Apocytochrome *b* and other mitochondrial DNA sequences are differentially expressed during the life cycle of *Trypanosoma brucei*

Jean E. Feagin, Douglas P. Jasmer and Kenneth Stuart*

Issaquah Health Research Institute, 1595 N.W. Gilman Blvd., Issaquah, WA 98027, USA

Received 11 March 1985; Revised 16 May 1985; Accepted 24 May 1985

ABSTRACT

Cytochromes and Krebs cycle enzymes are not detected in bloodstream forms of *Trypanosoma brucei* but are present in procyclic forms. We have analyzed transcription of mitochondrial sequences which contain the apocytochrome *b* gene and several other open reading frames (ORFs). Multiple transcripts map to individual DNA sequences located on both DNA strands. Larger low abundance transcripts map to multiple ORFs and may be precursor RNAs. Small abundant transcripts map to G+C rich sequences that do not have obvious protein coding functions. The larger of two presumptive apocytochrome *b* transcripts is strikingly more abundant in procyclic than bloodstream forms and other mitochondrial transcripts are also differentially abundant between these two forms. In addition, many mitochondrial transcripts appear to be differentially polyadenylated between bloodstream and procyclic forms. We suggest that the mechanisms which regulate the production of the mitochondrial respiratory system in *T. brucei* involve differential expression of mitochondrial genes.

INTRODUCTION

The cellular differentiations that constitute the life cycle of the protozoan parasite *Trypanosoma brucei* entail dramatic metabolic changes including facultative production of the mitochondrial cytochrome mediated respiratory system. Krebs cycle enzymes and cytochromes are not detected in bloodstream forms (BF) which exist in the mammalian host and are dependent on glycolysis for energy production. Upon passage to the insect host or appropriate culture conditions, BF differentiate into procyclic forms (PF) which contain a fully functional mitochondrial respiratory system (1, 2).

The unusual mitochondrial DNA of *T. brucei*, which is called kinetoplast DNA (kDNA), is a concatenated network that is composed of about 50 identical ~22 kilobase (kb) maxicircles and about 5000 divergent, 1 kb minicircles (1, 2). While the function of minicircles remains uncertain, recent studies have shown that the maxicircles have sequences related to those of other mitochondrial DNAs. The *T. brucei* maxicircle encodes mitochondrial rRNAs (3) and components of the mitochondrial respiratory system (4-8, Jasmer, *et al.*,

submitted), as does the maxicircle of the related trypanosomatid Leishmania tarentolae (9). The observation that mutants with natural (10) or laboratory induced (11) deletions of maxicircle sequences are unable to convert from BF to PF illustrates that the maxicircle is essential to the life cycle differentiation. Surprisingly, although kDNA is not essential to BF, since mutants devoid of kDNA survive well in the mammalian host (11), the maxicircle is transcribed in both BF and PF (12-14, Jasmer, *et al.*, submitted).

In order to investigate the role of maxicircle gene expression in the differentiation of BF to PF, we have examined transcription of a maxicircle region which contains the apocytochrome b (CYb) gene and several open reading frames (ORFs) with unknown functions (4-6). We find that major transcripts map to both DNA strands as do less abundant transcripts which may be processing precursors or intermediates. The CYb transcripts are differentially expressed between BF and PF. Other transcripts are also differentially expressed but with a pattern different from that exhibited by the CYb transcripts. In addition, a smaller proportion of each maxicircle transcript appears polyadenylated in BF than PF. These data indicate that maxicircle genes are differentially expressed during the life cycle of T. brucei.

METHODS

Organisms. T. brucei brucei clone IsTaR 1 from stock EATRO 164 was grown as BF and PF and isolated as described elsewhere (15). BF cells were harvested after three days infection in rats, and were all long slender forms. Cells were frozen in liquid N₂ and stored at -80°C.

Nucleic acid isolation. RNA was isolated according to Ross (16), with minor modifications. Briefly, cells were homogenized in 6 M urea, 2% SDS, 0.35 M NaCl, 1 mM EDTA, 10 mM Tris pH 8, extracted twice with an equal volume of phenol plus a half volume of CHCl₃, and once or twice with an equal volume of CHCl₃. CsCl was added to the final aqueous phase (1 gm/2.5 ml) which was then layered over 3 ml of 5.7 M CsCl and centrifuged in a Beckman SW27.1 rotor at 20,000 rpm for 20-24 hrs. The RNA pellet was dissolved in H₂O, ethanol precipitated, redissolved in H₂O, and stored at -80°C. Oligo(dT) cellulose chromatography was performed according to manufacturer's (Collaborative Research) instructions, except that binding buffer was 0.4 rather than 0.5 M NaCl. Plasmid DNA and single stranded M13 DNA were prepared according to Birnboim and Doly (17) and Messing (18), respectively. Restriction fragments of pTKHR67 and pTKHR40 (13), plasmid clones of the H1R2

and R2H2 maxicircle regions, respectively (see Fig. 2), were isolated from gels as previously described (19).

Gel electrophoresis, blotting, and hybridization. RNA was electrophoresed in 1.5% agarose gels containing 2.2 M formaldehyde (20) and was transferred to nitrocellulose filters (21) or to GeneScreen Plus (New England Nuclear) according to manufacturer's instructions. Plasmid probes were labelled by nick translation (22). M13 probes were prepared (18) by synthesizing the DNA strand complementary to M13 beyond the insert using M13 hybridization probe primer (Bethesda Research Laboratories, Inc.). Hybridization and washing conditions were as described previously (14).

DNA sequencing. The nucleotide sequence was determined by the dideoxy chain termination method (23) and the chemical degradation method (24) as diagrammed in Figure 1. Plasmid fragments were subcloned into M13 mp8 or 9 following digestion with restriction endonucleases or Bal 31 exonuclease, as described by Payne, *et al.* (8).

Nomenclature. ORFs with homology to mitochondrial genes of other organisms are given the name of the homologous gene, including the unassigned reading frames (URFs). The term URF is reserved for such sequences while ORFs lacking detectable homology to mitochondrial genes of other organisms are named ORF followed by a strand designation [I or II indicating the cytochrome *c* oxidase I (COI) or II (COII) coding strands, respectively (8)] and an alphabetic identification. COII is encoded on the same strand as CYb and the rRNAs. ORFs IIA, B, C, D, E, and ORF IA, correspond to URFs 6, 2, 2a, 3a, 3, and 4, respectively, in refs. 4 and 7.

RESULTS

The locations of genes identified by sequence analysis are indicated on the abbreviated restriction map of the *T. brucei* 164 maxicircle (Fig. 2).

Figure 1. Sequencing strategy for the region between the rRNA and CYb genes. The arrows below the restriction map indicate the direction and extent of sequencing. The * indicates sequence determined by the chemical degradation method (23) to resolve ambiguities due to secondary structure. Ah, *Aha* III; B, *Bam*H I; Cf, *Cfo* I; R, *Eco*R I; Rs, *Rsa* I.

Figure 2. Linearized restriction map of the *T. brucei* 164 maxicircle. The upper map shows the entire maxicircle (13), and some genes identified by sequence analysis (3, 7, 8, Jasmer, *et al.*, submitted). The lower map is an expansion of the ~4400 nt region between the B and H2 sites and shows the CYb (4, 6) and ORF IA (7, 8) genes. Two sets of overlapping reading frames, ORFs IIB and IIC (B/C) and ORFs IID and IIE (D/E) (4-6) are also shown. Arrowheads indicate the direction of transcription. Genes drawn above the line are transcribed from strand II, those below from strand I. URF and ORF designations are according to the nomenclature system described in *Methods*. Ah, *Aha* III; Al, *Alu* I; B, *Bam*H I; D, *Dde* I; H, *Hind* III; P, *Pst* I; R, *Eco*R I; Sa, *Sau*3A I.

These include the genes for the 12S and 9S rRNAs (3), CYb (4-6), COI, COII, and URFs 1, 4, and 5 (7, 8, Jasmer, *et al.*, submitted). ORFs (IA and IIA) not identified with other mitochondrial genes but which have corresponding transcripts (14, as discussed below) are also indicated. We examined transcription of the ~5 kb region between the rRNA genes and the H2 site. None of the transcripts described below were detected in a mutant devoid of kDNA (11), indicating they are maxicircle transcripts.

Apocytochrome b

A 1089 nucleotide (nt) ORF with homology to CYb is located downstream from and on the same strand as the rRNA genes (4, 6; see Fig. 2). Two transcripts measuring 1350 and 1200 nt map to the CYb gene as shown by northern blot analysis using probes containing the CYb coding sequence (Fig. 3a-c). Those probes containing only the CYb coding sequence (Fig. 3b,c) hybridize primarily to the 1350 and 1200 nt transcripts, but also detect 2000 and 1650 nt transcripts which may be precursor transcripts, as discussed later. The 1350 and 1200 nt transcripts both hybridize with single stranded M13 probes containing the CYb coding strand (Fig. 4c-e). Both are retained on oligo(dT) cellulose (Fig. 5c,d), suggesting they are polyadenylated. Both the 1350 and 1200 nt transcripts therefore have the appropriate size and map location for CYb mRNAs.

The CYb transcripts are the most abundant transcripts from this maxicircle region, based on the intensity of their hybridization with probes representing the whole region (data not shown), but their abundance differs

Figure 3. Localization of R2H2 maxicircle transcripts by northern blot analysis. The map shows the location of the probes (nick translated restriction fragments of pTKHR40) that were used for each panel. Ten μ g of total BF (left lane) and PF (right lane) RNA was electrophoresed in each lane; rehybridization of such filters with a nuclear cDNA gave similar signal strength for BF and PF lanes (data not shown). Denatured pBR322 restriction fragments were used as size markers; selected sizes are indicated in nt. Arrowheads in panels a-c indicate the 1350 and 1200 nt CYb transcripts; those in panel e indicate the 1650 and 1450 nt ORF IA transcripts. Gene locations, restriction sites and abbreviations are as given in Fig. 2.

between BF and PF (Fig. 3a-c, 4c-e). There is little, if any, 1350 nt transcript in BF compared to PF RNA. The apparent 1350 nt BF transcript in figure 3b may result from contamination of that particular probe with an ORF IA probe which detects a 1450 nt transcript. The 1200 nt transcript is also less abundant in BF than PF RNA, but the disparity is smaller. The difference in abundance of these transcripts between BF and PF is accentuated in polyadenylated RNA (Fig. 5c,d).

ORF IA

The 1335 nt ORF IA (8) is on strand I (see Fig. 2, nomenclature section in Methods). A probe completely contained within ORF IA hybridizes to 1650 and 1450 nt transcripts (Fig. 3e). A set of transcripts with sizes intermediate between these two is also seen; the 1450 nt transcript is

Figure 4. Identification of transcribed strand by northern blot analysis. The map shows the location of the single stranded M13 probes that were used for each panel. Probes with righthand arrows hybridize to transcripts from strand II while those with lefthand arrows hybridize with transcripts from strand I. Probes complementary to the d and g probes detect no transcripts (data not shown). Arrowheads in panels c-e indicate the 1350 and 1200 nt CYb transcripts; those in panels f and g indicate the 1650 and 1450 nt ORF IA transcripts. BF and PF lanes, size markers, gene locations, and abbreviations are as given in Figs. 2 and 3.

detected as two components, a discrete band and a more diffuse one. Single stranded probes reveal that both the 1650 and 1450 nt transcripts are transcribed from strand I (Fig. 4f,g). A single stranded probe complementary to probe 4e also detects 1650 and 1450 nt transcripts (unpublished data). ORF IA extends 107 nt into the adjacent maxicircle segment (7, 8) and probes containing that sequence also hybridize to 1650 and 1450 nt transcripts (14). Both ORF IA transcripts are retained on oligo(dT) cellulose (Fig. 5e), suggesting that both are polyadenylated. As with CYb, ORF IA probes detect some low abundance transcripts, as discussed below.

The 1650 and 1450 nt transcripts differ in abundance between BF and PF, but, in contrast to CYb, both are more abundant in BF than PF total RNA (Fig. 3e, 4f,g). However, ORF IA poly(A)⁺ transcripts are more abundant in PF RNA (Fig. 5e).

Figure 5. Polyadenylation of *T. brucei* maxicircle transcripts. The map shows the location of the probes (nick translated restriction fragments of pTKHR67 and pTKHR40) that were used for each panel. In panel a, the left and right lanes are oligo(dT) unbound (10 μ g) and bound (1 μ g) PF RNA, respectively. In panels b-e, both lanes contain 1 μ g of oligo(dT) bound RNA; the left lane is BF and the right lane PF RNA. In panel f, both lanes contain 10 μ g oligo(dT) unbound RNA; the left lane is BF and the right lane is PF RNA. Arrowheads in panel a indicate the rRNAs, those in panels c and d indicate the 1350 and 1200 nt CYb transcripts, and those in panels e and f indicate the 1650 and 1450 nt ORF IA transcripts. Size markers, gene locations, and abbreviations are as given in Figs. 2 and 3.

Sequence organization between the rRNA and CYb genes

The nucleotide sequence for the maxicircle region between the rRNA and CYb genes is presented in Fig. 6a. The sequences up to the BamH I site and from the EcoR I site to the end are identical with those presented by Eperon, et al. (3) and Benne, et al. (4), respectively, and the sequence between nt 1282 and the EcoR I site is consistent with amino acid sequence presented by Benne, et al. (5). Assuming that UGA specifies tryptophan (4-8, Jasmer, et al., submitted) and taking 100 amino acids as the lower size limit, we have found two clusters of ORFs in this sequence (Fig. 6b). No transcripts map to the ORFs from strand I. The ORFs from strand II would encode proteins rich in

Figure 6. Nucleotide sequence and ORFs between the rRNA and Cyb genes. The nucleotide sequence between the rRNA and Cyb genes is shown in 6a. The first nucleotide in 6a immediately follows the 3' terminal nucleotide of the 9S rRNA gene, as defined by Benne, *et al.* (5). ORFs IIB and IIC are indicated by hashmarks above the sequence, with arrows showing the direction of transcription. The BamH I and EcoR I sites are boxed for reference sites. Gs are underlined and Cs are underdotted to delineate G+C rich regions and illustrate G versus C strand bias. The final eleven nucleotides shown overlap the Cyb ORF. The sequences from nt 1 to 154, and from 2037 to 2135, are from refs. 3 and 4, respectively. A schematic map showing the location of ORFs which exceed 100 amino acids is shown in 6b.

Table 1. Nucleotide composition of the sequences between the rRNA and ORF IA genes.

nucleotides	A	T	G	C
51-363	25.2	24.6	<u>36.7</u>	13.4
392-630	17.6	28.0	<u>5.0</u>	<u>49.4</u>
631-970	43.1	35.6	12.8	<u>8.5</u>
971-2125	30.5	21.4	<u>39.4</u>	8.7
2126-3215 (CYb)	30.5	46.8	<u>16.4</u>	6.3
3216-3623	37.1	23.4	<u>34.3</u>	5.2
3624-4959 (ORF IA)	52.9	28.8	<u>10.9</u>	7.5

Single strand nucleotide composition (%) for the rRNA coding strand is given. Nucleotide numbering begins with the first nucleotide of the sequence given in Fig. 6a. For the G+C rich regions, the larger of the G or C composition is underlined.

glycine and charged amino acids, which is unusual for mitochondrial proteins (see 4, 5), and reflects the G+C richness of these sequences compared to other regions of the maxicircle (Table 1). These sequences also exhibit a strong G versus C strand bias. Two of the ORFs are named ORF IIB and IIC (see Fig. 2, 6b, and nomenclature section in Methods) since, as described below, transcripts map to those ORFs. Another G+C rich, G versus C strand biased region, containing ORFs IID and IIE (see Fig. 2), is located between the CYb and ORF IA genes (4, 6). Similar G+C rich sequences are found elsewhere in the maxicircle (Jasmer, *et al.*, submitted).

Transcripts from between the rRNA and CYb genes

At least eight transcripts map to the ~2100 nt region between the rRNA and CYb genes. Low abundance 1200 and 1000 nt transcripts that are more abundant in PF RNA are seen with a probe flanking the 9S rRNA (Fig. 7a). Probes covering the rest of this region also detect a low abundance 1200 nt transcript (Fig. 7b-e). Probes 7b-d show the 1200 nt transcript as a broad band more abundant in BF RNA (Fig. 7b-d); this is more apparent in longer exposures. The 1200 nt transcript detected with the 7e probe is seen as a more compact band which is similar in abundance in both stages. The differences in abundance pattern and the fact that the probes hybridizing to the transcript span more than 1200 nt suggest that multiple 1200 nt transcripts map to this region.

Probes covering ~1700 nt, between the BamH I site and the CYb gene, detect an abundant 850 nt transcript that is slightly more abundant in PF than BF RNA (Fig. 3a, 7b-e). It is possible for a single transcript of this size

Figure 7. Localization of maxicircle transcripts between the rRNA and CYb genes. The map shows the location of the probes used for each panel. Probe a was an M13 clone which hybridizes to strand II transcripts; probes b-e were nick translated plasmid fragments from pTKHR67 and pTKHR40. Arrowheads indicate the 1200 and 850 nt transcripts. BF and PF lanes, size markers, and abbreviations are as given in Figs. 2 and 3. Cf, *Cfo* I.

to hybridize to this set of probes but reduced hybridization to the end probes would be expected. While hybridization to probe 7b is reduced, that to 3a is quite strong. Again, this suggests distinct but similarly sized transcripts. The 1200 and 850 nt transcripts are transcribed from strand II, as determined by their hybridization with single stranded probes (Fig. 4a,c,d).

Transcripts of approximately 850, 700, and 580 nt map to the region containing the 672 and 369 nt ORFs IIB and IIC but the bands appear diffuse, giving the impression of a continuum of sizes (Fig. 3a, 4c,d). This is not due to RNA degradation since it is seen with numerous RNA preparations which give relatively compact bands with other probes. These transcripts are transcribed from strand II (Fig. 4c,d) and appear somewhat more abundant in PF than BF RNA. The largest of these transcripts may be one of the 850 nt transcripts discussed above.

Transcripts of approximately 620 and 480 nt are detected with probes

adjacent to both sides of the BamH I site (Fig. 7b, unpublished data); both are transcribed from strand I (Fig. 4b). The larger appears more abundant in PF RNA while the smaller is similar in abundance in both life cycle forms. In addition, a 460 nt transcript is detected with a probe flanking the 9S rRNA. It is distinct from the 480 nt transcript since it is more abundant in PF RNA and is transcribed from strand II (Fig. 7a). In other blots, the 480 and 460 nt transcripts each appear to be composed of two transcripts (data not shown). These small transcripts are transcribed from the C rich strand of these G+C rich sequences (Table 1). The larger low abundance transcripts in the BF lane of Fig. 4b have not been detected consistently.

Transcripts from between CYb and ORF IA

Four transcripts between 760 and 300 nt were detected using probes containing sequences from the ~400 nt region between the CYb gene and the ORF IA sequence. This region of the maxicircle contains the two overlapping ORFs IID (255 nt) and IIE (303 nt) (4). A probe containing this region detects a prominent 530 nt transcript, a less abundant 760 nt transcript, and low abundance 400 and 300 nt transcripts (Fig. 3d). The 760 and 530 nt transcripts were observed as diffuse bands connected by a continuum of transcript sizes, like the transcripts immediately upstream of CYb, and are transcribed from strand II (Fig. 4c,e). Again, the transcripts are encoded on the C rich strand (Table 1). The 530 nt transcript appears similar in abundance in both life cycle stages, the 760 nt transcript may be slightly more abundant in PF RNA, and the smaller transcripts are more abundant in BF RNA. Although we did not detect the less abundant 400 and 300 nt transcripts with single stranded probes, they show differential abundance patterns similar to the ORF IA transcripts, suggesting they may be transcribed from strand I.

Polyadenylation of transcripts

The major transcripts described above appear to be polyadenylated (Fig. 5). The presence of mitochondrial rRNA in oligo(dT) bound and unbound fractions was examined as a control. Some of the 1150 and 600 nt rRNA transcripts remains in the bound fraction, but at much reduced levels compared to the 850 nt and small transcripts which are presumably polyadenylated (Fig. 5a). Comparison of Fig. 5 with in Figs. 3, 4, and 7 shows that the same transcripts are seen in the poly(A)⁺ fraction as in total RNA, except those which are difficult to detect because of their low abundance. The poly(A)⁻ fraction contains reduced amounts of each transcript; in most cases, the transcripts appear the same size in both polyadenylated and nonpolyadenylated RNA (unpublished data). For the 1450 nt ORF IA transcript, however, the

Figure 8. Low abundance transcripts from the H1H2 maxicircle region. The map shows the location of the probes (nick translated plasmid fragments from pTKHR67 and pTKHR40) that were used for each panel. Arrowhead in panel a indicates the 1700 nt transcript; those in panels b and c indicate the 2000 and 1650 nt transcripts. BF and PF lanes, size markers, gene locations, and abbreviations are as given in Figs. 2 and 3.

discrete band seen in total and poly(A)⁻ RNA is missing in poly(A)⁺ RNA while the diffuse component is still present (compare Figs. 3e, 5e,f), suggesting that a specific subfraction of the 1450 nt transcript is not polyadenylated. The reverse is observed for a 1700 nt transcript which hybridizes to rRNA gene sequences; the poly(A)⁺ fraction contains a discrete higher molecular weight band compared to the wider band seen in total (Fig. 8a) and poly(A)⁻ RNA (Fig. 5a).

Figure 9. Transcription map of the *T. brucei* maxicircle between the rRNA genes and the H2 site. More abundant transcripts are shown as heavy lines and less abundant ones as thin lines. Those shown above the map are transcribed from strand II and those below from strand I. The 2600 nt transcript extends into the adjacent segment, as indicated by the dotted line. The 1200 and 850 nt transcripts marked with * have ambiguous map locations.

Maxicircle transcripts appear to be differentially polyadenylated between BF and PF. Within each life cycle stage, the relative abundance of maxicircle transcripts to each other is similar in total and poly(A)⁺ RNA. However, the relative abundance of BF transcripts to corresponding PF transcripts is reduced in poly(A)⁺ RNA compared to total RNA, suggesting that a greater proportion of each transcript is polyadenylated in PF. For example, the 1200 nt CYb transcript is somewhat less abundant in BF than PF total RNA (Fig. 3c, 4c); this disparity is accentuated in poly(A)⁺ RNA (Fig. 5c). The 1650 and 1450 nt ORF IA transcripts, which are more abundant in BF than PF in both total and poly(A)⁻ RNA (Fig. 3e, 5f), are more abundant in PF in poly(A)⁺ RNA (Fig. 5e). Other maxicircle transcripts exhibit similar patterns. Probes for nuclear genes reveal that their transcripts have similar BF versus PF abundance patterns in poly(A)⁺ and total RNAs (unpublished data). Thus, the differential polyadenylation is a mitochondrial phenomenon.

Potential precursor transcripts

Several large low abundance transcripts are seen with probes from this region of the maxicircle. These transcripts are not detected in RNA from mutants devoid of kDNA (data not shown) and thus are not the result of crosshybridization with nuclear transcripts. It is also unlikely that they represent crosshybridization with other maxicircle sequences. Since almost every probe reveals several specific transcripts, crosshybridization would generate recognizable and reciprocal patterns; we have not observed such patterns. The size and mapping data for these transcripts suggests they are possible precursors for some of the transcripts identified above. Probes containing mitochondrial rRNA coding sequences reveal a less abundant 1700 nt

transcript whose size approximates the sum of the two rRNAs. Its apparent absence in the BF lane of Fig. 8a illustrates the lower abundance of rRNA in BF RNA; the 1700 nt transcript is seen in longer exposures.

Two, and sometimes three, large low abundance transcripts (2000, 1650, and 2700 nt, respectively) are detected with probes containing CYb or adjacent sequences (Fig. 3b, 8b,c). Although we have not detected these transcripts using single stranded probes, based on their sizes, they could be precursors of CYb and perhaps one or more of the adjacent ORFs. The probe in Fig. 8c contains very little of the CYb coding sequence but detects the potential precursor transcripts as strongly as the 1350 and 1200 nt CYb transcripts. It also detects a 970 nt low abundance transcript which is not seen with adjacent probes.

A 2600 nt transcript may be a precursor to the ORF IA transcripts (Fig. 3e) since it is transcribed from the same strand as ORF IA (Fig. 4g) and a transcript of this size is also detected with probes from the ORF IA and ORF 1 coding sequences from the adjacent H2R3 region (14). A low abundance 1850 nt transcript is occasionally seen with ORF IA probes, although the strand from which it is transcribed has not been identified (14). In addition, a low abundance 1200 nt transcript is usually seen with ORF IA probes and is transcribed from strand I (Fig. 4g). This transcript is enriched in the poly(A)⁺ fraction (Fig. 5e), but is smaller than the ORF IA coding sequence. Low abundance transcripts that are smaller than the associated ORF have also been observed for other regions of the maxicircle (Jasmer, *et al.*, submitted). Their significance is uncertain.

From the mapping data presented above, we have constructed a schematic transcription map of the region of the maxicircle examined here (Fig. 9). The map indicates the approximate locations of transcripts, and shows only one transcript each for the ambiguous 1200 and 850 nt transcripts discussed above. Due to mapping uncertainties, the 1850, 1000, and 970 nt transcripts from the ORF IA region and the rRNA and CYb flanks, respectively, are not shown. The map shows that most of this region is represented by abundant transcripts from one strand or the other, and that multiple transcripts map to most, if not all, of the transcribed sequences.

DISCUSSION

Analysis of over 15 kb of continuous nucleotide sequence of the *T. brucei* maxicircle has permitted the identification of structural genes and an overview of the organization of this mitochondrial genome. The 12S and 9S

rRNA genes, sequences homologous to CYb, COI, COII, and ORFs 1, 4, and 5 from other mitochondrial systems, and ORFs IA and IIA, whose functions are not known, are closely spaced in the sequenced region (3-8, Jasmer, et al., submitted; see Fig. 2). Transcripts map to each of these elements (4, 5, 14, Jasmer, et al., submitted) and thus, except for the rRNAs, they are probably functional protein coding genes. The organization of this genome departs from that of other small mitochondrial DNAs (25) since three putative protein coding sequences, COI, URF 1, and ORF IA, are located on the strand opposite that encoding the rRNAs.

Sequence analysis and function

While it is likely that the CYb and ORF IA sequences encode functional proteins, the role of other sequences in the maxicircle region analyzed here is unclear. Transcripts map to the small ORFs that flank the CYb gene, ORFs IIB, IIC, IID, and IIE (4, 5; see Fig. 9 and nomenclature section in Methods), but it remains uncertain whether they encode proteins. None of these ORFs contain conventional initiation codons near their 5' ends, and the predicted proteins would be highly charged, which is unusual for mitochondrial systems (4, 5). In addition, all except ORF IIB are considerably smaller than the abundant transcripts which map to them. Thus, the association of transcripts with these ORFs may be fortuitous, although a protein coding role cannot be ruled out yet. Similarly, the 460 nt transcript maps to the ORFs flanking the 9S rRNA but again the ORFs lack conventional initiation codons and the predicted proteins would be highly charged. The other ORFs from the region between the 9S rRNA and CY b genes (Fig. 6b) appear nonfunctional since transcripts from this region do not map to them.

The sequences spanning the BamH I site are G+C rich and strongly strand biased, and produce several abundant small transcripts. As in other G+C rich maxicircle sequences (4, Jasmer, et al., submitted), it is the C rich strand which is transcribed. The same is true of the transcripts flanking the CYb gene. The role of these G+C rich sequences is unknown, but their locations on the flanks of a number of maxicircle genes is reminiscent of tRNA sequences which serve as processing signals in other mitochondrial genomes (25-27), implying that these sequences may serve a similar processing function.

The sequence between the 9S rRNA and CYb genes that produces 1200 and 850 nt transcripts contains no corresponding ORFs. RNA splicing or other translation correction mechanisms would therefore be required for the production of a protein. Most T. brucei maxicircle transcripts are not spliced but since the CO II coding sequence exists in two different reading

frames (7, 8), this possibility cannot as yet be excluded. However, the abundance of stop codons in this sequence makes a protein coding function unlikely.

Polyadenylation is sometimes employed as a criterion for assessing whether a transcript has a protein coding function. However, it has been suggested (26) that polyadenylation in mitochondrial systems may be linked to transcript processing and thus may not reflect protein coding functions. Our observation of virtually universal polyadenylation, by the criterion of oligo(dT) binding, of *T. brucei* maxicircle transcripts, including transcripts that are unlikely to code for proteins, is consistent with polyadenylation associated with processing.

Potential Precursor Transcripts

Several of the transcripts we have observed occur in lower abundance than the presumptive mature transcripts, hybridize to adjacent coding sequences, and approximate the sum of the sizes of the transcripts from these adjacent sequences. Although we have not yet demonstrated a precursor-product relationship, these characteristics are consistent with the possibility that the large transcripts are precursors for the smaller more abundant ones. Thus, the 2700 nt transcript is a possible precursor to the CYb, ORF IIB/C, and ORF IIC/D transcripts (Fig. 9). The 2000 and 1650 nt transcripts may reflect subsequent removal of one of the transcripts of the sequences that flank CYb and/or other processing functions. Similarly, the 2600 nt transcript is a possible precursor to the ORF IA and URF 1 transcripts, as discussed previously (14). Potential precursors have been observed for other *T. brucei* maxicircle transcripts (14, Jasmer, et al., submitted).

We and others (5) have mapped a 1700 nt transcript to the 12S and 9S rRNA genes. Interestingly, about half of this potential precursor transcript binds to oligo(dT); the unbound transcript is slightly smaller. Potential precursors of mitochondrial rRNAs have been identified in several organisms and are generally polyadenylated (27-29). In human mitochondria, a fraction of the presumptive precursor transcripts is not polyadenylated. It has been suggested that the unbound RNA is the rRNA precursor while the oligo(dT) bound RNA may be a processing intermediate from transcripts between a promoter and the rest of the downstream transcripts (29). A similar situation may exist in *T. brucei*. The maxicircle promoter location is unknown, but a noncoding region occurs just upstream of the rRNA genes, analogous to promoter location in other mitochondrial genomes (25). That potential precursor transcripts of 4-7 kb are occasionally detected (Jasmer, et al., submitted, unpublished data)

lends further credence to the possibility that the T. brucei maxicircle, again like other mitochondrial genomes (25), initiates transcription at a small number of promoters and processes the resulting large transcripts.

Multiple Transcripts

Multiple transcripts appear to be the rule rather than the exception for T. brucei maxicircle genes. Previous transcription studies noted that the transcripts mapping to the maxicircle exceeded the single strand coding capacity (5, 12, 13). We have shown, here and elsewhere (14, Jasmer, et al., submitted), that, with the exception of URF 5, two transcripts, differing by 150-200 nt, map to each potential protein coding gene indicated in Fig. 2, including ORFs IA and IIA (14, Jasmer, et al., submitted). The mechanism that results in the size difference between these transcripts is unknown but has been discussed in detail elsewhere and may include addition or deletion of sequences, probably at the termini of the transcripts (14, Jasmer, et al., submitted). Polyadenylation, an obvious possibility, appears not to be responsible for the size difference, since both transcripts in a set occur in poly(A)⁺ RNA (Fig. 5, Jasmer, et al., submitted, unpublished data) and mitochondrial transcripts generally have short (50-60 nt) poly(A) tails (30). Whatever the mechanism(s) responsible for the multiple transcripts, it seems to generate some transcripts intermediate in size between the predominant transcripts in a set. The transcript sets from the ORF IIB/C and IID/E sequences have a greater proportion of these intermediate size transcripts.

In contrast to our finding of multiple transcripts for most maxicircle genes, Benne, et al. (4) have reported single transcripts of approximately 1200, 550, and 400 nt that map to the CYb gene, ORF IIB/C, and ORF IID/E, respectively, of T. brucei strain 427. This probably does not reflect differences between strains 427 and 164 since we have detected both CYb transcripts in several strains of T. brucei (unpublished data). It is possible that the differences in results reflect differences in experimental procedures.

Multiple transcripts from a single gene arising by a variety of mechanisms have been reported in a number of systems. A single gene may express different tissue specific (31, 32) or developmental stage specific (32-34) transcripts. Multiple transcripts may result in related gene products with distinct functions (35) or two distinct products from the same gene (36). As an organism which resides in two different hosts in various differentiated states, T. brucei may utilize such subtle forms of control of gene expression to shift rapidly among various physiological and developmental states.

We do not know whether one or both predominant transcripts in a set are translated or if they are translated in only one or both stages of the life cycle. As discussed below, greater abundance of transcripts like the larger CYb transcript correlates with production of a functional mitochondrial respiratory system, implying that the larger transcripts are translated. However, we cannot exclude the possibility that both transcripts are functional.

Differential Expression

An intriguing aspect of the T. brucei mitochondrion is that the cytochrome mediated respiratory system that is present in PF is absent in BF (1, 2). It was somewhat surprising when we (13) and others (12) found that maxicircle genes are transcribed in both life cycle stages. We report here that a more subtle mechanism than transcriptional activation or inactivation of the maxicircle appears to regulate the expression of the mitochondrial genome during the life cycle of T. brucei. While each gene is transcribed in both BF and PF, specific maxicircle transcripts vary dramatically in abundance between these life cycle stages. We find different expression patterns among maxicircle transcripts, implying independent, although perhaps coordinated, regulatory mechanisms. Furthermore, as will be discussed below, the extent of maxicircle transcript polyadenylation varies between life cycle stages.

The dramatically lower abundance of the 1350 nt CYb transcript as opposed to the slightly lower abundance of the 1200 nt CYb transcript in BF compared to PF illustrates this differential expression. This resembles the pattern of expression of the COI and COII genes (14, Jasmer, et al., submitted) but is in contrast to ORF 1A transcripts, both of which are more abundant in BF than in PF. Many other transcripts from the region studied here are slightly more abundant in PF than BF, but some are similar in abundance or slightly more abundant in BF. The differential patterns observed, while sometimes subtle, are reproducible, and some probes detect several transcripts with different abundance patterns, providing internal controls. Thus, mechanisms exist which independently regulate the abundance of specific maxicircle transcripts during the life cycle.

In addition to their differential abundance during the life cycle, maxicircle transcripts are differentially polyadenylated between BF and PF. While the relative abundances of the poly(A)⁺ transcripts remain similar to those seen with total RNA for both life cycle stages, the abundance of all BF transcripts relative to PF transcripts is decreased. This differential polyadenylation may also play a role in regulating mitochondrial gene

expression during the life cycle. Since the BF transcripts are not substantially smaller than those in PF, the poly(A) tails may be short or the other processes responsible for the generation of different maxicircle transcript size classes may also be differentially active between BF and PF.

A coherent picture of *T. brucei* maxicircle transcription and genetic organization is now beginning to emerge. Several unprecedented and provocative phenomena have been reported here, including multiple transcripts for most maxicircle genes, differential expression of some transcripts, and differential polyadenylation of maxicircle transcripts between life cycle stages. It is likely that these phenomena have a role in regulating expression of the mitochondrial respiratory system during different life cycle stages. Elucidation of the mechanisms effecting these events remains a challenge for the future.

ACKNOWLEDGEMENTS

We thank Joanne West, Elke Gobright, and Mark Payne for excellent technical assistance, and Robert Aline, Jr. for critical reading of the manuscript. This work was supported by NRSA AI07130 to D.P.J. and NIH grant AI14102 to K.S.

Abbreviations: BF, bloodstream form(s); COI and COII, cytochrome c oxidase subunits I and II; CYb, apocytochrome b; kb, kilobase(s); kDNA, kinetoplast DNA; nt, nucleotide(s); ORF, open reading frame; PF, procyclic form(s); URF, unassigned reading frame.

*To whom reprint requests should be sent

REFERENCES

1. Englund, P.T., Hajduk, S.L. and Marini, J.C. (1982) *Ann. Rev. Biochem.* 51, 695-726.
2. Stuart, K. (1983) *Mol. Biochem. Parasitol.* 9, 93-104.
3. Eperon, I.C., Janssen, J.W.G., Hoeijmakers, J.H.J. and Borst, P. (1983) *Nucl. Acids Res.* 11, 105-125.
4. Benne, R., De Vries, B.F., Van den Burg, J. and Klaver, B. (1983) *Nucl. Acids Res.* 11, 6925-6941.
5. Benne, R., Agostinelli, M., De Vries, B.F., Van den Burg, J., Klaver, B. and Borst, P. (1983) in *Nucleo-mitochondrial interactions*, Schweyen, R.J., Wolf, K., and Kaudewitz, F. Eds., pp. 285-302, Walter de Gruyter and Co., Berlin.
6. Johnson, B.J.B., Hill, G.C. and Donelson, J.E. (1984) *Mol. Biochem. Parasitol.* 13, 135-146.
7. Hensgens, L.A.M., Brakenhoff, J., De Vries, B.F., Sloof, P., Tromp, M.C., Van Boom, J.H. and Benne, R. (1984) *Nucl. Acids Res.* 12, 7327-7344.
8. Payne, M., Rothwell, V., Jasmer, D.J., Feagin, J.E. and Stuart, K. (1985) *Mol. Biochem. Parasitol.* (in press).
9. de la Cruz, V.F., Neckelman, N. and Simpson, L. (1984) *J. Biol. Chem.* 259, 15136-15147.
10. Borst, P. and Hoeijmakers, J.H.J. (1979) *Plasmid* 2, 20-40.
11. Stuart, K. (1983) *J. Cell. Biochem.* 23, 13-26.

12. Hoeijmakers, J.H.J., Snijders, A., Janssen, J.W.G. and Borst, P. (1981) *Plasmid* 5, 329-350.
13. Stuart, K.D. and Gelvin, S.B. (1982) *Mol. Cell. Biol.* 2, 845-852.
14. Feagin, J.E. and Stuart, K. (1985) *Proc. Natl. Acad. Sci. USA* (in press).
15. Stuart, K., Gobright, E., Jenni, L., Milhausen, M., Thomashow, L. and Agabian, N. (1984) *Parasitology* 70, 747-754.
16. Ross, J. (1976) *J. Mol. Biol.* 106, 403-420.
17. Birnboim, H.C. and Doly, J. (1979) *Nucl. Acids Res.* 7, 1513-1523.
18. Messing, J. (1983) *Methods Enzymol.* 101, 20-78.
19. Feagin, J.E., Setzer, D.R. and Schimke, R.T. (1983) *J. Biol. Chem.* 258, 2480-2487.
20. Lehrach, H., Diamond, D., Wozney, J.M., and Boedtker, H. (1977) *Biochemistry* 16, 4743-4751.
21. Thomas, P.S. (1980) *Proc. Natl. Acad. Sci. USA* 77, 5201-5205.
22. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) *J. Mol. Biol.* 113, 237-251.
23. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
24. Maxam, A.M. and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* 74, 560-564.
25. Clayton, D.A. (1984) *Ann. Rev. Biochem.* 53, 573-594.
26. Ojala, D., Montoya, J. and Attardi, G. (1981) *Nature* 290, 470-474.
27. Ojala, D., Merkel, C., Gelfand, R. and Attardi, G. (1980) *Cell* 22, 393-403.
28. Cleaves, G.R., Jones, T. and Dubin, D.T. (1976) *Arch. Biochem. Biophys.* 175, 303-311.
29. Montoya, J., Gaines, G.L. and Attardi, G. (1983) *Cell* 34, 151-159.
30. Hirsch, M. and Penman, S. (1973) *J. Mol. Biol.* 80, 379-391.
31. Young, R.A., Hagenbüchle, O. and Schibler, U. (1981) *Cell* 23, 451-458.
32. Medford, R.M., Nguyen, H.T., Destree, A.T., Summers, E. and Nadal-Ginard, B. (1984) *Cell* 38, 409-421.
33. Benyajati, C., Spoerel, N., Haymerle, H. and Ashburner, M. (1983) *Cell* 33, 125-133.
34. Crossland, L.D., Rodermal, S.R. and Bogorad, L. (1984) *Proc. Natl. Acad. Sci. USA* 81, 4060-4064.
35. Carlson, M. and Botstein, D. (1982) *Cell* 28, 145-154.
36. Amara, S.G., Jonas, V., Rosenfeld, M.G., Ong, E.S. and Evans, R.M. (1982) *Nature* 298, 240-244.