hum Mol Genet **13:**935-44.

40. **Emelyanov, V. V.** 2003. Mitochondrial connection to the origin of the eukaryotic cell. Eur J Biochem **270:**1599-618.

41. **Fairlamb, A. H.** 2003. Chemotherapy of human African trypanosomiasis: current and future prospects. Trends Parasitol **19:**488-94.

42. **Falkenberg, M., M. Gaspari, A. Rantanen, A. Trifunovic, N. G. Larsson, and C. M. Gustafsson.** 2002. Mitochondrial transcription factors B1 and B2 activate transcription of human mtDNA. Nat Genet **31:**289-94.

43. **Feagin, J. E.** 1992. The 6-kb element of Plasmodium falciparum encodes mitochondrial cytochrome genes. Mol Biochem Parasitol **52:**145-8.

44. **Fernandez-Silva, P., J. A. Enriquez, and J. Montoya.** 2003. Replication and transcription of mammalian mitochondrial DNA. Exp Physiol **88:**41-56.

45. **Fey, J., and L. Marechal-Drouard.** 1999. Compilation and analysis of plant mitochondrial promoter sequences: An illustration of a divergent evolution between monocot and dicot mitochondria. Biochem Biophys Res Commun **256:**409-14.

46. **Fisher, R. P., and D. A. Clayton.** 1988. Purification and characterization of human mitochondrial transcription factor 1. Mol Cell Biol **8:**3496-509.

47. **Furger, A., N. Schurch, U. Kurath, and I. Roditi.** 1997. Elements in the 3' untranslated region of procyclin mRNA regulate expression in insect forms of Trypanosoma brucei by modulating RNA stability and translation. Mol Cell Biol **17:**4372-80.

48. **Gaspari, M., M. Falkenberg, N. G. Larsson, and C. M. Gustafsson.** 2004. The mitochondrial RNA polymerase contributes critically to promoter specificity in mammalian cells. Embo J.

49. **Grams, J., M. T. McManus, and S. L. Hajduk.** 2000. Processing of polycistronic guide RNAs is associated with RNA editing complexes in Trypanosoma brucei. Embo J **19:**5525-32.

50. **Gray, M. W., G. Burger, and B. F. Lang.** 2001. The origin and early evolution of mitochondria. Genome Biol **2:**REVIEWS1018.

51. **Gray, M. W., B. F. Lang, and G. Burger.** 2004. Mitochondria of Protists. Annu Rev Genet.

52. **Gunter, T. E., D. I. Yule, K. K. Gunter, R. A. Eliseev, and J. D. Salter.** 2004. Calcium and mitochondria. FEBS Lett **567:**96-102.

53. **Haag, J., C. O'HUigin, and P. Overath.** 1998. The molecular phylogeny of trypanosomes: evidence for an early divergence of the Salivaria. Mol Biochem Parasitol **91:**37-49.

54. **Hannaert, V., F. Bringaud, F. R. Opperdoes, and P. A. Michels.** 2003. Evolution of energy metabolism and its compartmentation in Kinetoplastida. Kinetoplastid Biol Dis **2:**11.

55. **Hatzack, F., S. Dombrowski, A. Brennicke, and S. Binder.** 1998. Characterization of DNA-Binding Proteins from Pea Mitochondria. Plant Physiol **116:**519-28.

56. **Hedtke, B., T. Borner, and A. Weihe.** 1997. Mitochondrial and chloroplast phage-type RNA polymerases in Arabidopsis. Science **277:**809-11.

57. **Hess, W. R., and T. Borner.** 1999. Organellar RNA polymerases of higher plants. Int Rev Cytol **190:**1-59.

58. **Hoffmann, M., and S. Binder.** 2002. Functional importance of nucleotide identities within the pea atp9 mitochondrial promoter sequence. J Mol Biol **320:**943-50.

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

60. **Ikeda, T. M., and M. W. Gray.** 1999. Identification and characterization of T3/T7 bacteriophage-like RNA polymerase sequences in wheat. Plant Mol Biol **40:**567-78.

61. **Jacobs, M. A., S. R. Payne, and A. J. Bendich.** 1996. Moving pictures and pulsed-field gel electrophoresis show only linear mitochondrial DNA molecules from yeasts with linear-mapping and circular-mapping mitochondrial genomes. Curr Genet **30:**3-11.

62. **Jang, S. H., and J. A. Jaehning.** 1991. The yeast mitochondrial RNA polymerase specificity factor, MTF1, is similar to bacterial sigma factors. J Biol Chem **266:**22671-7.

63. **Karlberg, O., B. Canback, C. G. Kurland, and S. G. Andersson.** 2000. The dual origin of the yeast mitochondrial proteome. Yeast **17:**170-87.

64. **Koslowsky, D. J., G. R. Riley, J. E. Feagin, and K. Stuart.** 1992. Guide RNAs for transcripts with developmentally regulated RNA editing are present in both life cycle stages of Trypanosoma brucei. Mol Cell Biol **12:**2043-9.

65. **Kurland, C. G., and S. G. Andersson.** 2000. Origin and evolution of the mitochondrial proteome. Microbiol Mol Biol Rev **64:**786-820.

66. **Lee, H. K.** 2004. Mitochondrial pathogenesis from genes and apoptosis to aging and disease. Overview. Ann N Y Acad Sci **1011:**1-6.

67. **Lee, M. G.** 1998. The 3' untranslated region of the hsp 70 genes maintains the level of steady state mRNA in Trypanosoma brucei upon heat shock. Nucleic Acids Res **26:**4025-33.

68. **Lukes, J., D. L. Guilbride, J. Votypka, A. Zikova, R. Benne, and P. T. Englund.** 2002. Kinetoplast DNA network: evolution of an improbable structure. Eukaryot Cell **1:**495-502.

69. **Lurin, C., C. Andres, S. Aubourg, M. Bellaoui, F. Bitton, C. Bruyere, M. Caboche, C. Debast, J. Gualberto, B. Hoffmann, A. Lecharny, M. Le Ret, M. L. Martin-Magniette, H. Mireau, N. Peeters, J. P. Renou, B. Szurek, L. Taconnat, and I. Small.** 2004. Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell **16:**2089-103.

70. **Mahmood, R., J. C. Hines, and D. S. Ray.** 1999. Identification of cis and trans elements involved in the cell cycle regulation of multiple genes in Crithidia fasciculata. Mol Cell Biol **19:**6174-82.

71. **Mangus, D. A., S. H. Jang, and J. A. Jaehning.** 1994. Release of the yeast mitochondrial RNA polymerase specificity factor from transcription complexes. J Biol Chem **269:**26568-74.

72. **Martinez-Calvillo, S., S. Yan, D. Nguyen, M. Fox, K. Stuart, and P. J. Myler.** 2003. Transcription of Leishmania major Friedlin chromosome 1 initiates in both directions within a single region. Mol Cell **11:**1291-9.

73. **Masters, B. S., L. L. Stohl, and D. A. Clayton.** 1987. Yeast mitochondrial RNA polymerase is homologous to those encoded by bacteriophages T3 and T7. Cell **51:**89-99.

74. **Matsunaga, M., and J. A. Jaehning.** 2004. Intrinsic promoter recognition by a "core" RNA polymerase. J Biol Chem **279:**44239-42.

75. **Matsushima, Y., R. Garesse, and L. S. Kaguni.** 2004. Drosophila mitochondrial transcription factor B2 regulates mitochondrial DNA copy number and transcription in schneider cells. J Biol Chem **279:**26900-5.

76. **McCulloch, V., B. L. Seidel-Rogol, and G. S. Shadel.** 2002. A human mitochondrial transcription factor is related to RNA adenine methyltransferases and binds S-adenosylmethionine. Mol Cell Biol **22:**1116-25.

77. **McCulloch, V., and G. S. Shadel.** 2003. Human mitochondrial transcription factor B1 interacts with the C-terminal activation region of h-mtTFA and stimulates transcription independently of its RNA methyltransferase activity. Mol Cell Biol **23:**5816-24.

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

79. **Michelotti, E. F., M. E. Harris, B. Adler, A. F. Torri, and S. L. Hajduk.** 1992. Trypanosoma brucei mitochondrial ribosomal RNA synthesis, processing and developmentally regulated expression. Mol Biochem Parasitol **54:**31-41.

80. **Myler, P. J., D. Glick, J. E. Feagin, T. H. Morales, and K. D. Stuart.** 1993. Structural organization of the maxicircle variable region of Trypanosoma brucei: identification of potential replication origins and topoisomerase II binding sites. Nucleic Acids Res **21:**687-94.

81. **Nok, A. J.** 2003. Arsenicals (melarsoprol), pentamidine and suramin in the treatment of human African trypanosomiasis. Parasitol Res **90:**71-9.

82. **Oda, K., K. Yamato, E. Ohta, Y. Nakamura, M. Takemura, N. Nozato, K. Akashi, T. Kanegae, Y. Ogura, T. Kohchi, and et al.** 1992. Gene organization deduced from the complete sequence of liverwort Marchantia polymorpha mitochondrial DNA. A primitive form of plant mitochondrial genome. J Mol Biol **223:**1-7.

83. **Osinga, K. A., M. De Haan, T. Christianson, and H. F. Tabak.** 1982. A nonanucleotide sequence involved in promotion of ribosomal RNA synthesis and RNA priming of DNA replication in yeast mitochondria. Nucleic Acids Res **10:**7993-8006.

84. **Parsons, M.** 2004. Glycosomes: parasites and the divergence of peroxisomal purpose. Mol Microbiol **53:**717-24.

85. **Pollard, V. W., and S. L. Hajduk.** 1991. Trypanosoma equiperdum minicircles encode three distinct primary transcripts which exhibit guide RNA characteristics. Mol Cell Biol **11:**1668-75.

86. **Pollard, V. W., S. P. Rohrer, E. F. Michelotti, K. Hancock, and S. L. Hajduk.** 1990. Organization of minicircle genes for guide RNAs in Trypanosoma brucei. Cell **63:**783-90.

87. **Priest, J. W., and S. L. Hajduk.** 1994. Developmental regulation of mitochondrial biogenesis in Trypanosoma brucei. J Bioenerg Biomembr **26:**179-91.

88. **Rantanen, A., M. Gaspari, M. Falkenberg, C. M. Gustafsson, and N. G. Larsson.** 2003. Characterization of the mouse genes for mitochondrial transcription factors B1 and B2. Mamm Genome **14:**1-6.

89. **Rehkopf, D. H., D. E. Gillespie, M. I. Harrell, and J. E. Feagin.** 2000. Transcriptional mapping and RNA processing of the Plasmodium falciparum mitochondrial mRNAs. Mol Biochem Parasitol **105:**91-103.

90. **Ridgley, E. L., Z. H. Xiong, and L. Ruben.** 1999. Reactive oxygen species activate a Ca2+-dependent cell death pathway in the unicellular organism Trypanosoma brucei brucei. Biochem J **340 (Pt 1):**33-40.

91. **Rodeheffer, M. S., B. E. Boone, A. C. Bryan, and G. S. Shadel.** 2000. Nam1p, a protein involved in RNA processing and translation, is coupled to transcription through an interaction with yeast mitochondrial RNA polymerase. J Biol Chem **15:**15.

92. **Sanchez, J. P., Y. Murakami, J. A. Huberman, and J. Hurwitz.** 1998. Isolation, characterization, and molecular cloning of a protein (Abp2) that binds

to a Schizosaccharomyces pombe origin of replication (ars3002). Mol Cell Biol **18:**1670-81.

93. **Santiago, J., and C. G. Vallejo.** 1998. Identification of a mitochondrial RNA polymerase in the crustacean Artemia franciscana. Arch Biochem Biophys **353:**276-84.

94. **Sasaki, N., H. Kuroiwa, C. Nishitani, H. Takano, T. Higashiyama, T. Kobayashi, Y. Shirai, A. Sakai, S. Kawano, K. Murakami-Murofushi, and T. Kuroiwa.** 2003. Glom is a novel mitochondrial DNA packaging protein in Physarum polycephalum and causes intense chromatin condensation without suppressing DNA functions. Mol Biol Cell **14:**4758-69.

95. **Schnaufer, A., G. J. Domingo, and K. Stuart.** 2002. Natural and induced dyskinetoplastic trypanosomatids: how to live without mitochondrial DNA. Int J Parasitol **32:**1071-84.

96. **Schubot, F. D., C. J. Chen, J. P. Rose, T. A. Dailey, H. A. Dailey, and B. C. Wang.** 2001. Crystal structure of the transcription factor sc-mtTFB offers insights into mitochondrial transcription. Protein Sci **10:**1980-8.

97. **Schurch, N., A. Furger, U. Kurath, and I. Roditi.** 1997. Contributions of the procyclin 3' untranslated region and coding region to the regulation of expression in bloodstream forms of Trypanosoma brucei. Mol Biochem Parasitol **89:**109-21.

98. **Seidel-Rogol, B. L., V. McCulloch, and G. S. Shadel.** 2003. Human mitochondrial transcription factor B1 methylates ribosomal RNA at a conserved stem-loop. Nat Genet **33:**23-4.

99. **Shadel, G. S., and D. A. Clayton.** 1996. Mapping promoters in displacement-loop region of vertebrate mitochondrial DNA. Methods Enzymol **264:**139-48.

100. **Shadel, G. S., and D. A. Clayton.** 1995. A Saccharomyces cerevisiae mitochondrial transcription factor, sc- mtTFB, shares features with sigma factors but is functionally distinct. Mol Cell Biol **15:**2101-8.

101. **Simpson, L.** 1997. The genomic organization of guide RNA genes in kinetoplastid protozoa: several conundrums and their solutions. Mol Biochem Parasitol **86:**133-41.

102. **Simpson, L.** 1986. Kinetoplast DNA in trypanosomid flagellates. Int Rev Cytol **99:**119-79.

103. **Smeitink, J., L. van den Heuvel, and S. DiMauro.** 2001. The genetics and pathology of oxidative phosphorylation. Nat Rev Genet **2:**342-52.

104. **Stuart, K., and J. E. Feagin.** 1992. Mitochondrial DNA of kinetoplastids. Int Rev Cytol **141:**65-88.

105. **Stuart, K., and A. K. Panigrahi.** 2002. RNA editing: complexity and complications. Mol Microbiol **45:**591-6.

106. **Taanman, J. W.** 1999. The mitochondrial genome: structure, transcription, translation and replication. Biochim Biophys Acta **1410:**103-23.

107. **Teixeira, S. M., and W. D. daRocha.** 2003. Control of gene expression and genetic manipulation in the Trypanosomatidae. Genet Mol Res **2:**148-58.

108. **Thomas, J. O., and A. A. Travers.** 2001. HMG1 and 2, and related 'architectural' DNA-binding proteins. Trends Biochem Sci **26:**167-74.

109. **Tiranti, V., A. Savoia, F. Forti, M. F. D'Apolito, M. Centra, M. Rocchi, and M. Zeviani.** 1997. Identification of the gene encoding the human mitochondrial RNA polymerase (h-mtRPOL) by cyberscreening of the Expressed Sequence Tags database. Hum Mol Genet **6:**615-25.

110. **Torri, A. F., K. I. Bertrand, and S. L. Hajduk.** 1993. Protein stability regulates the expression of cytochrome c during the developmental cycle of Trypanosoma brucei. Mol Biochem Parasitol **57:**305-15.

111. **Unseld, M., J. R. Marienfeld, P. Brandt, and A. Brennicke.** 1997. The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366,924 nucleotides. Nat Genet **15:**57-61.

112. **Vanhamme, L., and E. Pays.** 1995. Control of gene expression in trypanosomes. Microbiol Rev **59:**223-40.

113. **Wallace, D. C.** 1999. Mitochondrial diseases in man and mouse. Science **283:**1482-8.

114. **Wang, Y., and G. S. Shadel.** 1999. Stability of the mitochondrial genome requires an amino-terminal domain of yeast mitochondrial RNA polymerase. Proc Natl Acad Sci U S A **96:**8046-51.

115. **Ward, B. L., R. S. Anderson, and A. J. Bendich.** 1981. The mitochondrial genome is large and variable in a family of plants (cucurbitaceae). Cell **25:**793-803.

116. **Worthey, E. A., S. Martinez-Calvillo, A. Schnaufer, G. Aggarwal, J. Cawthra, G. Fazelinia, C. Fong, G. Fu, M. Hassebrock, G. Hixson, A. C. Ivens, P. Kiser, F. Marsolini, E. Rickell, R. Salavati, E. Sisk, S. M. Sunkin, K. D. Stuart, and P. J. Myler.** 2003. Leishmania major chromosome 3 contains two long convergent polycistronic gene clusters separated by a tRNA gene. Nucleic Acids Res **31:**4201-10.

117. **Xu, B., and D. A. Clayton.** 1992. Assignment of a yeast protein necessary for mitochondrial transcription initiation. Nucleic Acids Res **20:**1053-9.

118. **Young, D. A., R. L. Allen, A. J. Harvey, and D. M. Lonsdale.** 1998. Characterization of a gene encoding a single-subunit bacteriophage-type RNA polymerase from maize which is alternatively spliced. Mol Gen Genet **260:**30-7.

CHAPTER II

An intragenic guide RNA location suggests a complex mechanism for mitochondrial gene expression in *Trypanosoma brucei*

INTRODUCTION

Trypanosoma brucei is a parasitic protozoan with an extraordinarily complex mitochondrial genome. This genome, the kinetoplast, consists of a network of thousands of 1 kb minicircles unique to kinetoplasts and 50 copies of a 22 kb maxicircle that is analogous to mitochondrial DNA in other organisms (25). The mRNAs produced from most maxicircle genes must be edited by the insertion or deletion of uridylate residues to create mature transcripts. This process is mediated by small, complementary guide RNAs (gRNAs) encoded on minicircle DNA. The process of editing occurs post-transcriptionally and has been studied extensively (for recent reviews see (18, 26, 27). However, the initial step in mitochondrial gene expression, transcription, has been largely ignored.

While no kinetoplast (kDNA) promoter elements have been described to date, there are regions from which transcription has been shown to initiate. In *T. brucei*, most minicircle gRNA genes encode primary transcripts located within cassettes of imperfect inverted 18 bp repeats (15, 22). These inverted repeats have been implicated in playing a role in transcription. However, there have been no direct tests of their function in transcription initiation. Although precise maxicircle transcription start sites have not been identified, transcription of the major strand has been shown to begin at least 1200 nt upstream of the 5' end of the mature 12S rRNA and proceed polycistronically (19). Mature transcripts are rapidly processed from this precursor, often by mutually exclusive events as many of the coding regions overlap (17, 23). Transcripts corresponding to the maxicircle minor strand are also present (20), but the transcription start site has not been identified.

Observations from relatives of *T. brucei* indicate that not all maxicircle genes are transcribed as part of a polycistronic precursor. The maxicircles of both *Leishmania tarantolae* and *Crithidia fasciculata* encode numerous gRNAs located primarily within intergenic spaces (2, 29). Many of these mature gRNAs are primary transcripts (3), indicating that several transcription units exist on these maxicircle genomes. In contrast, the coding region of the *T. brucei* maxicircle is much more compact and is thought to encode only 2 gRNAs (29) (Figure 1). Interestingly, both of these gRNAs are located within mRNA coding sequences. The COII guide RNA, gCOII, is unusual in that it is encoded within the 3' end of the COII mRNA and is thought to function in *cis* as no discrete gCOII transcript has been detected (2, 17). The other gRNA, gMURF2-II has not been tested for expression. This gene is located within the 5' end of the ND4 message, indicating that the gRNA and the mRNA could be generated via mutually exclusive processing events from a precursor RNA. Intriguingly, this gRNA is also located downstream of one of three intergenic regions on the compact maxicircle, raising the possibility that the gMURF2-II gene is a discrete transcription unit regulated by a promoter located within this region. This locus therefore represents a unique region on the *T. brucei* maxicircle that may be regulated by alternative processing and/or transcription initiation events.

In order to further characterize potential transcription units in *T. brucei* kDNA, we sought to more closely investigate the product of the conserved maxicircle gMURF2-II gene. This gRNA is expressed exclusively from the maxicircle gMURF2-II locus and is contained entirely within the 5' end of the ND4 gene. Characterization of the structure of the 5' ends of the gMURF2-II and the ND4 transcripts demonstrates that the gMURF2-II

RNA is a primary transcript while ND4 mRNA is processed, likely from the polycistronic precursor. In addition, to examine unusual minicircle transcription units, we investigated the expression of gCYb(560), which is not located between inverted repeats. This gRNA is also a primary transcript indicating that minicircle gRNA transcription initiation signals exist outside the inverted repeat regions. These results imply that gRNA genes operate as individual transcription units, regardless of genomic context, and suggest a complex mechanism of transcription and processing of mitochondrial transcripts.

MATERIALS AND METHODS

Cell growth and mitochondrial isolation. Procyclic *Trypanosoma brucei*, EATRO 164, IsTaR 1 serodeme, derived from VAT 1.7 clone A was grown in SDM-79 medium supplemented with 5% Fetal Bovine Serum (Sigma) at 27°C. Mitochondrial isolations were prepared from cells harvested at a density of ~3 x 10^7 cells ml^{-1}. Cells were lysed in a dounce homogenizer under hypotonic conditions and mitochondrial vesicles were isolated on Nycodenz step gradients as described (13). In order to obtain precursor transcripts for the ligation experiments, isolated mitochondria were incubated in transcription buffer (5mM HEPES, pH 7.6, 3 mM potassium phosphate pH 7.7, 125 mM sucrose, 6mM KCl, 10 mM $MgCl_2$, 1 mM EDTA, 2 mM DTT) and 100 μM rNTPs for 30 minutes (12). Mitochondria were snap frozen in a dry ice ethanol bath, and stored at -80° C prior to RNA extraction described below.

Oligonucleotide probes and primers. All DNA oligonucleotides were synthesized by IDT. The RNA oligo was synthesized by Dharmacon. Forward primers (F) match the RNA sequence while reverse (R) primers are complementary.

gMURF2-II F:	5'-GAAAGCACAAAAATAAAATTAAATTAGAG-3'
gMURF2-II R:	5'-CATTCAATTACTCTAATTTAATTTTATTTTTGTGC-3'
gA6-14sU R:	5'TAATTATCATATCACTGTCAAAATCTGATTCGTTATC
	GGAGTTATAGCCCTATAGTGAGTCGTATT AAATT-3'
gND7(506) R:	5'-CACTAACTATACTACAGGTTATTTACATCG-3'
gCYb-(560A/B) R:	5'-CCTCCCYATTACTCAGAAATCTACATTGTC-3'
oligo dA-XBS R:	5'-GATCTAGAGGATCCCGGGAAAAAAAAAAAAAAA3'
XBS F:	5'-GATCTAGAGGATCCCGGG-3'
RNA oligo:	5' AGAUUUUGACAGUGAUAUGAUAAUUA 3'

Nucleic acid Isolation, Northern and Southern blot analysis. Total *T. brucei* genomic DNA was isolated as described previously (8). Mitochondrial RNA was isolated by the acid guanidinium-phenol-chloroform method (7). Guide RNA was gel purified on a 6% 8M urea acrylamide gel. For Northern blot analysis, mitochondrial RNA was electrophoresed on a 5% 37.5:1, 8 M urea acrylamide gel and transferred to Nytran (Schleicher & Schuell) using a semi-phor electroblotter (Hoeffer). Northern and Southern blots were hybridized with ^{32}P end-labeled oligonucleotide probes using standard conditions.

5' and 3' end mapping. Gene specific DNA oligonucleotides were labeled with γ-^{32}P ATP using T4 polynucleotide kinase (Invitrogen) and gel purified on a 12% 8M urea acrylamide gel. 2.5 μg of mitochondrial RNA was reverse transcribed as previously described (8) in the presence of 100 kCPM of primer. RNA sequencing ladders were generated by adding equal amounts of deoxy- and dideoxy- (ATP or TTP) in the extension reaction. Reactions were stopped by the addition of an equal volume of 95%

formamide gel-loading buffer and resolved on an 8%, 8M urea acrylamide sequencing gel. Gels were fixed in 7% acetic acid + 7% methanol, dried, and exposed on film.

Enzymatic treatments of RNA. 10 μg of mitochondrial RNA (mtRNA) or 2 μg of gel purified gRNA was dephosphorylated with alkaline phosphatase (New England Biolabs) following manufacturer's instructions. Both untreated and phosphatased RNA (5.0 μg mtRNA or 1.0 μg gRNA) was phosphorylated with T4 polynucleotide kinase (Invitrogen) supplemented with 1 mM ATP. The alkaline phosphatase was inactivated with proteinase K digestion prior to phenol/chloroform extraction. All enzymatic reactions were terminated by phenol/chloroform extraction followed by ethanol precipitation in the presence of linear acrylamide carrier.

RNA ligations and poisoned primer extension analysis. A 26 nt long synthetic RNA oligonucleotide (Dharmacon) was used as the acceptor RNA in all ligation reactions. This RNA oligo possesses a 5' OH and is consequently unable to serve as a donor molecule in the ligation reaction. 5.0 μg mtRNA or 0.5 μg gRNA was used in each ligation. Untreated, alkaline phosphatase treated, kinase treated, or alkaline phosphatase followed by kinase treated RNA was mixed with 200 pmol of the acceptor RNA oligo (>6-fold molar excess) and incubated overnight at room temperature with 20 units of T4 RNA ligase (New England Biolabs) in the supplied 1X buffer supplemented with 40% PEG-400. The ligations were phenol/chloroform extracted, ethanol precipitated, and resuspended in 1X reverse transcription buffer (50 mM KCl, 20 mM Tris, pH 8.5, 0.5 mM EDTA, 8 mM $MgCl_2$). Poisoned primer extension reactions were performed on each population of ligated RNA (1.0 μg of mtRNA or 0.1 μg of gRNA) with 100 kCPM (~ 50 fmol) of primer in the presence of 1.6 mM each dATP, dCTP, dTTP and ddGTP as

described previously (8). Reactions were resolved on an 8%, 8M urea acrylamide sequencing gel. Gels were fixed in 7% acetic acid + 7% methanol, dried, and exposed on a phosphorimager cassette (Molecular Dynamics). For each gel, the signal for the ligated and unligated products was quantified using ImageQuant software (Molecular Dynamics) and the percent ligation efficiency was represented as the amount of ligated product/(ligated product + unligated product)*100.

RESULTS

gMURF2-II maxicircle localization and expression.

In *T. brucei*, the gMURF2-II gene is located within the ND4 coding region downstream of one of three mapped regions of intergenic space on the maxicircle (Figure 1). Although the position of the gMURF2-II gene is conserved in *L. tarantolae* and *C. fasciculata*, in these organisms transcription initiates within an intergenic region upstream of ND4. The gMURF2-II transcript has been detected in both *Leishmania* and *Crithidia* as a discrete RNA (1, 2) and has been shown to be a primary transcript in *Leishmania* (3). Conservation of the gMURF2-II gene in *T. brucei* indicates the possibility of a promoter element within the coding region of maxicircle genome in spite of the extensive overlap between the gRNA and mRNA genes.

Linearized map of the coding region of the *T. brucei* maxicircle.

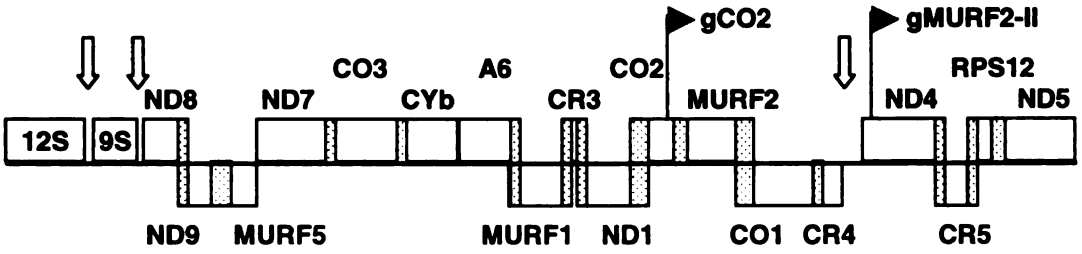


Figure 1 Major and minor strand genes are represented above or below the line, respectively. Overlapping regions are stippled. Intergenic spaces are indicated by white arrows. The gRNA sequences are represented by black flags. MURF, Maxicircle Unidentified Reading Frame; CR, "C" – rich; CO, cytochrome oxidase; A6, ATPase 6; ND, NADH dehydrogenase; CYb, apocytochrome b; RPS12, ribosomal protein subunit 12.

FIGURE 1

In order to confirm that the gMURF2-II RNA was indeed expressed in *T. brucei*, we performed a Northern blot on isolated mitochondrial RNA (mtRNA) using an oligonucleotide probe complementary to both the gMURF2-II gRNA as well as the 5' end of the ND4 message. As shown in Figure 2A (lanes 1-3), the gMURF2-II probe hybridizes to a small RNA of about 60-70 nt representing the gMURF2-II RNA, as well as to the ~1.3 kb ND4 transcript. The gRNA signal detected by the gMURF2-II probe is similar in size to the minicircle encoded gRNA, gA6(14), as seen in the blot probed with an oligo complementary to gA6(14) in lane 4.

Since the majority of the gRNAs are encoded in minicircle DNA in *T. brucei*, we performed a Southern blot of *T. brucei* DNA in order to confirm the maxicircle location of the gMURF2-II gene. As seen in the left panel in Figure 2B, the gMURF2-II probe hybridizes exclusively to the predicted maxicircle fragments and not to the ~1 kb minicircle DNA. In contrast, a probe against a minicircle encoded gRNA, gA6(14), hybridizes exclusively to the predicted minicircle fragments (15) (right panel, Fig. 2B). These results confirm that no additional copies of the gMURF2-II gene exist in the kinetoplast genome and that this gRNA must be transcribed from this region of maxicircle DNA.

Figure 2. Expression of the gMURF2-II gene. (A) Northern blot of 1 μg (lane 1), 2.5 μg (lane 2), 5 μg (lane 3), or 10 μg (lane 4)of mitochondrial RNA probed with an oligonucleotide complementary to the gMURF2-II (lanes 1-3) or an oligonucleotide complementary to the minicircle gRNA, gA6(14) (lane 4). The gMURF2-II probe detects both the ~65 nt gRNA transcript as well as the ~1.3 kb ND4 mRNA, while the gA6(14) probe hybridizes to a gRNA band only. (B) Restriction digest and Southern blot analysis of *T. brucei* genomic DNA. 3 μg of DNA was digested with *HindIII* (lane 1), *HindIII* + *EcoRI* (lane 2), or *EcoR1* (lane 3), separated on a 1% agarose gel, transferred to a nytran membrane and probed with an end-labeled gMURF2-II oligo (left panel). The blot was then stripped and re-hybridized with an end-labeled gA6(14) probe (right panel). (C) Restriction maps of the *T. brucei* maxicircle and the gA6(14) minicircle (*HindIII*, H; *EcoR1*, E).

gMURF2-II RNA expression and maxicircle localization

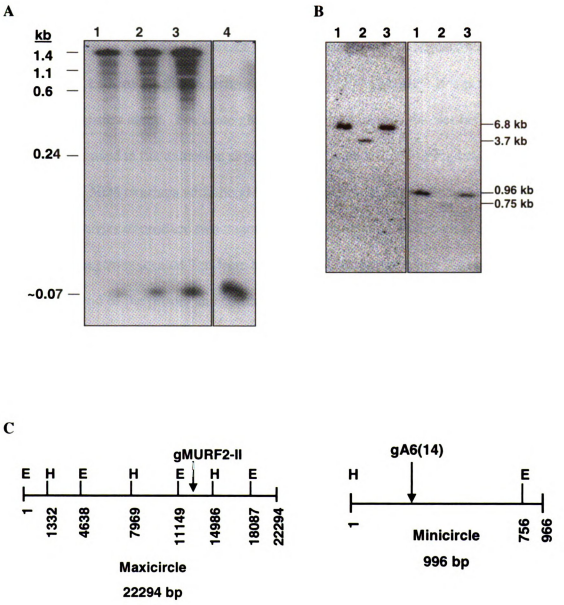


FIGURE 2

The gMURF2-II gene is located within the ND4 gene downstream of an intergenic region.

In order to determine the exact location of the gMURF2-II gene, we mapped both the 5' and 3' ends of the gRNA transcript. The 5' end of gMURF2-II was mapped using a primer extension assay. The same gMURF2-II probe used in the Northern blot in Figure 2 was used in the extension experiment and is underlined in Figure 3A. Because the 5' end of ND4 overlaps with the gMURF2-II gRNA sequence, this probe hybridizes to both transcripts to produce two major extension products. As shown in Figure 3B, the smaller of the two is located 7 nt from the 3' end of the primer while the larger is 13 nt from the 3' end of the primer and corresponds to the mapped 5' end of ND4 (9). Two additional minor products of 17 nt and 19 nt are also visible (Figure 3B, lane 1). In order to determine which of these bands corresponded to the 5' end of gMURF2-II, the primer extension analysis was performed on gel purified gRNAs (Figure 3B, lane 2). In this lane, only the 7 nt product is present, demonstrating that this band represents the 5' end of the gMURF2-II transcript while the two larger products in lane 1 represent alternate, minor mRNA 5' ends. The maxicircle sequence within this region and the location of the mapped 5' ends of the major and minor strand products are shown in Figure 3A. This result demonstrates that the 5' end of the gMURF2-II gene is located downstream of the 5' end of the ND4 gene, within the 5' UTR.

The 3' end of the gMURF2-II gRNA was mapped using 3' RACE. cDNA was synthesized from mtRNA using a 5' tagged oligo-d(A) primer which hybridizes to the poly(U) tail common to the 3' end of all gRNAs. The PCR product obtained using a gMURF2-II gene specific primer and a tag-specific primer was sequenced to identify the

Figure 3. (A) Nucleotide sequence of the gMURF2-II locus including the mapped 5' ends of gMURF2-II and ND4 transcripts as well as the minor strand CR4 gene (24). The 5' end of the gRNA was determined by primer extension analysis using the gMURF2-IIR primer indicated by the arrow underneath the sequence. The 3' end of the gRNA indicated was obtained by sequencing RT-PCR products. The start codon for ND4 is underlined. (B) Primer extension mapping of gMURF2-II and ND4 5' ends. End labeled gMURF2-IIR primer was extended in the presence of 1 μg mitochondrial RNA (lane 1) or gel-purified guide RNA (lane 2). Nucleotide sequence was obtained via primer extension of total mitochondrial RNA in the presence of ddA, or ddT (lanes 3 and 4, respectively). Extension products were run on an 8% acrylamide-urea gel. The 5' ends of the gRNA and mRNA are indicated by arrows. Two minor mRNA products are marked with asterisks.

Mapping the 5' ends of gMURF2-II and ND4

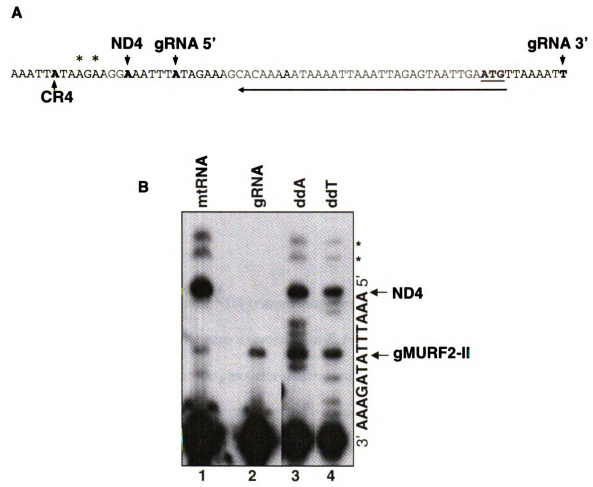


FIGURE 3

3' end of the gRNA(data not shown). As seen for many gRNAs, the 3' end of the gMURF2-II ends in a short stretch of T's. (Indicated in the sequence shown in Figure 3A) (28). Since gRNAs are post-transcriptionally modified by the addition of a short, poly(U) tail (3), it is unclear whether the 3' end of this gRNA is generated via transcription termination at this site or processing of a larger transcript. Together, the results of the 5' and 3' mapping of the gMURF2-II gRNA demonstrate that the gRNA is encoded completely within the ND4 gene. Although it is common for maxicircle genes to overlap in *T. brucei*, this is the first case in which discrete intragenic transcript has been detected.

The mature gMURF2-II and ND4 transcripts are generated via different events.

Studies of mitochondrial gene expression in *T. brucei* have shown that the maxicircle is transcribed as a polycistronic precursor which is rapidly processed into mature transcripts (19). Due to the close packing and overlap of genes on the maxicircle, the generation of mature transcripts often involves mutually exclusive events (17, 23). According to this model of polycistronic transcription, the intragenic location of the gMURF2-II gRNA suggests that either the gMURF2-II or the ND4 transcript could be made from a given precursor RNA, but not both. However, minicircle gRNAs in *T. brucei* are primary transcripts as are both the minicircle and maxicircle gRNAs from *Leishmania* and *Crithidia*. Consequently, gMURF2-II may in fact be a primary rather than a processed transcript. In addition, if a promoter does exist within the intergenic region upstream of the gMURF2-II/ND4 region, it is possible that ND4 and/or other downstream mRNAs are primary transcripts.

In order the distinguish among these possibilities, we employed a modified version of a technique described by Bruderer *et. al.* (4) to characterize the structure of the 5' ends of gMURF2-II and ND4. Mitochondrial transcripts do not possess the 5' cap structure seen in most nuclear RNAs. Consequently, primary transcripts possess tri-phosphate 5' ends while processed transcripts have monophosphate 5' ends. This technique takes advantage of the specificity of the enzyme T4 RNA ligase to ligate an RNA oligo to the 5' end of transcripts bearing a monophosphate moiety. Transcripts with a di- or tri-phosphate or hydroxyl group at the 5' end will not serve as substrates for this enzyme. Thus, a processed transcript will ligate to an exogenous RNA linker oligo, while primary transcripts will not. The negative control consists of a phosphatase treatment that results in all transcripts bearing unligatable 5' OH ends. Subsequent treatment of phosphatased RNAs with kinase results in 5' monophosphorylated molecules that serve as the positive control for ligation. It is the comparison of the ability of the various treated samples to ligate to the RNA oligo that allows one to infer the nature of the 5' end of the molecule. Primary transcripts, by nature of their polyphosphate ends, will remain unligated in all but the positive control, while processed transcripts will ligate in all but the negative control. Figure 4A shows the predicted ligation outcome for each treatment for a processed and a primary transcript.

We subjected mitochondrial RNA or gel-purified guide RNA to the various treatments and incubated them in the presence of T4 RNA ligase and a 26 nt synthetic RNA oligo containing a single C residue 17 nt from the 3' end (Figure 4B). We then performed a poisoned primer extension analysis in the presence of ddGTP to compare the

Figure 4. Ligation dependent poisoned primer extension assay on total mitochondrial RNA and gel purified gRNA. (A) Table of the expected ligation outcomes for processed and primary transcripts subjected to various enzymatic treatments. Treatments: U, untreated; K, kinase; AP, alkaline phosphatase; AP/K, alkaline phosphatase followed by kinase; no ligase, untreated RNA not subjected to ligation. (B) Nucleotide sequence of the synthetic RNA oligo (lowercase italics) and the 5' end of ND4/gMURF2-II (capital letters). The 3' ends of the primers used in the extension reactions are underlined and the sizes of the expected products are indicated by arrows. (C) Primer extension products were electrophoresed on 8% acrylamide-urea gels. The 7nt unligated gMURF2-II product is visible in all treatments but the positive control (AP/K), while the 13 nt unligated ND4 product is present only in the negative control (AP). (D) Nucleotide sequence of the synthetic RNA oligo (lowercase italics) and the 5' end of gND7-506 (capital letters). The minicircle e encoded gND7(506) remains unligated in all but the positive control (AP/K). Asterisks in (C) indicate the minor mRNA 5' ends seen in FIGURE 2A, and the corresponding ligation products.

Ligation-dependent poisoned primer assay for gMURF2-II and ND4

A

	Ligation? (Y/N)				
	U	**AP**	**K**	**AP+K**	**No ligase**
Processed	Y	N	Y	Y	N
Primary	N	N	N	Y	N

B

5' agauuuugaCagugauauugauaauua AAAUUUAUAGAAAGCACAA. . .

30 nt -mRNA
24 nt - gRNA
13 nt mRNA
7 nt gRNA

C

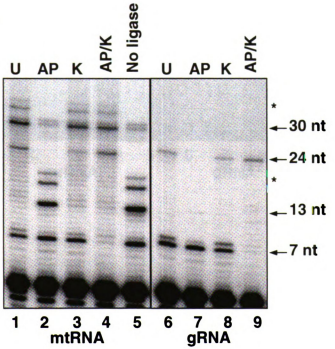


FIGURE 4

Ligation dependent poisoned primer assay for gND7-506

D

5'gND7-506 RNA

*5' agauuuuga***C***agugauaugauaauua* **A**AAUACGAUGU. . .

22 nt **5 nt**

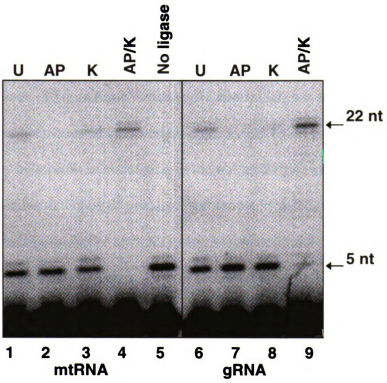


FIGURE 4 (cont'd)

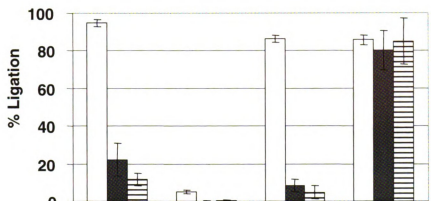
ability of various transcripts to ligate to the RNA oligo. This assay was sensitive enough to allow us to directly analyze the ligation results without the subsequent PCR step described in the original technique. A representative gel is shown in Figure 4C. As expected, neither the gRNA nor the mRNA ligates in the negative control sample treated with alkaline phosphatase (Figure 4C, lane 2). The predominant bands in this lane represent the 7nt gMURF2-II 5' end and the 13 nt ND4 5' end. The larger bands seen at 17 and 19 nt represent the minor mRNA 5' ends seen in Figure 3B. An identical result was obtained in the no ligase negative control (Figure 4C, lane 5). The faint doublet seen at ~30 nt in these two negative control lanes is difficult to explain. These bands were not observed in the original primer extension mapping of 5' ends (Figure 3B), and may be a result of a variation in the mitochondrial isolation procedure. In order to optimize detection of precursor transcripts, isolated mitochondria were incubated in transcription buffer and rNTPs for 30 minutes before RNA isolation (12). Consequently, this band may represent a minor precursor product that is typically not detected, and this signal was subtracted from subsequent quantifications.

In contrast to the negative control lanes, both the ND4 and the gMURF2-II RNAs do ligate in the samples that were kinased following phosphatase treatment (Figure 4C, lane 4) as evidenced by the ddGTP stops seen at 24 nt and 30 nt for the gRNA and mRNA, respectively, as well as the minor mRNA ligated bands seen at 34 and 36 nt. Significantly, virtually no unligated gRNA or mRNA remain in this positive control lane, indicating the efficiency of the enzymatic steps required for this experiment. Likewise, the 13 nt unligated ND4 mRNA band is not visible in the untreated and kinase-only lanes (lanes 1 and 3, respectively), although the ligated 30 nt mRNA band is present in these

lanes. This indicates that the majority of the ND4 mRNA possesses a 5' monophosphate at the 5' end as expected for a transcript processed from a polycistronic precursor. In contrast, the unligated 7 nt product corresponding the 5' end of gMURF2-II is present all lanes but the positive control, demonstrating that the majority of the gMURF2-II population possesses an unligatable triphosphate end expected for a primary transcript. The result for the guide RNA is more clearly visible in the right half of the panel in which the treatments, ligation, and primer extension were performed on gel-purified gRNA. Again, the 7 nt product corresponding to the 5' end of the unligated gRNA is the predominant product in all lanes but the positive control (Figure 4C, lane 9) in which only the ligated band is visible.

Although the majority of the gMURF2-II remains unligated in the untreated lanes, some ligation was observed for this treatment (Figure 4C). Consequently, to better compare the ligatability of the gMURF2-II and ND4 transcripts, the signals corresponding to the ligated and unligated products were quantified, and the ligation efficiency for each RNA was calculated. Comparison of these values shows that ND4 has a ligation efficiency > 80% in all lanes but the negative control, while the gRNA ligates appreciably only in the positive control lane (Figure 5). In addition, to confirm that the pattern obtained for gMURF2-II was similar to that seen for other mitochondrial primary transcripts, we performed this experiment using a primer complementary to the minicircle-encoded gRNA, gND7(506) (16). Minicircle gRNAs have been shown to be primary transcripts by their ability to be capped *in vitro* by the enzyme guanylyl transferase (21). As seen for gMURF2-II, the gND7(506) RNA does not serve as an efficient substrate for T4 RNA ligase unless first treated with alkaline

Relative Ligation Efficiencies



	U	AP	K	AP+K
ND4	94	2	85	84
gMURF2-II	22	0.1	8	80
gND7-506	11	0.5	5	85

FIGURE 5. Calculated RNA ligation efficiencies. Phosphor images of replicate gels were analyzed using ImageQuant software (Molecular Dynamics). Percent ligation for an RNA in a given treatment was calculated as ligated product/(ligated + unligated product) X 100. The percent ligation for each transcript in each treatment is shown in the bar graph and the values are shown below. ND4 values were calculated using only the major unligated and ligated products. Error bars represent the standard deviation obtained from the replicate gels.

FIGURE 5

phosphatase followed by kinase treatment (Figure 4D). The 5 nt band corresponding to the mapped 5' end of the unligated gND7(506) RNA is the predominant band in all lanes but the AP/K positive control (Figure 4D, lanes 4 and 9) in which only the 22 nt ligated product is visible. Likewise, the ligation efficiency calculated for gND7(506) for each treatment is similar to that observed for gMURF2-II (Figure 5).

Together, these results indicate that the majority of the gRNA populations tested contain a triphosphate 5' end. However, it is clear from the small amount of ligated gRNA seen in the untreated lanes that a minor population does contain a monophosphate 5' end. The 5' phosphates of in vitro transcribed molecules have been observed via NMR to be labile under normal storage and handling conditions resulting in a primary transcript population containing a mixture of tri-, di-, and monophosphates (Dr. Charles Hoogstraten, personal communication). However, we cannot rule out the possibility that the minor population of ligatable gRNAs in the untreated lane is a result of a 5' monophosphate generated by an endogenous mitochondrial ribonuclease. Despite the presence of this small population, the pattern of ligatability for the processed mRNA is clearly distinct from that of the primary gRNA transcripts, visibly reflecting the difference in the structure of their 5' ends.

An atypical minicircle gRNA not encoded within a cassette is also a primary transcript but shows heterogeneity of start site selection.

The gene for the gND7(506) gRNA is located between 18 bp imperfect inverted repeats, as are most minicircle gRNA genes. Each minicircle has on average 3 of these cassettes (22). Transcription of these minicircle gRNAs initiates a fixed distance

downstream of the 3' end of the upstream repeat, leading to speculation that the repeat sequences play a role in transcription. In addition, the stability of the resulting gRNA seems to depend upon transcription initiation at this location as downstream gRNAs transcribed when the mitochondrial RNA polymerase reads through the upstream cassette are degraded (10). However, some gRNA cassettes lack functional gRNA genes while other gRNA genes are not located within cassettes (14). The gRNAs that mediate the initial editing of the CYb message are not located within inverted repeats. These gRNAs (gCYb(558) and gCYb(560A/B) are redundant but not identical and are located on closely related minicircles that each also contain three gRNA cassettes (24). Interestingly, the analogous gRNA identified in both *Leishmania* and *Crithidia* (gCYb-I), is maxicircle encoded (2, 29). The *T. brucei* gCYb gRNAs have previously been shown to have heterogeneous 5' ends as determined in primer extension analysis (24). Likewise, these transcripts also have heterogeneous 3' ends that can extend into the downstream forward repeat (24), indicating that some read-through transcription occurs. This 5' and 3' heterogeneity could be due to lack of precise transcription initiation and termination signals and/or the result of inexact processing of this RNA from a precursor generated via read-through from the upstream gRNA cassette. In order to determine whether the unusual gCYb(560) gRNA is a primary transcript, we performed a similar ligation-dependent primer extension using a primer complementary to gCYb(560). As seen in Figure 6, poisoned primer extension analysis does not result in a single band, but rather a series of bands of ~13 nt for the unligated gRNA reflecting the 5' heterogeneity previously observed for these gRNAs. This primer is specific for gCYb(560).

FIGURE 6. Ligation-dependent poisoned primer extension analysis of an atypical minicircle-encoded gRNA. (A) Nucleotide sequence of the synthetic RNA oligo (lowercase italics) and the 5' ends of the redundant gCYb(560) and gCYb(558) gRNAs (capital letters). The gRNA sequence includes the most common 5' ends (24), and sequence corresponding to the 3' end of the primer used for the primer extension reaction is underlined. (B) 8% acrylamide-urea gel of the primer extension products. The arrows indicate the most common 5' ends shown in (A). The asterisk indicates a 7 nt synthetic RNA oligo-dependent artifact resulting from partial hybridization of the 3' end of the primer to the RNA oligo during the extension reaction. (C) Calculated ligation efficiency for gCYb(560) as described in Figure 5.

Ligation-dependent poisoned primer extension assay for gCYb(560)

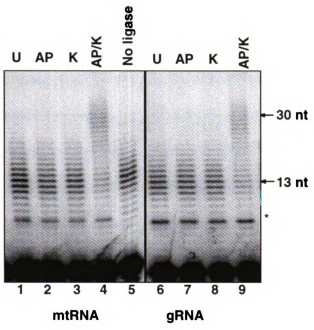
A

gCYb(560)

*5' agauuuuga**C**agugauaugauaauua* **G**GAGAUAGUAAAAGACAAUGUAGAUUUC. . .3'

gCYb558 5' GGGAGAUUAAAAGACAAUGUGAAUUUU. . . 3'

B



C

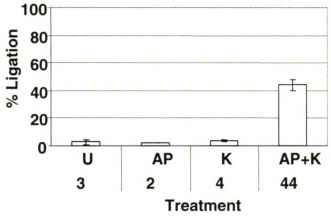


FIGURE 6

However, due to the similarity in sequence among the redundant CYb gRNAs, this primer can also anneal, albeit less efficiently, to the gCYb558 RNA (Figure 6A). The distance from the 3' end of the primer to the most common mapped 5' ends of these gRNAs is very similar (difference of 1 nt) and does not account for the extent of the observed heterogeneity. From these experiments, it is clear that the RNA ligation-dependent pattern shown in Figure 6 matches that seen for the maxicircle encoded gMURF2-II as well as the minicircle cassette encoded gND7(506) in that the gRNAs remain unligated in all but the positive control lane. In addition, the heterogeneity of the ligated product demonstrates that the heterogeneity seen at the 5' end of gCYb(560) is the result of imprecise transcription start site rather than processing. These results, taken together with the maxicircle encoded gMURF2-II result, indicate that mature gRNAs in *T. brucei* are primary transcripts regardless of their kDNA location and suggest important implications for potential transcription regulation of gRNA expression.

DISCUSSION

Although there are a core set of genes retained by all mitochondria, mitochondrial genomes vary dramatically in size, gene content, and organization (5). The *T. brucei* mitochondrial genome is unusual in that two distinct classes of transcripts are localized to two completely different genome components. The maxicircle component encodes primarily the ribosomal RNA and protein genes homologous to those encoded in the mtDNA of other organisms. There is little room for intergenic regulatory elements within this compact genome, and the rRNA and mRNA genes are transcribed as part of a polycistronic precursor originating within the variable region (19). Interestingly, the

maxicircle genomes of the related kinetoplastid organisms *Leishmania* and *Crithidia* have more potential for regulatory regions and encode additional transcription units in the form of multiple gRNA genes (2, 29). Analysis of sequences surrounding these maxicircle gRNA genes does not reveal an obvious consensus promoter element (data not shown).

We have demonstrated that one of the two conserved maxicircle encoded gRNA genes in *T. brucei*, gMURF2-II, represents an individual transcription unit located within the coding region of the maxicircle. This is the only defined maxicircle transcription initiation site reported for *T. brucei*. Its location is interesting for two reasons. The gRNA gene is contained entirely within the coding region of an mRNA that is processed from a polycistronic precursor. This is the first report of a discrete, primary transcript generated from a completely intragenic kinetoplast gene. Additionally, the gMURF2-II locus is downstream of one of only 3 intergenic regions on the maxicircle genome. Given the tight packing and in most cases overlap of adjacent genes, the retention of this intergenic region indicates a potential regulatory function. The identification of a primary transcript within the coding region of the *T. brucei* maxicircle gives us a small, stable transcript to measure in *in vitro* transcription assays as we work to define promoter elements in *T. brucei* kDNA. Intriguingly, replication bubbles have also been observed in electron microscopy studies near this region of the maxicircle (6). This is particularly interesting in light of evidence linking transcription to maxicircle maintenance in *T. brucei* (11).

Analysis of the structure of the 5' ends of the gMURF2-II and ND4 transcripts indicated that they are generated via distinctly different events. While the ND4 RNA must be precisely processed from the polycistronic precursor, the gMURF2-II RNA is

initiated *de novo* a mere 7 nt downstream of the ND4 5' end and is destined for termination and/or poly-uridylation 50 bp downstream from the gRNA start site. Although there are two minor ND4 processing sites upstream of the major 5' end of the mRNA, we did not detect any smaller minor products using an ND4 mRNA-specific primer located downstream of the 3' end of the gRNA (data not shown). This result suggests a complex model for maxicircle transcription in which distinct complexes may be involved in the specific generation of the gRNA and mRNA.

The majority of gRNAs in *T. brucei* are encoded on the distinct minicircle genome unique to these organisms, and are located between inverted imperfect 18 bp repeats that have been implicated as playing a role in transcription (21, 22). Our results indicate that the unusual minicircle-encoded gCYb(560) gene also produces a mature primary transcript, demonstrating that minicircle gRNA genes located outside of these repeats also function as independent transcription units. The heterogeneity at the 5' and 3' ends of gCYb(560) previously reported and detected in this study indicates that while the repeats are not essential for minicircle gRNA expression, they may contain signals for precise start site selection and termination. A recent analysis of minicircle genomes predicts the presence of other gRNA genes located outside of inverted repeats (14). It will be interesting to determine if these genes are indeed expressed and if so whether they similarly have heterogeneous ends.

Taken together, the data presented here indicate that gRNA genes in *T. brucei*, regardless of genome location, are individual transcription units. Although a common gRNA promoter sequence is not obvious, these results indicate that a gRNA specific determinant of transcription exists within kDNA and may reflect an as of yet unidentified

gRNA transcription factor. The unique location of the intragenic transcription unit for the gMURF2-II gRNA suggests the possibility of a distinct gRNA transcription complex as well as a maxicircle polycistronic transcription complex or a single complex capable of both activities. *In vitro* and/or *in organelle* transcription experiments using this maxicircle locus will help to distinguish between these possibilities and to identify such putative gRNA transcription factors.

LITERATURE CITED

1. **Arts, G. J., H. van der Spek, D. Speijer, J. van den Burg, H. van Steeg, P. Sloof, and R. Benne.** 1993. Implications of novel guide RNA features for the mechanism of RNA editing in *Crithidia fasciculata*. Embo J **12:**1523-32.

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

3. **Blum, B., and L. Simpson.** 1990. Guide RNAs in kinetoplastid mitochondria have a nonencoded 3' oligo(U) tail involved in recognition of the preedited region. Cell **62:**391-7.

4. **Bruderer, T., L. C. Tu, and M. G. Lee.** 2003. The 5' end structure of transcripts derived from the rRNA gene and the RNA polymerase I transcribed protein coding genes in *Trypanosoma brucei*. Mol Biochem Parasitol **129:**69-77.

5. **Burger, G., M. W. Gray, and B. F. Lang.** 2003. Mitochondrial genomes: anything goes. Trends Genet **19:**709-16.

6. **Carpenter, L. R., and P. T. Englund.** 1995. Kinetoplast maxicircle DNA replication in *Crithidia fasciculata* and *Trypanosoma brucei*. Mol Cell Biol **15:**6794-803.

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

8. **Clement, S. L., and D. J. Koslowsky.** 2001. Unusual organization of a developmentally regulated mitochondrial RNA polymerase (TBMTRNAP) gene in *Trypanosoma brucei*. Gene **272:**209-18.

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

10. **Grams, J., M. T. McManus, and S. L. Hajduk.** 2000. Processing of polycistronic guide RNAs is associated with RNA editing complexes in *Trypanosoma brucei*. Embo J **19:**5525-32.

11. **Grams, J., J. C. Morris, M. E. Drew, Z. Wang, P. T. Englund, and S. L. Hajduk.** 2002. A trypanosome mitochondrial RNA polymerase is required for transcription and replication. J Biol Chem **277:**16952-9.

12. **Harris, M. E., D. R. Moore, and S. L. Hajduk.** 1990. Addition of uridines to edited RNAs in trypanosome mitochondria occurs independently of transcription. J Biol Chem **265:**11368-76.

13. **Hauser, R., M. Pypaert, T. Hausler, E. K. Horn, and A. Schneider.** 1996. In vitro import of proteins into mitochondria of *Trypanosoma brucei* and *Leishmania tarentolae*. J Cell Sci **109 (Pt 2):**517-23.

14. **Hong, M., and L. Simpson.** 2003. Genomic organization of *Trypanosoma brucei* kinetoplast DNA minicircles. Protist **154:**265-79.

15. **Jasmer, D. P., and K. Stuart.** 1986. Sequence organization in African trypanosome minicircles is defined by 18 base pair inverted repeats. Mol Biochem Parasitol **18:**321-31.

16. **Koslowsky, D. J., G. J. Bhat, A. L. Perrollaz, J. E. Feagin, and K. Stuart.** 1990. The MURF3 gene of *T. brucei* contains multiple domains of extensive editing and is homologous to a subunit of NADH dehydrogenase. Cell **62:**901-11.

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

18. **Madison-Antenucci, S., J. Grams, and S. L. Hajduk.** 2002. Editing machines: the complexities of trypanosome RNA editing. Cell **108:**435-8.

19. **Michelotti, E. F., M. E. Harris, B. Adler, A. F. Torri, and S. L. Hajduk.** 1992. *Trypanosoma brucei* mitochondrial ribosomal RNA synthesis, processing and developmentally regulated expression. Mol Biochem Parasitol **54:**31-41.

20. **Myler, P. J., D. Glick, J. E. Feagin, T. H. Morales, and K. D. Stuart.** 1993. Structural organization of the maxicircle variable region of *Trypanosoma brucei*: identification of potential replication origins and topoisomerase II binding sites. Nucleic Acids Res **21:**687-94.

21. **Pollard, V. W., and S. L. Hajduk.** 1991. *Trypanosoma equiperdum* minicircles encode three distinct primary transcripts which exhibit guide RNA characteristics. Mol Cell Biol **11:**1668-75.

22. **Pollard, V. W., S. P. Rohrer, E. F. Michelotti, K. Hancock, and S. L. Hajduk.** 1990. Organization of minicircle genes for guide RNAs in *Trypanosoma brucei*. Cell **63:**783-90.

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

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

25. **Shapiro, T. A., and P. T. Englund.** 1995. The structure and replication of kinetoplast DNA. Annu Rev Microbiol **49:**117-43.

26. **Simpson, L., S. Sbicego, and R. Aphasizhev.** 2003. Uridine insertion/deletion RNA editing in trypanosome mitochondria: a complex business. Rna **9:**265-76.

27. **Stuart, K., and A. K. Panigrahi.** 2002. RNA editing: complexity and complications. Mol Microbiol **45:**591-6.

28. **Thiemann, O. H., and L. Simpson.** 1996. Analysis of the 3' uridylylation sites of guide RNAs from *Leishmania tarentolae*. Mol Biochem Parasitol **79:**229-34.

29. **van der Spek, H., G. J. Arts, R. R. Zwaal, J. van den Burg, P. Sloof, and R. Benne.** 1991. Conserved genes encode guide RNAs in mitochondria of *Crithidia fasciculata*. Embo J **10:**1217-24.

CHAPTER III

Unusual organization of a developmentally regulated mitochondrial RNA polymerase (TBMTRNAP) gene in *Trypanosoma brucei*

INTRODUCTION

The developmental regulation of mitochondrial gene expression is a critical part of the complex *Trypanosoma brucei* life cycle and occurs through the coordinated expression of both the nuclear and mitochondrial genomes (20). Mitochondrial function in *T. brucei* is highly regulated, with expression of a complete electron transport chain restricted to the procyclic (insect) stage of the life cycle. This regulation is complex and relies on the guide RNA-mediated editing of maxicircle mRNAs to form mature transcripts (25). Consequently, the expression of functional mitochondrial-encoded proteins involves not only the transcription of ribosomal and protein coding genes localized to the maxicircle DNA, but also the transcription of hundreds of small guide RNAs genes present on both maxicircles and minicircles. Although there has been extensive research into the process of RNA editing in *T. brucei*, the precise mechanism of regulation of mitochondrial activity remains unknown as is the nature of the mitochondrial transcription machinery.

The mitochondrial genome of a wide variety of organisms including plants, animals, and protozoa, is transcribed by a nuclear-encoded mitochondrial RNA polymerase belonging to a family of single-subunit RNA polymerases. These polymerases share significant homology to the bacteriophage T7 RNA polymerase (2, 26). Unlike the single-subunit T7 RNA polymerase, the catalytic core enzyme of the mitochondrial RNA polymerase requires at least one specificity factor for accurate and efficient transcription initiation (26). Despite the conservation of the core enzyme, the nature and number of these MTRNAP specificity factors vary among organisms.

To further our investigation of the role of mitochondrial transcription in the developmental regulation of mitochondrial gene expression in *T. brucei*, we have cloned and characterized the gene for the mitochondrial RNA polymerase (*TBMTRNAP*) in this organism. The full-length cDNA includes the *TBMTRNAP* open reading frame (ORF) of 3822 bp corresponding to 1274 amino acids. The expression of the *TBMTRNAP* gene is regulated during the *T. brucei* life cycle with a greater relative abundance of the *TBMTRNAP* mRNA in the procyclic form than in the bloodstream form trypanosomes. This regulation is quite complex and appears to involve both alternative polyadenylation of the *TBMTRNAP* mRNA as well as stage specific differences in the stability of the resulting transcripts.

MATERIALS AND METHODS

Cell culture and nucleic acid preparation

T. brucei brucei procyclic form cells (Istar clone from stock EATRO 164) were grown as previously described (11). Total genomic DNA was isolated by digestion of cells with Proteinase K in the presence of 1% SDS followed by RNase digestion and phenol/chloroform extraction (13). Total RNA was isolated as previously described (11). Poly(A+) RNA was purified from total RNA with oligo-dT polystyrene-latex resin using the Oligotex kit from Qiagen according to the recommendations of the manufacturer.

Degenerate PCR

COnsensus-**DE**generate **H**ybrid **O**ligonucleotide **P**rimers were designed with the aid of the web-based program CODEHOP (22). These primers were generated from two blocks of highly conserved amino acids produced from a multiple alignment of the

carboxyl termini of seven mitochondrial RNA polymerase genes and gene fragments (Figure 1). Each CODEHOP primer was designed using the *T. brucei* codon bias and consists of a degenerate 3' core region and a non-degenerate 5' consensus clamp. PCR amplification was performed in a 50 µl reaction volume with 5 µg of sheared genomic DNA, 1 µM of each primer, 0.25 mM each dNTP, and 2.5 U of Taq polymerase (Promega). Reactions were carried out in a thermocycler (MJ Research, Inc PTC-100) using the touchdown protocol recommended for the CODEHOP PCR. The resulting PCR product was sequenced from both ends using the non-degenerate clamp portion of the CODEHOP primers.

Genomic DNA cloning and gene copy number

A high-density microarrary filter (SM12#58) from the *T. brucei* TREU927/4 bacteriophage P1 genomic library was kindly provided by Vanessa Leech of the Laboratory for Parasite Genome Analysis at Cambridge. The filter was probed with a 40 nt high specific activity antisense riboprobe transcribed using the Uhlenbeck single-stranded T7 transcription protocol (18). Briefly, a template for transcription by T7 RNA polymerase (Ambion) was created by hybridizing the 61 nt oligonucleotide T7RP-1, (5'AATCACTTGG TATTACGTTT CGCGAGACGC AAAACTGCCC TATAGTGAGT CGTATTAAAT T3') designed using sequence obtained from the degenerate PCR product with a 22 nt T7 promoter oligo (5'AATTTAATAC GACTCACTAT AG3') to create a double stranded T7 promoter. A total of 5 positive P1 clones (1E5, 1G1, 2D11, 16G11, 5E12) were identified and ordered from the facility. P1 DNA was obtained from each clone using the Qiagen Midiprep kit following modifications suggested by the manufacturer for isolation of large constructs. P1 DNA

was digested with KpnI, EcoRI, BamHI and all pairwise combinations and resolved on a 0.75% agarose gel. The resolved DNA was transferred to Hy-bond Nytran membrane using downward capillary transfer and probed with the same riboprobe used to identify the clones. Positive fragments were subcloned into pBluescript SK- (Stratagene), propagated in *E. coli* DH5 α F' IQ (Gibco BRL) cells and sequenced. The copy number for the mitochondrial RNA polymerase gene was determined by probing a Southern blot of a restriction digest of 10 μg of *T. brucei* genomic DNA as described for the P1 clones.

cDNA cloning

The 3'end of the putative *T. brucei* mitochondrial RNA polymerase cDNA was obtained by a rapid amplification of cDNA ends (RACE) strategy using double-stranded adapter-ligated cDNA created with the Marathon cDNA Amplification Kit (Clontech) following the manufacturer's suggestions. Subsequent PCR reactions used the Clonetech Adapter Primer AP1 and a gene-specific primer 3'R1 (5'GGTATTACGT TTCGCGAGAC GCAAAACTGC CG3') with the following conditions: 94°C for 1 min, 5 cycles of 94°C for 30 sec, 72°C for 4 min, 5 cycles of 94°C for 30 sec, 70°C for 4 min, 20 cycles of 94°C 30 sec and 68°C for 4 min with a hold at 72°C for 10 min. Products were resolved on a 1% agarose gel, subcloned into the Clonetech vector, and sequenced.

The 5' end of the polymerase cDNA was determined using reverse transcriptase, PCR (RT-PCR). Three gene-specific, antisense primers were designed downstream of the first ATG in the *TBMTRNAP* open reading frame. Primer CS-1 (5'CGTAGAACCC GCAGATGATT CTGAGGC3'), CS-2 (5'TGTGAAGCAA GTGATAGCAA ACTACCCCGC TG3'), and 5'RT-1, (5'CTTTATCGGT GCCAGACTCA ACC3'). For each RT-PCR, cDNA was reverse transcribed from 1 μg of Poly(A+) RNA, using 40

fmol of gene-specific primer, 12.5 μM each dNTP, and 2 Units AMV reverse transcriptase (Seikagaku) in a 20 μl reaction volume. RNA and primer were denatured for 2 min at 70°C and cooled at a rate of 2°C /min to 50°C at which point dNTPs and enzyme were added and the reaction was incubated at 50°C for an additional 45 minutes. PCR was performed on 5.0 μl of each RT reaction with 0.2 μM each of the appropriate gene-specific primer and the spliced leader primer, TbSL-sal, (5'GTCGTCGACA ACTAACGCTA TTATTAGAACAG3'), 0.05 mM each dNTP, 2.5 U Taq (Promega) with initial denaturation at 94°C for 2 minutes followed by 30 cycles of 94°C for 30 sec, 60°C for 30 sec, 72°C for 30 seconds and a hold at 72°C for 10 minutes. Products were resolved on a 2% agarose gel and sequenced from both the 5' and 3' ends using the PCR primers.

Northern blots

4.5 μg of Poly (A+) RNA and 30 μg of total RNA were electrophoresed on a 0.66M formaldehyde-1% agarose gel. The resolved RNA was transferred to a Hy-Bond Nytran membrane using downward capillary transfer and hybridized with an internal 1.8 kb riboprobe generated by digestion of the BamH1 clone with KpnI and transcription from the T3 promoter using the Ambion T3 Maxiscript kit. The templates for the 5' probe and 3' UTR antisense riboprobes were created via PCR to incorporate a T7 promoter. Primer pairs were: T75'probe (5' GACTCACTAT AGGGTCAATA AGATTCATGA3') and 5'probe (5'GGCGGTTGAG TCTGGCACCG ATAAAGTGG3') for the 5' probe and primers T73'UTR (5'GACTCACTAT AGGGAAGAAT TAAATCCATG3') and 3'UTR (5'GATATTAGAA ACTCGCAGAT TACAAAGTGT3') for the 3'UTR probe. Riboprobes were transcribed using the T7

Maxiscript kit (Ambion). In each case, the blot was hybridized in 50% formamide, 5X SSPE, 1% SDS at 68°C for 16 hours and washed twice at room temperature with 0.5X SSPE, 0.5% SDS for 15 minutes and once in 0.1X SSPE, 0.1% SDS at 68°C for 1-2 hours. The relative abundance of the various transcripts was determined by phosphoimager analysis (Molecular Dynamics) and normalized to the signal obtained by rehybridizing the blot with an end-labeled oligo against the β-tubulin gene.

Sequencing

Sequencing was performed by the MSU Sequencing Facility using an ABI 377 sequencer. The 6182 bp BamHI fragment was sequenced using the Tn7 transposon-based GPS™-1 Genome Priming System from New England Biolabs following the manufacturer's instructions.

RESULTS

Genomic Cloning

COnsensus-**DE**generate **H**ybrid **O**ligonucleotide **P**rimers (22) (CODEHOP) were designed from an alignment of the highly conserved carboxyl termini of several mitochondrial RNA polymerase genes in the database (Figure. 1). Using these primers, a 519 bp fragment was amplified from *T. brucei* genomic DNA. A BLASTX search (www.ncbi.nlm.nih.gov/blast) using sequence obtained from this product demonstrated significant homology ($E=9 \times 10^{-25}$) to mitochondrial RNA polymerase genes and gene fragments in the database. Therefore, we assigned to this gene the name *TBMTRNAP*, in accordance with the suggested nomenclature for *T. brucei* genes (3). The remaining genomic sequence of the mitochondrial RNA polymerase gene was obtained by screening

a P1 bacteriophage genomic library with a riboprobe generated from sequence obtained from the degenerate PCR product. A restriction map of the *TBMTRNAP* locus was constructed from the positive P1 clones (Figure 2A), and the 6182 bp BamH1 fragment was sequenced to construct a genomic contig which was extended 2 kb with sequence obtained from the Kpn-EcoRI clone. A Southern blot of *T. brucei* genomic DNA indicated that the *TBMTRNAP* exists as a single-copy gene (Figure 2B). With only one noted exception, (12), the *T. brucei* genome contains no introns. Consequently, the genomic sequence contains the mitochondrial RNA polymerase open reading frame, which consists of 3822 bp corresponding to 1274 amino acids shown in Figure 3A. A BLASTP search on the entire open reading frame indicated that the TBMTRNAP bears the strongest similarity to plant mitochondrial RNA polymerases with the most significant homology to the *Arabidopsis thaliana* (Genbank™ accession number AL391144) MTRNAP. This homology is localized to amino acids of the TBMTRNAP with 37% identity and 52% similarity to the *Arabidopsis* protein. The significance of this specific relationship to the plant enzymes is unclear. Evidence suggests that trypanosomes once possessed a chloroplast lost some time in their distant evolutionary past (14). However, closer inspection of the alignment does not show stronger similarity of the TBMTRNAP with plants than with other mtRNAP sequences. Consequently, this affinity is likely due to a long-branch attraction, a common pitfall when using highly diverged sequences (9).

CODEHOP primer design

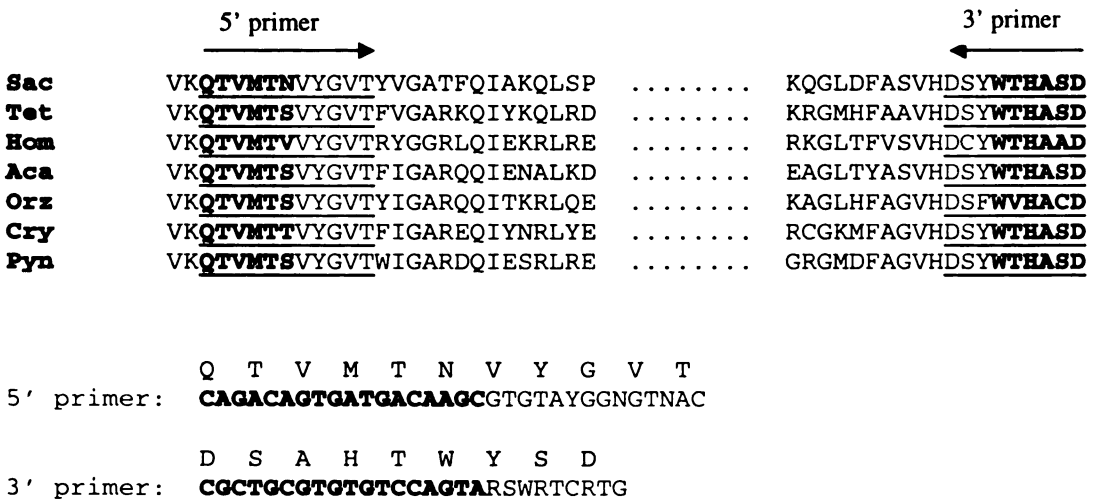


Figure 1. Portions of a multiple alignment of seven MTRNAPs generated using the program Clustal W. Amino acids from which primers were designed are underlined with the 5' consensus clamp in bold. Nucleotide sequence of the CODEHOP primers shown below the alignment: N=A/C/G/T, R=A/G, W=A/T, Y=C/T. GenBank™ accession numbers: Sac = *Saccharomyces cerevisiae* M17539, Tet = *Tetrahymena pyriformis* U34406, Hom = *Homo sapiens* U75370, Aca = *Acanthamoeba castellanii* U34405, Orz = *Oryza sativa* U34283, Cry = *Cryptomonas phi* U34404, Pyn = *Pycnococcus provasolii* U34286.

FIGURE 1

Figure 2. (A) Restriction map of the *TBMTRNAP* locus and cloning strategy. The *TBMTRNAP* ORF is represented by a long rectangle within which the degenerate PCR product is boxed. The bold lines above the restriction map indicate the subcloned P1 fragments. Arrows beneath the map indicate the cDNA clones. The 5' and 3' UTRs are represented by lines extending from the *TBMTRNAP* ORF. SAS indicates the splice acceptor site. The multiple polyadenylation sites are indicated by the downward arrows. Relative positions of the primers used in the 5' RT-PCR experiments are indicated by arrowheads above the map. (B) Southern blot analysis of the *T. brucei* genomic DNA. 10 μg of DNA was digested with BamHI (lane 1), EcoRI (lane 2), KpnI (lane 4), BamHI + EcoRI (lane 3), KpnI + BamHI (lane 5), and KpnI + EcoRI (lane 6), separated on a 0.75% gel, transferred to Nytran, and hybridized with a riboprobe spanning the KpnI site at 4330 and the BamH1 site at 6182 relative to the first nucleotide in the *TBMTRNAP* ORF.

TBMTRNAP locus restriction map and cloning strategy

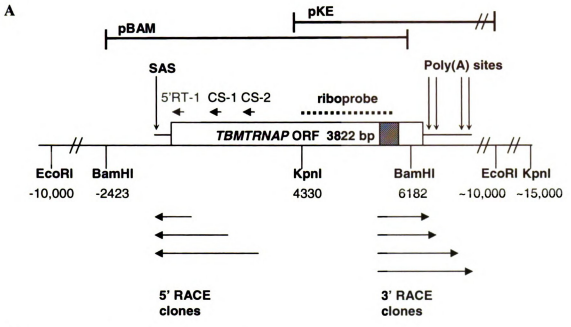


FIGURE 2

TBMTRNAP protein sequence and mitochondrial targeting signal

A

```
MRRLSLAPIK WPAIAQRWGT RGTDVPNRAK GAIKSPKSRK SRLPYVEVGE RTDRMKEFVS   60
SGTDDNSGDS DNNHNSVKSI NNTDCRNKGT DGGGEIDMDT VAGTDAFTPT VDMEEGYGAP  120
VMATTGSTSS ESVLSVIDVL CDDLFSCKTD VLKSFARTRS VSGSTKSELV MNLIDLANRE  180
AHGESQDLSI TKRVLGDVFN EEVRVSGVRS QWTNKSLEAS HIEMACDNHL KFEAMSYTQK  240
TLMVRRNASE SSRVLRKLIR PFQQDYNRVA EAAVGGVFRN FKAEADQLVN SMMASEIVTP  300
FSSFLVSYVR FATSGNLLTR TNVIERLEQL ELVQIAMDED SDAVLQRLMA LLDDVVAEVE  360
RRALCFGGKR TSWGTETPEV LRCSIQAFSD STAQQHEDLG ERVPAWVHDL VRMFNNWEAI  420
VRTDIPVLED VPRYVFDCIL SGVVCYHLLH KIYGDIISPT DVPPTVRGAV KRAMEAHFKD  480
EEKAGHALAK LFAYVQAPGD SGVVKAEFFP AVLSQMQGAF HDVGQSTLHN VRTAIVLREI  540
CATVSRCTRT CGLFHMLVKL TEDLRFQFRF STIFKKLNGR DRTTIRTKYS INMVKTLEYF  600
EQEYAVTPTS TKAVGFIVIL LLHMSMCSGA ASGHGTKAVL ERYISVTNKE EFSIISISDL  660
EATHVRDLRV AFPPQLSYNS WLHQSNTKDK SVYNVLPSVA VKGSTNQAMI ISQAPMMKAL  720
DAISRVPWRI SKYMLHVQEA IVREGYGFGK IRPAFYPLHY CAKSRGDISY ESTGMDDDDD  780
KTEVYNLQQR REYELQQDED WKNLSELRSN RIHYLQALRQ ARSLVQFSHI YFPNSMDFRG  840
RMYPLPGRLN HTGSDPFRAL LEYAEPKPLG KEGLYWLKVH LANKMGMSKL SFDERVHYVN  900
EHIDDVVCSA EQPLYGDKWW QEAAEPMQCL MACKELADAL KCSQGPENFL SRIPVAVDGS  960
YNGLQHYSAI GRDAFGATLV NLVPSERPAD AYTGILKEMM SSIKADAERD HPVAQRCLGT 1020
GKGQDRDHIK RKSIKRPIMT QVYGVTGYGM SELILDELVK QNKNHGLWTS TDMREMADYL 1080
REKVLESLGI TFRETQNCRR WITDVTNIIW EVQPAELRTA LCWTTPLGLV VRQPYKMRKE 1140
MMIFTVHGCA RVPANAFSAA SRKQLTAIAP NLIHSLDASH LAMTAIEMQN LGLSMMAVHD 1200
SYWTYACDLP TLSRVLREQF VTLYGKYDPL WELKEQWEEA YFMDLRRHGK VLPDPPKRGD 1260
LDLNVVLNSP YFFS                                                   1274
```

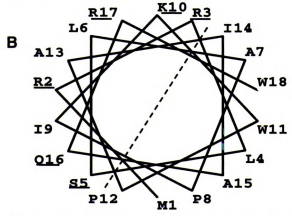


Figure 3. (A) Amino acid sequence of the *T. brucei* mitochondrial RNA polymerase with the putative mitochondrial targeting sequence in bold. (B) Helical wheel projection of the first 18 amino acids in the *TBMTRNAP* ORF. The dashed line divides the hydrophobic domain from the hydrophilic domain of the helix. Hydrophilic residues are underlined.

FIGURE 3

However, as indicated in Figure 4, the carboxyl-terminal portion of the protein bears strong homology to the bacteriophage-like MTRNAP characterized for organisms as phylogenetically diverse as humans and yeast. Interestingly, inspection of the full alignment indicates that even within the highly conserved blocks, the TBMTRNAP sequence deviates from that of the other mtRNAPs and T7 RNAP, including a small insert into domain III. However, residues that have been shown to be critical for activity of the T7 RNA polymerase (2, 24) are conserved. The amino terminus of this protein is predicted to form an amphipathic helix and to be targeted to the mitochondrion via the putative sequence underlined in Figure 3.

cDNA cloning

Nuclear transcripts in *T. brucei* are expressed as polycistronic precursors from which mature mRNAs are rapidly processed via the *trans*-splicing of a conserved, capped, 39 nt spliced leader RNA to the 5' end and the 3' cleavage/polyadenylation of the individual mRNAs (28). Therefore, to determine the 5' end of the *TBMTRNAP* mRNA, a modified RACE strategy using three different gene-specific primers (Figure 2A) and the conserved spliced leader primer was used. Three independent clones derived from the three separate PCR products mapped the splice acceptor site to an AG dinucleotide at a position 29 nt upstream of the first ATG codon in the *TBMTRNAP* open reading frame. The 3' end of the *TBMTRNAP* mRNA was determined by 3' RACE experiments. Four distinct PCR products of various sizes were obtained from the one pair of primers used for the 3' RACE (Figure 2A). Sequencing of these cloned products revealed two size classes of *TBMTRNAP* mRNA with 3' UTRs that differ by 500 nt.

Figure 4: Multiple alignment of representative mtRNAPs with the T7 RNAP. Cartoon schematic showing different lengths and motif conservation among the human, yeast, *Arabidopsis*, and *T. brucei* mtRNAPs with that of T7RNAP. The roman numerals below the diagram depict the conserved motif designations in the T7 RNAP. Motifs that are strongly conserved are shaded in black, those with moderate conservation are in gray, and minimal conservation is depicted by white boxes. The pinky specificity domain thought to be important in promoter recognition is indicated by the dotted oval.

Diagram of multiple alignment of representative mtRNAPs with T7RNAP

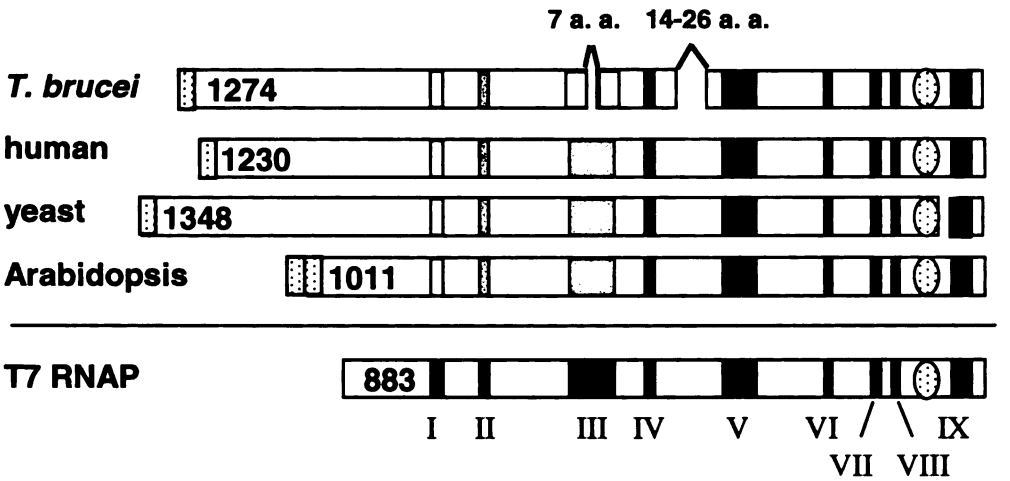


FIGURE 4

mRNA expression

In order to determine the relative levels of *TBMTRNAP* mRNA during the life cycle of *T. brucei*, RNA from the procyclic (insect) form and bloodstream form trypanosomes was examined. Additionally, RNA from dyskinetoplastid trypanosomes that lack mitochondrial DNA and therefore functional mitochondria, was included in this experiment. Northern blot analysis with a riboprobe specific for the 5' end of the ORF detected three distinct transcripts with apparent molecular weights of 4.4 kb, 4.7 kb, and 6.9 kb (Figure 5A). The 4.4 kb and 4.7 kb transcripts are visible in procyclic, bloodstream, and dyskinetoplastid trypanosomes, while the larger 6.9 kb transcript not visible in the procyclic lane has been detected in other Northern blots using larger amounts of poly(A) RNA (data not shown). An identical pattern was seen using the internal probe located within the coding region of the *TBMTRNAP* (data not shown). The difference in size between the 4.4 kb and 4.7 kb *TBMTRNAP* transcripts reflects the two distinct 3' UTRs mapped in the 3' RACE experiments as demonstrated by hybridization of a 3' UTR specific probe solely to the 4.7 kb band (Figure 5B). Whereas the amount of the 4.7 kb *TBMTRNAP* mRNA was not significantly different among the three forms, the 4.4 kb transcript was present in procyclic cells at levels 4 times that seen in bloodstream or dyskinetoplastid trypanosomes. The abundance of *TBMTRNAP* mRNA in trypanosomes lacking mitochondrial DNA is roughly equivalent to that in bloodstream form cells when normalized against tubulin mRNA (Figure 5C), which is expressed equally during the life cycle. Surprisingly, while both the internal and the 5' probes gave identical hybridization patterns, the riboprobe specific for the 3' UTR of the *TBMTRNAP* mRNA hybridized to two additional transcripts of 2.6 kb and 0.7 kb.

Figure 5. Developmental profile of steady-state transcripts from the *TBMTRNAP* locus. 4.5 μg of poly(A)+ or 30 μg of total RNA from procyclic (P), bloodstream (BS), or dyskinetoplastid (DK) trypanosomes was probed with the 5' probe riboprobe (A) and rehybridized with the 3'UTR riboprobe (B). Equal loading of samples was controlled by rehybridization of the blot with an end-labeled oligonucleotide detecting the β-tubulin gene (C). (D) Mature mRNAs processed from the 6.9 kb polycistronic transcript are indicated below the diagram of the *TBMTRNAP* locus. Filled circles represent splice acceptor sites. Filled rectangles represent polypyrimidine tracts.

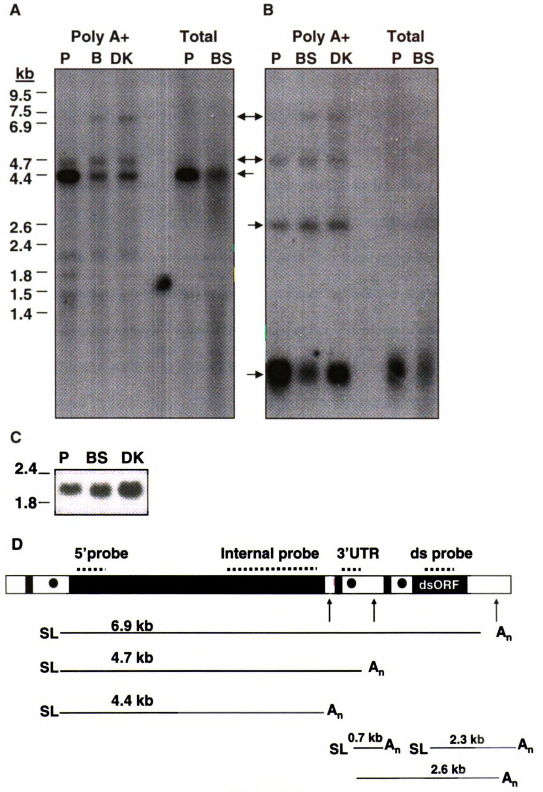

A
Poly A+
Total
P B DK
P BS
kb
9.5
7.5
6.9
4.7
4.4
2.6
2.4
1.8
1.5
1.4
B
Poly A+
Total
P BS DK
P BS
C
P BS DK
2.4
1.8
D
5'probe
Internal probe
3'UTR
ds probe
dsORF
SL
6.9 kb
An
SL
4.7 kb
An
SL
4.4 kb
An
SL
0.7 kb
An
SL
2.3 kb
An
2.6 kb
An

FIGURE 5

A Southern blot of *T. brucei* DNA probed with the 3' UTR riboprobe did not reveal any related sequences in the *T. brucei* genome (data not shown). Since these transcripts were not detected with the internal probe, which is located only 240 nt upstream of the 3' UTR riboprobe, it is clear that at least a portion of the sequence present in the 3'UTR of the 4.7 kb *TBMTRNAP* transcript is also part of the 5' ends of the 2.6 kb and 0.7 kb transcripts. Therefore, these transcripts must be produced by mutually exclusive processing events. Rather interestingly, the 2.6 kb transcript is present at relatively equal levels in all three forms while the 0.7 kb transcript is quite abundant in all forms relative to the other transcripts detected in the Northern hybridizations and is 4 times more abundant in procyclic than in bloodstream trypanosomes. We are unable to detect a product that extends beyond the mapped splice site acceptor shown in Figure 5 for the *TBMTRNAP* mRNA in 5' RACE experiments using cDNA from both procyclic and bloodstream trypanosomes. This together with our hybridization results leads us to believe that the 6.9 kb transcript may be a pre-mRNA that has not yet undergone the trans-splicing/polyadenylation processing events to produce the mature *TBMTRNAP* mRNA and the downstream transcripts diagrammed in Figure 5D. The *trans*-splice site for the 0.7 kb transcript maps to a location 120 bp downstream of the proximal poly(A) site of the TBMTRNAP mRNA. Although the 3' end of this transcript is quite heterogeneous, several 3' RACE and nested RACE experiments demonstrated that the 3' end maps to the same region as that of the 4.7 kb *TBMTRNAP* mRNA (data not shown). The 2.6 transcript detected with the 3' UTR probe proved more difficult to characterize. Repeated 5' RACE experiments using the spliced leader primer and a gene-specific primer located downstream of the distal poly(A) site (see Figure 5D) failed to detect a

product large enough to have been detected by the 3' UTR riboprobe. Rather, the 5' end of the product that was consistently obtained with this primer pair mapped to a site 126 nt downstream of the *TBMTRNAP* mRNA distal poly(A) site. Likewise, 3' RACE experiments using a gene specific primer located within the 3'UTR riboprobe region failed to detect any transcripts long enough to be the 2.6 kb transcript seen in this Northern blot. However, when the two gene-specific primers used in both of these RACE experiments were used together in RT-PCR a product was obtained that would traverse this area. This PCR product is expected based upon our proposed 6.9 kb precursor, but may also account for the 2.6 kb transcript.

DISCUSSION

We have identified and cloned a single mitochondrial RNA polymerase gene in *Trypanosoma brucei* that encodes two differentially expressed mRNAs. While the significant similarity of the predicted TBMTRNAP sequence to MTRNAPs in the database is confined to the catalytic carboxyl terminus, the lack of sequence conservation in the amino terminus of the TBMTNRAP is not surprising. The variation in the length and composition of the non-conserved region reflects species-specific interactions with other proteins involved in transcription and RNA processing. For example, the amino terminal portion of the yeast mitochondrial RNA polymerase interacts with proteins that directly link transcription to both RNA processing and translation (1, 21). Other sequence differences also likely reflect species or group specific functions. For instance, the "pinky specificity loop" is located within the C-terminal conserved region and shows group-specific differences (Figure 1). Although there is sequence similarity within each

major group (i.e., plants, fungi, kinetoplast, vertebrates), there is little similarity across groups of organisms. This region in the T7 RNAP is important for promoter recognition via the insertion of these amino acids into the DNA major groove (24). Interestingly, recent data suggest that the mitochondrial RNA polymerase has innate promoter recognition ability (15). Consequently, these differences likely reflect co-evolution with respective promoter sequences.

In spite of this intrinsic promoter-recognition property, all mitochondrial RNA polymerases studied to date require at least one specificity factor for accurate transcription initiation. The nature and number of these factors varies among organisms, but can be divided into two classes: mtTFB, which is required in both animals and yeast (10, 17), and mtTFA, a 25 kDa high mobility group (HMG) protein (5), required only in animals. These are the only two systems for which the protein machinery required for mitochondrial transcription has been dissected. Fungi and animals are more closely related to each other than either of them are to other phyla of organisms. However, they show distinct differences in the composition of the mtRNAP holoenzyme, making it likely that other variations will become evident as more systems are studied. It is evolutionarily interesting that the catalytic subunit of the mtRNAP is conserved in *T. brucei* as it is one of the deepest branching organisms to possess mitochondria. Given the remarkable nature of the kinetoplast genome and the early divergence of *T. brucei*, one might expect to find unique protein components comprising the mitochondrial transcription machinery. There may be guide RNA-specific, minicircle-specific, and/or maxicircle-specific transcription factors not found in higher eukaryotes.

The mitochondrion of *T. brucei* is developmentally regulated during its complex life cycle (20), making the regulation of the *TBMTRNAP* mRNA intriguing. While living in the insect midgut, the mitochondrion of the procyclic form *T. brucei* is fully functional. In the glucose-rich bloodstream, however, mitochondrial activity is reduced, and energy is derived through glycolysis. This regulation occurs through the control of both nuclear and mitochondrial genes primarily through post-transcriptional and post-translational processes (20). Our finding that the *TBMTRNAP* gene is transcribed into two distinct mRNAs subject to differential regulation during the life cycle suggests that mitochondrial differentiation might be achieved, in part, through the regulated expression of this gene. It is significant that the *TBMTRNAP* mRNA with the shorter (50 nt) 3' UTR is more abundant in procyclic cells than in bloodstream or dyskinetoplastid cells, while the mRNA with the longer 3' UTR (550 nt) seems relatively equal in all three forms. Stage-specific differences in the length of the 3' UTR as a result of alternative polyadenylation site selection has been observed for other *T. brucei* mRNAs (7). As in other eukaryotic transcripts, *T. brucei* 3'UTRs are known to contain information that affects such processes as mRNA turnover and translatability (6, 8, 28). Thus, elements within the 3' UTR of the TBMTRNAP mRNA are likely to be important for the regulation of its expression during the *T. brucei* life cycle.

Polyadenylation in *T. brucei* is dependent upon a downstream polypyrimidine track followed by a 3' trans-splice site (16, 23). Trans-splicing is rapid and seems to occur as soon as the 3' splice site is available to the processing machinery (27). This posses an interesting question for the choice of poly(A) site on the newly synthesized *TBMTRNAP* transcript. In order for the proximal poly(A) site to be favored by the

processing machinery in procyclic cells, the splice site immediately downstream of the polypyrimidine track following this poly(A) site must be favored over downstream splice sites. Likewise, in bloodstream form cells, this 3' splice site must be used less efficiently in favor of a downstream splice site. The 0.7 kb transcript, which contains a trans splice site located just downstream of the polypyrimidine track following the proximal poly(A) site, is likely the product of such a decision. This 0.7 kb transcript is 4 times more abundant in the procyclic cells than in the bloodstream and dyskinetoplastid cells, which have relatively equal amounts when normalized against the tubulin probe.

It is important to note, however, that the proximal poly(A) site is favored over the distal site in all three forms as the 4.4 kb transcript is more slightly more abundant in all forms than is the 4.7 kb transcript. Although the 4.4 kb transcript is dramatically "upregulated" in procyclic form cells, it does not result in a concomitant decrease in the amount of the 4.7 kb transcript which leads one to believe that this phenomenon is not solely a result of a shift in the choice of poly(A) site. However, given the coordinated decrease in the amount of the 6.9 kb transcript, we propose that regulation of the transcripts originating from this locus is a result of alternative polyadenylation in concert with differential stability. One should note that if there is a destabilizing sequence in the longer 3' UTR, it must either be transcript-dependent, or exist in the 120 nt located between the proximal poly(A) site and the trans-splice site of the 0.7 kb transcript, since this RNA shares considerable sequence with the 3' UTR yet is clearly quite stable.

Detection of this stable 0.7 kb transcript was intriguing. This RNA contains no ORFs larger than 35 amino acids and may merely represent an RNA processing by-product as discussed above. The stability of this small transcript, however, suggests that

it may have an important non-coding function, as most dead end RNAs are rapidly degraded. Searches of sequence databases from *Leishmania* major and *T. cruzi* found no homologous sequences suggesting that the RNA sequence is not conserved among these related organisms. However, the physiological importance of non coding and small ORF-coding transcripts has become more apparent with the analysis of whole genome sequences. Therefore, it will be interesting to determine the significance of this small, stable transcript in *T. brucei.*

Given the polycistronic nature of transcription units in *T. brucei*, post-transcriptional events such as trans-splicing, polyadenylation, and RNA turnover can serve as potent regulators for the individual genes within a given pre-mRNA (4, 6). This type of regulation could be especially important for genes such as the TBMTRNAP, whose function is likely critical for differentiation of the trypanosomes as they progress through the dramatically different environments they encounter during their complex life cycle. Interestingly, the regulation of the expression of the transcripts from the TBMTRNAP locus in dyskinetoplastid trypanosomes appears quite similar to that seen for bloodstream form cells. Mitochondrial genome deletion (rho0) mutants in HeLa cells have been shown to demonstrate defects in nuclear-organelle cross-talk which can result in a higher level of expression of some mitochondrially targeted genes including the mitochondrial transcription factor mtTFA (19). Consequently, further investigation into the protein levels of TBMTRNAP during the life cycle and in dyskinetoplastid trypanosomes are warranted and will provide insight into the significance of the regulation of *TBMTRNAP* mRNA processing and stability. Likewise, further studies are necessary to determine how the *TBMTRNAP* gene product functions in the context of the

unique kinetoplast genome. Of particular interest will be the identification of accessory factors required for transcription initiation and of the promoter regions recognized by the *T. brucei* mitochondrial RNA polymerase complex.

LITERATURE CITED

1. **Bryan, A. C., M. S. Rodeheffer, C. M. Wearn, and G. S. Shadel.** 2002. Sls1p is a membrane-bound regulator of transcription-coupled processes involved in *Saccharomyces cerevisiae* mitochondrial gene expression. Genetics 160:75-82.

2. **Cermakian, N., T. M. Ikeda, P. Miramontes, B. F. Lang, M. W. Gray, and R. Cedergren.** 1997. On the evolution of the single-subunit RNA polymerases. J Mol Evol 45:671-81.

3. **Clayton, C., M. Adams, R. Almeida, T. Baltz, M. Barrett, P. Bastien, S. Belli, S. Beverley, N. Biteau, J. Blackwell, C. Blaineau, M. Boshart, F. Bringaud, G. Cross, A. Cruz, W. Degrave, J. Donelson, N. El-Sayed, G. Fu, K. Ersfeld, W. Gibson, K. Gull, A. Ivens, J. Kelly, L. Vanhamme, and et al.** 1998. Genetic nomenclature for *Trypanosoma* and *Leishmania*. Mol Biochem Parasitol 97:221-4.

4. **Clayton, C., and H. R. Hotz.** 1996. Post-transcriptional control of PARP gene expression. Mol Biochem Parasitol 77:1-6.

5. **Dairaghi, D. J., G. S. Shadel, and D. A. Clayton.** 1995. Human mitochondrial transcription factor A and promoter spacing integrity are required for transcription initiation. Biochim Biophys Acta 1271:127-34.

6. **D'Orso, I., J. G. De Gaudenzi, and A. C. Frasch.** 2003. RNA-binding proteins and mRNA turnover in trypanosomes. Trends Parasitol 19:151-5.

7. **Erondu, N. E., and J. E. Donelson.** 1992. Differential expression of two mRNAs from a single gene encoding an HMG1-like DNA binding protein of African trypanosomes. Mol Biochem Parasitol 51:111-8.

8. **Furger, A., N. Schurch, U. Kurath, and I. Roditi.** 1997. Elements in the 3' untranslated region of procyclin mRNA regulate expression in insect forms of *Trypanosoma brucei* by modulating RNA stability and translation. Mol Cell Biol 17:4372-80.

9. **Gribaldo, S., and H. Philippe.** 2002. Ancient phylogenetic relationships. Theor Popul Biol 61:391-408.

10. **Jang, S. H., and J. A. Jaehning.** 1991. The yeast mitochondrial RNA polymerase specificity factor, MTF1, is similar to bacterial sigma factors. J Biol Chem 266:22671-7.

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

12. **Mair, G., H. Shi, H. Li, A. Djikeng, H. O. Aviles, J. R. Bishop, F. H. Falcone, C. Gavrilescu, J. L. Montgomery, M. I. Santori, L. S. Stern, Z. Wang, E. Ullu, and C. Tschudi.** 2000. A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA. Rna 6:163-9.

13. **Maizels RM, B. M., Robertson BD, Selkirk ME.** 1991. Parsite Antigens, Parasite Genes A Laboratory Manual for Molecular Parasitology. Cambridge University Press, Cambridge.

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

15. **Matsunaga, M., and J. A. Jaehning.** 2004. Intrinsic promoter recognition by a "core" RNA polymerase. J Biol Chem 279:44239-42.

16. **Matthews, K. R., C. Tschudi, and E. Ullu.** 1994. A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. Genes Dev 8:491-501.

17. **McCulloch, V., B. L. Seidel-Rogol, and G. S. Shadel.** 2002. A human mitochondrial transcription factor is related to RNA adenine methyltransferases and binds S-adenosylmethionine. Mol Cell Biol 22:1116-25.

18. **Milligan, J. F., D. R. Groebe, G. W. Witherell, and O. C. Uhlenbeck.** 1987. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. Nucleic Acids Res 15:8783-98.

19. **Miranda, S., R. Foncea, J. Guerrero, and F. Leighton.** 1999. Oxidative stress and upregulation of mitochondrial biogenesis in mitochondrial DNA-depleted HeLa cells. Biochem Biophys Res Commun 258:44-9.

20. **Priest, J. W., and S. L. Hajduk.** 1994. Developmental regulation of mitochondrial biogenesis in *Trypanosoma brucei*. J Bioenerg Biomembr 26:179-91.

21. **Rodeheffer, M. S., B. E. Boone, A. C. Bryan, and G. S. Shadel.** 2000. Nam1p, a protein involved in RNA processing and translation, is coupled to transcription through an interaction with yeast mitochondrial RNA polymerase. J Biol Chem 15:15.

22. **Rose, T. M., E. R. Schultz, J. G. Henikoff, S. Pietrokovski, C. M. McCallum, and S. Henikoff.** 1998. Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. Nucleic Acids Res 26:1628-35.

23. **Schurch, N., A. Hehl, E. Vassella, R. Braun, and I. Roditi.** 1994. Accurate polyadenylation of procyclin mRNAs in *Trypanosoma brucei* is determined by pyrimidine-rich elements in the intergenic regions. Mol Cell Biol 14:3668-75.

24. **Sousa, R.** 1996. Structural and mechanistic relationships between nucleic acid polymerases. Trends Biochem Sci 21:186-90.

25. **Stuart, K., T. E. Allen, S. Heidmann, and S. D. Seiwert.** 1997. RNA editing in kinetoplastid protozoa. Microbiol Mol Biol Rev 61:105-20.

26. **Tracy, R. L., and D. B. Stern.** 1995. Mitochondrial transcription initiation: promoter structures and RNA polymerases. Curr Genet 28:205-16.

27. **Ullu, E., K. R. Matthews, and C. Tschudi.** 1993. Temporal order of RNA-processing reactions in trypanosomes: rapid trans splicing precedes polyadenylation of newly synthesized tubulin transcripts. Mol Cell Biol 13:720-5.

28. **Vanhamme, L., and E. Pays.** 1995. Control of gene expression in trypanosomes. Microbiol Rev 59:223-40.

CHAPTER IV

A conserved mtTFB in the early eukaryote *Trypanosoma brucei*: Implications for the origin and evolution of mitochondrial transcription machinery

Introduction

Mitochondrial genomes vary considerably in gene content, structure and patterns of expression (5). In spite of this dramatic variation, these genomes are transcribed by what appears to be a universally conserved mitochondrial RNA polymerase (mtRNAP) homologous to the phage T7 RNAP (7). Although mitochondrial transcription has been characterized for only a few systems, the nuclear encoded mtRNAP gene is present in all mitochondrial-bearing eukaryotes for which sequence data is available. Likewise, we have shown that this gene is also conserved in kinetoplastids, some of the most ancient organisms to contain mitochondria (9). Mechanistically, the mtRNAP functions somewhat differently than its phage homologue in that it requires at least one additional factor for transcription initiation. In yeast, the mitochondrial transcription factor B (mtTFB, also known as MTF1) is necessary and sufficient to reconstitute transcription in association with the mtRNAP (RPO41), on mitochondrial promoters (19, 31). In contrast, early studies of mammalian mitochondrial transcription showed a requirement for a distinct factor belonging to the HMG family of proteins, mtTFA or TFAM, in association with partially purified mtRNAP (14-16, 29). Although an mtTFB-like activity was identified biochemically in *X. laevis*, the corresponding gene was never subsequently identified (4). These differences in the nature and number of transcription factors are likely due to co-evolution of the transcription apparatus with its respective genome.

Searches of the rapidly expanding genomic databases for homologues to the yeast transcription factor did not reveal similar genes in other organisms. This was not surprising given the lack of sequence conservation even among fungal relatives (6). In

addition, early sequence comparisons had likened the yeast mtTFB to sigma factors, themselves a diverse group of proteins. Consequently, the degenerate PCR technique used to identify the mtRNAP homologue in a wide variety of organisms was unsuitable for the purpose of finding homologues to mtTFB. A major breakthrough in the field was the crystallization of the Sc-mtTFB which decisively demonstrated that the protein was not homologous to sigma factors as had previously been speculated (32). Analysis of the protein structure indicated that it belonged to large family of ribosomal RNA methyltransferases. Specifically mtTFB is homologous to the eubacterial KsgA, a protein responsible for the nearly universal dimethylation of two adjacent adenine residues in the highly conserved 3' tetraloop of small subunit rRNA (28). The news of this surprising relationship was followed rapidly by the identification of the mtTFB homologue, TFBM1, in humans by two separate groups (13, 24). Interestingly, these homologues, and those subsequently identified in the sequence databases, bear much greater similarity to their KsgA ancestor than does the yeast mtTFB. Humans and other "higher" animals also possess a paralogue to mtTFB, TFBM2 (13, 23, 30).

Trypanosoma brucei is an ancient protist with an unusual mitochondrial genome and is one of the earliest eukaryotes to harbor mitochondria (18). Our lab has previously identified a developmentally regulated mitochondrial RNA polymerase in *T. brucei* homologous to the mtRNAPs of other organisms (9) As was the case with the vertebrate genome databases, however, initial searches of the *T. brucei* genomic sequence database did not reveal any putative mitochondrially targeted proteins with homology either to transcription factor TFAM or to the yeast mtTFB. However, looking for rRNA methyltransferase homologues we have identified the homologue to the mtTFB-1 in *T.*

brucei. This gene encodes a 467 amino acid protein with a putative mitochondrial targeting sequence. Although clearly divergent from the animal mtTFB homologues in the database, TbmtTFB bears much stronger similarity to these enzymes than does the yeast MTF1 gene. RNAi knockdowns of the TbmtTFB transcript resulted in a moderate decrease of target transcript and a slow growth phenotype. However, this was accompanied by only slight decreases in the steady state levels of various mitochondrial transcripts. Over-expression of TbmtTFB fused to a C-terminal tandem affinity purification (TAP) tag in *T. brucei* resulted in an increase in trypanosome growth rate and a slight increase in some mt-mRNA levels but was unexpectedly accompanied by slight decreases of mitochondrial rRNA. The finding of the TbmtTFB homologue in *T. brucei* suggests that the core components of the mitochondrial transcription machinery in yeast and animals was set in place millions of years ago, early in the course of eukaryotic evolution. Although our initial data are insufficient to conclusively support or negate a role as a transcription factor, we are currently working on obtaining conditional double knockouts that will allow us to decisively determine its function in *T. brucei*. Given the phylogenetic placement of *T. brucei*, either outcome will be interesting and significant.

Methods and Materials

Database searches and Sequence analysis:

The *S. cerevisiae* MTF1p (Genbank accession number NP_013955) protein sequence was used to query the *T. brucei* TIGR database (http://www.tigr.org/tdb/e2k1/tba1/) using tblastn. The resulting *T. brucei* nucleotide sequence was then used in a tblastx search against the NCBI protein databases for further

identification. TIGR sequences homologous to the KsgA family of methyltransferases were assembled into a contig that corresponded to the 3' 500 bp of the gene. Mitochondrial targeting sequences were predicted using the following programs: MITOPROT, http://www.mips.biochem.mpg.de/cgi-bin/proj/medgen/mitofilter, Predotar, http://www.inra.fr/predotar/, TargetP, http://www.cbs.dtu.dk/services/TargetP/ and iPSORT, http://hypothesiscreator.net/iPSORT/. Multiple alignments were generated by Clustal X 1.81.

Cloning of the TbmtTFB:

The portion of the TBMTTFB open reading frame not contained in the TIGR database as well as the 5' splice acceptor site was determined by sequencing 5'RACE (rapid amplification of cDNA ends) products as described previously (9). The *T. brucei mtTFB* open reading frame was then amplified from *T. brucei* genomic DNA (which lacks introns) using *Pfu* polymerase (Stratagene) and sequenced by the MSU sequencing facility.

Nucleotide sequence accession number: The TbmtTFB homologue has been submitted to the GenBank database under accession no. AY133948.1

Trypanosome growth, transfections, and RNAi inductions:

Procyclic form *T. brucei brucei* strain 29-13 (generated by the George Cross lab), which contains both the T7 RNA polymerase and the Tet Repressor integrated within the genome, were grown in SDM-79 supplemented with 10% fetal bovine serum in the presence of Geneticin (15μg/ml) and hygromycin (50 μg/ml) (39). TbmtTFB RNAi vectors were generated by amplifying a 477 bp fragment (position 430 to 907 in the 1401 ORF) with oligonucleotides 5'XhoI ErmC RNAi and 3' HindIII ErmC RNAi, and

another 460 bp fragment (position 892 to 931 in the 1401 ORF) with oligonucleotides 5'XhoIRNAi-2 and 3'HindIII RNAi-2. A third RNAi vector was made by amplifying the mtTFB ORF with 5'ApamtTFB and 3'HindII ErmC RNAi-2, digesting both this PCR product and the RNAi-1 vector with HindIII, followed by ligation to generate the 1157 +477 = 1634 dsRNA fragment.

For transfections, mid-log phase cells between 2 and 8 x 10^6 cells/ml were centrifuged at 1000 x g, washed with ice cold Cytomix+phosphate buffered sucrose and resuspended in fresh cytomix/PBS at a concentration of 2 x 10^8 cells/ml. 1 x 10^8 cells in 500 µl was combined with 30 µg of digested plasmid DNA on ice and transfected using a Bio-Rad Gene Pulser electroporator at 1500 V and 25 µF. After transfection, cells were resuspended in 10 mls of fresh medium and allowed to recover for 16-20 hours. Cells were then spun down and resuspended in fresh medium containing geneticin, hygromycin, and 2.5 µg/ml phelomycin. Resistant cells appeared 3-4 weeks after selection.

Growth Curves:

For induction of dsRNA or TbmtTFB-TAP expression, 1 µg/ml tetracycline was added to the medium. Inductions were started with cells at a density of 1 x 10^6 cells/ml and were harvested between 5 and 10 x 10^6 cells/ml, and cells were diluted back down to 1 x 10^6 cells/ml. Growth curves were generated by plotting the cell count multiplied by the total dilution over a period of 2-3 weeks. Inductions were performed in duplicate on three separate occasions.

Nucleic acid Isolation, Northern and Southern blot analysis. Total *T. brucei* genomic DNA was isolated as described previously (9). Mitochondrial RNA was isolated by the

acid guanidinium-phenol-chloroform method (8). Northern and Southern blots were hybridized with ^{32}P end-labeled oligonucleotide probes using standard conditions.

Western blot analysis

At each time point, ~5 x 10^7 cells were harvested, washed once with PBS+6mM glycine, resuspended in 50 µl of 2XSDS-PAGE loading buffer and stored at -80°C. 1 x 10^7 cells were electrophoresed on an 8% SDS-PAGE gel, transferred to PVDF, and probed with peroxidase-antiperoxidase (PAP, Sigma P-2026) which reacts with the protein A portion of the TAP fusion protein. The signal was developed with the SuperSignal West Pico chemiluminescent system (Pierce)

Conditional Double Knockout Vectors

Drug resistance vectors with the blasticidin resistance gene (pHD889) and puromycin drug resistance gene (pHD996) were kindly provided by Dr. Christine Clayton. Sequence flanking the TbmtTFB ORF was amplified from 29-13 genomic DNA using pfu polymerase (Stratagene). An 880 bp 5' fragment was amplified with 5'UTR NotIlong and 5' UTR XhoIRev and the 1200 bp 3' fragment was amplified with 3'UTR NheIlong and 3'UTR NgoMIVRev. The 5' fragment was inserted into the pHD996 NotI and XhoI sites, and sequenced, followed by the insertion of the 3' fragment between the NheI and NgoMIV sites to create the vector, pKOpur. The puromycin drug resistance cassette was removed from pKOpur and replaced with the blasticidin resistance cassette from pHD889 to create pKObsd.

The tandem affinity purification vector available from Euroscarf/Cellzome contained a hygromycin selectable marker, making it inappropriate for use in 29-13 cells. Consequently, the 586 bp (TAP) expression cassette was purified from the vector

pHD918 (Cellzome/Euroscarf) digested with BamH1 and HindIII, ligated into the gel-purified backbone of vector pLEW100 (3) from which the 1877 HindIII-BamH1 luciferase fragment had been removed to generate pLEW-TAP. The TbmtTFB ORF was amplified with the kinased primers 5' ApaImtTFB and 3'mtTFBnostop, and ligated into HpaI-digested, phosphatase-treated pLEW-TAP to create pLEW-mtTFB-TAP. Transfections were performed as described above for the RNAi vectors, and cells were selected in the presence of 2.5 μg/ml phelomycin (pLEW TAP)

Results:

Identification of a putative mtTFB in *T. brucei*:

Although previously likened to sigma factors, the recent crystal structure of the yeast mitochondrial transcription factor mtTFB revealed structural homology with a family of RNA methyltransferases (32). With this new structural information, we investigated a low-significance hit (E>0.01) to the yeast mtTFB in the *T. brucei* TIGR database. The yeast mtTFB protein sequence was blasted against the *T. brucei* database with 2 resulting partially overlapping, low probability hits, 37D9.TR M13 Rev and 50N15.TR M13 Rev. A tblastp search with these DNA sequences against the NCBI NR protein database returned several hypothetical eukaryotic proteins including the subsequently identified human mtTFB1 (24), as well as the genes for the prokaryotic methyltransferase KsgA. Because the TIGR database contained only 450 bp of the 3' end of the gene, the remaining sequence was obtained by sequencing 5' RACE products. The putative TbmtTFB open reading frame of 1401 nt encodes a 467 amino acid polypeptide predicted to be targeted to the mitochondrion by the web-based programs

iPSORT, MitoProtII, Predotar, and TargetP. Southern blot analysis indicates that this is a single copy gene (data not shown), and analysis of contigs at the *T. brucei* genome project website indicates it is located on chromosome X. Homologues to TbmtTFB are also present in the genomes of the related kinetoplastid organisms, *Leishmania major* and *T. cruzi* as determined by searching their genomic databases (www.sanger.ac.uk, and http://tcruzidb.org, respectively).

The kinetoplast homologues are significantly similar to vertebrate mitochondrial transcription factors and retain several of the methyltransferase domains also conserved in these proteins (Figure 1). It is interesting that the yeast mtTFB, while clearly related, is not significantly similar to either the human or trypanosome homologues. For example, a Blastp search of the NCBI NR database using the *S. cerevisiae* mtTFB protein sequence yields significant results only for other fungal mtTFBs including those of *Schizosaccharomyces pombe*, *Kluyveromyces lactis*, and *Saccharomyces kluyveri* with E values $<10^{-23}$. The human mtTFB homologue is much less similar with an E value of 1.4. In contrast, the *T. brucei* protein is quite similar to the mouse and human mtTFB proteins (E values of 10^{-11} and 10^{-8}, respectively). The divergence of the yeast protein sequence is likely the result of the greater evolution rate of yeast in general (12, 36).

Schematic of alignment of kinetoplast mtTFBs with animal and yeast homologues

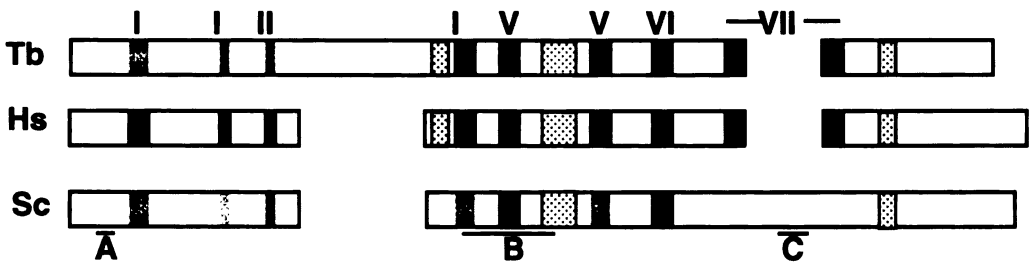


Figure 1: Alignment of kinteoplast mtTFBs with animal and yeast homologues. Schematic diagram of the structure of the mtTFBs. Black boxes indicate conserved methyltransferase domains. Gray boxes indicate conserved regions with less homology. Breaks in the boxes indicate absence of sequence homology caused by an insertion in one of more of the sequences relative to the others. The stippled boxes represent insertions conserved in the mtTFB homologues that are not present in the eubacterial KsgA proteins. Regions underneath the Sc schematic represent deleterious mutations in yeast mtTFB. The majority of these mutations occur in regions that are not conserved among the other mtTFB predicted protein sequences. Tb, *Trypanosoma brucei*; Hs, *Homo sapiens*, Sc, *Saccharomyces cerevisiae*.

FIGURE 1

Interestingly, the mitochondrial mtTFB sequences also have domains in common that are not shared with the KsgA homologues. Comparison of the available mtTFB sequences in an alignment with KsgA sequences reveal 3 blocks of sequence unique to the mtTFBs (Figure 1). In addition, the kinetoplast sequences contain an insertion relative to the other aligned mtTFB sequences. This insertion occurs between domains III and IV, and does not disturb regions known to be important for SAM binding or catalysis. Of the clusters of mutations performed on the yeast mtTFB sequence (10, 20, 35) shown in Figure 1, only cluster B is highly conserved among the other mtTFB sequences.

RNAi knockdown of TbmtTFB

In order to determine whether this putative mitochondrial transcription factor plays a role in transcription in *T. brucei* mitochondria, we attempted to use RNA interference (RNAi) to knockdown expression of this gene. RNAi has been used effectively to study the functions of several *T. brucei* genes including the *T. brucei* mitochondrial RNA polymerase, *TBMTRNAP* (17). Briefly, the vector pZJM contains a target PCR gene fragment inserted between two opposing tetracycline-inducible T7 promoters (Figure 2A). Upon transfection into *T. brucei*, pZJM integrates into the non-transcribed rDNA spacer region of the *T. brucei* genome to create a stable cell line expressing double stranded RNA (dsRNA) corresponding to the insert region (38).

Figure 2: **(A)** Vector for expression of double stranded RNA (dsRNA) **(B)** Schematic of TbmtTFB ORF with regions targeted by dsRNA shown as boxes. Probes used in Northern blots and RPAs are indicated by numbers above the diagram. **(C)** Northern blot analysis of RNA from cells expressing dsRNA-1. TbmtTFB mRNA (top panel) or dsRNA-1 (bottom panel) using probe 1. **(D)** Ribonuclease protection assay of TbmtTFB mRNA (top panel) or Northern blot of dsRNA-2 (bottom panel) of RNA from cells expressing dsRNA-2. Lanes loaded with RNA from uninduced cells are indicated by minus (-) signs, while those from induced cells are indicated by plus (+) signs.

Targeting portions of the TbmtTFB *ORF* with RNAi has no effect on TbmtTFB mRNA abundance

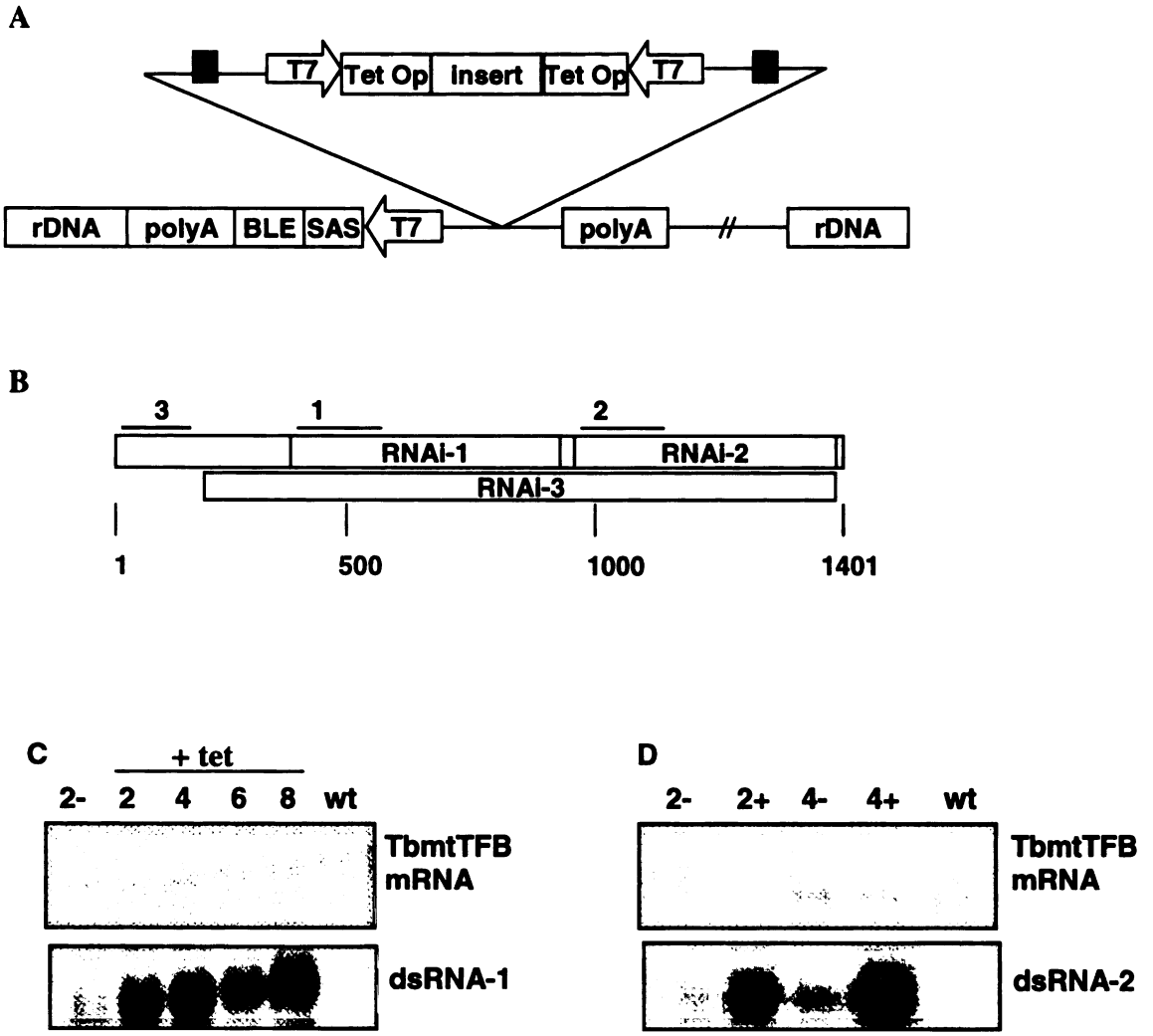


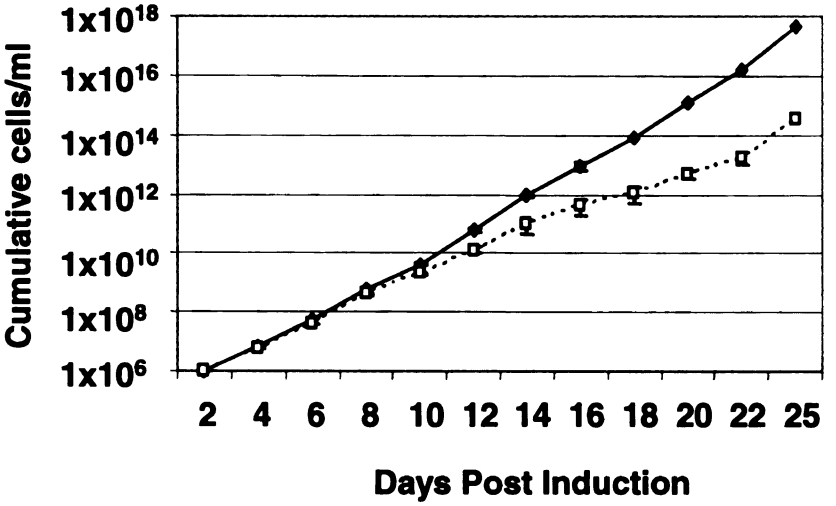
FIGURE 2

Our first two attempts targeted the middle 477 nt (RNAi-1) and a non-overlapping 460 nt region of the 3' end (RNAi-2) of the transcript (Figure 2B). Northern blot analyses indicate that the dsRNA is produced in small amounts in the absence of tetracycline, indicating some leaky expression of the T7 driven cassette (Figure 2, C and D, bottom panels). Addition of 1 μg/ml of tetracycline induces the production of large amounts of dsRNA easily observed in the Northern blots. However, both Northern blot analysis and ribonuclease protection assays (RPAs) indicate that induction with tetracycline did not lead to a decrease in the detectable TbmtTFB mRNA signal (Figure 2 C and D, respectively). In addition, the growth rate of cells expressing RNAi-1 or RNAi-2 dsRNA was identical to that of uninduced cells (data not shown). Because a correlation between dsRNA size and efficiency of RNAi has been described (21), a larger fragment corresponding to 1157 of the ORF was cloned into the vector already containing the middle fragment to generate RNAi-3 (Figure 2B). Upon induction with tetracycline, a slow growth phenotype was observed in cells transfected with this construct. This phenotype was observed in each of three independent experiments, and a representative graph is shown in Figure 3A. Measurement of the slopes of these growth curves indicates that induction results in a 30% decrease in average growth rate. Northern blot analysis using a riboprobe specific to the first 185 nt (not shared with the dsRNA) of the TbmtTFB ORF demonstrated an immediate, but slight decrease in the amount of TbmtTFB transcript (Figure 3B). Unfortunately, the TbmtTFB mRNA is quite rare and difficult to detect on Northern blots. However, quantification of the TbmtTFB signal indicates that it is reduced to 50% of uninduced levels by day 8 followed by a recovery in transcript levels; a pattern often seen in trypanosomes (Figure 3C).

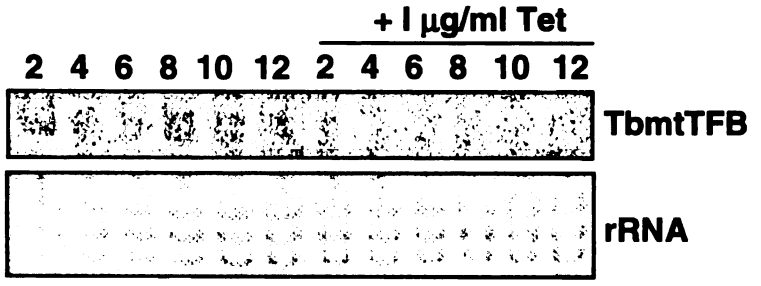
Figure 3. (A) Effect of dsRNAi-3 on cell growth. Growth of RNAi cells induced (doted line/open symbol) or uninduced (solid line/filled symbol). Cumulative cell density was determined by multiplying the cell count by the cumulative dilution factors over the course of the experiment as described in the materials and methods. (B) Northern blot analysis of TbmtTFB mRNA for uninduced and induced cells (top panel). rRNA intensity (bottom panel) was used as a loading control. (C) Quantitation of TbmtTFB signal in induced cells relative to uninduced signal (induced signal/uninduced signal)*100.

Effects of dsRNAi-3 expression

A



B



C

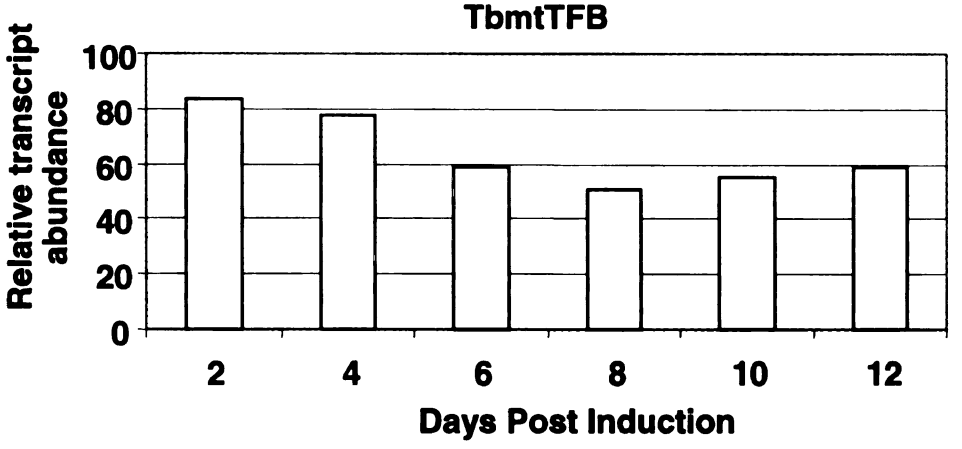


FIGURE 3

In order to determine if the decrease in TbmtTFB mRNA affected the level of steady state mitochondrial transcripts, total RNA was analyzed via northern blot analysis. Several Northern blots were analyzed for many of the transcripts and representative blots are shown in Figure 4. Overall, the reduction in TbmtTFB mRNA level appears to have only slight effects on mitochondrial transcript levels. There is a general trend toward reduction in RNA levels. However, there are transcript specific differences in these effects. The 9S rRNA shows the most pronounced decrease with RNA levels in induced cells reduced to ~60% of that seen in the uninduced lanes. Interestingly, this reduction in the 9S rRNA signal from induced cells is reversible, as this signal appears to increase to levels above the uninduced cells after day 8. This effect was seen on multiple blots, but as shown in the quantitation in Figure 5A, this signal was subject to considerable variability. The large subunit rRNA also showed a slight and repeatable decrease relative to uninduced cells, but in contrast to the 9S rRNA, did not show a pronounced recovery.

In addition to the ribosomal RNAs encoded on the major strand of the maxicircle, mRNAs encoded on both the major and minor strand were also analyzed. While the abundant minor strand ND1 transcript appears relatively unchanged, the less abundant COI transcript remains at ~70-80% of uninduced levels. The major strand mRNAs ND4 and Cyb are reduced farther to levels ~60-70% of the uninduced signal. Several of the transcripts appear as doublets as a result of differing polyA tail lengths (2). Interestingly, in each case the reduction in transcript abundance in induced cells is more dramatic for the upper band than the lower band. This effect is summarized in the graph in Figure 5 D for the ND4 transcript. Although there is a general trend towards reduction of mitochondrial transcripts in cells expressing dsRNA, the slight decrease in the

Northern blot analysis of mitochondrial transcripts

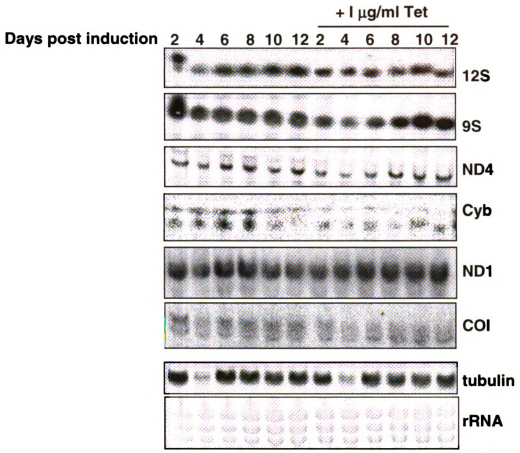


Figure 4: A Northern blot analysis of mitochondrial transcripts from induced and uninduced cells. 10 μg of total RNA from each time point was loaded per lane. Transcripts were detected with end-labeled oligonucleotide probes, stripped and re-hybridized. Ethidium bromide staining of rRNA and tubulin signal is shown to indicate loading.

FIGURE 4

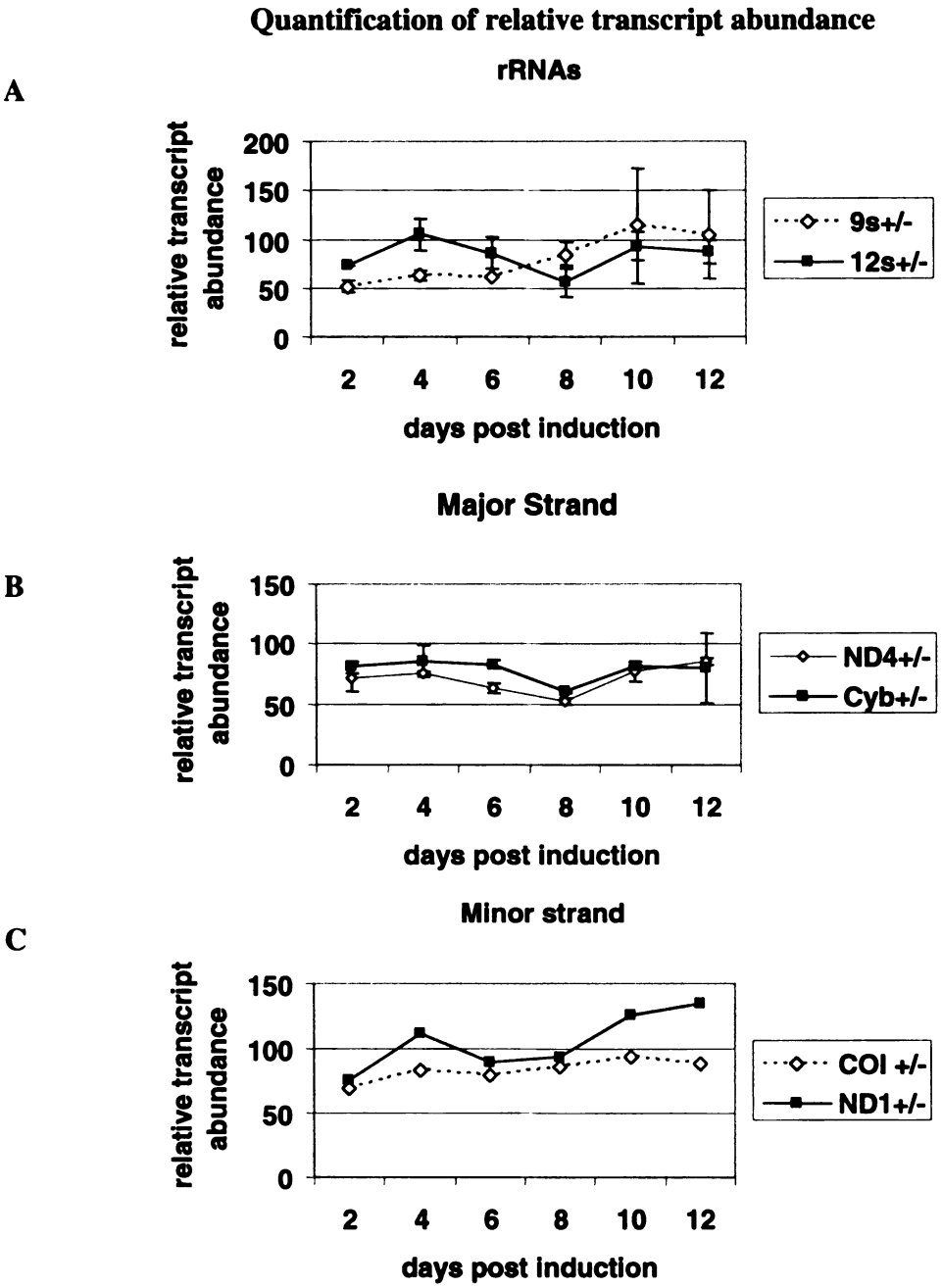


Figure 5: Quantification of relative transcript abundance for various transcripts from induced and uninduced cells. Relative transcript abundance was calculated as (induced signal/uninduced signal)*100.

FIGURE 5

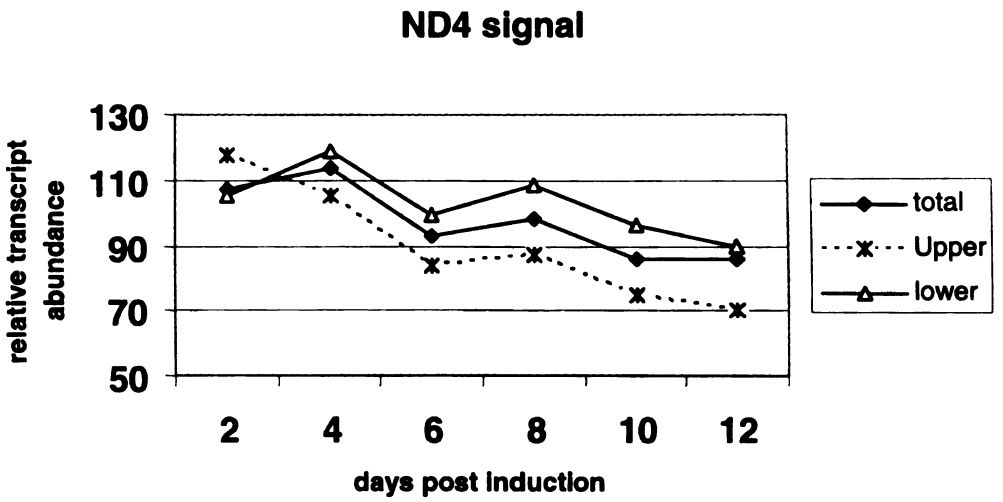


Figure 5D. Quantification of relative transcript abundance of upper and lower ND4 transcripts. The decrease in abundance for the transcript with the longer polyA tail is greater than that seen for the transcript with the shorter polyA tail.

FIGURE 5 (cont'd)

TbmtTFB mRNA caused by RNAi has little overall effects on steady state mitochondrial transcript levels.

Links between mitochondrial transcription and DNA replication have been observed for both yeast and mammals. Likewise, knockdown of TbmtRNAP via RNAi results in a decrease in maxicircle abundance but not minicircle abundance, suggesting a role for transcription in maxicircle maintenance. However, no significant difference was seen in maxicircle or minicircle levels in cells expressing dsRNA against the TbmtTFB as determined by Southern blot analysis (data not shown).

Gene Replacement

Given the subtle decrease in TbmtTFB mRNA upon induction of dsRNA, it was unclear whether the slight but consistent effects seen on mitochondrial RNA levels were genuine and/or a result of the RNAi mediated gene knockdown. Although RNAi is a well documented system in *T. brucei*, there have been other transcripts for which RNAi has not worked. However, these instances are generally undocumented and it is unclear what factors may contribute to RNAi failure. Consequently, we decided to perform the relatively more labor intensive process of double allelic replacement as shown in Figure 6. Sequence flanking the 5' and 3' of the ends of TbmtTFB ORF was cloned into a vector bearing either the blasticidin (BSD) or puromycin (PUR) drug resistance cassette. Cells were transfected with the knockout plasmids and those surviving selection were analyzed via PCR and Southern blotting to confirm TbmtTFB gene replacement. Several independent transfections were performed. Although each knockout plasmid was able to target the TbmtTFB locus when transfected individually, cells surviving selection after

Figure 6. Conditional Double Knockout Strategy **(A)** Diagram of TbmtTFB locus and drug resistance vectors used to replace the wild type allele. **(B)** The inducible, ectopic expression vector for the C-terminal TAP tagged TbmtTFB. **(C)** Optimization of expression of TbmtTFB-TAP. Cells at 2 x 106 cells/ml were induced with increasing amount of tetracycline, harvested 48 hours later. Whole cell lysate was electrophoresed on an 8% SDS-PAGE gel, transferred to PVDF and probed with PAP reagent which detects the protein A moiety of the TAP tag.

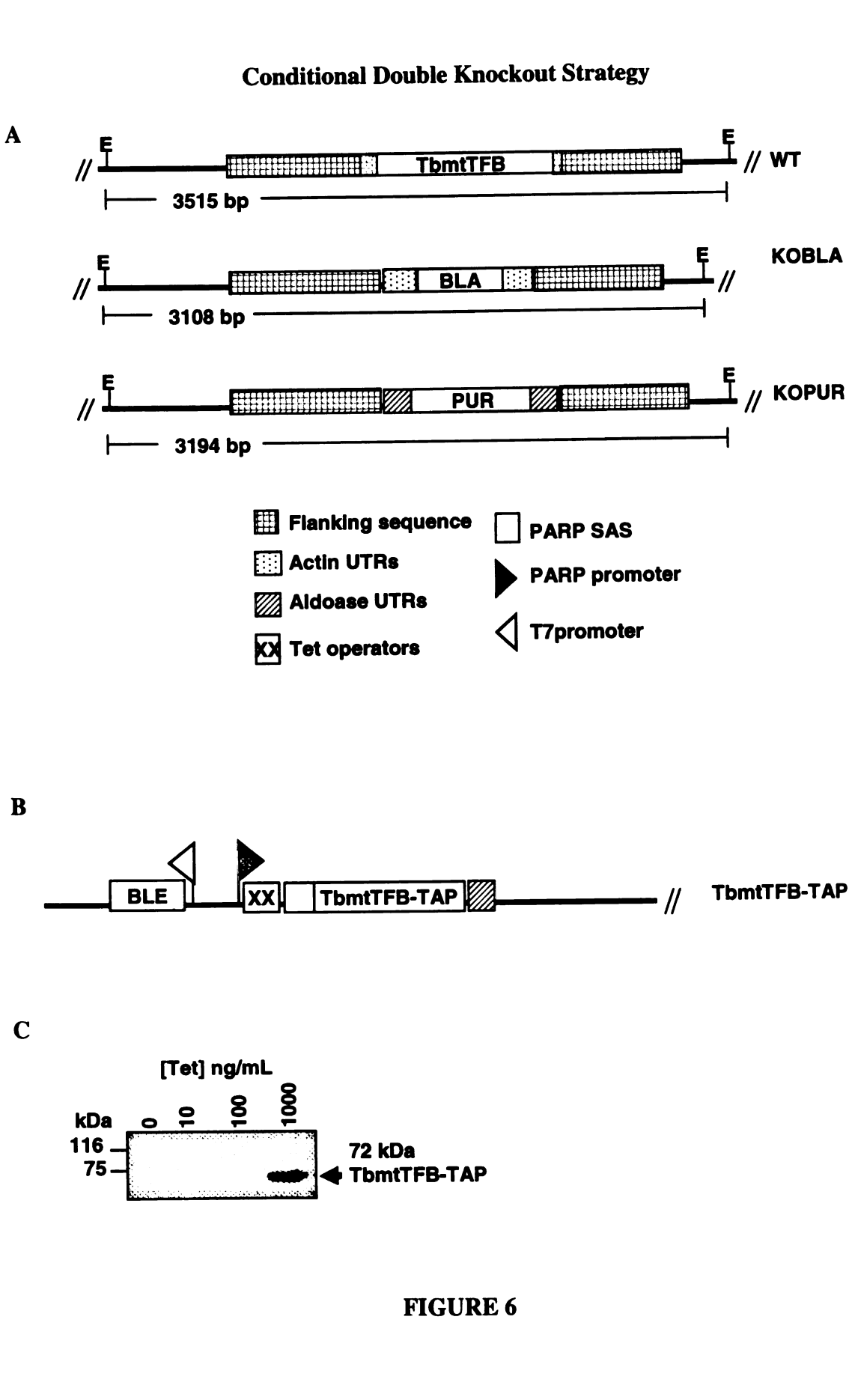

Conditional Double Knockout Strategy
A
E
TbmtTFB
E
// WT
3515 bp
E
BLA
E
KOBLA
3108 bp
E
PUR
E
KOPUR
3194 bp
Flanking sequence
Actin UTRs
Aldoase UTRs
Tet operators
PARP SAS
PARP promoter
T7promoter
B
BLE
XX
TbmtTFB-TAP
TbmtTFB-TAP
C
[Tet] ng/mL
kDa
0
10
100
1000
116
75
72 kDa
TbmtTFB-TAP

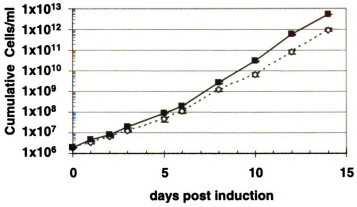
FIGURE 6

sequential or co- transfection with both plasmids were observed to have mis-targeted at least one of the drug resistance genes (data not shown). Because this result indicated that the product of the TbmtTFB gene is essential in *T. brucei*, we transfected a single knockout (TbmtTFB-/+) strain with a tetracycline-inducible *T. brucei* expression vector containing an ectopic copy of the TbmtTFB ORF fused at the C-terminus with a tandem affinity purification (TAP) tag to create strain TbmtTFB-/+::TbmtTFBTAP. Stably transfected cells were screened for their ability to express TbmtTFBTAP upon induction of tetracycline, and expression was optimized (Figure 6C).

During the initial TbmtTFBTAP optimization experiments, a slight increase in growth rate was observed for the cells expressing TbmtTFBTAP relative to uninduced controls. Because this effect complemented the decrease seen in the RNAi experiments described above, we decided to investigate this phenotype further. Growth curves were performed using the TbmtTFB-/+::TbmtTFBTAP cell line in the presence or absence of 1 μg/ml tetracycline. Cells were counted and harvested every other day for RNA, DNA and protein for a period of 2 weeks. Three independent experiments were performed in duplicate and a representative growth curve is shown in Figure 7A. The Western blot shown in Figure 7B indicates that tight regulation of TbmtTFB-TAP expression was maintained throughout the course of the experiment.

Expression of TbmtTFB results in a growth rate increase

A



B

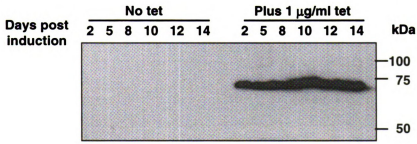


Figure 7: Effect of TbmtTFBTAP expression on cell growth. **(A)** Growth of *T. brucei* cells with tet (solid symbol) or without (open symbol). Total cells/ml determined as in Figure 3. (B) Western blot of whole cell lysates with PAP reagent. Tight regulation of TbmtTFBTAP expression was maintained throughout the experiment.

FIGURE 7

In order to determine if the growth rate increase seen upon expression of the TbmtTFBTAP protein was accompanied by a change in steady state mitochondrial transcript levels, total RNA was analyzed via Northern blot (Figure 8). As seen for the RNAi strain, differences between RNA from induced cells versus uninduced controls were slight. The increase seen for the mRNA levels in cells expressing the TbmtTFB-TAP protein does complement the slight decreases seen in the cells expressing dsRNA. There appears to be no significant difference in the abundance of transcripts with long polyA tails versus those with short polyA tails, however. Unexpectedly, a decrease in the amount of both 12S and 9S ribosomal RNAs was observed upon induction of TbmtTFBTAP. As seen in the RNAi experiment no significant differences were seen for kDNA levels (data not shown).

Northern blot analysis of various mitochondrial transcripts from mtTFB-TAP induction experiment

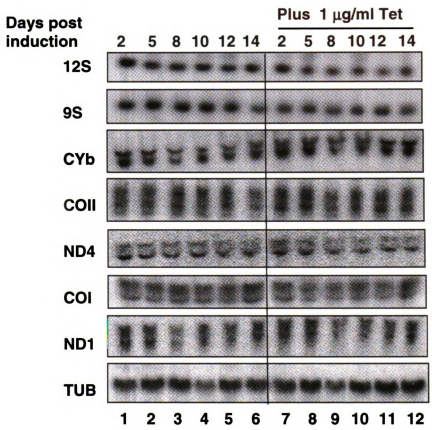
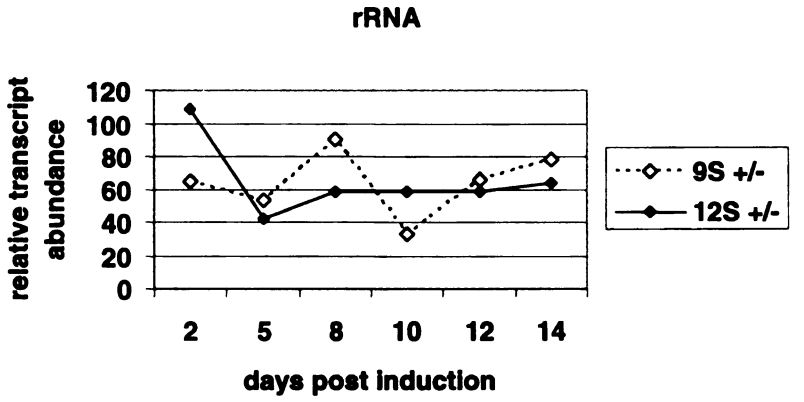


Figure 8. Analysis of the effect of TbmtTFBTAP expression on mtRNA levels. 10 mg of total RNA from each time point was loaded per lane. Transcripts were detected with end-labeled oligonucleotide probes. Tubulin signal is shown to indicate loading.

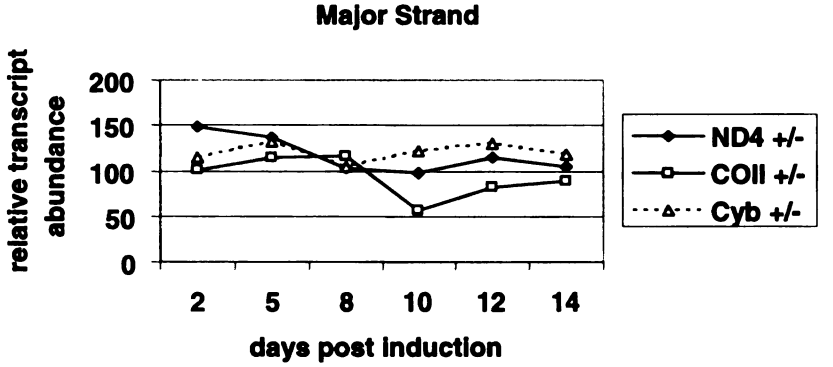
FIGURE 8

Quantification of relative transcript abundance

A



B



C

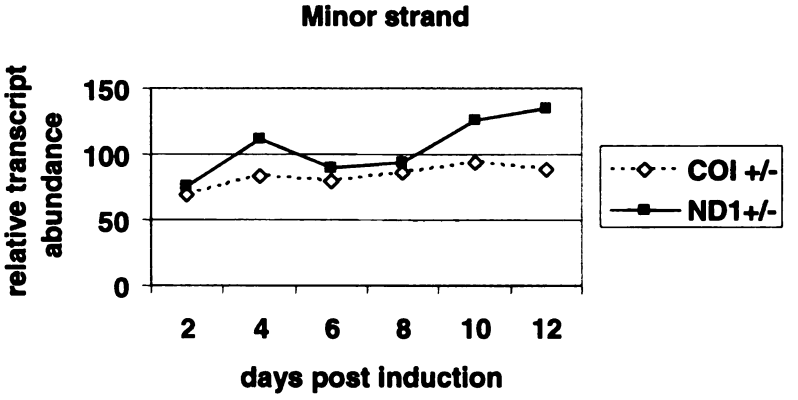


Figure 8: Quantification of relative transcript abundance for various transcripts from induced and uninduced cells. Relative transcript abundance was calculated as (induced signal/uninduced signal)*100.

FIGURE 8 (cont'd)

Discussion:

Role of TbmtTFB in *T. brucei*

Our RNAi attempts result in only a slight decrease in the target TbmtTFB transcript level with unknown effects on the level of the TbmtTFB protein. Interestingly, it has been shown that about 20% of siRNA is found associated with polyribosomes in *T. brucei*, suggesting an interaction with the translational machinery (11). It is theoretically possible that expression of the TbmtTFB protein is reduced due to an siRNA inhibition of translation. However, without antibodies to TbmtTFB, we cannot know for sure. A simpler explanation for the growth decrease consistently seen after day 7 post-induction is the stress on the cell caused by excessive dsRNA production. However, the lack of any growth defect in the other RNAi experiments producing similar amounts of dsRNA would argue against this trivial explanation.

It does appear that several mitochondrial transcripts decrease upon induction with tetracycline, however these effects are too slight and variable to be convincing given the incomplete knockdown of the TbmtTFB mRNA and unknown effect on the protein level. In contrast to the growth decrease seen during the RNAi knockdown of TbmtTFB, the expression of a C-terminal TAP tagged TbmtTFB results in an increased trypanosome growth rate. Again, however, effects on steady state mitochondrial transcript levels are slight and inconclusive.

If TbmtTFB is a transcription factor in *T. brucei* it is not necessarily surprising that a moderate increase/decrease in the expression of this gene would not have an effect on the steady state levels of various mitochondrial transcripts. In yeast, estimates for yeast mitochondria indicate that MTF1 is limiting with only about 20 molecules per cell

in contrast to the 50 molecules of RPO41 and 10-30 copies mtDNA (22). However, in spite of this, overexpression in yeast of RPO41 and/or MTF1 separately or together did not result in an increase in mitochondrial transcript levels (22)(and references therein). Consequently, slight perturbations of the TbmtTFB level may have little impact on transcription. Additionally, as is the case for mitochondrial gene expression in other organisms, post-transcriptional processing and stability appear to play a dominant role in mitochondrial gene expression in *T. brucei*. For example, work from the Hajduk lab has shown that while transcription rates are identical for bloodstream and procyclic form trypanosomes, the rRNAs are 30-fold more abundant in procyclics suggesting that regardless of transcription rate, stability has the final say, at least for these transcripts (26). Consequently, experiments looking at the rate of transcription in isolated organelles may be needed to determine if TbmtTFB affects the rate of transcription in *T. brucei*.

One important feature of mitochondrial transcripts in *T. brucei* that is correlated with developmental regulation is polyA tail length (2). Interestingly, two distinct pathways for RNA turnover exist in *T. brucei* mitochondria; a slow, polyA tail-independent pathway, and a rapid turnover specific for transcripts bearing polyA tails (27). Consequently, transcripts with longer polyA tails may be less stable and more sensitive to slight perturbations in transcription rate. Therefore, it is interesting for several of the transcripts, those with longer polyA tails appear be more sensitive to RNAi against TbmtTFB, although this differential sensitivity was not observed in the overexpression experiments.

Analysis of Alignment:

Much of the early work describing interactions of the yeast mitochondrial RNA polymerase with the transcription factor mtTFB and the holoenzyme with yeast promoters worked under the assumption that the yeast mtTFB was homologous to the RpoD sigma factor. Consequently, many of the mutagenesis studies were performed on regions with proposed homology to sigma, and are not conserved in the other mtTFBs (10, 20, 35). Indeed, it is now known that these regions are in fact not homologous to sigma factors. With the identification of homologues in other organisms, it is worth while to re-evaluate this mutagenesis data and reinterpret it in light of recent findings. Although many of the mtTFB mutations described in these studies were not deleterious, a few were essential and several exhibited a dependence of the holoenzyme on supercoiled templates. Interestingly, among the 3 clusters of mutations identified that impair interaction with RPO41(10), cluster A is not conserved among other mtTFB genes which instead show a stronger similarity to KsgA (Figure 1). Likewise, cluster C is an insertion that is specific to fungal mtTFBs, and mutations in cluster C, specifically S218, I221K and D225G exist in an unstructured loop-like region which extends away from the rest of the protein in the crystal structure of sc-mtTFB (32). Consequently, given the lack of conservation of these residues in the other mtTFB homologues, interactions between mtRNAP and mtTFB have likely co-evolved differently to recognize divergent promoters. In contrast, the residues in cluster B are conserved in all mtTFBs and span KsgA domains IV, and V as well one of the domains specifically shared only by the mtTFB proteins. In addition, mutations in regions between cluster B and C (178-189) result in a dependence on supercoiled templates *in vitro*(34), and these residues are more

highly conserved. Consequently, the conservation of residues in this region may be important for interact with the mtRNAP and/or promoter in a manner that is shared among all eukaryotes.

Interestingly, mammalian transcription requires TFAM which bends DNA to decrease the energy required for promoter melting. Supercoiled DNA has a similar effect on decreasing the energy of promoter melting, perhaps explaining why residues in the yeast mtTFB that result in a dependence on supercoiled templates are not conserved in the mammalian mtTFBs. Interestingly, *T. brucei* kDNA is not supercoiled *in vivo*, however database searches of the *T. brucei* genome indicate that there are no TFAM homologues (data not shown). Instead, there is a small histone like protein associated with kDNA (40). RNAi against this protein results in an increase in mitochondrial transcripts, probably due to increased accessibility to the DNA, suggesting that this does not function analogously to TFAM (1). Given recent evidence indicting that the mtRNAP has an innate ability to recognize promoters in mitochondrial DNA, the associated proteins therefore assist in other steps of the process, such as promoter opening.

Possible other functions for TbmtTFB

The original function of mtTFB was likely to modify the small subunit ribosomal RNA (SSUrRNA). Interestingly, the methyltransferase, and the modification for which it is responsible, is conserved across all domains of life as well as in organelles (28). One notable exception is the *S. cerevisiae* mitochondrial 12S rRNA. Given the divergence of the MTF1 gene from the methyltransferase homologues, it is not surprising that this modification does not occur. Mammalian mitochondrial SSU rRNA is at least partially

modified, and TFBM1 and its paralogue, TFBM2 are presumably both capable of this activity (33). Interestingly, methyltransferase domains are less conserved in TFBM2 and it is at least 2 orders of magnitude more efficient than TFBM1 in stimulating transcription (13). Mutagenesis studies have shown that the methyltransferase activity in TFBM1 and 2 is not required for activity as a transcription factor (25). It is not known, however, if the methyltransferase activity is required for mitochondrial biogenesis *in vivo*. It is possible that the methylation of SSU rRNA in mammalian mitochondria is important for mitoribosome biogeneis.

KsgA mutants in bacteria are viable, although their doubling time increases and their ribosomes require a greater number of factors for translation *in vitro*. They also show more translation defects such as frameshift and nonsense mutations. It is possible that TbmtTFB is not a transcription factor but rather serves only to modify the highly degenerate and diminutive mitochondrial 9S rRNA. All mitochondrial SSU rRNAs are reduced in length relative to their nuclear encoded homologues, but kinetoplast rRNA is among the shortest (37). Such a function for TbmtTFB would explain the lack of effect on steady state RNA levels in spite of the slow growth phenotype observed in the dsRNA producing strains and the growth increase seen in cells expressing TbmtTFBTAP. These growth effects could be the result of changes in translation efficiency. Unfortunately, there are no antibodies to mitochondrially encoded proteins in *T. brucei*, so we were unable to test this hypothesis. If mitochondrial RNA levels remain unaffected in the conditional double knockout strain when expression of the the ectopic TbmtTFBTAP is downregulated, it would interesting to determine whether or not the 9S rRNA is found

associated with mitochondrial polysomes or if mitochondrial translation rates differ between wild-type and knockout cells.

LITERATURE CITED

1. **Avliyakulov, N. K., J. Lukes, and D. S. Ray.** 2004. Mitochondrial histone-like DNA-binding proteins are essential for normal cell growth and mitochondrial function in *Crithidia fasciculata*. Eukaryot Cell **3:**518-26.

2. **Bhat, G. J., A. E. Souza, J. E. Feagin, and K. Stuart.** 1992. Transcript-specific developmental regulation of polyadenylation in *Trypanosoma brucei* mitochondria. Mol Biochem Parasitol **52:**231-40.

3. **Biebinger, S., L. E. Wirtz, P. Lorenz, and C. Clayton.** 1997. Vectors for inducible expression of toxic gene products in bloodstream and procyclic *Trypanosoma brucei*. Mol Biochem Parasitol **85:**99-112.

4. **Bogenhagen, D. F., and N. F. Insdorf.** 1988. Purification of *Xenopus laevis* mitochondrial RNA polymerase and identification of a dissociable factor required for specific transcription. Mol Cell Biol **8:**2910-6.

5. **Burger, G., M. W. Gray, and B. F. Lang.** 2003. Mitochondrial genomes: anything goes. Trends Genet **19:**709-16.

6. **Carrodeguas, J. A., S. Yun, G. S. Shadel, D. A. Clayton, and D. F. Bogenhagen.** 1996. Functional conservation of yeast mtTFB despite extensive sequence divergence. Gene Expr **6:**219-30.

7. **Cermakian, N., T. M. Ikeda, R. Cedergren, and M. W. Gray.** 1996. Sequences homologous to yeast mitochondrial and bacteriophage T3 and T7 RNA polymerases are widespread throughout the eukaryotic lineage. Nucleic Acids Res **24:**648-54.

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

9. **Clement, S. L., and D. J. Koslowsky.** 2001. Unusual organization of a developmentally regulated mitochondrial RNA polymerase (TBMTRNAP) gene in *Trypanosoma brucei*. Gene **272:**209-18.

10. **Cliften, P. F., J. Y. Park, B. P. Davis, S. H. Jang, and J. A. Jaehning.** 1997. Identification of three regions essential for interaction between a sigma-like factor and core RNA polymerase. Genes Dev **11:**2897-909.

11. **Djikeng, A., H. Shi, C. Tschudi, S. Shen, and E. Ullu.** 2003. An siRNA ribonucleoprotein is found associated with polyribosomes in *Trypanosoma brucei*. Rna **9:**802-8.

12. **Doolittle, R. F., D. F. Feng, S. Tsang, G. Cho, and E. Little.** 1996. Determining divergence times of the major kingdoms of living organisms with a protein clock. Science **271:**470-7.

13. **Falkenberg, M., M. Gaspari, A. Rantanen, A. Trifunovic, N. G. Larsson, and C. M. Gustafsson.** 2002. Mitochondrial transcription factors B1 and B2 activate transcription of human mtDNA. Nat Genet **31:**289-94.

14. **Fisher, R. P., and D. A. Clayton.** 1988. Purification and characterization of human mitochondrial transcription factor 1. Mol Cell Biol **8:**3496-509.

15. **Fisher, R. P., and D. A. Clayton.** 1985. A transcription factor required for promoter recognition by human mitochondrial RNA polymerase. Accurate initiation at the heavy- and light-strand promoters dissected and reconstituted in vitro. J Biol Chem **260:**11330-8.

16. **Fisher, R. P., J. N. Topper, and D. A. Clayton.** 1987. Promoter selection in human mitochondria involves binding of a transcription factor to orientation-independent upstream regulatory elements. Cell **50:**247-58.

17. **Grams, J., J. C. Morris, M. E. Drew, Z. Wang, P. T. Englund, and S. L. Hajduk.** 2002. A trypanosome mitochondrial RNA polymerase is required for transcription and replication. J Biol Chem **277:**16952-9.

18. **Gray, M. W., G. Burger, and B. F. Lang.** 2001. The origin and early evolution of mitochondria. Genome Biol **2:**REVIEWS1018.

19. **Jang, S. H., and J. A. Jaehning.** 1991. The yeast mitochondrial RNA polymerase specificity factor, MTF1, is similar to bacterial sigma factors. J Biol Chem **266:**22671-7.

20. **Karlok, M. A., S. H. Jang, and J. A. Jaehning.** 2002. Mutations in the yeast mitochondrial RNA polymerase specificity factor, Mtf1, verify an essential role in promoter utilization. J Biol Chem **277:**28143-9.

21. **LaCount, D. J., and J. E. Donelson.** 2001. RNA interference in African trypanosomes. Protist **152:**103-11.

22. **Mangus, D. A., S. H. Jang, and J. A. Jaehning.** 1994. Release of the yeast mitochondrial RNA polymerase specificity factor from transcription complexes. J Biol Chem **269:**26568-74.

23. **Matsushima, Y., R. Garesse, and L. S. Kaguni.** 2004. *Drosophila* mitochondrial transcription factor B2 regulates mitochondrial DNA copy number and transcription in schneider cells. J Biol Chem **279:**26900-5.

24. **McCulloch, V., B. L. Seidel-Rogol, and G. S. Shadel.** 2002. A human mitochondrial transcription factor is related to RNA adenine methyltransferases and binds S-adenosylmethionine. Mol Cell Biol **22:**1116-25.

25. **McCulloch, V., and G. S. Shadel.** 2003. Human mitochondrial transcription factor B1 interacts with the C-terminal activation region of h-mtTFA and stimulates transcription independently of its RNA methyltransferase activity. Mol Cell Biol **23:**5816-24.

26. **Michelotti, E. F., M. E. Harris, B. Adler, A. F. Torri, and S. L. Hajduk.** 1992. *Trypanosoma brucei* mitochondrial ribosomal RNA synthesis, processing and developmentally regulated expression. Mol Biochem Parasitol **54:**31-41.

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

28. **O'Farrell, H. C., J. N. Scarsdale, and J. P. Rife.** 2004. Crystal structure of KsgA, a universally conserved rRNA adenine dimethyltransferase in *Escherichia coli.* J Mol Biol **339:**337-53.

29. **Parisi, M. A., and D. A. Clayton.** 1991. Similarity of human mitochondrial transcription factor 1 to high mobility group proteins. Science **252:**965-9.

30. **Rantanen, A., M. Gaspari, M. Falkenberg, C. M. Gustafsson, and N. G. Larsson.** 2003. Characterization of the mouse genes for mitochondrial transcription factors B1 and B2. Mamm Genome **14:**1-6.

31. **Schinkel, A. H., M. J. Koerkamp, E. P. Touw, and H. F. Tabak.** 1987. Specificity factor of yeast mitochondrial RNA polymerase. Purification and interaction with core RNA polymerase. J Biol Chem **262:**12785-91.

32. **Schubot, F. D., C. J. Chen, J. P. Rose, T. A. Dailey, H. A. Dailey, and B. C. Wang.** 2001. Crystal structure of the transcription factor sc-mtTFB offers insights into mitochondrial transcription. Protein Sci **10:**1980-8.

33. **Seidel-Rogol, B. L., V. McCulloch, and G. S. Shadel.** 2003. Human mitochondrial transcription factor B1 methylates ribosomal RNA at a conserved stem-loop. Nat Genet **33:**23-4.

34. **Shadel, G. S., and D. A. Clayton.** 1995. A *Saccharomyces cerevisiae* mitochondrial transcription factor, sc- mtTFB, shares features with sigma factors but is functionally distinct. Mol Cell Biol **15:**2101-8.

35. **Shadel, G. S., and D. A. Clayton.** 1995. A *Saccharomyces cerevisiae* mitochondrial transcription factor, sc-mtTFB, shares features with sigma factors but is functionally distinct. Mol Cell Biol **15:**2101-8.

36. **Sipiczki, M.** 2000. Where does fission yeast sit on the tree of life? Genome Biol **1:**REVIEWS1011.

37. **Sloof, P., J. Van den Burg, A. Voogd, R. Benne, M. Agostinelli, P. Borst, R. Gutell, and H. Noller.** 1985. Further characterization of the extremely small mitochondrial ribosomal RNAs from trypanosomes: a detailed comparison of the 9S and 12S RNAs from *Crithidia fasciculata* and *Trypanosoma brucei* with rRNAs from other organisms. Nucleic Acids Res **13:**4171-90.

38. **Wang, Z., J. C. Morris, M. E. Drew, and P. T. Englund.** 2000. Inhibition of *Trypanosoma brucei* gene expression by RNA interference using an integratable vector with opposing T7 promoters. J Biol Chem **275:**40174-9.

39. **Wirtz, E., M. Hoek, and G. A. Cross.** 1998. Regulated processive transcription of chromatin by T7 RNA polymerase in *Trypanosoma brucei*. Nucleic Acids Res **26:**4626-34.

40. **Xu, C. W., J. C. Hines, M. L. Engel, D. G. Russell, and D. S. Ray.** 1996. Nucleus-encoded histone H1-like proteins are associated with kinetoplast DNA in the trypanosomatid *Crithidia fasciculata.* Mol Cell Biol **16:**564-76.

CHAPTER V

Summary and Future Research

SUMMARY

Relatively little is known about the scope of mitochondrial transcription in eukaryotes because this has only been elaborated for two relatively close clades: animals and fungi. Whole genome sequence comparisons suggest that new components of the mitochondrial proteome have continued to evolve after the divergence of the major taxa. It is clear from sampling these relatively few taxa that group specific evolution has resulted in distinct differences in the mechanisms for transcription of the mitochondrial genome, both in *cis* elements and *trans*-acting factors. The precise origin of the mitochondrial transcription apparatus is uncertain. Since *T. brucei* belongs to one of the earliest branches of eukaryotes that possess mitochondria, learning more about the differences and similarities between trypanosomes and other organisms will provide evolutionary insight into the development of this unusual transcription machinery. Additionally, the potential to discover novel components and/or mechanisms of transcription in this ancient organism opens up the possibility of finding much needed chemotherapeutic agents against this pathogen.

Mitochondrial transcription in *T. brucei* is complex and developmentally regulated. This regulation involves the transcription of ribosomal and protein coding genes localized to the maxicircle DNA as well as the transcription of hundreds of minicircle-encoded guide RNAs. Although there has been extensive research into the process of RNA editing in *T. brucei*, there has been little research published regarding the transcription of the kinetoplast genome. The objective of this study has been to learn more about mitochondrial transcription in *T. brucei* by identifying and characterizing

protein components and promoter elements required for RNA synthesis in this unusual organelle.

Mapping transcription start sites

Transcription start sites have been mapped for several minicircle gRNA genes (12), but no maxicircle transcription initiation sites had been identified prior to this work. In *T. brucei*, transcription has been shown to initiate within the variable region of the maxicircle at least 1.2 kb upstream of the 5' end of the 12S rRNA and proceed polycistronically through the genome (10). This transcript is processed rapidly, as precursors spanning more than two genes have not been detected. Since this precursor is so unstable, and because the initiation site has not been precisely mapped, this region does not make an ideal area to use as a promoter for *in vitro* transcription assays.

In *T. brucei*, two classes of transcripts are produced from two distinct mitochondrial genome components. Guide RNAs (gRNAs) are usually minicircle encoded and exist as primary transcripts, while the maxicircle encoded rRNAs and mRNAs are processed from a polycistronic precursor. The genes for the gRNAs gMURF2-II and gCYb(560) each have uncommon kDNA locations not typically associated with transcription initiation events. I have demonstrate that the conserved maxicircle gRNA, gMURF2-II, has an unusual location within the ND4 gene. This is the first report of a completely intragenic gene in kinetoplast DNA. In addition, the gMURF2-II and ND4 transcripts are generated by distinctly different events; the ND4 mRNA is processed from a polycistronic precursor, while transcription of the gRNA initiates downstream of the 5' end of the ND4 gene. The gCYb(560) gene has an atypical

minicircle location in that it is not flanked by the inverted repeat sequences that surround the majority of minicircle gRNA genes. These data show that the mature gCYb(560) gRNA is also a primary transcript and that the 5' end heterogeneity previously observed for this gRNA is a result of multiple transcription initiation sites and not imprecise 5' end processing. Together, these data indicate that gRNA genes represent individual transcription units regardless of genomic context and suggest a complex mechanism for mitochondrial gene expression in *T. brucei.* In addition, this region thus provides an excellent area to study the mechanism by which the *T. brucei* transcriptional and/or processing machinery discriminates between these events. This work was described in a publication this year in *Eukaryotic Cell* (5).

Conserved mitochondrial RNA polymerase subunits in *T. brucei* mtRNAP

A previous survey had indicated that the gene for a single subunit catalytic core mtRNAP homologous to the bacteriophage T7 was widespread in many eukaryotic lineages, but was not detected in close relatives of *T. brucei* (2). Using a modified degenerate PCR technique, I was able to identify the conserved catalytic subunit in *T. brucei* (4). In addition, I found that the *TbmtRNAP* mRNA is one of several mature mRNA species that are post-transcriptionally processed from a stable, polycistronic precursor. This finding that the *TbmtRNAP* gene is transcribed into two distinct mRNAs subject to differential regulation during the *T. brucei* life cycle suggests that mitochondrial differentiation might be achieved in part through the regulated expression of this gene. Further investigation of the *TbmtRNAP* protein levels are required to

determine the significance of the regulation of the processing and stability of this transcript.

mtTFB

In addition to identifying the *TbmtRNAP*, I have also found a homologue to the dual-function mitochondrial transcription factor mtTFB in *T. brucei*. Initial attempts using RNA interference indicate that the *TbmtTFB* belongs to a growing number of under-reported genes in *T. brucei* that appear to be "immune" to RNAi as the presence of *TbmtTFB* dsRNA does not result in a decrease of the target *TbmtTFB* transcript. Although a final RNAi attempt resulted in a decreased growth rate, the reduction in the target *TbmtTFB* mRNA was minimal, and only minor and variable differences were seen in the steady-state levels of mitochondrial transcripts. Given the inconclusive nature of these data, we cannot be certain of the precise function of the *TbmtTFB*.

Evolutionary significance: a potential intermediate?

The conservation of the T7 like mtRNAP in the deeply branching *T. brucei* linear supports the supposition that the mtRNAP was recruited early in the course of eukaryotic evolution (3), as trypanosomes are one of the first organisms to have mitochondria. A puzzling and significant difference between mtRNAP and T7 RNAP is the requirement of the mtRNAP for additional factors for mitochondrial transcription initiation. Although, as with the T7 RNAP, promoter specificity appears to reside within the mtRNAP itself (6, 8), during the course of evolution it has become dependent on other factors to initiate

transcription. An interesting question arises from these facts: If the mtRNAP was acquired early in evolution, when did the requirement for accessory factors arise?

Phylogenetic analyses suggest that the mtTFB groups specifically with α-proteobacterial KsgA protein (data not shown), suggesting that this gene originated from the original endosymbiont. However, database searches indicate that this gene is not encoded in the genome of even the most eubacterial mitochondrion of the protist *Reclinomonas americana*. The mtDNA of *R. americana* instead still encodes the subunits for eubacterial RNA polymerase, and retains putative eubacterial promoters. Presumably, this organism utilizes the eubacterial multi-subunit RNAP to transcribe its genome, and the T7-like RNAP was recruited for this function sometime after Jakobids split from the last common ancestor. The absence of an mtTFB homologue in the *R. americana* mtDNA suggests that the gene for the KsgA may have been transferred to the nucleus before the recruitment of the T7 RNAP. Given the universality of the modification performed by KsgA and its homologues, including in most mitochondria, this protein was likely also re-imported into the mitochondrion to methylate the small subunit rRNA sometime before the recruitment of the ssRNAP. So, it is possible that there exists intermediate organisms among the early branching eukaryote for which the mtRNAP behaves like its phage cousins and does not require any accessory factor for transcription. Kinetoplastids would be an excellent candidate for this intermediate given their branching order. Consequently, should we determine that the *TbmtTFB* does not affect transcription in kDNA, it will be an evolutionarily significant discovery. The development of the conditional TbmtTFB knockout strain will allow us to determine whether or not *TbmtTFB* is a transcription factor like its homologues, which would

suggest that its dual function was set very early in eukaryotic evolution, or whether it evolved the ability to assist the mtRNAP in transcription initiation at a later time.

Neither the protein components nor the potential promoter elements for mitochondrial transcription in *T. brucei* were known when I began this project. Consequently, this work has made a significant contribution to the field of mitochondrial transcription in *T. brucei*. Likewise, it has laid the foundation for the development of an *in vitro* transcription system which will allow us to learn more about the regulation of mitochondrial gene expression in trypanosomes. Given the remarkable nature of the kinetoplast genome and the early divergence of this ancient protist, it would not be surprising to discover that the transcription apparatus contains unique enzymes and/or mechanisms of transcription. In particular, we propose that there may be maxicircle or minicircle specific transcription factors, and/or transcription factors that specifically regulate gRNA transcription. Additionally, links between transcription and other aspects of mitochondrial genome expression such as replication and RNA processing have been described for other organisms. Consequently, a better understanding of the mitochondrial transcription apparatus will also allow us to examine similar links in *T. brucei* and may expand drug targets considerably.

FUTURE WORK

Mapping transcription start sites

On of the objectives of this study has been to find transcription start sites on the maxicircle genome as these regions are candidates for promoters. Similar to T7 RNAP promoters, mitochondrial promoter regions lie just upstream of and encompass the

transcription start site. Currently, the only transcription start site known on the *T. brucei* maxicircle is that for the gRNA gMURF2-II (5). The majority of the genes on the maxicircle, however, are thought to be transcribed as part of a polycistronic precursor, the start site for which is unknown. Using a ribonuclease protection assay, Michelotti et al, demonstrated that an unstable RNA precursor originates within the variable region (VR) at least ~1200 nt upstream of the mapped 5' end of the processed 12S rRNA (10). It is unknown how much farther upstream the precursor may originate as the upstream portions of the VR had yet to be cloned at the time of this study and were not tested. Interestingly, given the link between the *TbmtRNAP* and maxicircle maintenance, an origin of replication has also been mapped to this region (1). It is known from other studies that both strands of the VR region are transcribed at low levels (11).

Our lab is currently using the "Differential Display of RNA Ligase Mediated Rapid Amplification of cDNA Ends" (DDRLACE) technique to determine the start site of both the major and minor strand precursors. Total RNA is treated, ligated, and reverse transcribed into cDNA that is then PCR amplified with a gene specific primer and a ligated-linker specific primer. Amplification products that appear only in the positive control lanes reflect primary transcripts and will be cloned and sequenced to localize the start site. Once a putative 5' end is known, poisoned primer extensions as in Chapter II will be performed to confirm these ends are the result of transcription initiation. The regions surrounding these mapped start sites, as well as the region surrounding the gMURF2-II start site, will then be tested in *in vitro* studies to define promoter elements.

In vitro transcription

Initial recombinant *TbmtRNAP* expression attempts:

The ability to create an *in vitro* transcription system is critical to furthering our understanding of the mechanism of mitochondrial transcription initiation in *T. brucei* and the characterization of elements in mitochondrial DNA required for promoter recognition. Due to its low abundance in cells, *in vitro* mitochondrial transcription assays in the literature have depended in part on the use of recombinant mtRNAP (7, 9). Unfortunately, progress in our lab towards the recombinant expression of the *TbmtRNAP* and consequently the *in vitro* transcription assay has been problematic. The polymerase is insoluble in a variety of systems including baculovirus mediated expression in insect cells (data not shown). Although a significant amount of time was invested into optimizing the expression system, including attempts at co-expression with the putative mtTFB homologue, soluble protein was never obtained. Given this difficulty, at this point it would be prudent to develop an assay with partially purified mitochondrial extracts from *T. brucei*. This assay would take advantage of the gRNA transcription start sites identified in Chapter II and the surrounding kDNA sequence to define the precise nucleotide sequence requirements for transcription initiation.

Function of the *TbmtTFB*

Much to the dismay of researchers in the molecular parasitology community, trypanosomes are adept at surmounting negative selection pressures. This makes them very successful organisms, but often difficult to manipulate. It is clear that the only definitive answer as to the function of the *TbmtTFB* homologue in *T. brucei* mitochondria

will come as a result of either biochemical characterization of this protein or through analysis of double knockouts. The failure to obtain double knockouts so far perhaps suggests that *TbmtTFB* is an essential gene, or at least that those cells lacking the gene grow too slowly and are out-competed by those still retaining a copy during the selection process. Unfortunately, however, this does nothing to clarify its role. Cloning of *T. brucei* cells is difficult as a result of density dependent growth, but our lab has recently improved the survival rate dramatically and experiments are currently underway to determine whether true conditional knockouts have been obtained.

LITERATURE CITED

1. **Carpenter, L. R., and P. T. Englund.** 1995. Kinetoplast maxicircle DNA replication in Crithidia fasciculata and *Trypanosoma brucei*. Mol Cell Biol **15:**6794-803.

2. **Cermakian, N., T. M. Ikeda, R. Cedergren, and M. W. Gray.** 1996. Sequences homologous to yeast mitochondrial and bacteriophage T3 and T7 RNA polymerases are widespread throughout the eukaryotic lineage. Nucleic Acids Res **24:**648-54.

3. **Cermakian, N., T. M. Ikeda, P. Miramontes, B. F. Lang, M. W. Gray, and R. Cedergren.** 1997. On the evolution of the single-subunit RNA polymerases. J Mol Evol **45:**671-81.

4. **Clement, S. L., and D. J. Koslowsky.** 2001. Unusual organization of a developmentally regulated mitochondrial RNA polymerase (TBMTRNAP) gene in *Trypanosoma brucei*. Gene **272:**209-18.

5. **Clement, S. L., M. K. Mingler, and D. J. Koslowsky.** 2004. An intragenic guide RNA location suggests a complex mechanism for mitochondrial gene expression in *Trypanosoma brucei*. Eukaryot Cell **3:**862-9.

6. **Gaspari, M., M. Falkenberg, N. G. Larsson, and C. M. Gustafsson.** 2004. The mitochondrial RNA polymerase contributes critically to promoter specificity in mammalian cells. Embo J.

7. **Mangus, D. A., and J. A. Jaehning.** 1996. Transcription in vitro with *Saccharomyces cerevisiae* mitochondrial RNA-polymerase. Methods Enzymol **264:**57-66.

8. **Matsunaga, M., and J. A. Jaehning.** 2004. Intrinsic promoter recognition by a "core" RNA polymerase. J Biol Chem **279:**44239-42.

9. **Matsunaga, M., S. H. Jang, and J. A. Jaehning.** 2004. Expression and purification of wild type and mutant forms of the yeast mitochondrial core RNA polymerase, Rpo41. Protein Expr Purif **35:**126-30.

10. **Michelotti, E. F., M. E. Harris, B. Adler, A. F. Torri, and S. L. Hajduk.** 1992. *Trypanosoma brucei* mitochondrial ribosomal RNA synthesis, processing and developmentally regulated expression. Mol Biochem Parasitol **54:**31-41.

11. **Myler, P. J., D. Glick, J. E. Feagin, T. H. Morales, and K. D. Stuart.** 1993. Structural organization of the maxicircle variable region of *Trypanosoma brucei:* identification of potential replication origins and topoisomerase II binding sites. Nucleic Acids Res **21:**687-94.

12. **Pollard, V. W., S. P. Rohrer, E. F. Michelotti, K. Hancock, and S. L. Hajduk.** 1990. Organization of minicircle genes for guide RNAs in *Trypanosoma brucei*. Cell **63:**783-90.