Volume 11 Number 20 1983 Nucleic Acids Research

The nucleotide sequence of a segment of *Trypanosoma brucei* mitochondrial maxi-circle DNA that contains the gene for apocytochrome *b* and some unusual unassigned reading frames

Rob Benne, Berend F.De Vries, Janny Van den Burg and Bep Klaver

Section for Medical Enzymology, Laboratory of Biochemistry, University of Amsterdam,
P.O.Box 60.000, 1005 GA Amsterdam, The Netherlands

Received 5 September 1983; Accepted 30 September 1983

ABSTRACT

The nucleotide sequence of a 2.5-kb segment of the maxi-circle of Trypanosoma brucei mtDNA has been determined. The segment contains the gene for apocytochrome b, which displays about 25% homology at the amino acid level to the apocytochrome b gene from fungal and mammalian mtDNAs. Northern blot and S_1 nuclease analyses have yielded accurate map positions of an RNA species in an area that coincides with the reading frame. The segment also contains two pairs of overlapping unassigned reading frames, which lack homology with any known mitochondrial gene or URF. The DNA sequence in these areas is AG-rich (70%), resulting in URFs with an unusually high level of glycine and charged amino acids (60%). They may not encode proteins, in spite of their size and the fact that abundant transcripts are mapped in these areas.

INTRODUCTION

Kinetoplast DNA (kDNA) is the unusual mitochondrial DNA of trypanosomes. It is found in a specialized portion (the kinetoplast) of the single mitochondrion of these organisms. It consists of a complex network of catenated circles, which can be divided into two size classes [1-3]. In T. brucei the network consists of about 10^4 mini-circles (0.3 µm, 1 kb) and 10^2 maxi-circles (6 µm, 20 kb). In most trypanosome species the mini-circles are heterogeneous in sequence [4-8], they evolve rapidly [9], they are not transcribed [10-14] and play an as yet unresolved role in mitochondrial housekeeping. The maxi-circles are probably the equivalent of mtDNA found in other organisms, since they are homogeneous in sequence and are transcribed [10-15]. As yet, however, few trypanosome mitochondrial genes have been characterized. The tentative identification of two major poly(A)$^-$ RNAs as the ribosomal RNAs (rRNAs) [2,10,12] has been subsequently substantiated by sequence analysis of T. brucei maxi-circle segments [16]. Moreover, the poly(A)$^+$ transcripts are sufficient both in size and number to account for a set of proteins normally encoded by the mitochondrial genome [14,15,17]. Cross-hybridization studies [18] between trypanosome mtDNA fragments and a probe for subunit II of cytochrome c oxidase from Zea mais and yeast tentatively localize the cox II gene on a large EcoRIxHindIII fragment of the maxi-circle (R1-D1; see ref. 9). The use of heterologous probes at low stringency of hybridization, however, does not appear to give conclusive results in all cases, since e.g. yeast probes for ATPase subunit 6 and cytochrome c oxidase subunit III hybridize with 10 kb or more of maxi-circle DNA from Leishmania tarentolae [19]. Furthermore, attempts to unambiguously demonstrate the existence of a separate mitochondrial protein-synthesizing system have been unsuccessful in trypanosomes [17,20]. Therefore, the most rapid approach to unequivocally identify the mitochondrial protein genes in trypanosomes seems to undertake nucleotide sequence analysis of the maxi-circle.

In this report we present the sequence of 2520 nucleotides of the above-mentioned R1-D1 fragment on which we identify the gene for apocytochrome b and a number of URFs. The URFs are unusual and may not represent protein-coding genes.

MATERIALS AND METHODS

Materials - Restriction endonucleases were from New England Biolabs or Boehringer Mannheim; DNA polymerase (large fragment), calf intestine phosphatase and T4 DNA ligase from Boehringer; Avian Myeoblastosis virus reverse transcriptase from Dr. J.Beard (Life Sciences Inc., St. Petersburg, Fa, USA); exonuclease Bal-31 from New England Biolabs or Bethesda Research Laboratories; low-melting agarose from Bethesda Research Laboratories.

DNA and RNA - Trypanosome RNA (T. brucei 427, culture form) was isolated with the LiCl precipitation method [21]. Poly(A)$^+$ RNA was isolated according to ref. 22; RNA was stored at -20°C as an ethanol precipitate. Plasmid DNA and M13 RF DNA were isolated according to ref. 23.

Assays - Restriction enzyme digestion, agarose gel electrophoresis and blot analysis of RNA or DNA fragments, nick translation and hybridization were performed as described [10]. S_1 nuclease protection experiments were performed as described in ref. 24, outlined in detail in ref. 16. 0.5 µg of poly(A)$^+$ RNA was used. S_1 nuclease incubation was at 30°C for 30 min. Bal-31 digestion was performed at 30°C for the periods of time indicated; routinely 0.5 U of Bal-31 was used per µg of DNA. Incubations were stopped by the addition of phenol.

Cloning in bacteriophage M13 and sequence analysis - 1. Cloning of the EcoRIxHindIII fragment: Recombinant plasmids or recombinant bacteriophage M13 RF DNA containing the R1 R2 maxi-circle DNA fragment [10,25] were used as starting material for kDNA sequence determination. M13 clones containing a varying part of the EcoRI-HindIII fragment (R1-D1) in two orientations were generated with Bal-31 using non-random cloning procedures [26], except that the gel isolation step of Bal-31-treated fragments was omitted. Plasmid (or M13 RF) DNA was cut either with SstI (procedure A) or with HindIII (procedure B) (see Fig. 1) and treated with Bal-31 for different time periods with the aim of successive removal of 3 kb from all ends in steps of about 250 nucleotides. The DNA from a number of different Bal-31 digestions was pooled, cut with HindIII (procedure A) or EcoRI (procedure B) and ligated into M13 mp9 cut with HindIII and HindII (procedure A) or into M13 mp8 cut with EcoRI and HindII (procedure B) [27].

The ligation mixtures were used to transform Escherichia coli strain JM101 and recombinant plaques containing parts of the R1-D1 fragment were picked following plaque lifting [28] and hybridization of the filter with a Q1-Q2 probe (see ref. 10). The approach resulted in a large number of recombinant DNA clones, which are listed in Fig. 1: the clones obtained with procedure A (MCB clones) all have the part adjacent to the HindIII site in common, whereas the part progressively shortened with Bal-31 is oriented to the vector's priming site. The sequence reads away from the EcoRI site; the starting positions are given. The BCM clones contain inserts of progressively smaller sizes of opposite polarity, their sequence reads towards the EcoRI site . All cloning reactions were performed with calf-intestine phosphatase-treated vector with 0.5 µg of M13 DNA, a vector/insert ratio varying from 3:1 to 1:3 at a total DNA concentration of 30 µg/ml.

2. DNA sequence analysis: DNA sequence analysis was carried out by the dideoxynucleotide chain-termination technique [29]. The nature of the procedure followed provided an ample source of overlapping clones of both orientations. All clones mentioned in Fig. 1 have been submitted to sequence analysis.

RESULTS

Sequence analysis

A detailed restriction map of the maxi-circle of T. brucei has been published by Hoeijmakers et al. [10]. The nucleotide sequence of a 3-kb EcoRIxHindIII fragment (R1-D1 [10]) was determined using a non-random M13 sequencing method [26]. The sequence strategy followed is shown in Fig. 1 as discussed extensively above.

Fig. 2 shows the sequence of 2520 nucleotides starting with the EcoRI site. Open reading frames longer than 75 amino acid residues are also indicated. These were obtained by translating the nucleotide sequence into amino acids with a genetic code in which UGA specifies tryptophan, a modification of the universal code found in mitochondria from all species studied so far [30-37]. The longest reading frame (nucleotide 626-1714) has been identified as the T. brucei gene for apocytochrome b, the others have remained unassigned (see below). They vary in size from 85 (URF 3a) to 363 amino acid residues (the gene for apocytochrome b). The sequence of URF 2a is completely contained within URF 2, whereas 90% of URF 3a overlaps with URF 3.

The identification of the apocytochrome b gene

The nucleotide and amino acid sequences of the reading frames were compared to sequences of previously identified mitochondrial genes from Saccharomyces cerevisiae [39-46], Aspergillus nidulans [34,35] and beef [32]. The longest frame (see Fig. 2) displays significant homology with the gene for apocytochrome b from these organisms: about 45% at the nucleotide and 25% at the amino acid level. The highest degree of homology at the amino acid

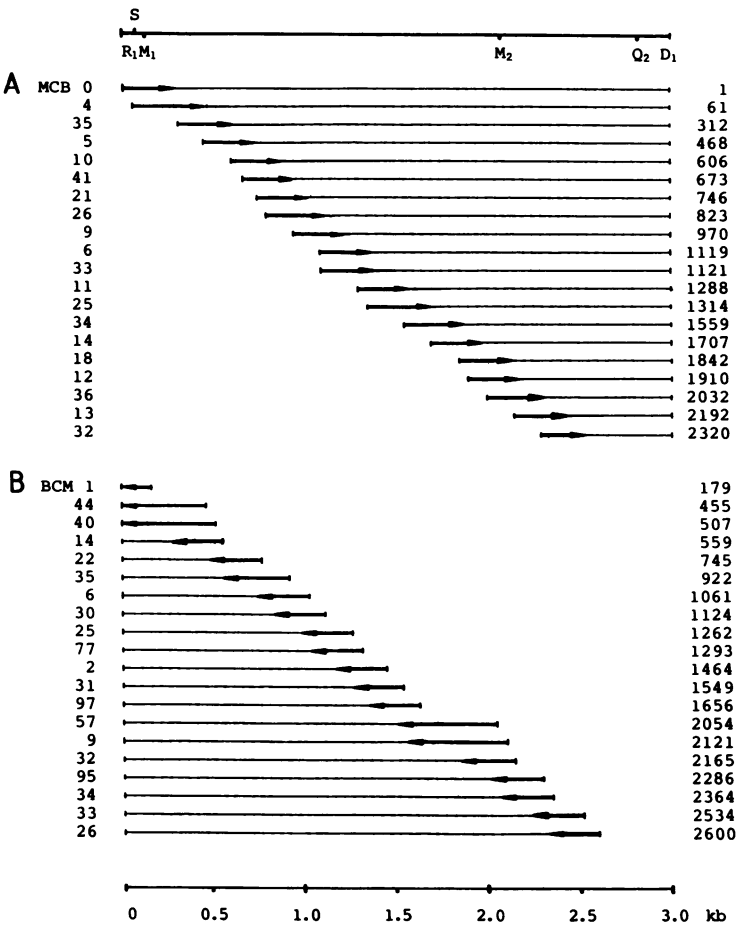

Fig. 1. Sequence strategy. The map is derived from ref. 10 and represents the R1-D1 area. Phage M13 clones containing R1-D1-derived inserts in two orientations were constructed and submitted to sequence analysis as described. A: Clones obtained with procedure A are called MCB. B: Clones of the opposite orientation, obtained following procedure B, are called BCM. The arrow indicates the region of each clone used in the sequence determination. The number at the right hand side indicates the position in nucleotides from the EcoRI site at which the insert sequence starts. Abbreviations: R, EcoRI; M, MboI; D, HindIII; S, SstI; Q, TaqI.

level occurs in an alignment assuming only a few internal insertions/deletions, as shown in Fig. 3. The T. brucei gene is shorter both at the N- and the C--termini (22 and 18 amino acid residues, respectively) and the protein most probably starts at the methionine at position 14. The alignment of Fig. 3

```
URF 2 ----->
 E  F  S  F  V  E  H  P  K  G  I  W  G  I  F  G  G  A  R  R  R  A  E  H  Y  L  R  R  A  G  A  E  G  F  L  R  K  E  G  T
GAATTCAGCTTTGTGGAACACCCCAAAGGGATTTGAGGAATTTTTGGGGGAGCTCGACGGCCGGCCGGAGCATTATTTGAGGAGGGCGGGAGCAGAAGCCTTTCTGAGGAAAGAGGGGACC
        10        20        30        40        50        60        70        80        90       100       110       120
                                                         URF 2a ----->
 E  I  D  E  G  Y  F  L  V  I  E  D  C  L  K  L  N  K  K  A  F  W  K  Q  I  F  G  G  H  R  Q  R  R  R  V  L  E  E  F  V
                                                        K  G  F  L  E  G  D  F  W  G  T  P  P  E  E  E  G  F  G  R  V  C  F
CACATCCATCAAGGTTATTTTTTGGTTATTGAGGATTGTTTAAAATTGAATAAAAAGGCTTTTTGGAAGCGGATTTTTGGGGGACACCGCCAGACGAGGAGGGTTTTCCAAGAGTTTGTT
       130       140       150       160       170       180       190       200       210       220       230       240

 L  R  Q  G  F  E  Q  R  G  E  R  E  R  E  R  N  Q  P  E  R  R  V  E  E  G  Q  F  E  Q  E  Q  R  L  S  D  Q  Q  K  E  G
  E  R  R  F  W  G  E  G  R  E  G  T  G  E  E  R  T  R  E  E  S  W  G  R  R  F  W  R  R  G  E  A  F  G  P  R  E  G  R  E
TTGAGACGAGGTTTTGAGGGGAGGCGAGAGAGGGAACGGGAGAGGAACGGACCAGAGAGGAGAGTTGAGGAAGGCGGTTTTGAAGGAGAGGGGAGGCTTTCCGACCAAGGGAAGGAACGG
       250       260       270       280       290       300       310       320       330       340       350       360

 R  L  R  K  G  K  T  I  C  E  G  E  G  F  L  E  Q  F  W  E  E  R  Q  F  Q  E  T  R  W  D  C  L  Q  K  Q  R  G  F  W  A
  V  K  K  R  K  N  N  L  W  Q  R  R  V  F  G  G  V  L  G  R  E  G  F  W  G  N  Q  M  R  L  F  A  E  T  K  G  F  L  G  K
AGGTTAAGAAAAGGAAAAACAATTTGTGAGGGAGAAGGGTTTTTGGAGGGGTTTTGGGAAGAGAGGGGTTTTGGGGAAACCAGATGAGATTGTTTGCAGAAACAAAGGGGTTTTTGGGGA
       370       380       390       400       410       420       430       440       450       460       470       480

 K  E  Y  N  L  Q  R  G  E  S  Q  R  R  N  T  G  Q  K  T  G  F  R  K  R  E  R  G  E  G  K  G  F  S  W  N  E  E  V  V  C
  G  I  Q  F  A  E  G  G  E  R  K  E  E  H  G  R  E  D  R  I  *  (end URF 2a)
AAGGAATACAATTTGCAGAGGGGGGAGAGCGGAAGGAGGAACACGGGAGGGAAGACAGGATTTAGGAAGGGAGAGAGAGGAGAGGGGAAAGGGTTTAGTTGGAATGAAGAGGTAGTTTGT
       490       500       510       520       530       540       550       560       570       580       590       600
                                 Cyt. b ----->
 R  K  L  R  I  M  V  I  N  F  I  *
                              I  L  Y  K  S  G  E  K  R  K  G  L  L  M  S  G  C  L  Y  R  I  Y  G  V  G  F  S  L  G  F  F  I
AGGAAGTTAACAATAATGGTTATAAATTTTATATAAAAGGGGAGAAAAAAAGAAAGGGTCTTTTAATGTCAGGTTGTTTATATAGAATATATGGGGTAGGTTTTAGTTTAGGATTTTTTAT
       610       620       630       640       650       660       670       680       690       700       710       720

 A  L  Q  I  I  C  G  V  C  L  A  W  L  F  F  S  C  F  I  C  S  N  W  Y  F  V  L  F  L  W  D  F  D  L  G  F  V  I  R  S
AGCATTGCAAATAATTTGTGGAGTGTGTTTAGCTTGATTATTTTTTAGTTGTTTTATTTGTTCAAATTGATATTTTGTATTATTTTTATGAGATTTTGATTTGGGTTTTGTGATAAGAAG
       730       740       750       760       770       780       790       800       810       820       830       840

 V  H  I  C  F  T  S  L  L  Y  L  L  L  Y  I  H  I  F  K  S  I  T  L  I  I  L  F  D  T  H  I  L  V  W  F  I  G  F  I  L
TGTACATATATGTTTTACATCTTTATTATATTTACTATTATATATCCATATATTTAAGTCAATAACGTTAATAATATTGTTTGACACACATATATTAGTATGATTTATAGGTTTTATATT
       850       860       870       880       890       900       910       920       930       940       950       960

 F  V  F  I  I  I  I  A  F  I  Q  Y  V  L  P  C  T  M  M  S  Y  W  G  L  T  V  F  S  N  I  I  A  T  V  P  I  L  G  I  W
GTTTGTATTTATAATAATAATAGCTTTTATAGCATATGTACTGCCTTGTACAATGATGTCATACTGAGGTTTAACGGTGTTTAGTAATATTATAGCAACAGTACCAATTTTAGGTATATG
       970       980       990      1000      1010      1020      1030      1040      1050      1060      1070      1080

 L  C  Y  W  I  W  G  S  E  F  I  N  D  F  T  L  L  K  L  H  V  L  H  V  L  L  P  F  I  L  L  I  I  L  I  L  H  L  F  C
ATTATGTTATTGAATTTGGGGAAGTGAATTTATAAACGATTTTACATTATTAAAGTTACATGTATTACATGTGTTATTACCATTTATATTACTAATAATATTAATTTTACATTTATTTTG
      1090      1100      1110      1120      1130      1140      1150      1160      1170      1180      1190      1200

 L  H  Y  F  M  S  S  D  A  F  C  D  R  F  A  F  Y  C  E  R  L  S  F  C  M  W  F  Y  L  R  D  M  F  L  A  F  S  I  L  L
TCTACATTATTTTATGACTTCTGATGCATTTTGTGATAGGTTTGCATTTTATTGTGAAAGATTAAGTTTTTGTATGTGGTTTTATTTGAGAGATATGTTTTTAGCATTTTCAATATTATT
      1210      1220      1230      1240      1250      1260      1270      1280      1290      1300      1310      1320

 C  M  M  Y  V  I  F  I  N  W  Y  F  V  F  H  E  E  S  W  V  I  V  D  T  L  K  T  S  D  K  I  L  P  E  W  F  F  L  Y  L
ATGTATGATGTATGTTATATTTATAAATTGGTATTTTGTATTTCATGAGGAATCTTGAGTTATAGTAGATACACTAAAAACATCAGATAAAATATTACCAGAATGATTTTTTTTTGTATTT
      1330      1340      1350      1360      1370      1380      1390      1400      1410      1420      1430      1440

 F  Q  F  L  K  A  I  P  D  K  F  M  G  L  F  L  M  V  I  L  L  F  S  L  F  L  F  I  L  N  C  I  L  W  F  V  Y  C  R  S
ATTCCGTTTTTTAAAGCCAATCCCAGATAAGTTTATGGGTTTGTTTTTAATGGTTATTTTATTATTCTCATTATTTTTATTTATATTGAATTGTATATTATGATTTGTGTATTGTAGAAG
      1450      1460      1470      1480      1490      1500      1510      1520      1530      1540      1550      1560

 S  L  L  W  L  T  Y  S  L  I  L  F  Y  S  I  W  M  S  G  F  L  A  L  Y  V  V  L  A  Y  P  I  W  M  E  L  Q  Y  W  V  L
TTCATTATTATGATTAACATATTCGTTAATATTATTTTATAGTATATGAATGAGTGGTTTTTTAGCATTATATGTAGTATTAGCATATCCAATATGAATGGAATTACAATACTGAGTATT
      1570      1580      1590      1600      1610      1620      1630      1640      1650      1660      1670      1680
                                                                                  URF 3 ----->
 L  L  F  L  L  I  V  C  R  L  D  *                URF 3a ----->                  S  K  R  Q  I  L  Q  G  R  E  G  D  R  R  G  N
                                                    K  N  K  Y  F  D  I  I  K  V  K  E  E  F  W  A  E  E  K  E  T  G  E  E  M
ATTATTATTTTTGTTGATAGTGTGTAGGTTAGATTAGTTTAGAATAAAAAAATAAGTATTTTGATATTATTAAAGTAAAAGAGGAATTTTGGGGGGAAGAGAAGGAGAGAGGAGAGGAAA
      1690      1700      1710      1720      1730      1740      1750      1760      1770      1780      1790      1800

 E  G  E  R  F  W  E  G  Q  F  L  R  Q  G  K  R  I  L  N  L  N  Y  L  F  K  L  W  E  R  S  K  E  E  K  S  R  Q  I  L  R
  K  E  K  G  F  E  R  G  V  F  W  G  E  E  K  E  F  W  I  W  T  I  C  L  S  Y  G  R  E  A  R  R  R  K  V  G  E  F  W  G
TGAAGGAGAAAGGTTTTGAGAGGGGGGTTTTTTGAGGGGAGGAAAAAGAATTTTGAATTTGAACTATTTGTTTAAGTTATGGGAGAGAAGCAAGGAGGAGAAAAGTAGGGGAATTTTGAG
      1810      1820      1830      1840      1850      1860      1870      1880      1890      1900      1910      1920

 R  F  L  Q  R  Q  Q  R  A  T  A  V  L  K  T  P  I  F  R  R  I  R  Q  E  K  R  G  N  G  I  G  N  C  L  C  Q  T  F  R  R
  D  S  W  G  E  A  G  G  R  R  R  F  W  K  H  P  F  L  G  G  *  (end URF 3a)
GACATTCTTGGGGAGACGCCGGGCGGGCGACGGCGGTTTTGAAAACACCCATTTTTAGGACCATAACAGGGGAGAAAAGGGGAAATGGAATTGGGAATTGCCTTTGCCAAACTTTTAGAAG
      1930      1940      1950      1960      1970      1980      1990      2000      2010      2020      2030      2040

 K  S  R  K  Q  *  (end URF 3)
AAAGAGCAGGAAAGGTTAGGGGGAGGAGAGAAGAAAGGGAAAGTTGTGATTTTGGAGTTATAGAATAAGATCAAATAAGTTAATAATATCAAAGAAAAGTATATATACGCTAGAACAAAT
      2050      2060      2070      2080      2090      2100      2110      2120      2130      2140      2150      2160

GAAGAATAATAAATTTTTTAATATTGATAAAAGATAATTTTACAACTCAAAAACCAAGAAATTGATAAGAAAAAAATAAATATATTAACAATTAATCTAAAATAAAAAATATAAATGATAAT
      2170      2180      2190      2200      2210      2220      2230      2240      2250      2260      2270      2280

AAGTCATATTATAAAGAAAAAGCCAATACAAATACAAAGGTAACTTAGTTGTAATAATAGACAGAAAACTTTGATAAAAAATCCAAATACAATTGGAATAGCTCCAATGCAAAGAAACAG
      2290      2300      2310      2320      2330      2340      2350      2360      2370      2380      2390      2400

ACATGCAAGTAGTAAACTTATTAAAAAGTTATTAAAAAAAGAAAAAAATATGAAGTTGATTAAAAAATAGTTTTCATTGTATTTAAAGTCAAAAATATTATATATAATAAAAAAATAGTA
      2410      2420      2430      2440      2450      2460      2470      2480      2490      2500      2510      2520
```

Fig. 2. The nucleotide sequence of the R1-D1 fragment. The sequence of the coding strand (see Figs 5 and 6) is shown for nucleotide 1-2520. Map coordinates of the reading frames are: URF 2, 1-633; 2a, 174-542; 3, 606-1714; 4, 1754-2056; 4a, 1728-1982. URF 1 has been discussed elsewhere [17]. The sequence was assembled using the computer programmes of Staden [38].

also shows that a high percentage (about 50%) of the amino acid substitutions between the *T. brucei* sequence and the other genes is conservative (e.g. arginine for lysine at positions 100, 179, 221 and 331; glutamate for aspartate at positions 220 and 258; numerous substitutions involving leucine, isoleucine, alanine, valine).

Invariant histidines in the N-terminal moiety of apocytochrome *b* are thought to play a role in the attachment of heme groups [47]. In this light it is striking that of the seven invariant histidines in beef, yeast and *Aspergillus*, five are also conserved in trypanosomes. This would be sufficient for the binding of the two heme groups. Moreover, a very high percentage of α-helix breaking residues has been conserved: 6 out of 6 proline and 9 out of 13 glycine residues of the *T. brucei* sequence are invariant in all four species. We have calculated the segmental membrane solubility of apocytochrome *b* proteins according to Von Heijne [48]. As shown in Fig. 4, the distribution of membrane (in)soluble residues of the *T. brucei* sequence is very similar to that of yeast and beef apocytochrome *b*.

All these observations identify a protein of this sequence as a functional homologue of apocytochrome *b*. Direct verification at the protein level has not been possible due to the lack of cross-reactivity between the *T. brucei* protein and antibodies directed against yeast apocytochrome *b* and the practical difficulties of purification of sufficient amounts of mitochondrial proteins from *T. brucei*.

The genetic code

The alignment shown in Fig. 3 also permits a number of codon assignments to be made as shown in Table I. Apart from the UGA-tryptophan assignment, no other deviations from the standard code occur in trypanosome mitochondria:

1. AGA (and AGG; not shown) specifies arginine and not serine (as in *Drosophila* [36,37]) or a stop (as in mammals [49]), since this codon occurs in positions corresponding to conserved arginine or lysine residues in three out of five cases.

2. AUA codes for isoleucine and not for methionine as in *Drosophila* [36,37], mammals [49] and yeast [50]. In about 70% of the cases an AUA codon lines up with isoleucine or with residues which result from conservative substitutions such as leucine, valine and alanine, whereas the average content of these amino acids in an apocytochrome *b* gene is only about 40%.

Moreover, 34 out of the 43 possible isoleucine-specifying codons are AUA in the *T. brucei* sequence. Translation of AUA by methionine would result in a rather unusual amino acid composition for *T. brucei* apocytochrome *b*:

```
1
   ---------- ---------- --MSGCLYRI YGVGFSLGFF IALQIICGVC LAWLFFSCFI
                                       *  **     ***  *   **    *
   MTNIRKSHPL IKIVNNAFID LPAPSNISSW WNFGSLLGIC LILQILTGLF LAIHYTSDTT
                                       *  **      **  *    *    *
   -MAFRKSNVY LSLVNSYIID SPQPSSINYW WNMGSLLGLC LVIQIVTGIF MAMHYSSNIE
                               *       *  *       **  **  **
   -MRILKSHPL LKIVNSYIID SPQPANLSYL WNFGSLLALC LGIQIVTGVT LAMHYTPSVS

61
   CSNWYFVLFL WDFDLGFVIR SVHICFTSLL YLLLYIHIFK SITLIILFDT HILVWFIGFI
               *   *  **   *    *      ** *                   * ** *
   TAFSSVTHIC RDVNYGWIIR YIHANGASMF FICLYMHVGR GLYYGSYTF- -LETWNIGVI
               *   *   *   *    *         *  *                *  * *
   LAFSSVEHIM RDVHNGYILR YLHANGASFF FMVMFMHMAK GLYYGSYRSP RVTLWNVGVI
               *   *   *   *    *     * * **                * * **
   EAFNSVEHIM RDVNNGWLVR YLHSNTASAF FFLVYLHIGR GLYYGSYKTP RTLTWAIGTV

121
   LFVFIIIIAF IGYVLPCTMM SYWGLTVFSN IIATVPILGI WLCYWIWGSE FINDFTLLKL
   *    *  ** ******     * ** **  *      *  *   *  ****        **
   LLLTVIATAF IGYVLPWGQI SFWGATVITN LLSAIPYIGT NLVEWIWGGF SVDKATLTRF
    *   *  **  **      * * ** **  *      *  *      * **        *
   IFILTIATAF LGYCCVYGQM SHWGATVITN LFSAIPFVGN DIVSWLWGGF SVSNPTIQRF
           **  *****   * * ** **  *      *  *       ***     *  **
   ILIVMMATAF LGYVLPYGQM SLWGATVITN LMSAIPWIGQ DIVEFIWGGF SVNNATLNRF

181
   HVLHVLLPFI LLIILILHLF CLHYFMSSDA FCDRFAFYCE RLSFCMWFYL RDMFLAFSIL
      *  ****    * * **   **    *                 *        *   *
   FAFHFILPFI IIAIAIVHLL FLHET-GSN- NPTGISSDVD KIPFHPYYTI KDILGALLLI
     ** * ***      * **   **    *              *      *    *    *
   FALHYLVPFI IAAMVIMHLM ALHIM-GSS- NPLGITGNLD RIPMHSYFIF KDLVTVFLFM
     ** ****  *      **    *    *              ** *   *    *    *
   FALHFLLPFV LAALALMHLI AMHDTVGSG- NPLGISANYD RLPFAPYFIF KDLITIFIFF

241
   LCMMYVIFIN WYFVFHEESW VIVDTLKTSD KILPEWFFLY LFGFLKAIPD KFMGLFLMVI
   *                          * *    * *** **      *  **  *  *
   LALILLVLFA PDLLGDPDNY TPANPLNTPP HIKPEWYFLF AYAILRSIPN KLGGVLALAF
   *      *        *          * *    * ***  *      *  *** *  *   *
   LILALFVFYS PNTLGHPDNY IPGNPLVTPA SIVPEWYLLP FYAILRSIPD KLLGVITMFA
          *          *   *      *    * ***  *      *  **  *  *   *
   IVLSIFVFFM PNALGDSENY VMANPMQTPP AIVPEWYLLP FYAILRSIPN KLLGVIAMFA

301
   LLFSLFLFIL NCILWFVYCR SSLLWLTYSL ILFYSIWMSG FLALYVVLAY PIWMELQYWV
       * *  *               *        *               *    *     *
   SILILALIPL LHTSKQRSII FRPLSQCLFW ALVADLLTLT WIGGQPVE-H PYITIGQLAS
       *                    *       *        *       *    *     *
   AILVLLVLPF TDRSVVRGNT FKVLSKFFFF IFVFNFVLLG QIGACHVE-V PYVLMGQIAT
       *                    *       *                *    *     *
   AILALMVMPI TDLSKLRGVQ FRPLSKVVFY IFVANFLILM QIGAKHVE-T PFIEFGQIST

361
   LLLFLLIVCR LD-------- ----------           T.brucei
    * ****
   VLYFLLILVL IPTAGTIENK LLKW------           beef
      *
   FIYFAYFLII VPVISTIENV LFYIGRVNK-           S.cerevisiae
      *
   IIYFAYFFVI VPVVSLIENT LVELGTKKNF           A.nidulans
```

Fig. 3. Sequence comparison between apocytochrome b genes. The nucleotide sequence from position 626-1714 (see Fig. 2) was translated into amino acids and lined up with the apocytochrome b genes from beef, S. cerevisiae and A. nidulans. (-) indicates the position at which a deletion is assumed. (*) indicates homology with one of the apocytochrome b genes.

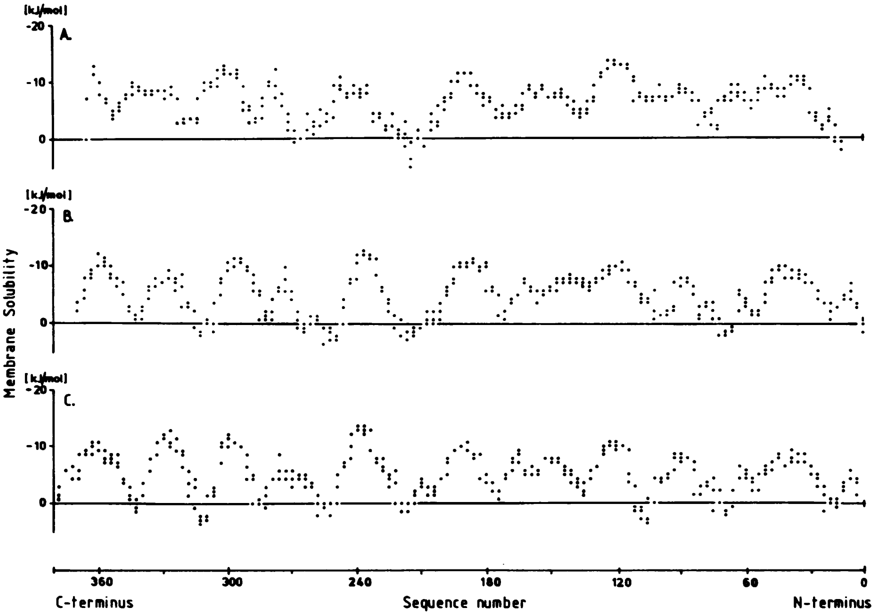

Fig. 4. Membrane solubility of apocytochrome b. The membrane solubility of blocks of 10 amino acids as a function of sequence position was calculated according to Von Heijne [48]. Values represent the average free energy gained upon the transfer of these blocks from a random coil in H_2O to an α-helical structure in the membrane. A: T. brucei; B: beef; C: S. cerevisiae strain D273-10B.

13.1% methionine and 2.6% isoleucine. The values obtained assuming that AUA specifies isoleucine in T. brucei mitochondria (3.4% methionine and 12.3% isoleucine) are in much better agreement with the average methionine and isoleucine content of the apocytochrome b genes from beef, yeast and Aspergillus (4.2 and 9.7%, respectively).

3. Similar arguments apply to AUU and the CUN family. AUU codes for isoleucine and not for methionine (as in mammals [49]), whereas the CUN codons line up mostly with leucine and not with threonine for which they code in yeast [51].

The URFs

The nucleotide composition of the URFs of the R1-D1 fragment is compared to that of the previously described URF 1 [17] and to that of the apocytochrome b gene. Table II shows a marked difference in the content of G and T residues between the apocytochrome b gene and the URFs: 16% G, 47% T

TABLE I

AMINO ACID ASSIGMENTS FOR UGA, AGA, AUA, AUU AND CUN

The frequency is given with which the codons in the *T. brucei* apocytochrome *b* gene align with a given amino acid in the sequences from beef (B), *S. cerevisiae* (S), *A. nidulans* (A) or in all three (B/S/A). The alignment of Fig. 3 was used.

Codon	Number	Amino acid	B	S	A	B/S/A
UGA	14	Tryp	4	4	3	3
		Tyr	1	1	1	1
		Others	9	9	10	2
AGA	5	Arg	1	2	2	1
		Lys	2	1	1	1
		Ser	1	0	0	0
		Others	1	2	2	0
AUA	34	Ile	10	6	5	2
		Leu	6	4	6	2
		Val	4	4	4	1
		Ala	4	3	2	2
		Met	0	2	4	0
		Others	10	15	13	4
AUU	7	Ile	2	1	1	0
		Leu	2	2	2	0
		Val	0	1	1	0
		Ala	0	1	1	0
		Met	0	0	0	0
		Others	3	2	2	0
CUN	5	Leu	3	2	2	0
		Thr	0	0	0	0
		Others	2	3	3	0

and 40% G, 22% T, respectively, whereas the A and C content is the same. Moreover, comparison of the DNA sequences of the URFs reveals the presence of small imperfectly repeated elements of relatively simple sequence, resulting in a low but significant homology (about 40%) between the DNA of different (non-overlapping) URFs. The statistical value calculated for DNAs of this nucleotide composition lies around 30% [52]. These properties of the DNA sequence result in URFs with very similar amino acid compositions, even though the reading frames overlap. Table II and Fig. 2 show that they all contain an extremely high content of glycine (about 18%) and of charged amino acids (about 40%, predominantly arginine and glutamate), whereas long stretches of hydrophobic residues which are typical for mitochondrial proteins are lacking. Not surprisingly, they lack homology with any of the known mitochondrial genes or URFs. Instead, the homology between them is quite high: e.g. 20% between URF 1 and URF 2a, assuming a few insertions and deletions. Whether or not we are dealing with protein genes is discussed below.

TABLE II

NUCLEOTIDE AND AMINO ACID COMPOSITION OF *T. brucei* MAXI-CIRCLE URFs

	Nucleotide composition				Amino acid composition				
	G	A	T	C	Gly	Lys	Arg	Glu	Others
URF 1	40.2	28.0	21.2	10.6	18.3	7.9	12.7	11.9	49.2
URF 2	40.0	30.3	22.1	7.6	17.5	7.1	14.7	15.2	45.5
URF 2a	43.6	30.1	19.0	7.3	20.3	5.7	16.3	17.1	40.6
URF 3	37.3	34.7	21.5	6.6	18.8	8.9	19.8	7.9	44.6
URF 3a	35.3	34.9	24.3	5.5	14.1	11.8	9.4	17.6	47.5
Cyt. *b*	16.3	30.5	46.8	6.3	4.4	2.5	2.2	3.0	87.9

Precise mapping of RNAs

Previous studies [10] have localized an RNA of about 570 nucleotides in the area corresponding to URF 2a, whereas a second, larger RNA of about 1120 nucleotides was tentatively mapped around a TaqI site (Q2; Fig. 1) at a distance of approximately 3 kb from the EcoRI site R1. In order to localize these RNAs more precisely and to determine a possible connection with the reading frames observed, their exact position was determined by Northern blots and S_1 nuclease analysis with poly(A)$^+$ RNA isolated from culture form *T. brucei*.

The results of the Northern blot analysis are shown in Fig. 5. The filters were probed with single-stranded M13 DNA, containing a varying part of the EcoRIxHindIII fragment in two orientations, as indicated. Depending on the orientation and map position of the inserts, three (lane 1), two (lane 2) or one (lane 3) RNA species appear, the positions of which are indicated in the figure (RNA 2, 550; 3, 1200; 4, 400 nucleotides, resp.). The results reveal the order from left to right in which these RNAs occur on the map of Fig. 1 (RNA 2-3-4). This experiment shows further that the direction of transcription is the same for all three of these RNAs (also from left to right), since they hybridize to M13 clones with kDNA inserts of the same polarity. The same direction of transcription has been found for the ribosomal genes [10]. No RNA species were found when the filters were probed with phage M13 DNA containing kDNA inserts of opposite polarity (Fig. 5, lanes 4 and 5), indicating that at least in this area the maxi-circle is asymmetrically transcribed.

In order to map these RNAs more precisely, S_1 nuclease analysis was performed (Fig. 6). *T. brucei* poly(A)$^+$ RNA was hybridized to phage M13 clone DNAs containing a varying portion of the EcoRIxHindIII fragment and S_1

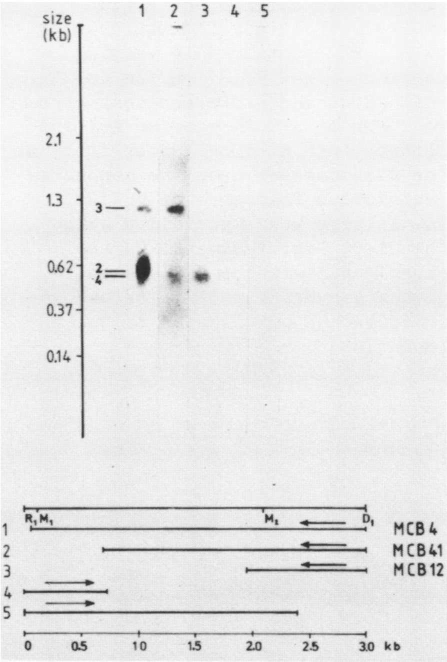

Fig. 5. Northern blot analysis. Glyoxylated poly(A)$^+$ RNA was run on a 2.0% agarose gel and blotted onto nitrocellulose filters, which were hybridized with phage M13 DNA containing a varying part of the R1-D1 region as indicated (see also Fig. 1). The DNAs were labelled to 10^7 cpm per μg by synthesizing the complementary DNA starting from the 'hybridization probe' primer (New England Biolabs), which primes beyond the inserts. Size markers used are denatured restriction fragments of pBR322. The arrows indicate the orientation of the insert.

nuclease analysis was performed as indicated. Since the starting position of the inserts is known exactly from sequence analysis, the size of the protected fragments fairly accurately predicts the map coordinates of the RNAs involved. In lane 1, the S_1 nuclease probe virtually contains the entire R1-D1 fragment, resulting in protected fragments with the size corresponding to the value obtained from Northern blot analysis (500 and 1110 nucleotides for RNA2 and RNA 3, respectively). This indicates that no major RNA rearrangements such as splicing have occurred. The slightly lower values of the S_1 nuclease experiment may be taken as an indication of the length of the poly(A) tail. In lane 2, the fragment protected by RNA 2 is 150 nucleotides shorter than that in lane 1. The observation that the clone used starts at nucleotide 309 from the EcoRI site puts the 5' start of the RNA around nucleotide 160 and the 3'-end around nucleotide 660, confirming the assumption that RNA 2 is localized entirely within the R1-D1 fragment. Indeed, preliminary RNA sequencing experiments using synthetic oligonucleotide primers (kindly provided by Prof. J.H.Van Boom, Leiden), map the 5'-end of RNA 2 at nucleotide 146 (not shown).

The same approach was used for RNA 3 (lanes 3-5), which was localized from about position 670 to 1780. The bands derived from RNA 4 appear in all lanes, confirming both the size and predicted position. We have mappped this

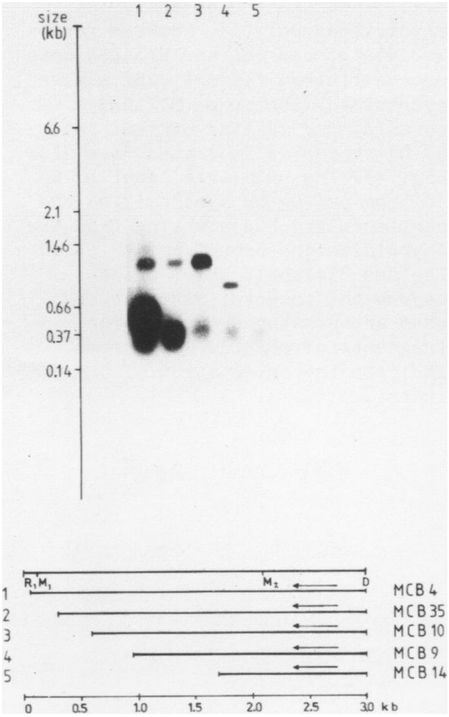

Fig. 6. S_1 nuclease protection analysis. Poly(A)$^+$ RNA was hybridized to M13 DNA containing a varying portion of the R1-D1 region as indicated. The protected fragments were run on a 0.7% agarose gel and blotted onto nitrocellulose filters. The size markers used are denatured restriction fragments of pBR322. The filters were hybridized with the Q1-Q2 TaqI fragment. Control experiments without RNA did not give any hybridization. The arrows indicate the orientation of the insert.

RNA from nucleotide 1800-2200, following essentially the same procedure (not shown). The experiments of Figs 5 and 6 show that, in general, the map positions of RNA 2, 3 and 4 coincide with those of URF 2a, the gene for apocytochrome b and URFs 3 and 3a, respectively. The first 49 amino acid residues of URF 2, however, are not contained within the area covered by RNA 2 (see also Discussion).

DISCUSSION

The gene for apocytochrome b

We have sequenced a 2.5-kb segment of the maxi-circle of T. brucei (Figs 1, 2). This analysis has revealed the presence of the gene for apocytochrome b and a number of URFs, some of which are quite long. The identification of the apocytochrome b gene is relatively straightforward and follows from comparison of the T. brucei sequences with known sequences from other organisms (Fig. 3). The T. brucei protein is slightly shorter than the other apocytochrome b's (350 vs about 380 amino acids), the degree of direct homology with the others is rather low (compare 50% homology between the beef and the yeast genes), but the functional homology is unmistakable (Figs 3 and 4). The gene is not split as it is in yeast [45] or Aspergillus [34] and the map positions of the messenger RNA (mRNA) (RNA 3, Figs 5 and 6) coincide with those of the reading frame. This indicates that very few non-coding bases occur on this RNA, a situation very similar to that observed in mammalian mitochondria [30,32,33,53,54].

The unusual URFs

It is difficult at this point in time to assign a protein coding function to the unusual URFs. They lack homology with known mitochondrial proteins and URFs, they have a highly unusual amino acid composition (see Table II), which does not differ much between them and none seems to contain an AUG codon in the N-terminal moiety. Furthermore, the evidence presented in Table I is much in favour of a standard translation of AUU and AUA (and other) codons although bizarre codon usage, such as recently postulated [36] to occur in Drosophila mitochondria (ATAA specifying methionine), cannot completely be ruled out. So far we have found no evidence for RNA splicing, which makes it unlikely that an AUG-containing fragment is spliced onto the remainder of the sequence. As a consequence, none of the putative proteins encoded by the unusual URFs start with methionine. (NB: The N-terminal sequence of URF 2 beyond the 5'-end of RNA 2 is disregarded. The complete sequence, not containing any methionine, has been presented in ref. 17.)

Before completely discarding the possibility that we are dealing with protein genes, however, it should be emphasized that stable transcripts map in areas that coincide with the URFs: RNA 1 with URF 1 [17], RNA 2 with URFs 2 and 2a, and RNA 4 with URFs 3 and 3a. These RNAs are presumably polyadenylated by virtue of the fact that they bind to oligo(dT) cellulose. We have obtained the nucleotide sequence of a complementary DNA (cDNA) derived from RNA 2 (unpublished). Indeed, we have found an A-rich sequence at the 3'-end of the RNA, which is not present in the genomic sequences obtained so far. We assume this sequence to be post-transcriptionally added. Surprisingly, however, it is not a classical poly(A) tail, since the A residues are interspersed with blocks of 3 or 4 T residues at a number of places.

However, although the combination of abundance and polyadenylation is indicative of a messenger function, other roles for such an RNA seem possible: the abundant 7S poly(A)$^+$ RNA in HeLa cell mitochondria may be involved in the process of DNA replication [54]. It is clear, therefore, that further experimentation is needed to firmly establish the function (if any) of the URF RNAs in T. brucei mitochondria.

Cross-hybridization of the R1-D1 area with heterologous probes

The nucleotide sequences obtained so far (Fig. 2; unpublished) fail to support the experiments of Johnson *et al*. [18] who observed cross-hybridization of Zea mais and yeast *cox*II probes with the R1-D1 fragment, since no homology with *cox*II genes could be found. It should be pointed out that some of the cross-hybridization experiments reported [19] are difficult to interpret since aspecific signals may occur at a higher stringency of hybridization than the specific ones. As mentioned above, G-rich areas occur on this fragment. Spurious hybridization signals, due to GC clusters, have been observed with yeast mtDNA [55] and may also occur with DNA fragments from this area. It is noteworthy that we have not been able to detect specific cross-hybridization of the R1-D1 area with specific probes for the apocytochrome *b* gene from yeast (not shown). This result is not surprising if one takes into account the low degree of homology at the nucleotide level (48%) that exists between the two apocytochrome *b* genes.

Gene organization in T. brucei mitochondria

The coding capacity of the *T. brucei* mitochondrial genome is similar to that of mammalian mtDNA if one takes into account that a 5-kb region, which varies in size between different strains [9], is not transcribed [10,14,17] and consists of repetitive DNA [56]. One would expect, therefore, a rather compact gene organization on the remainder of the *T. brucei* mtDNA, provided that it contains the usual set of mitochondrial genes. To date the contiguous sequence of approximately 6.6 kb of maxi-circle DNA is available (the D3-D1 fragment [16,17; unpublished; this paper]). So far, however, only the genes for the rRNAs and for apocytochrome *b* have been identified. The remainder of this segment (circa 3.7 kb) predominantly consists of the above-mentioned AG-rich DNA on which no known mitochondrial genes or URFs have been found. In particular the absence of tRNA genes is conspicuous in view of the fact that they occur as processing points between the other genes on mammalian DNA [30,32,33]. Hybridization studies have also failed to reveal the existence of mitochondrial tRNA genes [10,14,15,17].

It is unlikely that all of the missing genes are packed on that part of the maxi-circle remaining to be sequenced. Other solutions invoke a nuclear location for some of them (compare ATPase subunit 9, nuclear in mammals [30,32,33], mitochondrial in yeast [31]) or they may reside on a small subset of mini-circles. Alternatively, their sequence and structure may be very unusual, in which case they would be hard to detect in the type of analysis used routinely (compare the rRNA genes [17] and the unusual URFs [this paper]). In either case, a further study of trypanosome mtDNA seems rewarding.

ACKNOWLEDGEMENTS

We thank Prof. P.Borst for his continuing interest and valuable suggestions and Prof. L.A.Grivell and Dr. H.F.Tabak for critical reading of the manuscript. We are indebted to Prof. E.C.Slater for his comments on the apocytochrome b sequence, to Mrs G.J.M.Scholts and Mrs M.A.ten Bosch-Tilanus for expert typing of the manuscript and to Mr. E.Tigchelaar and Mr. C.E.Gravemeijer for skillful technical assistance. This work was supported in part by a grant to P.Borst from the Foundation for Fundamental Biological Research (BION), which is subsidized by The Netherlands Organization for the Advancement of Pure Research (ZWO).

Abbreviations: cDNA, complementary DNA; kb, kilo base pair(s); kDNA, kinetoplast DNA; mRNA, messenger RNA; rRNA, ribosomal RNA; URF, unassigned reading frame.

REFERENCES

1 Simpson, L. (1972) Intern.Rev.Cytol. 32, 139-207.

2 Borst, P. and Hoeijmakers, J.H.J. (1979) Plasmid 2, 20-40.

3 Englund, P.T. (1981) in: Biochemistry and Physiology of Protozoa (Levandowsky, M. and Hutner, S.H., Eds), 2nd edn, Vol. 4, Academic Press, New York, pp. 334-383.

4 Steinert, M. and Van Assel, S. (1980) Plasmid 3, 7-17.

5 Chen, K.K. and Donelson, J.E. (1980) Proc.Natl.Acad.Sci.U.S. 77, 2445-2449.

6 Challberg, S.S. and Englund, P.T. (1980) J.Mol.Biol. 138, 447-472.

7 Kleisen, C.M., Borst, P. and Weijers, P.J. (1976) Europ.J.Biochem. 64, 141-151.

8 Riou, G.F. and Yot, P. (1977) Biochemistry 16, 2390-2396.

9 Borst, P., Fase-Fowler, F., Hoeijmakers, J.H.J. and Frasch, A.C.C. (1980) Biochim.Biophys.Acta 610, 197-210.

10 Hoeijmakers, J.H.J., Snijders, A., Janssen, J.W.G. and Borst, P. (1981) Plasmid 5, 329-350.

11 Hoeijmakers, J.H.J. and Borst, P. (1978) Biochim.Biophys.Acta 521, 407-411.

12 Simpson, L. and Simpson, A.M. (1978) Cell 14, 169-178.

13 Simpson, A.M. and Simpson, L. (1980) Mol.Biochem.Parasitol. 2, 93-108.

14 Stuart, K.D. and Gelvin, S.B. (1982) Mol.Cell.Biol. 2, 845-852.

15 Simpson, A.M., Simpson, L. and Livingston, L. (1982) Mol.Biochem. Parasitol. 6, 237-252.

16 Eperon, I.C., Janssen, J.W.G., Hoeijmakers, J.H.J. and Borst, P. (1983) Nucl.Acids Res. 11, 105-125.

17 Benne, R., Agostinelli, M., De Vries, B.F., Van den Burg, J., Klaver, B. and Borst, P. (1983) in: Nuclear Cytoplasmic Interactions (Wolf, K., Schweyen, R.J. and Kaudewitz, F., Eds), De Gruyter, Berlin, in press.

18 Johnson, B.J.B., Hill, G.C., Fox, T.D. and Stuart, K. (1982) Mol. Biochem.Parasitol. 5, 381-390.

19 Simpson, L., Spithill, T.W. and Simpson, A.M. (1982) Mol.Biochem. Parasitol. 6, 253-264.

20 Kleisen, C.M. and Borst, P. (1975) Biochim.Biophys.Acta 390, 78-81.

21 Auffray, C. and Rougeon, F. (1980) Europ.J.Biochem. 107, 303-314.

22 Bernards, A., Van der Ploeg, L.H.T., Frasch, A.C.C., Borst, P., Boothroyd, J.C., Coleman, S. and Cross, G.A.M. (1981) Cell 27, 497-505.

23 Birnboim, H.C. and Doly, J. (1979) Nucl.Acids Res. 7, 1513-1523.

24 Berk, A.J. and Sharp, P.A. (1977) Cell 12, 721-732.

25 Brunel, F., Davison, J., Merchez, M., Borst, P. and Weijers, P.J. (1980) in: DNA - Recombination Interactions and Repair (Zadrazil, S. and Sponar, J., Eds), Pergamon, Oxford, pp. 45-54.

26 Poncz, M., Solowiejczyk, D., Ballantine, M., Schwartz, E. and Surrey, S. (1982) Proc.Natl.Acad.Sci.U.S. 79, 4298-4302.

27 Messing, J. (1981) in: Third Cleveland Symposium on Macromolecules: Recombinant DNA (Walton, A., Ed.), Elsevier, Amsterdam, pp. 143-153.

28 Benton, W.D. and Davis, R.W. (1977) Science 196, 180-182.

29 Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc.Natl.Acad. Sci.U.S. 74, 5463-5467.

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 Macino, C., Coruzzi, G., Nobrega, F.G., Li, M. and Tzagoloff, A. (1979) Proc.Natl.Acad.Sci.U.S. 76, 3784-3785.

32 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.

33 Bibb, M.J., Van Etten, R.A., Wright, C.T., Walberg, M.W. and Clayton, D.A. (1981) Cell 26, 167-180.

34 Waring, R.B., Davies, R.W., Lee, S., Grisi, E., McPhail Berks, M. and Scazzocchio, C. (1981) Cell 27, 4-11.

35 Davies, R.W., Scazzocchio, C., Waring, R.B., Lee, S., Grisi, E., McPhail Berks, M. and Brown, T.A. (1981) in: Mitochondrial Genes (Slonimski, P., Borst, P. and Attardi, G., Eds), Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pp. 405-410.

36 De Bruyn, M.H.L. (1983) Nature 304, 234-241.

37 Clary, D.O. and Wolstenholme, D.R. (1983) Nucl.Acids Res. 11, 4211-4227.

38 Staden, R. (1982) Nucl.Acids Res. 10, 4731-4751.

39 Bonitz, S.G., Coruzzi, G., Thalenfeld, B.E., Tzagoloff, A. and Macino, G. (1980) J.Biol.Chem. 255, 11927-11941.

40 Coruzzi, G. and Tzagoloff, A. (1979) J.Biol.Chem. 254, 9324-9330.

41 Coruzzi, G., Bonitz, S.G., Thalenfeld, B.E. and Tzagoloff, A. (1981) J.Biol.Chem. 256, 12780-12787.

42 Fox, T.D. (1979) Proc.Natl.Acad.Sci.U.S. 76, 6534-6538.

43 Macino, G. and Tzagoloff, A. (1980) Cell 20, 507-517.

44 Nobrega, F.G. and Tzagoloff, A. (1980) J.Biol.Chem. 255, 9828-9837.

45 Thalenfeld, B.E. and Tzagoloff, A. (1980) J.Biol.Chem. 255, 6173-6180.

46 Tzagoloff, A., Nobrega, M., Akai, A. and Macino, G. (1980) Current Genet. 2, 149-157.

47 Saraste, M. and Wikström, M. (1983) in: Structure and Function of Membrane Proteins (Palmieri, F. and Quagliariello, E., Eds), Elsevier, Amsterdam, in press.

48 Von Heijne, G. (1981) Europ.J.Biochem. 120, 275-278.

49 Barrell, B.G., Bankier, A.T. and Drouin, J. (1979) Nature 282, 189-194.

50 Hudspeth, M.E.S., Ainley, W.M., Shumard, D.S., Butow, R.A. and Grossman, L.I. (1982) Cell 30, 617-626.

51 Bonitz, S.G., Berlani, R., Coruzzi, G., Li, M., Macino, G., Nobrega, F.G., Nobrega, M.P., Thalenfeld, B.E. and Tzagoloff, A. (1980) Proc.Natl.Acad.Sci.U.S. 77, 3167-3170.

52 Shen, S.-H., Slightom, J.L. and Smithies, O. (1981) Cell 26, 191-203.

53 Montoya, J., Ojala, D. and Attardi, G. (1981) Nature 290, 465-470.

54 Ojala, D., Montoya, J. and Attardi, G. (1981) Nature 290, 470-474.

55 Tabak, H.F., Van der Laan, J.C., Landegent, J.E., Evers, R.F. and Wassenaar, G.M. (1982) Plasmid 8, 261-275.

56 Borst, P., Weijers, P.J. and Brakenhoff, G.J. (1982) Biochim. Biophys.Acta 699, 272-280.