Published in final edited form as:

Gene. 2001 July 11; 272(1-2): 209–218.

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

Sandra L. Clement and Donna J. Koslowsky *

Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824, USA

Abstract

We report here the characterization of a developmentally regulated mitochondrial RNA polymerase transcript in the parasitic protozoan, *Trypanosoma brucei*. The 3822 bp protein-coding region of the *T. brucei* mitochondrial RNA polymerase (*TBMTRNAP*) gene is predicted to encode a 1274 amino acid polypeptide, the carboxyl-terminal domain of which exhibits 29–37% identity with the mitochondrial RNA polymerases from other organisms in the molecular databases. Interestingly, the *TBMTRNAP* mRNA is one of several mature mRNA species post-transcriptionally processed from a stable, polycistronic precursor. Alternative polyadenylation of the *TBMTRNAP* mRNA produces two mature transcripts that differ by 500 nt and that show stage-specific differences in abundance during the *T. brucei* life cycle. This alternative polyadenylation event appears to be accompanied by the alternative splicing of a high abundance, non-coding downstream transcript of unknown function. Our 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.

Keywords

Polycistronic mRNA; Post-transcriptional regulation; Non-coding RNA; Alternative polyadenylation

1. 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 (Priest and Hajduk, 1994). The mitochondrial genome of *T. brucei* exists as an unusual structure of several thousand topologically linked circular DNA molecules called a kinetoplast. The kinetoplast contains a heterogeneous population of ~10,000 minicircles of 1 kb as well as 50 copies of the 22 kb maxicircle (Shapiro and Englund, 1995). The maxi-circle genome is analogous to the mitochondrial DNA of other organisms in that it encodes several components required for mitochondrial biogenesis and function. Mini-circle DNA, however, is found only in *T. brucei* and other members of the order Kinetoplastida. This unique genome encodes small guide RNA molecules that provide the sequence information during the post-transcriptional process of U-insertional/deletional editing of mitochondrial

mRNAs (for recent reviews see Stuart et al., 1997; Estevez and Simpson, 1999). 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 (Stuart et al., 1997). 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 localized to both maxicircles and minicircles. Although the past decade has seen 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 (Tracy and Stern, 1995; Cermakian et al., 1997). 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 (Tracy and Stern, 1995). Despite the conservation of the core enzyme, the nature and number of these mitochondrial RNA polymerase (MTRNAP) specificity factors vary among organisms.

As a first step in 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.

2. Materials and methods

2.1. Cell culture and nucleic acid preparation

Trypanosoma brucei brucei procyclic form cells (Istar clone from stock EATRO 164) were grown as previously described (Koslowsky and Yahampath, 1997). 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 (Maizels et al., 1991). Total RNA was isolated as previously described (Koslowsky and Yahampath, 1997). 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.

2.2. Degenerate PCR

Consensus-DEgenerate Hybrid Oligonucleotide Primers were designed with the aid of the web-based program CODEHOP (Rose et al., 1998). 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 (Fig. 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 of each dNTP, and 2.5 units 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.

2.3. Genomic DNA cloning and gene copy number

A high-density microarray 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 (Milligan et al., 1987). 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 T-3') designed using sequence obtained from the degenerate PCR product with a 22 nt T7 promoter oligo (5'-AATTTAATAC GACTCACTAT AG-3') to create a double-stranded T7 promoter. A total of five 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 Hybond 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 *Escherichia 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 μ g of *T. brucei* genomic DNA as described for the P1 clones.

2.4. 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 Clontech Adapter Primer AP1 and a gene-specific primer 3' R1 (5'-GGTATTACGT TTCGCGAGAC GCAAACTGC CG-3') with the following conditions: 94°C for 1 min, five cycles of 94°C for 30 s, 72°C for 4 min, five cycles of 94°C for 30 s, 70°C for 4 min, 20 cycles of 94°C for 30 s 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 Clontech 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* ORF: primer CS-1 (5'-CGTAGAACCC GCAGATGATT CTGAGGC-3'), CS-2 (5'-TGTGAAGCAA GTGATAGCAA ACTACCCCGC TG-3'), and 5' RT-1 (5'-CTTTATCGGT GCCAGACTCA ACC-3'). For each RT-PCR, cDNA was reverse-transcribed from 1 μ g of procyclic poly(A +) RNA, using 40 fmol of gene-specific primer, 12.5 μ M of each dNTP, and 2 units AMV reverse transcriptase (Seikagaku) in a 20 μ l 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 min. PCR was performed on 5.0 μ l of each RT reaction with 0.2 μ M each of the appropriate gene-specific primer and the spliced leader primer, TbSL-sal (5'-GTCGTCGACA ACTAACGCTA TTATTAGAACAG-3'), 0.05 mM of each dNTP, and 2.5 units Taq (Promega) with initial denaturation at 94°C for 2 min followed by 30 cycles of 94°C for 30 s, 60°C for 30 s, 72°C for 30 s and a hold at 72°C for 10 min. Products were resolved on a 2% agarose gel and sequenced from both the 5' and 3' ends using the PCR primers.

2.5. Northern blots

Poly(A +) RNA (4.5 µg) and 30 µg of total RNA were electrophoresed on a 0.66 M formaldehyde 1% agarose gel. The resolved RNA was transferred to a Hybond Nytran membrane using downward capillary transfer and hybridized with an internal 1.8 kb riboprobe generated by digestion of the *Bam*HI clone with *Kpn*I 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 AGATTCATGA-3') and 5' probe (5'-GGCGGTTGAG TCTGGCACCG ATAAAGTGG-3') for the 5' probe and primers T73' UTR (5'-GACTCACTAT AGGGAAGAAT TAAATCCATG-3') and 3' UTR (5'-GATATTAGAA ACTCGCAGAT TACAAAGTGT-3') for the 3' UTR probe. Riboprobes were transcribed using the T7 Maxiscript kit (Ambion). In each case, the blot was hybridized in 50% formamide, 5 × SSPE, 1% SDS at 68°C for 16 h and washed twice at room temperature with 0.5 × SSPE, 0.5% SDS for 15 min and once in 0.1 × SSPE, 0.1% SDS at 68°C for 1–2 h. The relative abundance of the various transcripts was determined by phosphorimager analysis (Molecular Dynamics) and normalized to the signal obtained by rehybridizing the blot with an end-labeled oligo against the β-tubulin gene.

2.6. Sequencing

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

3. Results

3.1. Genomic cloning

CODEHOP primers (Rose et al., 1998) were designed from an alignment of the highly conserved carboxyl termini of several mitochondrial RNA polymerase genes in the database (Fig. 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 (Clayton et al., 1998). 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 (Fig. 2A), and the 6182 bp *Bam*HI fragment was sequenced to construct a genomic contig which was extended by 2 kb with sequence obtained from the *Kpn*-*Eco*RI clone. A Southern blot of *T. brucei* genomic DNA indicated that the *TBMTRNAP* exists as a single-copy gene (Fig. 2B). With only one noted exception (Mair et al., 2000), the *T. brucei* genome contains no introns. Consequently, the genomic sequence contains the mitochondrial RNA polymerase ORF, which consists of 3822 bp corresponding to 1274 amino acids shown in Fig. 3A. The amino terminus of this protein is predicted to form an amphipathic helix (Fig. 3B) and to be targeted to the mitochondrion via the putative sequence in bold in Fig. 3A. A BLASTP search on the entire ORF 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) MTNRAP. This homology is localized to amino acids 689–1274 of the *TBMTRNAP* with 37% identity and 53% similarity to the *Arabidopsis* protein. However, as demonstrated in the alignment in Fig. 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, within one of the highly conserved blocks from which we designed our

degenerate primers, the *TBMTRNAP* sequence deviates from the consensus sequence. However, it is important to note that the residues within this block that have been shown to be critical for activity of the T7 RNA polymerase (Sousa, 1996; Cermakian et al., 1997) are conserved.

3.2. 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 (Vanhamme and Pays, 1995). Therefore, to determine the 5' end of the *TBMTRNAP* mRNA, a modified RACE strategy using three different gene-specific primers (Fig. 2A) and the conserved spliced leader primer was used. Three independent clones derived from the three separate procyclic RACE products mapped the splice acceptor site to an AG dinucleotide at a position 29 nt upstream of the first ATG codon in the *TBMTRNAP* ORF. PCR products amplified from the bloodstream form cDNA library using these same primers did not yield any products extended at the 5' end. The 3' end of the *TBMTRNAP* mRNA was determined by 3' RACE experiments on the bloodstream form cDNA library. Four distinct PCR products of various sizes were obtained from the one pair of primers used for the 3' RACE. Sequencing of these cloned products revealed two size classes of *TBMTRNAP* mRNA with 3' UTRs that differ by 500 nt.

3.3. 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, 4.7, and 6.9 kb (Fig. 5A). The 4.4 and 4.7 kb transcripts are visible in procyclic, bloodstream, and dyskinetoplastid trypanosomes. The larger, 6.9 kb transcript is not visible in the procyclic lane but has been detected at low levels in other Northern blots with 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). Hybridization of a 3' UTR-specific probe solely to the 4.7 kb band demonstrates that the difference in size between the 4.4 and 4.7 kb *TBMTRNAP* transcripts reflects the two distinct 3' UTRs mapped in the 3' RACE experiments (Fig. 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 four 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 (Fig. 5C), which is expressed equally during the life cycle (Clayton, 1988). 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 and 0.7 kb. A Southern blot of *T. brucei* DNA probed with the 3' UTR riboprobe did not reveal any additional related sequences in the *T. brucei* genome (data not shown). These transcripts were not detected with the internal probe located only 220 nt upstream of the 3' UTR riboprobe. Therefore, it is clear that a portion of the sequence present in the 3' UTR of the larger *TBMTRNAP* transcript is also part of the 5' ends of the 2.6 and 0.7 kb transcripts. Consequently, 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 four times more abundant in procyclic than in bloodstream trypanosomes. We are unable to detect a product that extends beyond the mapped splice site acceptor for the *TBMTRNAP* mRNA in 5' RACE experiments using cDNA

from either procyclic or 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 Fig. 5D. Further characterization of these downstream transcripts demonstrates that 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 distal poly(A) site of the *TBMTRNAP* mRNA (data not shown). The 2.6 kb 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 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. A riboprobe located downstream of this distal poly(A) site detected a low abundance transcript of approximately 2.3 kb (data not shown). This suggests that the 2.6 kb transcript is a precursor of the 2.3 kb mature transcript. RT-PCR utilizing gene-specific primers on procyclic poly(A) mRNA suggests that a precursor that spans this region does exist (data not shown). It is unclear why we were unable to amplify this transcript with the spliced leader-specific primer. Our data suggest that this precursor is stable enough to detect via Northern blot, yet stable precursors lacking *trans*-splice leaders are uncommon in *T. brucei*.

4. Discussion

We have identified and cloned a single mitochondrial RNA polymerase gene in *T. 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 *TBMTRNAP* is not surprising. The variation in the length and composition of the non-conserved region has been speculated to play a role in species-specific interactions with other proteins involved in transcription and RNA processing (Cermakian et al., 1997). 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, a 40 kDa protein with limited similarity to eubacterial sigma factors (Jang and Jaehning, 1991; Antoshechkin and Bogenhagen, 1995), and mtTFA, a 25 kDa high mobility group (HMG) protein (Dairaghi et al., 1995). 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 (for review see Priest and Hajduk, 1994), 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 (Priest and Hajduk, 1994). 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 (Erondy and Donelson, 1992). As with other eukaryotic transcripts, *T. brucei* 3' UTRs are known to contain information that affects such processes as mRNA turnover and translatability (Vanhamme and Pays, 1995; Furger et al., 1997). Thus, elements within the 3' UTR of the *TBMTRNAP* mRNA are likely to be important for the developmental regulation of its expression during the *T. brucei* life cycle.

Polyadenylation in *T. brucei* is dependent upon a downstream polypyrimidine track followed by a *trans*-splice site (Matthews et al., 1994; Schurch et al., 1994). *Trans*-splicing is rapid and seems to occur as soon as the splice site is available to the processing machinery (Ullu et al., 1993). This poses 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 splice site must be used less efficiently. 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 relatively abundant 0.7 kb transcript is four 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. This pattern of expression corresponds to that seen for the shorter *TBMTRNAP* mRNA.

It is important to note, however, that the proximal poly(A) site is favored over the distal site in all three forms analyzed as the 4.4 kb transcript is slightly more abundant in each. Although the shorter 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. This leads us to believe that the increase in abundance of the smaller *TBMTRNAP* transcript in procyclic cells is not solely a result of a shift in the choice of poly(A) site. However, given the coordinated increase in the abundance of the 0.7 kb transcript in the procyclic cells and the decrease in abundance 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 mRNA 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 it 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 byproduct 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 the expressed sequence tag (EST) databases available for both *Leishmania* and *Trypanosoma 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 is becoming more apparent with the analysis of whole genome sequences (Olivas et al., 1997). 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 (Teixeira et al., 1995; Clayton and Hotz, 1996; Hausler and Clayton, 1996; Hotz et al., 1997). This type of regulation could be especially important for genes such as the *T. brucei* mitochondrial RNA polymerase, 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 (Miranda et al., 1999). 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.

Abbreviations

CODEHOP, CONsensus-DEgenerate Hybrid Oligonucleotide Primers; ORF, open reading frame; RACE, rapid amplification of cDNA ends; RT-PCR, reverse transcription-polymerase chain reaction; *TBMTRNAP*, *Trypanosoma brucei* mitochondrial RNA polymerase gene; TbSL, *Trypanosoma brucei* spliced leader; UTR, untranslated region.

Acknowledgements

The authors wish to thank Vanessa Leech of the Laboratory for Parasite Genome Analysis at Cambridge University for providing the *T. brucei* P1 genomic library filter and clones. Tom Newman of the MSU Sequencing Facility refined the New England Biolabs Tn7 transposon-based GPSTTM-1 Genome Priming System which was used to sequence the *Bam*HI subclone. Ken Stuart kindly supplied both bloodstream and dyskinetoplastid RNA. This work was supported by NIH grant number AI34155 to D.J.K.

References

- Antoshechkin I, Bogenhagen DF. Distinct roles for two purified factors in transcription of *Xenopus* mitochondrial DNA. *Mol. Cell. Biol* 1995;15:7032–7042. [PubMed: 8524270]
- Cermakian N, Ikeda TM, Miramontes P, Lang BF, Gray MW, Cedergren R. On the evolution of the single-subunit RNA polymerases. *J. Mol. Evol* 1997;45:671–681. [PubMed: 9419244]
- Clayton CE. The molecular biology of the Kinetoplastidae. *Genet. Eng* 1988;7:1–56. [PubMed: 2483088]
- Clayton C, Hotz HR. Post-transcriptional control of PARP gene expression. *Mol. Biochem. Parasitol* 1996;77:1–6. [PubMed: 8784766]
- Clayton C, et al. Genetic nomenclature for *Trypanosoma* and *Leishmania*. *Mol. Biochem. Parasitol* 1998;97:221–224. [PubMed: 9879900]
- Dairaghi DJ, Shadel GS, Clayton DA. Human mitochondrial transcription factor A and promoter spacing integrity are required for transcription initiation. *Biochim. Biophys. Acta* 1995;1271:127–134. [PubMed: 7599198]
- Erondu NE, Donelson JE. Differential expression of two mRNAs from a single gene encoding an HMG1-like DNA binding protein of African trypanosomes. *Mol. Biochem. Parasitol* 1992;51:111–118. [PubMed: 1565127]
- Estevez AM, Simpson L. Uridine insertion/deletion RNA editing in trypanosome mitochondria – a review. *Gene* 1999;240:247–260. [PubMed: 10580144]
- Furger A, Schurch N, Kurath U, Roditi I. 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* 1997;17:4372–4380. [PubMed: 9234695]
- Hausler T, Clayton C. Post-transcriptional control of hsp70 mRNA in *Trypanosoma brucei*. *Mol. Biochem. Parasitol* 1996;76:57–71. [PubMed: 8919995]
- Hotz HR, Hartmann C, Huober K, Hug M, Clayton C. Mechanisms of developmental regulation in *Trypanosoma brucei*: a polypyrimidine tract in the 3'-untranslated region of a surface protein mRNA affects RNA abundance and translation. *Nucleic Acids Res* 1997;25:3017–3026. [PubMed: 9224601]

- Jang SH, Jaehning JA. The yeast mitochondrial RNA polymerase specificity factor, MTF1, is similar to bacterial sigma factors. *J. Biol. Chem* 1991;266:22671–22677. [PubMed: 1939277]
- Koslowsky DJ, Yahampath G. Mitochondrial mRNA 3' cleavage/polyadenylation and RNA editing in *Trypanosoma brucei* are independent events. *Mol. Biochem. Parasitol* 1997;90:81–94. [PubMed: 9497034]
- Mair G, Shi H, Li H, Djikeng A, Aviles HO, Bishop JR, Falcone FH, Gavrilescu C, Montgomery JL, Santori ML, Stern LS, Wang Z, Ullu E, Tschudi C. A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA. *RNA* 2000;6:163–169. [PubMed: 10688355]
- Maizels, RM.; Blaxter, ML.; Robertson, BD.; Selkirk, ME. *Parasite Antigens, Parasite Genes: A Laboratory Manual for Molecular Parasitology*. Cambridge: Cambridge University Press; 1991.
- Matthews KR, Tschudi C, Ullu E. A common pyrimidine-rich motif governs *trans*-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. *Genes Dev* 1994;8:491–501. [PubMed: 7907303]
- Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res* 1987;15:8783–8798. [PubMed: 3684574]
- Miranda S, Foncea R, Guerrero J, Leighton F. Oxidative stress and upregulation of mitochondrial biogenesis in mitochondrial DNA-depleted HeLa cells. *Biochem. Biophys. Res. Commun* 1999;258:44–49. [PubMed: 1022232]
- Olivas WM, Muhrad D, Parker R. Analysis of the yeast genome: identification of new non-coding and small ORF-containing RNAs. *Nucleic Acids Res* 1997;25:4619–4625. [PubMed: 9358174]
- Priest JW, Hajduk SL. Developmental regulation of mitochondrial biogenesis in *Trypanosoma brucei*. *J. Bioenerg. Biomembr* 1994;26:179–191. [PubMed: 8056785]
- Rose TM, Schultz ER, Henikoff JG, Pietrokovski S, McCallum CM, Henikoff S. Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Res* 1998;26:1628–1635. [PubMed: 9512532]
- Schurch N, Hehl A, Vassella E, Braun R, Roditi I. Accurate polyadenylation of procyclin mRNAs in *Trypanosoma brucei* is determined by pyrimidine-rich elements in the intergenic regions. *Mol. Cell. Biol* 1994;14:3668–3675. [PubMed: 7910942]
- Shapiro TA, Englund PT. The structure and replication of kinetoplast DNA. *Annu. Rev. Microbiol* 1995;49:117–143. [PubMed: 8561456]
- Sousa R. Structural and mechanistic relationships between nucleic acid polymerases. *Trends Biochem. Sci* 1996;21:186–190. [PubMed: 8871404]
- Stuart K, Allen TE, Heidmann S, Seiwert SD. RNA editing in kinetoplastid protozoa. *Microbiol. Mol. Biol. Rev* 1997;61:105–120. [PubMed: 9106367]
- Teixeira SM, Kirchhoff LV, Donelson JE. Post-transcriptional elements regulating expression of mRNAs from the amastin/tuzin gene cluster of *Trypanosoma cruzi*. *J. Biol. Chem* 1995;270:22586–22594. [PubMed: 7673251]
- Tracy RL, Stern DB. Mitochondrial transcription initiation: promoter structures and RNA polymerases. *Curr. Genet* 1995;28:205–216. [PubMed: 8529266]
- Ullu E, Matthews KR, Tschudi C. Temporal order of RNA-processing reactions in trypanosomes: rapid trans splicing precedes polyadenylation of newly synthesized tubulin transcripts. *Mol. Cell. Biol* 1993;13:720–725. [PubMed: 8417363]
- Vanhamme L, Pays E. Control of gene expression in trypanosomes. *Microbiol. Rev* 1995;59:223–240. [PubMed: 7603410]

	5' primer →		← 3' primer
Sac	VK <u>QTVMTN</u> VYGVTVYGATFQIAKQLSP		KQGLDFASVHDSY <u>WTHASD</u>
Tet	VK <u>QTVMTS</u> VYGVTFVGARKQIYKQLRD		KRGMHFAAVHDSY <u>WTHASD</u>
Hom	VK <u>QTVMTV</u> VYGVTRYGGRLQIEKRLRE		RKGLTFVSVHDCY <u>WTHAAD</u>
Aca	VK <u>QTVMTS</u> VYGVTFIGARQQIENALKD		EAGLTYASVHDSY <u>WTHASD</u>
Orz	VK <u>QTVMTS</u> VYGVTYIGARQQITKRLQE		KAGLHFAGVHDSF <u>WVHACD</u>
Cry	VK <u>QTVMTT</u> VYGVTFIGAREQIYNRLYE		RCGKMFAGVHDSY <u>WTHASD</u>
Pyn	VK <u>QTVMTS</u> VYGVTWIGARDQIESRLRE		GRGMDFAGVHDSY <u>WTHASD</u>

5' primer: Q T V M T N V Y G V T
CAGACAGTGATGACAAGCGTGTAYGGNGTNAC

3' primer: D S A H T W Y S D
CGCTGCGTGTGTCCAGTARSWRTCRTG

Fig. 1.

Degenerate PCR. 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.

Fig. 2.

(A) Restriction map of the *TBMTRNAP* locus and cloning strategy. Numbers are given relative to the first nt in the *TBMTRNAP* ORF. 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' RTPCR experiments are indicated by arrowheads above the map. (B) Southern blot analysis of the *T. brucei* genomic DNA. DNA (10 µg) was digested with *Bam*HI (lane 1), *Eco*RI (lane 2), *Kpn*I (lane 4), *Bam*HI + *Eco*RI

(lane 3), *KpnI* + *Bam*HI (lane 5), and *KpnI* + *Eco*RI (lane 6), separated on a 0.75% gel, transferred to Nytran, and hybridized with a riboprobe spanning the *KpnI* site at 1907 and the *Bam*HI site at 3759.

A.

MRRLSLAPIK	WPAIAQRWGT	RGTDVPNRAK	GAIKSPKSRK	SRLPYVEVGE	RTDRMKEFVS	60
SGTDDNSGDS	DNNHNSVKSI	NNTDCRNKGT	DGGGEIDMDT	VAGTDAFTPT	VDMEEGYGAP	120
VMATTGSTSS	ESVLSVIDVL	CDDLFSCKTD	VLKSFARTRS	VSGSTKSELV	MNLIDLANRE	180
AHGESQDL SI	TKRVLGDVFN	EEVRVSGVRS	QWTNKSLEAS	HIEMACDNHL	KFEAMSYTQK	240
TLMVRRNASE	SSRVLRKLIR	PFQQDYNRVA	EAAGGVFRN	FKAADQLVN	SMMASEIVTP	300
FSSFLVSYVR	FATSGNLLTR	TNVIERLEQL	ELVQIAMDED	SDAVLQRLMA	LLDDVVAEVE	360
RRALCFGGKR	TSWGTETPEV	LRC SIQAFSD	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	EFISIISIDL	660
EATHVRDLRV	AFPPQLSYNS	WLHQSN TKDK	SVYNVLPSVA	VKGSTNQAMI	ISQAPMMKAL	720
DAISRVPWRI	SKYMLHVQEA	IVREGYGFGK	IRPAFYPLHY	CAKSRGDISY	ESTGMDDDDD	780
KTEVYNLQQR	REYELQQDED	WKNLSELRSN	RIHYLQALRQ	ARSLVQFSHI	YFPNSMDFRG	840
RMYPPLPGR LN	HTGSDPFRAL	LEYAEPKPLG	KEGLYWLKVH	LANKMGM SKL	SFDERVHYVN	900
EHIDDVVCSA	EQPLYGDKWW	QEAAEPMQCL	MACKELADAL	KCSQGPNFL	SRIPVAVDGS	960
YNGLQHYS AI	GRDAFGATLV	NLVPSEPAD	AYTGILKEMM	SSIKADAERD	HPVAQRCLGT	1020
GKGQDRDHIK	RKSIKRPIMT	QVYGV TGYGM	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

B.

Fig. 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 is characteristic of mitochondrial targeting sequences. The line divides the hydrophobic and hydrophilic domains of the helix. Hydrophilic residues are underlined.

Tb	ISDLEATHVRDLRVAFPQLSYNSWLHQSNTKDKSVYNVLPSSVAVKGSTNQAHITSQPHHQAALDAISVPWRIS--KYH	734
Hs	YRKYLCLLASDQREVPBPCLPRQYWEELGAPEALREQPWPLPVQHSLCKLLAEHLVQATQHPCSLDKCHSSRLV---EVL	606
Sc	SQAERDVMKKSPEFKKLVMNAKSVFRSSQIEQSK-ILWQSIERATGSVLI SHLI QVVKVSVQGVDVTKAKVHGEAPAF	634
Tb	LHVQEAIVREGYCFGKIRPPEYPLHY-----CKK-----SRGD-----ISYES	772
Hs	YHVYSFRLVQIQIGILKMPAPYVQLLEKAAEPTLTFRADVDPHLCPPLPWTSPHSGAELLSPTKLKATVEGATQHQLLET	686
Sc	AHGYQYHNGSKLGLVKIKTKTLIRQLN-GERLIA SVQPQLLPHLVBP KQWVNWRSGGYHYT QSTLLETKDSPEQVAYLKAA	713
Tb	TGMDDDDKTEVYMLQ-----QR--REYELQ-----QDSQ---W	801
Hs	CPPTALHGLDRLTLQLGNCAWVNGRVLDLVLOLFQKGCPLQLGVPPPPSEAPQPPEAHLPKSAAPPKAELELRSLAHCQ	766
Sc	SDNGDI DRVYDGLNVLGATPWTUNRKVEDVVSQVWN-KCEGFLDIPGAQDERVLP--APPKNSDPSILRAWKLVQVKTIA	790
Tb	KMLSELRSNRIHYLQALRQARESLVQFSHIYFPNSHOFRGMHYPLPGELMHTGSDPFRALLEYAEPEPLGKEGLFWLKQHL	881
Hs	KVAREHMSLRERELYRLSLRQHLDLR-VEWLPNHDQFRGTYPCPPHEMHLGSDVARELLERQGEPLGPHGLDWLKINL	845
Sc	NKESSDRSNRCDTNYKLEIRAEFLGE-KLYEPHMLDFRGRYPLSPHEMHLGNCHSGELLLEWHGKLGPSGLKWLKINL	869
Tb	AMKCHGSK-LSFDERVHYVNERIDDVVCSAEQPLYGDKWQEEBEPHMOCLHACKSLADQLKCSQGPENELSRIPVAVDGS	960
Hs	VMLTGLKKEPLRKRLAESEVHDDILDSDQPLTGKQWVHGAESEPWQTLACCHSVAMVVRASD-PARYVSHLPVHQDGS	924
Sc	SMLEGFDR-LPLKORVARETESHLQDIKDSAEMLPTGDRWTTEDKPWQQLATCFELMEVHKHON-PBEETSHQPVHQDGT	947
Tb	YNGLQHYSAUGEDAE GATLVMLVPSEPPADRYT GILKEKHSISKDAREEDHPVAQRCLGTGKGQDDNHKHSIKRPIHT	1040
Hs	CNGLQHYAALGEDSVGASVNLLEP SDVP QDVYS GAVEA QVEVFRRQDAQRCHVQAQ-----VLESFITERKVVKQT VHT	996
Sc	CNGLQHYAALGQDVE GATQVNLVP SDKP QDVYAHVARELVQKRLERAEKGDENAK-----ILKKKITERKVVKQT VHT	1019
Tb	QVYGVTYGCHSELILDELVKQNKNHGLWTSTDHREHADVREEVLRSGITERRETQNCERMTDVTNITWEVQPAELR--	1118
Hs	VVYGVTRYGGRLOIEKRLRELSDFP----QEFVWEQSHMLVRQVEKSLQEHESGTRRIQHMLTESRELISHHG-----	1065
Sc	NVYGVTFVETATQIRAKQLSPIFDD-----RKESLDFSKYLTKHVESATRELEHSAHLIQDWLGE SAKRLSKSIRLDVDEK	1094
Tb	-----TALCWTTPGLGVVRQPKYKKEKHEFTVH---GCAVPAANFSAASEKQLTATAPMLIHSLDASHLAK	1183
Hs	-----SVVEWVTPLGVPVVIQPYRLDSKVVKQIGGGIQSITYTHNGDISRKPNTRKQKNGEPPNFEINSLDASHML	1134
Sc	SFKNGNKPDEFHSSVITTPGLGLPIVQPYREESKKQVETN--LQTVFISDPFAVNPVITARRQKAGLPPNFEINSLDASHML	1172
Tb	TALIEHQNLGLSHGRVHDSYWTTPACDLPLSRVLREQFVTLYGKYDPLWELKEQWEEMFHDLLR-----	1247
Hs	TALHCYRKGLTEVSVDQYWTHTADVSVMNQVREQFVRLH-SEPILOQLSRFLVKECSEPPQK-----	1197
Sc	SEAECKGKGLDERSVHDSYWTHTASDIDTMMVVLREQFVRLH-EVDLVRLKEEFDQRYKQYVKI GKLRSTDLAQKIRI	1251
Tb	-----H---GKVLFD	1254
Hs	-----ILEASQLKETIQQA	1210
Sc	RKDLRKLGRTTLADEIYFEKKRQELLNSPLIEDRVGEKHVTTVSLFEDITDLDALELENGGDENS GHSVL LPLRLPE	1331
Tb	PPKRGDLDMVVLNMPYFES	1274
Hs	VPKPGAFDLEQVKESTYFES	1230
Sc	IPPKGDFDVTVLNMSQYFES	1351

Fig. 4.

Alignment of the carboxy-terminal portion of the predicted amino acid sequence of the *T. brucei* MTRNAP (Tb) with that of *Homo sapiens* (Hs) and *Saccharomyces cerevisiae* (Sc). Residues that are important for catalytic activity of the T7 RNA polymerase are conserved in the *T. brucei* sequence and are indicated by asterisks placed below the amino acids.

Fig. 5.

Developmental profile of steady-state transcripts from the *TBMTRNAP* locus. Poly(A) + (4.5 μ g) or total RNA (30 μ g) 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) Schematic of the *TBMTRNAP* locus with polypyrimidine tracts marked as striped boxes upstream of the mapped *trans*-splice sites which are indicated by black circles. Long black boxes represent ORFs greater than 100 amino acids. The locations of the riboprobes used in the Northern blots

are indicated with dotted lines. A transcript map of the RNAs processed from the 6.9 kb polycistronic transcript is indicated below the diagram of the *TBMTRNAP* locus.