The nucleotide sequence of a 3.2 kb segment of mitochondrial maxicircle DNA from *Crithidia fasciculata* containing the gene for cytochrome oxidase subunit III, the N-terminal part of the apocytochrome *b* gene and a possible frameshift gene; further evidence for the use of unusual initiator triplets in trypanosome mitochondria

Paul Sloof*, Janny van den Burg, Arthur Voogd and Rob Benne

Section for Molecular Biology, Laboratory of Biochemistry, University of Amsterdam, AMC, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands

Received September 29, 1986; Revised and Accepted December 3, 1986

SUMMARY

A 3.2 kb segment of the maxicircle of Crithidia fasciculata mitochondrial (mt) DNA contains the gene for cytochrome oxidase subunit III (coxIII), the N-terminal portion of the gene for apocytochrome b (cytb) and two partially overlapping Unassigned Reading Frames (C.URF2/1). Transcript analysis of the segment reveals that both the coxIII gene and the C.URF2/1 area are transcribed into a pair of RNA products. With the C.fasciculata gene version as a probe, a coxIII gene could not be detected in nuclear and mtDNA of Trypanosoma brucei, indicating that the cytochrome oxidases of these two closely related trypanosome species may differ.

The nucleotide homology in the N-terminal region of the coxIII and cytb genes in T.brucei, Leishmania tarentolae and C.fasciculata starts at a UUA leucine codon, which adds further support to the hypothesis that apart from AUG, other initiator triplets are used in trypanosomal mitochondria: UUG, CUG and UUA, all triplets coding for leucine in the universal code.

Finally, the possibility is discussed that the two overlapping URFs (C.URF2/1) in fact represent a single, frameshift containing, gene.

INTRODUCTION

Trypanosomal mtDNA consists of a compact network of thousands of catenated circular DNA molecules, known as kinetoplast DNA (kDNA). In the kDNA network two types of circular DNAs occur (reviewed in 1,2): minicircles of 1-3 kilobasepairs (kb) in length depending on the species, which constitute 90-95% of the network, and 20-40 kb maxicircles, of which 25-50 copies are present. Minicircles are not transcribed and their function is unclear (3). In contrast, maxicircles are transcribed and nucleotide sequence analysis has revealed typical mitochondrial genes in T.brucei and L.tarentolae maxicircle DNA (reviewed in 4). So far, genes coding for the mitochondrial ribosomal RNAs (rRNAs) apocytochrome b (cytb), cytochrome oxidase (cox) subunits I, II and III, NADH dehydrogenase (ND) subunits I, IV and V and a number of unassigned reading frames (URFs) have been identified. Gene content and organization of the maxicircle proved to be largely identical in the sequenced area of the two trypanosome species, except in

the region flanked by the 9S mitochondrial rRNA gene and the cytb gene. In T.brucei this segment measures 2,1 kb and contains two URFs which would encode proteins rich in glycine and charged amino acids, generally highly unusual characteristics for mitochondrial proteins (5,6). On the L.tarentolae maxicircle this region is 1 kb larger and it contains the gene for subunit III of cytochrome oxidase and a number of URFs which are not homologous to the corresponding T.brucei URFs (7). The difference in gene content of this segment of the two otherwise very similar mitochondrial genomes is remarkable, particularly since the gene for an essential and evolutionarily conserved subunit of cytochrome c oxidase (8) appears to be missing in T.brucei. No difference in gene content and organization is found, for example, in the mitochondrial genome of mammalian species (9,10), which are further apart in evolution than the two trypanosomes, as judged by the degree of nucleotide conservation of homologous mitochondrial genes (see ref.4). This area of the trypanosome maxicircle is interesting for two other reasons: i) in L.tarentolae only one of the (putative) protein genes (ORF4) has an N-terminal AUG-codon. Comparison of other AUG-less mitochondrial genes from trypanosomes has suggested the possible use of leucine codons as initiator triplets (4,11). ii) Some of the ORFs are overlapping (e.g. LtORF1B/1A; ORF3/4). The gene for coxII is composed of two overlapping reading frames in L.tarentolae and T.brucei (7,12,13) and in the insect trypanosome C.fasciculata (11). Recently, we have found that in T.brucei and C.fasciculata this frameshift is restored at the RNA-level by the insertion into the transcript of four, non-DNA encoded, nucleotides during or after transcription ("RNA-editing", see refs.11,14). The possibility arises, therefore, that overlapping URFs in fact represent other frameshift-genes that are made continuous at the RNA-level.

In order to further assess the possible variability of the trypanosome cytochrome oxidase subunit composition and to use comparative sequence analysis as a tool for the identification of unusual initiation triplets and putative frameshift genes, we have determined the nucleotide sequence of this area in C.fasciculata.

MATERIALS AND METHODS

Restriction endonucleases were from New England Biolabs or Boehringer Mannheim; DNA polymerase (large fragment), calf intestine phosphatase and T4 DNA ligase from Boehringer; exonuclease Bal-31 from New England Biolabs or Bethesda Research Laboratories; low melting agarose from Bethesda Research Laboraties.

DNA and RNA. C.fasciculata cultivation and kDNA isolation are described in ref.15 or refs therein. Total cellular RNA was isolated according to

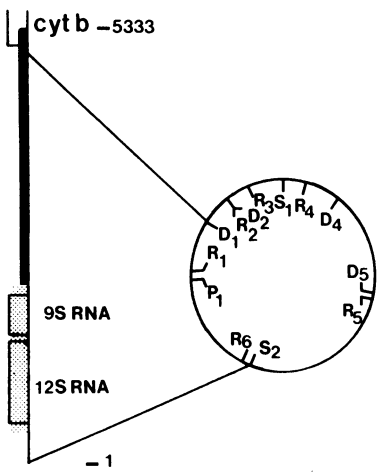

Fig.1: Sequenced area of C.fasciculata maxicircle DNA. The restriction map of C.fasciculata maxicircle DNA is taken from ref.15. The S_2D_1 region, obtained from the recombinant plasmid pD_5D_1 containing the D_5D_1 HindIII fragment (see ref.18), has been subjected to nucleotide sequence determination. The nucleotide sequence of the 12S and 9S mitochondrial rRNA genes, position 254-2110, has been reported previously (shaded region; ref.18); the nucleotide sequence presented here encompasses the area between positions 2111 and 5333 (black region). Relevant restriction sites are indicated: P = PstI; R = EcoRI; D = HindIII; S = SalI.

ref.16, and stored at -20°C as an ethanol precipitate. Plasmid and M13RF DNAs were isolated according to ref.17.

Cloning in plasmid and bacteriophage M13 DNA. The construction of the recombinant plasmid pD5D1, containing the D5D1 HindIII fragment of C.fasciculata maxicircle DNA has been described elsewhere (18). This plasmid has been used as starting material for the sequence determination (see Fig.1). From this plasmid fragments S2-P1, P1-D1 and R6-R1 were subcloned in bacteriophage M13. M13 clones containing a varying part of either S2-P1, R6-R1 or P1-D1 in either orientation were generated using Bal-31 exonuclease and non random cloning procedures according to ref.19, except that gel isolation of Bal-31 treated fragments was omitted (see ref.12). Ligation reactions were performed with vectors treated with calf-intestine phosphatase according to ref.12. This approach yielded a series of nested inserts from both strands. These clones were subjected to DNA sequence analysis.

From a clone bank of MboI restricted C.fasciculata kDNA in bacteriophage M13, a clone was selected which extended the sequence 124 residues beyond the D1 site. The start of this clone is at position 4948 (Fig.1).

DNA sequence analysis. DNA sequence determination was carried out by the dideoxynucleotide chain-termination technique (20). A number of overlapping sequences were obtained from both strands. In regions where the sequence was ambiguous, reliable sequences were obtained by priming with synthetic oligonucleotides (21).

Northern blotting and hybridization conditions. Northern blot analysis of glyoxylated total cellular RNA was performed according to ref.22, except that 80 µg RNA per lane was electrophoresed in 1.75% agarose gels in 10 mM phosphate buffer, pH 7.0. M13 probes were prepared by synthesizing the DNA strand complementary to M13 beyond the insert, using M13 hybridization probe primer (New England Biolabs) and ^{32}P-labeled dNTPs (Amersham, England). Hybridization conditions were as described (22).

RESULTS

Sequence analysis

Figure 2 presents the nucleotide sequence of the C.fasciculata maxicircle region from nucleotide 1981 up to 5333. Begin and endpoints of reading

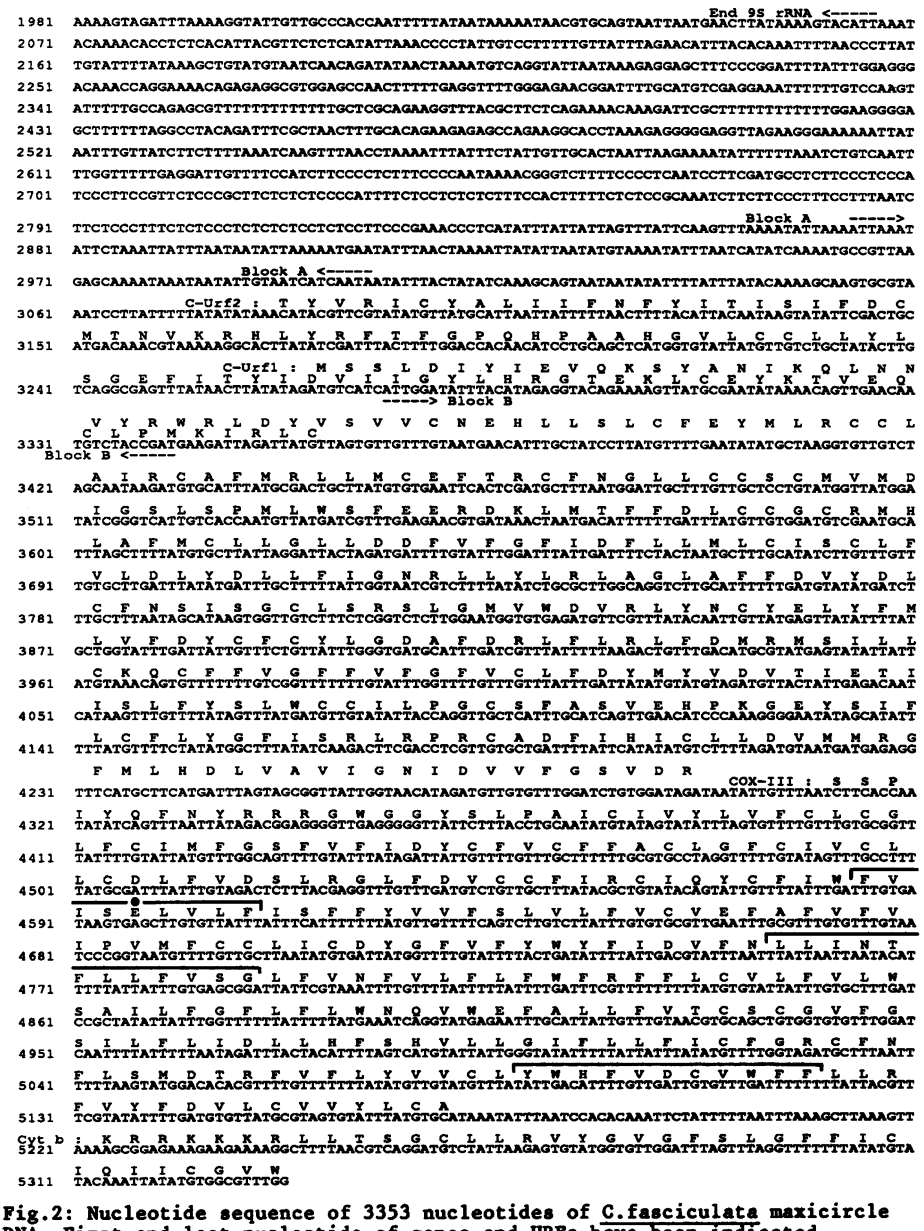

Fig.2: Nucleotide sequence of 3353 nucleotides of C.fasciculata maxicircle DNA. First and last nucleotide of genes and URFs have been indicated. Nucleotide number 1981 corresponds to nucleotide 1750 in ref.18. Genes were identified by comparison with aminoacid sequences in L.tarentolae mitochondrial genes (7). The aminoacid sequences in coxIII, which are conserved in Aspergillus (27), Neurospora (8), Saccharomyces (28), Drosophila (29), Bovine (10) and Human (30) are indicated by a bar. The invariant glutamic acid residue which reacts with DCCD in beef heart cytochrome c oxidase (31) is indicated by a dot.

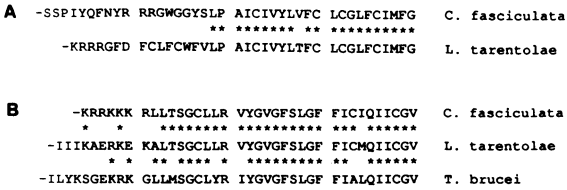

Fig.3: N-terminal amino acids of the C.fasciculata and L.tarentolae (7) coxIII genes (A) and the C.fasciculata, L.tarentolae (7) and T.brucei (23) cytochrome b genes (B). * indicates homology between compared sequences; - indicates a stopcodon.

frames are indicated. These reading frames were obtained by translating the nucleotide sequence into amino acids with a genetic code in which UGA encodes tryptophan as in most mitochondrial genetic systems, including that from trypanosomes (4). The region also contains two short segments, which are strongly conserved in T.brucei, L.tarentolae and C.fasciculata. These are indicated as blocks A and B in Fig.2. Several aspects of these reading frames and sequence elements are discussed below. We have also mapped transcripts in this maxicircle region by Northern blot analysis.

The genes for cytochrome c oxidase subunit III and apocytochrome b; evidence for unusual initiator triplets

A reading frame of 287 amino acids, in the C.fasciculata maxicircle segment from which the nucleotide sequence is given in Fig.2, is highly homologous to the coxIII gene in the corresponding area of the L.tarentolae maxicircle. The homologous region encompasses 269 residues, extending from a leucine at position 19 (see Fig.3A) to the stopcodon and showing an amino acid conservation of 92% with the L. tarentolae coxIII reading frame (85% at the nucleotide level).

The start of the amino acid homology at a leucine residue has also been observed in the apocytochrome b gene, at positions 9, 12 and 13 of the C.fasciculata, L.tarentolae and T.brucei reading frames respectively (see Fig.3B). We have previously reported that two other maxicircle reading frames, ND1 and URF5, lack methionine residues at their N-termini (12). Analysis of the corresponding reading frames in other trypanosome species shows that also in these cases the homology starts at leucine residues. Three different leucine triplets are found: TTG, TTA and CTG. Table I lists the "AUG-less" genes compared so far. The consistent occurrence of leucine-codons at the beginning of the amino acid homology of genes that lack an N-terminal AUG, suggests that they may serve as initiation codons (see also Discussion).

TABLE I: Possible initiation codons of trypanosomal mitochondrial genes for which the amino acid sequence homology starts at a leucine residue.

GENE	SPECIES	CODON	REF
URF 5	C.fasc.	CUG	(11)
	L.tar.	CUG	(7)
	T.brucei	UUG	(12)
ND I	C.fasc.	UUA	(11)
	L.tar.	UUA	(7)
	T.brucei	UUA	(12)
cyt.b	C.fasc.	UUA	
	L.tar.	UUA	(7)
	T.brucei	UUA	(23)
coxIII	C.fasc.	UUA	
	L.tar.	UUA	(7)

A coxIII gene has not been found on the sequenced part of the maxicircle from T.brucei. De la Cruz et al. report (7) that the coxIII gene is not present on cloned T.brucei maxicircle fragments that cover the non-sequenced region, as judged by hybridization experiments with the L.tarentolae gene as probe. Before it can be decided that a coxIII gene homologous to the L.tarentolae version, does not exist in T.brucei it should be ruled out that it is present either in: (i) nuclear DNA, (ii) a class of maxicircles which is different from the sequenced circle, or (iii) minicircle DNA. In order to

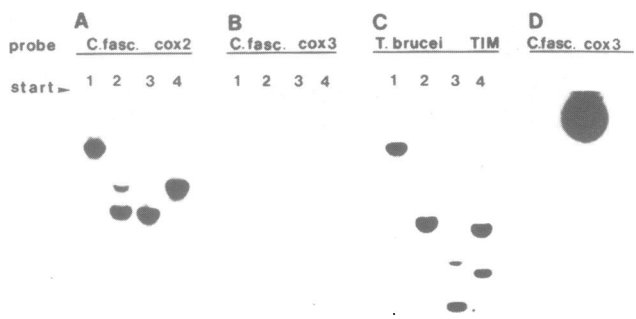

Fig.4: Hybridization of C.fasciculata coxII and coxIII probes to T.brucei DNA. 5.0 µg of total DNA from T.brucei has been restricted with HindIII (lane 1), TaqI (lane 2), AluI (lane 3) and Sau96 (lane 4), electrophoresed, blotted to nitrocellulose and hybridized with a C.fasciculata coxII probe (Panel A), C.fasciculata coxIII probe (Panel B) and a triosephosphate isomerase (T.I.M.) probe from T.brucei (24) (Panel C). Panel D gives the hybridization of the C.fasciculata coxIII probe to 50 ng of C.fasciculata kDNA, spotted on nitrocellulose. Hybridizations of panels A and B were performed in 6x SSC, 60°C and of panels C and D in 3x SSC, 65°C. (1x SSC: 0.15 M NaCl, 0.015 M Na-citrate, pH 7.0).

investigate this question, blots of restricted total DNA from T.brucei were hybridized with a C.fasciculata coxIII probe. The results of this experiment are given in Fig.4. No hybridization with the coxIII probe is observed, under conditions (6x SSC, 60°C) that permit a strong signal with a probe containing the C.fasciculata coxII gene which is 73% homologous to the T.brucei coxII gene (compare panel B and A, respectively). Single copy nuclear T.brucei genes like the one coding for the triosephosphate isomerase (TIM) (24) also show strong signals (panel C). The lack of hybridization of the coxIII probe suggests that the gene is highly diverged in T.brucei (in contrast to the other cox genes) or, alternatively, that it may even be missing (see also Discussion).

Overlapping reading frames C.URF1 and C.URF2

A set of two overlapping reading frames, C.URF1 and C.URF2, is located upstream from the coxIII gene (see Fig.2), at a position which corresponds to that of two overlapping reading frames in the L.tarentolae maxicircle, ORF3 and ORF4 (7). C.URF1 is 344 amino acids long and a stretch of the C-terminal 322 residues shows a 97% homology to the ORF4 reading frame of L.tarentolae. The C.URF2 reading frame is 92 amino acids long and contains a continuous region of 63 residues with 100% homology to ORF3 of L.tarentolae. Apart from the high degree of conservation at the amino acid level between C.URF1/ORF4 and C.URF2/ORF3 also the distribution of nucleotide substitutions over the three codons positions indicates that these genes code for protein. The second codon position appears to be the most constant whereas the third codon position is the most variable (see Table 2), a distribution which is characteristic for protein coding genes. Both sets of genes code for highly hydrophobic proteins (see Fig.2). Neither C.URF1/ORF4 nor C.URF2/ORF3 are homologous to the T.brucei URFs 1 and 2. The question how these overlapping reading frames are expressed is addressed below (see Discussion).

Homology in the remainder of the sequenced fragment

From the results described so far we conclude that the gene content and organization in the maxicircle region between the genes for the 9S mitochon-

TABLE II: Distribution of base substitutions over codon positions

	Pairs of homologous genes		
CODON POS.	C.URF2/L.ORF3	C.URF1/L.ORF4	COX 3 C.FASC./L.TAR.
1	24%	24%	24%
2	5%	7%	8%
3	71%	69%	68%

drial rRNA and apocytochrome b is well conserved between C.fasciculata and L.tarentolae but not in T.brucei. However, two short nucleotide sequences have been identified which are highly conserved between all three trypanosome species. They are indicated in Fig.2 as blocks A and B.

Block A is 130 bp long and its sequence is for 72% conserved between the three trypanosome species; block B is 80 bp in length and exhibits a similar level of nucleotide sequence conservation. The function of these sequence elements is unclear. Block B is a part of the sequence coding for C.URF2 in C.fasciculata and ORF3 in L.tarentolae, but no long URF is found in the corresponding T.brucei maxi-circle area. Analysis of the distribution of base substitutions between the T.brucei sequence on the one hand and C.fasciculata and/or L.tarentolae on the other hand argues against a possible protein coding function in T.brucei, since no bias in substitution for any of the three possible codon positions is found. Nucleotide substitution patterns of block A in the three different trypanosome species also fail to provide evidence for a protein coding function.

We have also considered the possibility that these highly conserved nucleotide blocks code for tRNAs by investigating their potential to form tRNA-like structures. However, a consistent set of tRNA structures from these regions in the three trypanosome species was not found. It is not clear therefore, what the function of these conserved elements could be. The degree of homology between the remainder of the fragment in C.fasciculata (the area flanked by A and the 9S gene) and the corresponding area in L.tarentolae is low and hardly exceeds background level. No long open reading frames occur in this region.

Transcription of the maxicircle region between the genes for the 9S rRNA and apocytochrome b

In order to investigate whether the identified reading frames are transcribed, we have probed total RNA from C.fasciculata on northern blots with single stranded M13 clones, containing different parts of the maxicircle region between the 9S mitochondrial rRNA and apocytochrome b genes in either orientation. The probes and their orientation are depicted in Fig.5A. With probe 1, which spans the SalI-PstI area (see Fig.1), the 12S and 9S rRNAs of 1141 and 612 nucleotides in length respectively, are detected (Fig.5B lane a). In a longer exposure (lane b) an additional transcript of about 430 nucleotides is seen which is also detected with probe 3 (lane d). Using probe 2, transcripts from the other strand in this region can be detected. Only an RNA of about 380 nucleotides in length is transcribed from this strand (lane

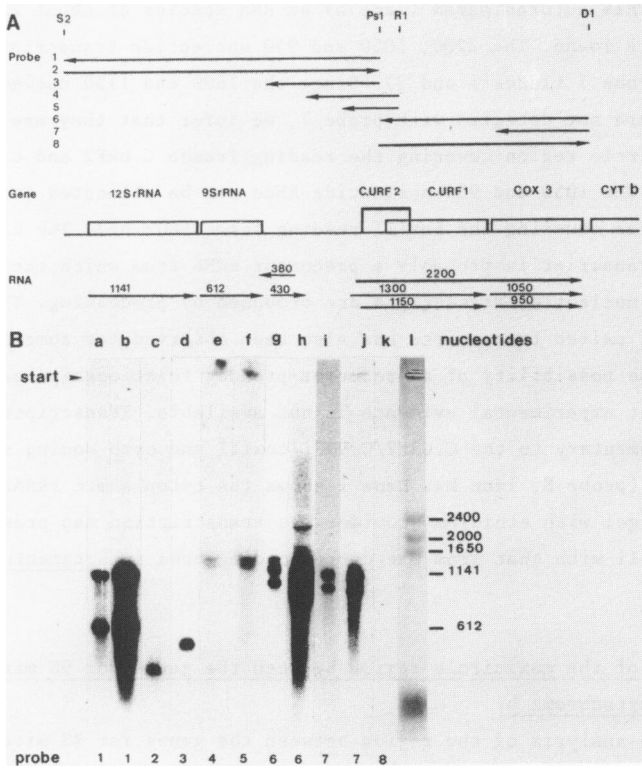

Fig.5: Transcript analysis. Total RNA from C.fasciculata was glyoxylated, electrophoresed and blotted onto nitrocellulose as described in Methods. The blots were hybridized with ss M13 clones depicted in panel A. Arrows indicate the orientation of the insert within the clone. Autoradiograms of the different hybridizations are shown in panel B. The scale of molecular weight markers has been derived from the 9S and 12S mitochondrial rRNAs (lanes a,b) and the cytoplasmic rRNAs (26, lane 1). The localization of the RNA species detected is given in panel A. Numbers refer to the number of nucleotides, using the molecular weight scale depicted in B. Autoradiograms shown in lanes b, h, j are long exposures of autoradiograms in lanes a, g and i respectively. The following probes are used: probe 1, lanes a,b; probe 2, lane c; probe 3, lane d; probe 4, lane e; probe 5, lane f; probe 6, lanes g,h; probe 7, lanes i,j; probe 8, lane k; ethidium bromide stained gel of total RNA shows the cytoplasmic rRNAs, lane 1.

c), its exact location within the SalI-PstI region has not been further determined. With probes 4 and 5 (lanes e and f) two RNAs are detected, of 1300 and 1150 nucleotides in length respectively. These RNAs are also found with probe 6, which spans the entire PstI-HindIII area (Fig.1), in addition to two smaller RNA species of 1050 and 950 nucleotides (lane g). In a longer

exposure of this autoradiogram (lane h) an RNA species of about 2200 nucleotides is found. The 2200, 1050 and 950 nucleotide transcripts are also found with probe 7 (lanes i and j). Since the 1300 and 1150 nucleotide transcripts are not detected with probe 7, we infer that they are transcribed from a maxicircle region covering the reading frames C.URF2 and C.URF1 (see Discussion). The 1050 and 950 nucleotide RNAs can be allocated to the maxicircle area covering the coxIII reading frame (807 bp). The 2200 nucleotide transcript is probably a precursor mRNA from which the 1300, 1150, 1050 and 950 nucleotide transcripts are produced by processing. The occurrence of paired transcripts has also been observed for some T.brucei genes (6). The possibility of a precursor-product relationship has been suggested, but experimental evidence is not available. Transcripts from the strand complementary to the C.URF2/C.URF1/coxIII and cytb coding strand, are not detected (probe 8, lane k). Lane 1 shows the cytoplasmic rRNAs after staining the gel with ethidium bromide. The transcription map presented here correlates well with that from the corresponding area in L.tarentolae (25).

DISCUSSION

Conservation of the maxicircle region between the genes for 9S mitochondrial rRNA and apocytochrome b

Sequence analysis of the region between the genes for 9S mitochondrial rRNA and apocytochrome b of maxicircle DNA from C.fasciculata, revealed three reading frames. One of these is 92% homologous to the coxIII gene, identified in the nucleotide sequence of the L.tarentolae maxicircle at a corresponding position. The remaining two reading frames are partially overlapping: the C-terminal 31 amino acids of C.URF2 overlap with the N-terminal region of C.URF1. Reading frames C.URF1 and C.URF2 are highly homologous to ORFs 4 and 3 of the L.tarentolae maxicircle (7), which overlap in a similar way. The three identified reading frames in this area of the C.fasciculata maxicircle are different from URFs 1 and 2, located in the corresponding area of the T.brucei maxicircle. With respect to gene-content, -organization and -homology most of this area of the maxicircle is highly conserved between L.fasciculata and L.tarentolae but not in T.brucei.

The only portion of this maxicircle segment that is conserved in all three trypanosome species is limited to two small regions: block A of 130 bp and block B of 80 bp. These elements are not likely to code for a protein or tRNA and their function is unknown (see Results).

coxIII

The coxIII amino acid sequences from C.fasciculata and L.tarentolae show a high level of conservation, but homology to the coxIII amino acid sequences from Aspergillus (27), Neurospora (8), Saccharomyces (28), Drosophila (29), Bovine (10) and Human (30) mitochondria, which show a mutual conservation varying from 60-90%, is rather low (20%). The conserved residues in the C.fasciculata and L.tarentolae coxIII sequences are limited to three hydrophobic regions: residues 94-104; 251-262 and 263-275 (indicated in Fig.2). The glutamic acid residue from beef heart coxII (31) which reacts with DCCD (dicyclohexylcarbodiimide; an inhibitor of proton-pumping activity), is invariant in the coxIII sequences of all species investigated so far (8), and is also present in C.fasciculata and L.tarentolae (E97, indicated by a dot in Fig.2). Knowledge about the function and functional domains of subunit III of the cytochrome oxidase complex is rather limited. However, it is likely that the three regions that are conserved in the two trypanosome species represent important structural and functional domains.

The coxI and coxII genes in the three trypanosome species are strongly conserved (73-79% at the nucleotide level, see ref.4). Since the coxIII genes in C.fasciculata and L.tarentolae are also highly homologous (85% at the nucleotide level), it is remarkable that no homologous coxIII gene could be detected in T.brucei, neither in kDNA, using a L.tarentolae coxIII gene-containing probe (7,32), nor in total DNA, using a C.fasciculata coxIII probe (Fig.4). This would imply that in cytochrome oxidase of T.brucei either a drastically divergent version of subunit III is present or that this subunit is lacking altogether. Characterization of trypanosomal cytochrome oxidase which will eventually solve this problem, is underway.

Translation initiation codons

Many of the genes in animal mtDNA do not appear to posses a conventional AUG-initiation codon, by virtue of the fact that interspecies homology between corresponding N-terminal gene sequences starts at AUA, AUU and AUC triplets (30). This observation prompted the suggestion that these triplets could serve as translational start codons in the animal mitochondrial genetic system, although direct experimental evidence is lacking. Following the same approach (comparison of N-terminal gene sequences) we provide evidence for the usage of still other initiation triplets in trypanosome mitochondrial genes: homology between corresponding T.brucei, C.fasciculata and L.tarentolae genes starts at leucine codons (UUA, CUG and UUG). The

suggestion that AUA may serve as an initiation codon in L.tarentolae (7) is, therefore, not supported by this analysis.

A possible frameshift gene

Two transcripts of 1150 and 1300 nucleotides respectively, are found in the region of the overlapping reading frames C.URF2/1. Both URFs map within the area covered by these transcripts, no other transcripts are found here. In trypanosome mitochondria the maxicircle appears to be transcribed into long multigene precursor RNAs, which are subsequently split into smaller RNA species (this paper; 6,25,33). In general, the length of these transcripts does not exceed the size of the corresponding reading frame by more than 200-300 nucleotides. No separate (small) transcript for C.URF2 was found, implicating one (or both) of the long RNAs mapped in this area as mRNA. In this case, however, exceptionally long C-terminal untranslated segments would occur (900 or 1050 nucleotides, respectively). Although a smaller URF2 covering transcript could have been missed in the analysis, we favour an alternative possibility, as outlined below. Other examples are found in trypanosome mitochondria in which a major transcript covers two overlapping reading frames: the coxII gene in T.brucei (11,14) C.fasciculata (11) and L.tarentolae (25), ORF5/ORF6 in L.tarentolae (25) and also ORF3/ORF4 in L.tarentolae (25) (corresponding to C.URF2/1) appears to be transcribed into an RNA of 1200 nucleotides covering both the ORFs. In an attempt to establish the mode of expression of the frameshifted coxII gene we have determined the sequence of T.brucei and C.fasciculata coxII-mRNA at the frameshift position. We found that the reading frame is restored at the RNA level by the presence of four extra nucleotides that are not encoded in the DNA ("RNA-editing", 9,14). Although the way in which this is achieved is unclear at the present time, the possibility arises that the other overlapping URFs mentioned, in fact, represent other examples of frameshift genes that are somehow "edited" during or after transcription. A further indication that C.URF2/1 could be such a frameshift gene comes from a close comparison between the nucleotide sequence of the area of reading frame overlap in L.tarentolae and that in C.fasciculata (see Fig.6). In the C.URF2/Lt.ORF3 area the amino acid homology is 100%, 10 silent third codon position substitutions occur. 4 amino acids before the end of C.URF2 this bias in nucleotide conservation changes frames, giving rise to a 100% amino acid conservation in the C.URF1/Lt.ORF4 reading frame. The remainder of the C.URF2 amino acid sequence is no longer homologous to the L.tarentolae ORF3 sequence and is 4 residues shorter. No other examples are known in which corresponding maxicircle genes in different

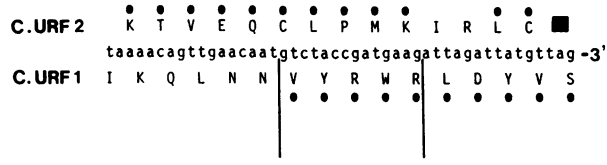

Fig.6: Aminoacid homology between C.fasciculata C.URF2/URF1 and L.tarentolae ORF3/ORF4. The C-terminal aminoacid sequence of C.URF2 is depicted at the top (stopcodon is represented by ■) and a part of the aminoacid sequence from the N-terminal region of C.URF1 at the bottom (residues 1 to 31 are overlapping with C.URF2; the sequence of residues 17-32 is shown). Marked residues (•) indicate conserved aminoacids in either Lt.ORF3 and Lt.ORF4, the reading frames in L.tarentolae that correspond to C.URF2 and C.URF1, respectively. The region in which the bias in nucleotide conservation between C.fasciculata and L.tarentolae switches frames encompasses 14 nucleotides. It is indicated in the figure as "frameshift area".

trypanosomes differ in length at the C-terminus, making it further unlikely that the C.URF2/Lt.ORF3 reading frame represents a separate gene. We are currently investigating whether the sequence in the region where C.URF2/1 overlap, is altered at the RNA level. Preliminary experiments, indeed, show that the RNA contains extra nucleotides in the expected position (frameshift area Fig.6) that restore the reading frame. These studies and a demonstration that this RNA is not transcribed from a second version of the C.URF2/1 gene, will be presented elsewhere.

ACKNOWLEDGEMENTS

We thank Profs P.Borst and L.A.Grivell and Dr.H.F.Tabak for critical reading of the manuscript. We are greatly indebted to Prof.J.H.van Boom, Gorlaeus Laboratory, University of Leiden, for the synthesis of oligonucleotides, to Mr.B.W.Swinkels, Netherlands Cancer Institute, for a TIM probe and to Ms G.J.M.Scholts for expert typing.

Abbreviations: kDNA, kinetoplast DNA; mtDNA, mitochondrial DNA; rRNA, ribosomal RNA; kb, kilobasepairs; nt, nucleotides; cytb, apocytochrome b; cox, cytochrome oxidase; C.URF, Unidentified Reading Frame of C.fasciculata mtDNA; Lt.ORF, Open Reading Frame of L.tarentolae mtDNA.

*To whom correspondence should be addressed

REFERENCES

1. Englund, P.T. (1981) In: Biochemistry and Physiology of Protozoa, 2nd Edn. Vol.4 (Levandowsky, M. and Hutner, S.A., Eds), Acad.Press, pp.334-383.

2. Borst, P. and Hoeijmakers, J.H.J. (1979) Plasmid 2, 20-40.
3. Hoeijmakers, J.H.J., Snijders, A., Janssen, J.W.G. and Borst, P. (1981) Plasmid 5, 329-350.
4. Benne, R. (1985) Trends in Genetics 1, 117-121.
5. Benne, R., Agostinelli, M., De Vries, B.F., Van den Burg, J., Klaver, B. and Borst, P. (1983) In: Mitochondria 1983: Nucleo-mitochondrial Interactions (Schweyen, R., Wolf, K. and Kaudewitz, F., Eds), De Gruyter, Berlin, pp.285-302.
6. Feagin, J.E., Jasmer, D.P. and Stuart, K. (1985) Nucl.Acids Res. 12, 4577-4596.
7. De la Cruz, V.F., Neckelmann, N. and Simpson, L. (1984) J.Biol.Chem. 259, 15136-15147.
8. Browning, K.S. and RajBhandary, U.L. (1982) J.Biol.Chem. 257, 5253-5256.
9. Jacq, C., Pajot, P., Lazowska, J., Dujardin, G., Claisse, M., Groudinsky, O., De la Salle, H., Grandchamp, C., Labouesse, M., Gargouri, A., Guiard, B., Spyridakis, A., Dreyfus, M. and Slonimski, P.P. (1982) In: Mitochondrial Genes (Slonimski, P.P., Borst, P. and G. Attardi, Eds) Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. pp.155-183.
10. Anderson, S., De Bruyn, M.H.L., Coulson, A.R., Eperon, I.C., Sanger, F. and Young, I.G. (1982) J.Mol.Biol. 156, 683-717.
11. Benne, R., Van den Burg, J., Brakenhoff, J., De Vries, B.F., Nederlof, P., Sloof, P. and Voogd, A. (1985) In: Achievements and perspectives of mitochondrial research. Vol.II: Biogenesis (Quagliariello, E., Slater, E.C., Palmieri, F., Saccone, C. and A.M. Kroon, Eds) Elsevier, Amsterdam, pp.325-336.
12. 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.
13. Payne, M., Rothwell, V., Jasmer, D.P., Feagin, J.E. and Stuart, K. (1985) Mol.Biochem.Parasitol. 15, 159-170.
14. Benne, R., Van den Burg, J., Brakenhoff, J.P.J., Sloof, P., Van Boom, J.H. and Tromp, M.C. (1986) Cell, 46, 819-826.
15. Hoeijmakers, J.H.J., Schoutsen, B. and Borst, P. (1982) Plasmid 7, 199-209.
16. Auffray, C. and Rougeon, F. (1980) Eur.J.Biochem. 107, 303-314.
17. Birnboim, H.C. and Doly, J. (1979) Nucl.Acids Res. 7, 151-152.
18. Sloof, P., Van den Burg, J., Voogd, A., Benne, R., Agostinelli, M., Borst, P., Gutell, R. and Noller, H.F. (1985) Nucl.Acids Res. 13, 4171-4190.
19. Poncz, M., Solowiejczyk, D., Ballantine, M., Schwartz, E. and Surrey, S. (1982) Proc.Natl.Acad.Sci.USA 79, 4298-4302.
20. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc.Natl.Acad.Sci.USA 74, 5463-5467.
21. Van der Marel, G.A., Marugg, J.E., De Vroom, E., Wille, G., Tromp, M., Van Boeckel, C.A.A. and Van Boom, J.H. (1982) Recl.Trav.Chim.Pays-Bas 101, 234-241.
22. Thomas, P.S. (1980) Proc.Natl.Acad.Sci.USA 77, 5201-5205.
23. Benne, R., De Vries, B.F., Van den Burg, J. and Klaver, B. (1983) Nucl.Acids Res. 11, 6925-6941.
24. Swinkels, B.W., Gibson, W.C., Osinga, K.A., Kramer, R., Veeneman, G.H., Van Boom, J.H. and Borst. P. (1986) EMBO J. 5, 1291-1298.
25. Simpson, A.M., Neckelmann, N., De la Cruz, V.F., Muhich, M.L. and Simpson, L. (1985) Nucl.Acids Res. 13, 5977-5993.
26. Cordingley, J.S. and Turner, M.J. (1980) Mol.Biochem.Parasitol. 1, 91-96.

27. Netzker, R., Köchel, H.G., Basak, N., Küntzel, H. (1982) Nucl.Acids Res. 10, 4783-4794.
28. Thalenfeld, B.E. and Tzagoloff, A. (1980) J.Biol.Chem. 255, 6173-6180.
29. De Bruijn, M.H.L. (1983) Nature 304, 234-241.
30. Anderson, S., Bankier, A.T., Barrell, B.G., De Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G. (1981) Nature 290, 457-465.
31. Prochaska, L.J., Bisson, R., Capaldi, R.A., Steffens, G.C.M. and Buse, G. (1981) Biochim.Biophys.Acta 637, 360-370.
32. Muhich, M.L., Simpson, L., Simpson, A.M. (1983) Proc.Natl.Acad.Sci.USA 80, 4060-4064.
33. Feagin, J.E. and Stuart, K. (1985) Proc.Natl.Acad.Sci.USA 82, 3380-3384.