The major transcripts of the kinetoplast DNA of *Trypanosoma brucei* are very small ribosomal RNAs

I.C Eperon+, J.W.G. Janssen*, J.H.J. Hoeijmakers* and P. Borst*

Laboratory of Molecular Biology, The MRC Centre, Hills Road, Cambridge CB2 2QH, UK, and *Section of Medical Enzymology and Molecular Biology, Laboratory of Biochemistry, University of Amsterdam, Jan Swammerdam Institute, P.O. Box 60.000, 1005 GA Amsterdam, The Netherlands

Received 15 September 1982; Revised and Accepted 1 December 1982

SUMMARY

The nucleotide sequence has been determined of a 2.2 kb segment of kinetoplast DNA, which encodes the major mitochondrial transcripts (12S and 9S) of Trypanosoma brucei. The sequence shows that the 12S RNA is a large subunit rRNA, although sufficiently unusual for resistance to chloramphenicol to be predicted. The 9S RNA has little homology with other rRNAs, but a possible secondary structure is not unlike that of the 2.5-fold larger E. coli 16S rRNA. We conclude that the 12S RNA (about 1230 nucleotides) and the 9S RNA (about 640 nucleotides) are the smallest homologues of the E. coli 23S and 16S rRNAs yet observed.

INTRODUCTION

An extraordinary structure is formed by the DNA in the single mitochondrion of the unicellular flagellates, the Kinetoplastida. In the case of the bloodstream parasite, Trypanosoma brucei, 10^4 minicircles of 0.3 μm and 50-100 maxicircles of 6 μm are interwoven into a network with a molecular weight of 4×10^9 (reviewed in refs. 1, 2 and 3). The minicircles are extremely diverse, with 100 or more different sequence classes within the network (4); the rate of evolution is very high (1,5). In contrast, maxicircles are homogeneous within each strain (6) and the sequence is quite well conserved among different species (7).

Cloned fragments of part of the maxicircle have been used to show that at least two major and six minor transcripts derive from the maxicircle in T. brucei, whereas the minicircles are not transcribed (8). The two major species closely resemble the 9S and 12S RNAs found in Leishmania mitochondria (9,10). A number of properties suggest that these major RNAs might be ribosomal RNAs. They are the only abundant high molecular weight RNAs found in trypanosome mitochondria, are equimolar and they are not retained on oligo (dT) cellulose (8,11). These are characteristics of mitochondrial rRNAs in other organisms. Further support comes from the high conservation of the sequences and sizes of these transcripts among kinetoplastid flagellates (12)

0305-1048/83/1101-0105$2.00/0

and from the observation that they derive from single adjacent genes (8).

However, certain features cast doubt on this suggestion. The 12S and 9S RNAs are very short compared to other rRNAs [1080 and 590 nucleotides in length, respectively (8), compared to 2904 and 1542 for Escherichia coli rRNAs], and the genes encoding these RNAs are arranged in the unprecedented order (5' to 3') of larger RNA to smaller RNA. Further, all attempts to isolate a ribosomal fraction containing these RNAs have been unsuccessful. It has even proved to be difficult to demonstrate that protein synthesis takes place at all in these mitochondria as it does in others.

It was therefore expected that the sequence of the genes encoding the 12S and 9S RNAs would resolve their function, and perhaps also illustrate the minimum requirements for functional ribosomal RNAs.

MATERIALS AND METHODS

The isolation of trypanosome RNA, restriction enzyme digestion of DNA, gel electrophoresis of DNA and nick-translation or end-labelling of DNA were done as described in reference 8. An EcoRI fragment of the maxicircle of T. brucei (RR3) cloned in bacteriophage lambda (13) was used as the source of maxicircle segments for sequence analysis and hybridisation. Restriction fragments were size-fractionated in agarose gels (8) and individual fragments were recovered by adsorption to Sepharose-lysine (14) or activated glass (15).

S1 nuclease protection experiments were done as described by Berk and Sharp (16). In brief, 1-5 ng of DNA and a 10-fold excess of RNA were co-precipitated with 25 µg tRNA from E. coli. The pellet was washed with 70% ethanol, vacuum-dried, dissolved in 20 µl hybridisation mix [40 mM Pipes, pH 6.4, 400 mM NaCl, 1 mM sodium EDTA in 80% (v/v) recrystallised formamide] and incubated for 4-6 h at 37°C. The DNA was then denatured for 15 min at 85°C and incubated at 37°C for 16 h. Then 300 µl S1 buffer (280 mM NaCl, 20 mM sodium acetate, pH 4.4, 4.5 mM zinc acetate, 20 µg denatured salmon sperm DNA and 15 units S1 nuclease per ml (Sigma Chemical Co.) were added. After an incubation of 30 min at 37°C the reaction was stopped with 75 µl stop-mix (2.5 M ammonium acetate, 50 mM EDTA and 20 µg E. coli tRNA per ml) and the nucleic acids were precipitated with an equal volume of isopropanol. After a 70% ethanol wash, the nucleic aids were dissolved in alkaline buffer (50 mM NaOH, 1 mM EDTA, 2% w/v Ficoll and 0.015% bromocresol green) and run for 4 h at 250 mA (75mV) in a 2% agarose gel. The DNA was blotted onto nitrocellulose by the Southern procedure (17) and hybridised with EcoRI

maxicircle fragment RR3 labelled by nick-translation (see ref. 8).

The sequence was established using the EcoRI fragment RR3 of the maxicircle, which had been recloned into pAT153 (18). A HindIII-BamHI fragment of 2.2 kb within RR3 had been shown to contain the genes for the 12S and 9S RNAs (8; see Results). Five to ten µg of this 2.2 kb fragment were prepared from the recombinant plasmid and used for all of the subsequent sequence analysis.

The sequence was deduced by cloning restriction fragments into bacteriophage M13 for analysis by the chain termination method (19,20). Fragments generated by TaqI cleavage were isolated by gel electrophoresis, and aliquots were further digested by RsaI and AluI (all restriction enzymes were purchased from New England Biolabs, Beverly, Mass., USA). These and the products of MnlI digestion were made flush-ended by treatment with DNA polymerase, where appropriate, and 0.05-0.2 µg were then ligated to unphosphorylated EcoRI linkers (Collaborative Research, Inc., Waltham, Mass., USA, or a gift to F. Sanger from T.P. Wang) using 1-5 units of T4 DNA ligase (Bethesda Reseach Laboratories, Inc., Rockville, Maryland, USA). The enzyme was inactivated by incubation at 70°C for 10 min. EcoRI (New England Biolabs, Beverly, Mass., USA) (5 units) and 0.1 volume of 1 M Tris-HCl (pH 7.9) were added and the mixture was incubated at 37°C for 3 h. dGTP was added to 0.1 mM and 0.5 units of DNA polymerase I (nach Klenow; Boehringer Mannheim, F.R.G.) were added. This reaction was intended to regenerate the termini normally produced after cleavage with EcoRI, for with unphosphorylated linkers the 3'-terminal GG dinucleotide produced by cleavage probably dissociates. After 15 min at room temperature, the reaction mixture was extracted with phenol and the DNA in the aqueous layer was precipitated with ethanol. The products were ligated to an EcoRI vector. The use of unphosphorylated linkers obviated the requirement for separation of the fragments with termini produced by the action of EcoRI from the products of digestion of polymerised, phosphorylated linkers. EcoRI (a gift from J. Karn) was also used to cut 0.1 µg of DNA directly with reduced stringency in 9 mM Tris-HCl (pH 8.5), 2.5 mM $MgCl_2$, 0.5 mM DTT (21). After extraction with phenol and precipitation with ethanol, 0.01-0.05 µg of DNA were ligated directly to the vector. The vector used in these experiments was M13mp2 (22). M13mp7 (23) was used in further experiments, having been cut with HincII to give flush ends and treated with calf alkaline phosphatase (the vector was a gift from G. Winter). HincII, AluI and HaeIII fragments were ligated directly to this vector. In all these experiments an M13-directed

primer was used, prepared from pSP16 (24) or synthesized chemically (25), a gift from M.J. Gait.

Some sequences were inaccessible using this approach, due to a lack of restriction enzyme cleavage sites. Whilst MnlI sites were not far removed from some of these sequences, the MnlI fragments containing these sequences were large compared to some of the other MnlI fragments and they were not cloned. Sequences from this region were therefore obtained in a different manner: 2 µg of the BamHI-HindIII fragment was digested with MnlI and the termini were repaired with [α-^{32}P]deoxynucleoside triphosphates (Amersham International, England) and E. coli DNA polymerase (Klenow fragment). After fractionation on a non-denaturing polyacrylamide gel, the predominant fragments were isolated and used as primers on the recombinant bacteriophage DNAs containing TaqI fragments which had been obtained as described earlier.

The sequences were compiled using the computer program DBUTIL (26) and analysed using a number of approaches (27,28; R. Staden, unpublished results).

RESULTS

Hoeijmakers et al. (8) have previously located the genes for the 12S and 9S mitochondrial RNAs on the 2.2-kb HindIII-BamHI fragment of the maxicircle and they demonstrated that both genes are transcribed from the same DNA strand in the order 12S-9S. To locate the genes more precisely, segments of the HindIII-BamHI fragment were hybridised with total RNA from culture-form trypanosomes, the hybrids were treated with S1 nuclease (16), the digest was size-fractionated on an alkaline gel and the protected DNA segments were detected by blotting onto nitrocellulose and hybridisation with a suitable DNA probe. Figure 1A shows that the two DNA fragments protected against S1 nuclease by hybridisation with RNA were estimated to be of 1230 and 640 nucleotides (range 1160-1230 and 630-650 in eight experiments). These values are higher than the 1080 and 590 determined for the corresponding 12S and 9S RNAs by gel electrophoresis after glyoxylation (8), but in view of the difficulties encountered in completely denaturing these A+U rich RNAs with glyoxal we consider the values for the DNA copies more accurate. These results show that the genes for the 9S and 12S RNAs lack any intervening sequences that could be detected in S1 nuclease experiments.

Figure 1B shows an S1 nuclease protection experiment with the 1170-bp TaqI fragment. The protected segment is marginally smaller than the 1000-bp HindIII-TaqI fragment, which places the 5' end of the 12S RNA approximately

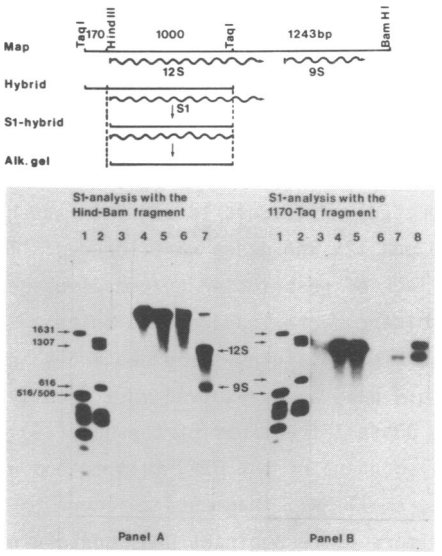

Fig. 1. Maxicircle segments protected against degradation by S1 nuclease by hybridisation with mitochondrial RNA. The top part of the figure shows a restriction map of the maxicircle segment containing the 9S and 12S RNA genes and schematically indicates the DNA fragments used for hybridisation together with (dotted lines) the sizes of the fragments produced after S1 digestion in the reaction of Panel B. 5 ng of purified TaqI fragment or HindIII/BamHI fragment were hybridised at 37°C with 5 µg of total culture-form RNA of T. brucei 427-60. (Corresponds with about 50 ng 9S and 12S RNA.) The hybrids were treated with 50 U S1 nuclease per ml and electrophoresed through a 2% alkaline agarose gel as described in Materials and Methods. The size markers used are restriction fragments of plasmid pBR322, end-labelled with ^{32}p. The control and protected DNA fragments were transferred to nitro-cellulose filters by blotting and hybridised with in vitro labelled RR3 clone DNA (18×10^6 dpm), as described in Materials and Methods.
PANEL A: S1 analysis using the 2.2 kb HindIII/BamHI fragment. (1) pBR322 x HinF marker; (2) pBR322 x TaqI marker; (3) RNA control; (4) Hind/Bam DNA input (undenatured, shorter exposure); (5) Hind/Bam DNA control (no RNA or nuclease digestion; shorter exposure); (6) Hind/Bam DNA control (digested with S1 nuclease); (7) Hind/Bam DNA + total culture-form RNA. The samples in lanes 3, 5, 6 and 7 were taken through the denaturation-hybridisation-S1 nuclease incubation procedure, but in lane 5 S1 nuclease was not added.
PANEL B: S1 analysis using the 1170 bp TaqI fragment. (1) pBR322 x HinF marker; (2) pBR322 x TaqI marker; (3) RNA control; (4) 1170 TaqI DNA input; (5) 1170 TaqI DNA control (not digested with S1 nuclease); (6) 1170 TaqI DNA control; (7) 1170 TaqI DNA + total culture-form RNA; (8) Hind/Bam x TaqI marker. The samples were handled as in Panel A.

20 bp to the right of the HindIII site. Further attempts to obtain more precise locations for the 12S and 9S RNAs by using smaller DNA fragments in the hybridisation were unsuccessful, although in parallel control experiments

with other hybrids, protected fragments down to 140 bp were easily detected. We have also not succeeded in end-labelling the purified RNAs for RNA sequencing and thus at present the exact termini of these RNAs remains uncertain. This uncertainty is large for the 640-nucleotide 9S RNA which must be located between the end of the 12S RNA gene (about position 1250) and the BamHI site at position 2241.

It proved to be unexpectedly difficult to sequence the HindIII-BamHI fragment on which the 9S and 12S RNA genes were located. The principal difficulty arose from a lack of restriction enzyme cleavage sites which could be used to produce recombinant bacteriophage DNAs suitable for verifying some of the longer sequences or for obtaining these sequences in the opposite sense. The situation would have been resolved more readily had it been possible to clone random DNase I fragments of the DNA (29) but, perhaps because of the high (A+T) content of the DNA, this method was unsuccessful. In some cases, therefore, small MnlI fragments adjacent to the uncertain sequence were used as primers on recombinant DNAs containing large inserts. Figure 2 shows the extent and direction of the sequences used in deducing the result shown in Figure 3. There are two regions (nucleotides 1-100 and 1163-1437) where the sequence was deduced two or three times from different cloned fragments, but always in one direction. These sequences were very clear and these A+T rich regions gave none of the indications of artefacts that are usually observed when difficulties arise. In contrast, a number of problems arose with the very G-rich 3'-terminal portion, the sequence of which was deduced in both orientations. Where sequences were deduced in the

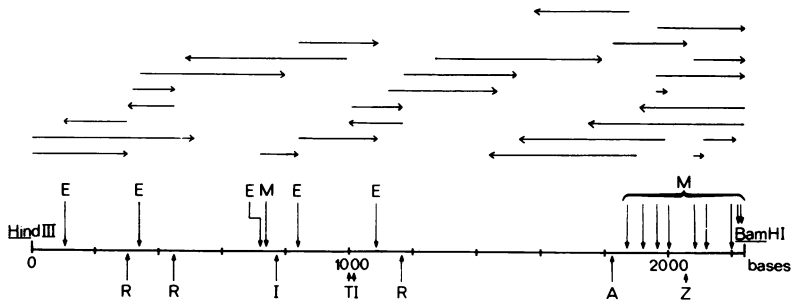

Fig. 2. The extent and direction of the sequences used in deducing the result shown in Figure 3. Overlapping gel readings from the sequence of each cloned fragment or from one priming reaction are shown by a single arow. The restriction endonuclease cleavage sites are indicated: A = AluI, E = EcoRI* (cleavage at GAATT), I = HincII, M = MnlI, R = RsaI, T = TaqI, Z = HaeIII.

opposite sense to Figure 3, a number of positions gave rise to artefact bands across all four tracks of a gel. These positions usually preceded a sequence of $C_{4-5}T_{1-2}$ (in the sense of the sequence deduced). The use of reverse transcriptase gave rise to similar problems. Sequences in the sense of Figure 3 were completely clear. There are two positions (1544 and 1739) at which the nucleotide is not clear from any single sequence, but was assigned with confidence on the basis of the multiple sequences crossing each point.

The approximate positions of the genes are marked in Figure 3. The genes can be identified as large and small subunit rRNA genes by similarities of some sequences to those of other rRNAs (see later), but the termini are difficult to assign.

The large subunit rRNA of animal mitochondria appears to be characterised by a 3'-terminal sequence which includes a small potential A-rich hairpin loop and a sequence similar to Pu-G-G-U-U-U_{OH} (30,31,32,33,34). The position marked in Figure 3 represents the sequence in the correct region which matches this most closely. The 5'-termini of mammalian mitochondrial large subunit rRNAs also share some sequences (34) but these are absent in the region around nucleotide 20 of Figure 3. However, nucleotides 1-8 are complementary to 1268-1275 (allowing for 2 G-T base-pairs), just before the proposed 3'-terminal nucleotide. This would form a secondary structure akin to that proposed for other large subunit rRNAs (35,36). The nuclease protection experiments suggest that the 5' terminus is probably near nucleotide 20, although this is only an approximate location with a high level of uncertainty. It is formally possible that there is a small intervening sequence between nucleotides 8 and 20, the 5' terminal section of the rRNA being too small to form stable hybrids with the gene.

Neither the 5' terminus nor, more surprisingly, the 3' terminus of the small subunit rRNA could be identified. The positions marked represent tentative suggestions based on a limited degree of homology with other rRNAs within the sequence and on a secondary structure analogous to that of prokaryotes and mitochondria (see later).

DISCUSSION

A. The 12S RNA

Figure 4 shows that part of the 12S RNA gene sequence can be aligned, after the introduction of gaps, with the sequence of the *E. coli* 23S rRNA (37) and that a high proportion of nucleotides are identical. Of the sequence between nucleotides 1008 and 1121 in Figure 3, and that between


12S RNA
12S RNA
9S RNA
9S RNA

nucleotides 2445 and 2608 of the E. coli rRNA, 50% and 34% respectively of all the nucleotides in these regions match. This region of the E. coli rRNA is even more highly conserved in rRNAs from mammalian mitochondria (31,32,38), yeast mitochondria (39), chloroplasts (40,41) and eukaryotic cytoplasm (42,43).

Correspondingly, several observations point to an important role for this region in the A site of the ribosome. A puromycin analogue, used as an affinity label for RNA sequences near the 3' terminus of amino-acyl-tRNA, was found to react with one nucleotide in the 23S rRNA of intact ribosomes (44). The sequence around the site of attachment was deduced (45), and one of the several positions where this sequence may be located is in the conserved region (37). The region may be at the subunit interface, for it is protected against kethoxal modification by the association of E. coli ribosomal subunits (46). One protected methylated oligonucleotide was positioned here by comparison with the location of a methylated oligonucleotide in hamster mitochondrial rRNA (38). The presence of much of this highly conserved region in the 12 S RNA strongly indicates that this RNA is a functional rRNA.

An interesting point is the partial absence in 12S RNA of a region that is thought to be involved in the sensitivity of ribosomes to chloramphenicol (CAP) (Fig. 4 and 5). This region consists of two decamers separated by a stem and loop. Fig. 5 shows that the sequence of both decamers is completely conserved in bacterial, chloroplast and mitochondrial ribosomes. All mutations that give rise to chloramphenicol resistance of mitochondrial ribosomes result in point mutations in either of these decamers (39,47,48,49). Moreover, the chloramphenicol-resistant cell-sap ribosomes from Physarum (42) and yeast (43) contain several changes in the decamers, even though the over-all sequence and apparent secondary structure of this region is conserved. Figs. 4 and 5 show that one of the decamers is altered in the 12S RNA whereas the other is absent altogether. We predict therefore that the ribosome in which this 12S RNA is presumably contained will lack the sensitivity to chloramphenicol which is characteristic of mitochondrial ribosomes in general.

Secondary stuctures, based on experimental work on E. coli 23S rRNA and

Fig. 3. The sequence of the HindIII/BamHI fragment of the T. brucei maxicircle used in the hybridisation experiments. The possible termini of the 12S and 9S RNA genes are marked, based on the results shown in Figure 1 and the similarities of sequences to those of E. coli rRNAs (37,86) and mammalian mitochondrial rRNAs (31,32,33,34) discussed in the text and marked in Figures 4 and 5.

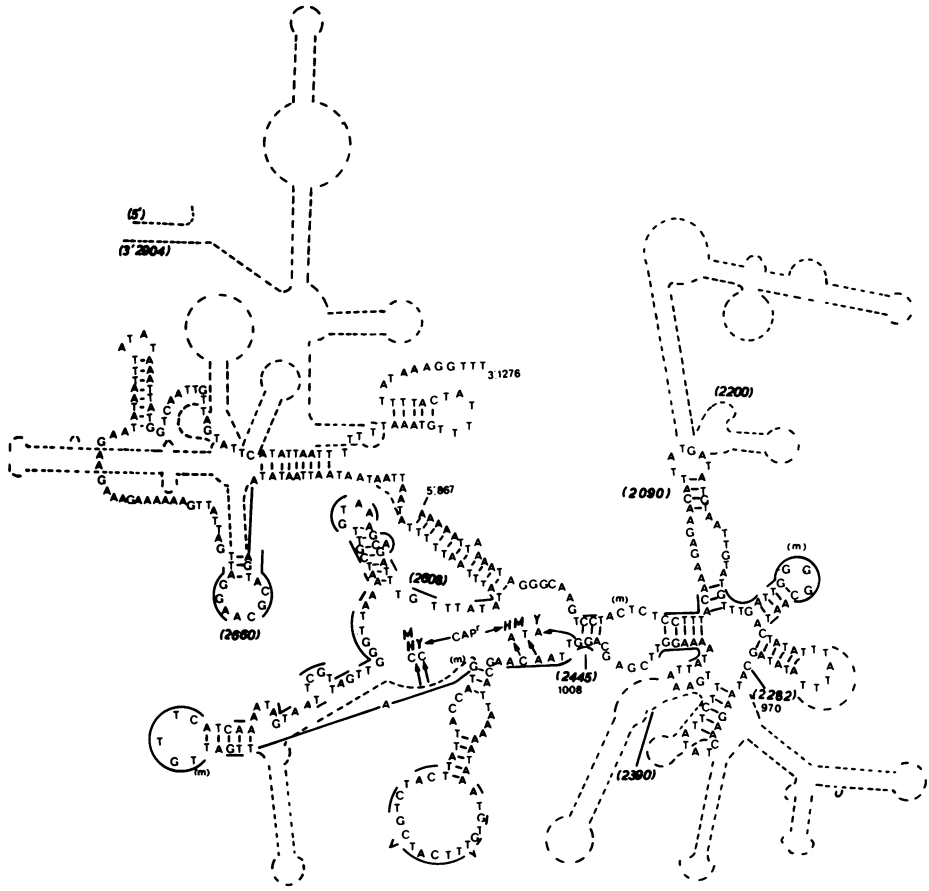

Fig. 4. The sequence of nucleotides 867-1276, arranged in a structure analogous to that for nucleotides 2042-2904 of E. coli 23S rRNA proposed by Branlant et al. (36). The E. coli structure to the same scale is marked by dashed lines and bold numbers. The methylated nucleotides in the E. coli sequence (36,38) are indicated (m), as are the nucleotides implicated in chloramphenicol resistance of yeast (Y; 39), mouse (M) and human (H; 47,48, 49) mitochondria. Lines show those nucleotides in the trypanosome sequence that are identical to those from E. coli.

comparative sequences, have been deduced for large subunit rRNAs of E. coli and mammalian mitochondria (35,36). The structure of the highly conserved region is preserved in yeast and mammalian mitochondria and, probably, chloroplasts (35,36,38,41). It is possible to form an analogous structure with the sequence in Figure 3 between nucleotides 867 and 1138. Figure 4

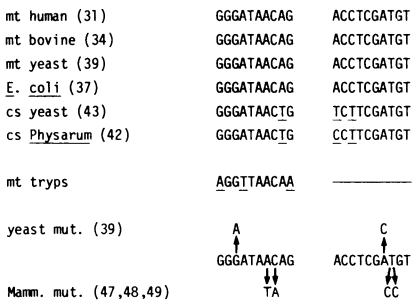

Fig. 5. The two blocks of sequences in large subunit rRNAs that are absolutely conserved in all chloramphenicol-sensitive ribosomes. Sequences that differ are underlined at the point of difference, and mutations that give rise to resistance to CAP are shown by arrows. The sequences are those of human (31), bovine (34) and yeast (39) mitochondrial rRNAs, E. coli 23S rRNA (37), yeast (43) and Physarum (42) cytoplasmic (cs) rRNAs, the trypanosome 12S RNA (this paper), and mutants in yeast (39) and mammalian (47,48,49) mitochondria.

shows the proposed structure from nucleotide 867 to the putative 3' terminus of the gene, and a diagram representing the E. coli 23S rRNA structure according to Branlant et al. (36).

The secondary structure between nucleotides 867 and 1138 corresponds quite well with that proposed for domain VI of the E. coli structure (36) and the underlining indicates that there are also a number of conserved sequences. However, it is rather curious that one strand of a helix proposed by Baer and Dubin (38) has been lost whilst the other remains by the criteria of sequence homology and the number of nucleotides separating two other helices.

Figure 4 shows that the region preceding the highly conserved structures is radically altered. For example, the binding site for protein L1 in E. coli 23S rRNA, around nucleotides 2090-2200 (50,51), has been reduced considerably, even beyond the already shortened structure proposed for the mammalian mitochondrial rRNAs (35,36). E. coli protein L1 may play a role in binding aminoacyl-tRNAs (52,53). The most dramatic change represents a complete loss of structures that are already reduced in mammalian mitochondria relative to E. coli rRNA. The region between nucleotides 2282 and 2390 in E. coli 23S rRNA appears to form a complex with large subunit proteins L5, L8 and L25 and 5S RNA (54). This region is reduced to 31 nucleotides in rRNAs from mammalian mitochondria (35,36) which seem to lack any equivalent to a 5S RNA. In the trypanosome 12S RNA, these sequences may have been completely lost (Fig. 4).

Beyond the long range interaction of nucleotides 1127-1138 with 867-879, the 3'-terminal region of the 12S rRNA does not seem to be able to form structures analogous to the 3' termini of non-mitochondrial rRNAs (35,36). However, as in mammalian mitochondrial rRNAs, there is an almost identical sequence corresponding to the loop of the 2650-2670 hairpin in the E. coli 23S rRNA (Figure 4). This hairpin, the site of α-Sarcin cleavage, may also be part of the A site of the ribosome (43). The 3'-terminal region is perhaps slightly longer in the trypanosome mitochondrial 12S rRNA compared with mammalian mitochondrial 16S rRNAs. This is the only segment of the trypanosome rRNA which may be as long as or longer than the corresponding segment of mammalian mitochondrial rRNAs, which is rather surprising where there is (except for the above hairpin) so little obvious conservation of sequence or structure. However, our present assignment of termini predicts that the 12S rRNA would be 1256 nucleotides in length, which agrees quite well with the value of 1240 nucleotides deduced from the hybridization experiments.

There is very little similarity to other rRNAs in the 5'-terminal 800 nucleotides of this gene, which correspond to approximately 2000 nucleotides of E. coli 23S rRNA and 1050 of mammalian mitochondrial rRNAs. Even the otherwise well-conserved regions of large subunit rRNAs cannot be certainly identified, which impedes attempts to extend convincing possible secondary structures and thereby analyse the minimum requirements for a ribosomal RNA.

B. The 9S RNA

Figure 6a shows those sequences in the region of the 9S RNA gene that are identical to sequences in the 16S rRNA of E. coli. Some of these sequences are also found in mammalian mitochondrial rRNAs, and an extra region of identity is shared with the bovine mitochondrial 12S rRNA. If these matches are significant, the 9S RNA would be even less similar to other rRNAs than are the mammalian mitochondrial rRNAs to prokaryotic or cytoplasmic rRNAs (31,55; H.F. Noller and C.R. Woese, personal communication).

The three largest blocks of identity comprise 8, 9 and 13 nucleotides. The first block is within a region in which 19 of the 26 nucleotides (from 1433 to 1458) match the sequence in E. coli 16S rRNA of nucleotides 509 to 534 (compare Figures 6a and 6b). This sequence and structure is highly conserved among bacteria (56,57) and can be identified in the sequences of mammalian mt rRNA genes (34), the yeast mt 18S rRNA gene (58) and the Xenopus laevis and yeast cytoplasmic rRNA genes (59,60). This region of the E. coli 16S rRNA appears to be involved in binding ribosomal protein S4, based on the

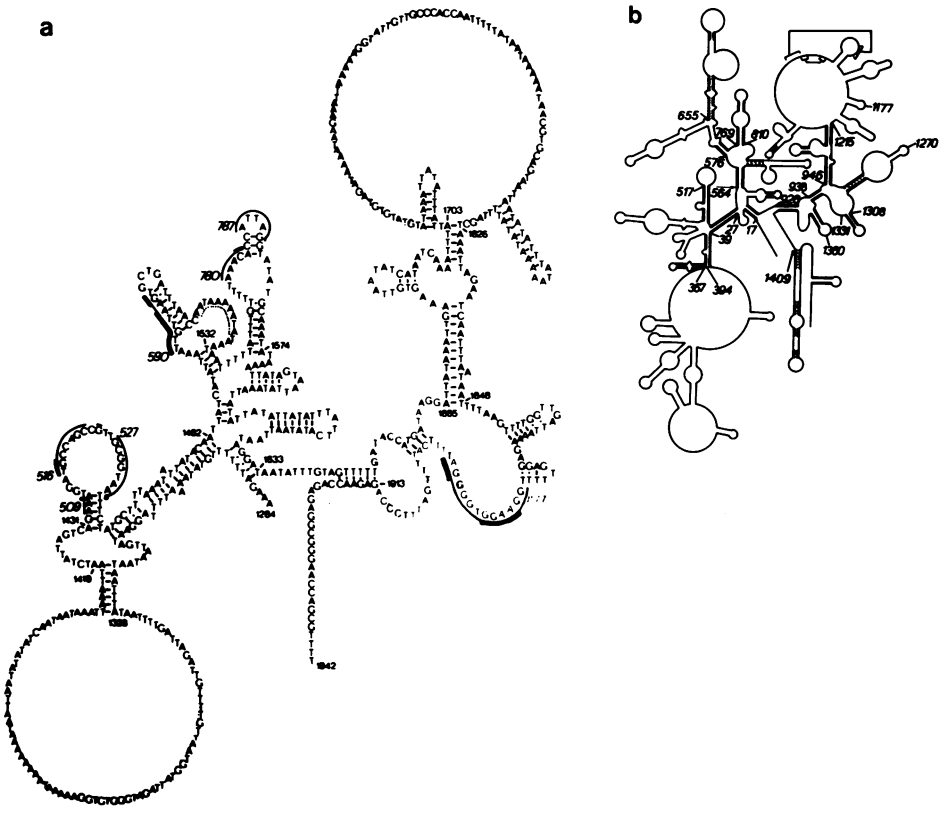

Fig. 6. (a) The sequence of nucleotides 1284-1942, arranged in a structure analogous to that of E. coli 16S rRNA (56,57,63,64) and that proposed for mammalian mitochondrial 12S rRNA (55; H.F. Noller and C.R. Woese, personal communication). The larger sequences identical to those in E. coli 16S rRNA are underlined. Where the sequence in E. coli is single-stranded the match is underlined once; double underlining refers to sequences that are base-paired in the E. coli rRNA. The dotted lines show an additional match with bovine 12S rRNA. (b) A structure of E. coli 16S rRNA according to ref. 56. Heavy lines indicate that homologous structures may be formed by the human and bovine 12S rRNAs, the dots representing less certain assignments (H.F. Noller and C.R. Woese, personal communication).

recovery of oligonucleotides cross-linked to S4 after irradiation with ultra-violet light (61) and on homology between the 16S rRNA and the site of translational inhibition by S4 of the α operon (62). The local structure around 1433 to 1460 does not quite match that of the E. coli rRNA, but the S4 binding site on the α operon mRNA also shows a rather dissimilar secondary structure to that of the rRNA (62). It has been suggested that one role of

S4 is to stabilise the long-range interactions between nucleotides 27-37 and 547-556 and between 564-570 and 880-886 (56). (Figure 6b represents diagrammatically the structure of E. coli 16S rRNA according to these authors.) The homology with the E. coli sequence may therefore predict that the 9S RNA can form analogous long-range interactions.

The second block of identical sequence is between nucleotides 1497-1505 of Figure 3 and 590-598 of the E. coli 16S rRNA. Using provisional secondary structures for mammalian mt rRNA (H.F. Noller and C.R. Woese, personal communication; 55) and eukaryotic cytoplasmic rRNAs to optimise alignments of rRNA sequences, it seems that this region is very poorly conserved. However, there appears to be some similarity between the E. coli and yeast mitochondrial rRNAs at this point, despite a different putative secondary structure (58). It is, therefore, possible that this represents a genuine homology rather than a spurious match in the case of the 9S RNA.

The largest match is found between 1882-1894 of Figure 3 and 1177-1189 of E. coli 16S rRNA. The human and bovine 12S rRNA sequences (31,34) also match this sequence, but their other sequences in this region and an alignment based on secondary structures suggest that the mammalian mitochondrial sequence is homologous with 1331-1339 of E. coli (H.F. Noller, personal communication). Thus, despite the fact that this is the largest continuous block of similarity between the trypanosome and E. coli rRNA, this similarity is probably also fortuitous or at least not a useful basis for an alignment of the two sequences. However, the match of a block of 8 nucleotides with human mitochondrial rRNA may be significant.

There appeared to be, therefore, only one sequence (1433-1460) that was clearly homologous to other rRNAs. In order to analyse the sequence in more detail, this region was used as a basis for attempting to form a secondary structure akin to that of prokaryotic 16S rRNAs (56,57,63,64,65) and that predicted for the small mammalian 12S rRNAs (H.F. Noller and C.R. Woese, personal communication; 55). Although the low content of C in the sequence reduces the stringency of the test when G-T base-pairs are permitted, such an exercise does demand consistency with other rRNA sequences. It also serves an heuristic purpose, with regard to the function of such a small rRNA. The result is shown in Figure 6a. It should be stressed that sequences from the termini of the rRNAs and sequences from 9S RNAs of closely related organisms will be required to test a suggested structure.

The structure was expected to show whether smaller sequences matching the E. coli rRNA were homologous with that rRNA. Only two such sequences were

found, shown in Figure 6a at nucleotides 1550 to 1561. These appear to be homologous with a sequence and structure that is well conserved in small-subunit rRNAs. Although the 1883-1890 block does not appear to be homologous with the E. coli sequence in Figure 6a, it is striking that the matching sequence in human mitochondrial 12S rRNA is in exactly the same place in its postulated secondary structure (34).

The 5' terminus shown was suggested on the grounds that the weak interactions formed by nucleotides 1288-1291 and 1293-1312 are analogous to the long-range interactions formed by nucleotides 17-20 and 27-37 of the E. coli 16S rRNA. This is preceded by the sequence ATTTTTG, which also occurs 660 nucleotides downstream. This length is similar to the estimated size for the RNA of 640 nucleotides. ATTTTG may be an RNA processing site and was therefore used to define one possible 3' terminus of the 9S RNA.

Although the secondary structures and some sequences near the 3' termini of small subunit RNAs are very highly conserved (31,55,57,59,60,66), not even a tentative match is found with these sequences within the sequence in Figure 3. This may imply that this part of the rRNA has been completely lost, as predicted by the secondary structure in Figure 6a. There are G-G-A-A sequences, usually found methylated within a highly conserved stem-loop structure near rRNA 3' termini, at positions 1930 and 2042. In neither case is a reasonable structure formed around the sequence. It is possible that the sequence at 1930 plays a compensating or vestigial role with regard to the sequences normally found in rRNAs. Most of the rRNA sequences that are necessary for subunit association (67), connected with codon-anticodon interaction (56,68), or methylated (38) in various organisms are unrecognizable.

The comparison of Figures 6a and 6b shows that there are analogues of many of the most well-conserved secondary structure interactions of small subunit rRNAs. These include both "universal" (57) and principal (56) interactions which form the very highly-conserved hairpins (such as those around nucleotides 520, 790 and 970 in E. coli) and define the major domains of the structure.

The greatest changes from the E. coli 16S rRNA are found in regions corresponding to nucleotides 50-510, 590-650, 990-1214 and 1400-1542 of the E. coli rRNA (Figure 6b). The pattern of simplification and loss of features in the first three regions extend the observations on mammalian mt 12S rRNAs (H.F. Noller, unpublished; 34,55), but the changes at the 3' terminus are radical.

Two methods were used to test the statistical significance of the matches with E. coli 16S rRNA shown in Figure 6a. One was based on the number of ways of arranging the observed number of matches in the given order (A.R. Crowther, personal communication), whilst the other calculated the probability of an observed match occurring between the preceding and following matches and proceeded sequentially to all observed matches. Comparing two random sequences of the length of the E. coli 16S rRNA and of the region wherein the 9S RNA has been located, it was found to be highly probable that matches of such length and in the same order in both sequences would occur. However, what is significant is that some of the matches involve sequences that are amongst the most highly conserved in all small subunit rRNAs examined previously, and that they can fall in an analogous position in the tentative secondary structure. Thus, although this structure and sequence differ greatly from those of other rRNAs, they may be merely extending the trends observed earlier with mammalian mitochondial rRNAs (H.F. Noller and C.R. Woese, personal communication; 34,55).

C. Protein Synthesis in Trypanosome Mitochondria

The results discussed in sections (A) and (B) show that the 12S RNA manifests homology with the region of other large subunit rRNAs that is best conserved and has a number of important functions related to the A site of the ribosome, whilst the 9S RNA is at least consistent with a small subunit rRNA.

However, there may seem to be alternatives to the conclusion that these are functional rRNAs. The rRNAs could be supplanted in the ribosome by transcripts imported from the nucleus, or there may in fact be no protein synthesis in the trypanosome mitochondria. However, in either case the 9S and 12S RNAs would be expected to perform essential functions, for the corresponding genes are among the most conserved in kinetoplast DNA (see Introduction) and in such rapidly evolving molecules as mitochondrial genomes (34,69) it is very unlikely that redundant genes would be retained for long. Two alternative functions for the RNAs are as mRNAs or as factors in RNA processing enzymes. Although the genetic code in the kinetoplast is unknown, the use of a mitochondrial code does not show reading frames longer than a few hundred nucleotides [cf. the extensive expression of defined proteins from cloned maxicircle DNA in E. coli reported by Brunel et al. (14)]. Further, neither alternative function explains the similarities with known rRNAs. Since no kinetoplast ribosomes have been isolated per se, it is quite possible that the ribosome contains other small transcripts, which need not

be mitochondrially encoded. These could compensate for the deficit at the 3' terminus of the 9S rRNA which we have proposed. Such small, extra transcripts would parallel those found in the cytoplasmic ribosomes of Trypanosoma brucei (70) and Crithidia fasciculate (71).

The reduction in size of vertebrate mitochondrial rRNAs compared with those of eukaryotic cytoplasm, prokaryotes and fungal mitochondria is to some extent compensated for by an increase in the number of ribosomal proteins (72,73,74,75,76,77,78). A consequence of this is that the nucleus may be seen as slowly supplanting mitochondrial genetic funtions. If the kinetoplast ribosome is found to contain an even higher number or mass of proteins, it would be possible to regard these rRNAs as demonstrating those functions of rRNAs that are too specific or complex to be replaced by proteins.

With such an unusual rRNA complement the mitochondrial ribosomes of trypanosomes might have unusual properties. A high protein/RNA ratio might make the ribosomes unstable in standard isolation proedures, explaining the problems encountered in their isolation. The 12S rRNA lacks the sequence that appears to be required for CAP resistance and this could explain some of the unusual results obtained with mitochondrial protein synthesis in trypanosomes. Cellular protein synthesis in the insect trypanosome Crithidia is insensitive to inhibitors of mitochondrial protein synthesis like D-chloramphenicol or ethidium bromide and completely inhibited by inhibitors of cell-sap protein synthesis, like cycloheximide or anisomycin (79; R. Benne, personal communication). In contrast, two papers have reported CAP-sensitive protein synthesis in T. brucei (80,81), but both present results that are open to alternative interpretations. The origin of the CAP-sensitive ribosomes isolated by Hanas et al. (80) was not verified and we doubt whether they come from the mitochondria. Spithill et al. (81) isolated a mitochondrial fraction from T. brucei and found the amino acid incorporation into acid-soluble material of this fraction to be partly sensitive to CAP and other inhibitors of prokaryotic protein synthesis. Since the level of incorporation in these experiments was about two orders of magnitude lower than in mitochondria from other organisms, the significance of this activity for mitochondrial biosynthesis is doubtful. We maintain, therefore, our prediction that mitochondrial ribosomes of trypanosomes will turn out to be resistant to CAP. Unlike any other mitochondrial ribosomes, they might even be sensitive to cycloheximide, which would provide the simplest explanation for the lack of cycloheximide-insensitive protein

synthesis in trypanosomes.

The sequence reported here reveals several other interesting features. The kinetoplast genome is clearly unlike that of fungal mitochondria, in that the rRNA genes are not "split" (cf.:39,82,83,84,85) or widely dispersed, but it is also unlike that of mammalian mitochondria where tRNA genes are interspersed among rRNA and mRNA genes (34). No tRNAs have yet been located in this sequence, which covers 12% of the genome. Indeed, no 4S RNAs have been detected that might be encoded by the maxicircle, based on experiments where maxicircle fragments have been used as probes for hybridization with fractionated total cellular RNA or where labelled total cellular 4S RNAs have been hybridized to the maxicircle (M. Agostinella, A. Snijders, J.H.J. Hoeijmakers and P. Borst, unpublished). It is possible that all the tRNAs involved in protein synthesis are imported from the nucleus, or that their properties are as exceptional as the kDNA and rRNAs in the trypanosome mitochondria.

Acknowledgements

We would like to thank F. Sanger for advice and encouragement, A.T. Bankier for communicating the optimum conditions for digestion of DNA with EcoRI* activity and H.F. Noller for communicating a possible secondary structure for the human and bovine 12S rRNAs.

This work was supported in part by a grant to P.B. from the North Atlantic Treaty Organisation (NATO-559).

+Present address: Department of Molecular Biophysics and Biochemistry, Yale University, CT 06510, USA

References

1. Borst, P. and Hoeijmakers, J.H.J. (1979) in ICN-UCLA Symp. of Extrachromosomal DNA, Borst, P., Dawid, I.D., Weissmann, S.M. and Fox, C.F., Eds., Vol. 15, pp. 515-531 Academic Press, New York
2. Borst, P. and Hoeijmakers, J.H.J. (1979) Plasmid 2, 20-40
3. Englund, P.O. (1981) in Biochemistry and Physiology of Protozoa, Levandowsky, M. and Hutner, S.H., Eds., 2nd edn., Vol. 4, pp.334-383, Academic Press, New York
4. Steinert, M., Van Assel, S., Borst, P. and Newton, B.A. (1976) in The Genetic Function of Mitochondrial DNA, Saccone, C. and Kroon, A.M., Eds., pp. 71-81, North-Holland, Amsterdam
5. Borst, P., Fase-Fowler, F., Hoeijmakers, J.H.J. and Frasch, A.C.C. (1980) Biochim. Biophys. Acta 610, 197-210
6. Borst, P., Fairlamb, A.H., Fase-Fowler, F., Hoeijmakers, J.H.J. and Weislogel, P.O. (1976) in The Genetic Function of Mitohondrial DNA, Saccone, C. and Kroon, A.M., Eds., pp. 59-69, North-Holland, Amsterdam
7. Borst, P., Hoeijmakers, J.H.J. and Hajduk, S.O. (1981) Parasitology 82, 81-93

8. Hoeijmakers, J.H.J., Snijders, A., Janssen, J.W.G. and Borst, P. (1981) Plasmid 5, 329-350
9. Simpson, L. and Simpson, A.M. (1978) Cell 14, 169-178
10. Simpson, A.M. and Simpson, L. (1980) J. Mol. Parasitol., in press
11. Borst, P., Hoejmakers, J.H.J., Frasch, A.C.C., Snijders, A., Janssen, J.W.G. and Fase-Fowler, F. (1980) in The Organization and Expression of the Mitochondrial Genome, Kroon, A.M. and Saccone, C., Es., pp. 7-20, North-Holland, Amsterdam
12. Cheng, D. and Simpson, L. (1978) Plasmid 1, 297-315.
13. Brunel, F., Davison, J., Thi, V.H. and Merchez, M. (1980) Gene 12, 223-234
14. Osinga, K.A., Evers, R.F., van der Laan, J.C. and Tabak, H.J. (1981) Nucl. Acids Res. 9, 1351-1364
15. Vogelstein, B. and Gillespie, D. (1979) Proc. Nat. Acad. Sci. USA 76, 615-619.
16. Berk, A.J. and Sharp, P.A. (1977) Cell 12, 721-732
17. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517
18. Twigg, A.J. and Sherratt, D. (1980) Nature 283, 216-218
19. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J. Mol. Biol. 143, 161-178
20. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Nat. Acad. Sci. USA 74, 5463-5467
21. Polisky, B., Greene, P., Garfin, D.E., McCarthy, B.J., Goodman, H.M. and Boyer, H.W. (1975) Proc. Nat. Acad. Sci. USA 72, 3310-3314
22. Gronenborn, B. and Messing, J. (1978) Nature 272, 375-377
23. Messing, J., Crea, R. and Seeburg, P.H. (1981) Nucl. Acids Res. 9, 309-321
24. Anderson, S., Gait, M.J., Mayol, L. and Young, I.G. (1980) Nucl. Acids Res. 9, 1731-1743
25. Duckworth, M.-L., Gait, M.J., Goelet, P., Hong, G.F., Singh, M. and Titmas, R.C. (1981) Nucl. Acids Res. 9, 1691-1706
26. Staden, R. (1980) Nucl. Acids Res. 4, 4037-4051
27. Staden, R. (1977) Nucl. Acids Res. 4, 4037-4051
28. Staden, R. (1980) Nucl. Acids Res. 8, 817-825
29. Anderson, S. (1981) Nucl. Acids Res. 9, 3015-3027
30. Dubin, D.T., Timko, K.D. and Baer, R.J. (1981) Cell 23, 271-278
31. Eperon, I.C., Anderson, S. and Nierlich, D.P. (1980) Nature 286, 460-467
32. Van Etten, R.A., Walberg, M.W. and Clayton, D.A. (1980) Cell 22, 157-170
33. Dubin, T.D., Timko, K.D., Ksu Chen, C.C. α Azzolina, T. (1981) in Mitochondrial Genes, Slonimski, P., Borst, P. and Attardi, G., Eds., Cold Spring Harbor Laboratory, New York, in press.
34. Anderson, S., Barrell, B.G., de Bruijn, M.H.L., Coulson, A.R., Eperon, I.C., Sanger, F. and Young, I.G. (1982) J. Mol. Biol. 156, 683-717.
35. Glotz, C., Zwieb, C., Brimacombe, R., Edwards, K. and Kossel, H. (1981) Nucl. Acids Res. 9, 3287-3306
36. Branlant, C., Krol, A., Machatt, M.A., Pouyet, J. and Ebel, J.-P. (1981) Nucl. Aids Res. 9, 4303-4324
37. Brosius, J., Dull, T.J. and Noller, H.F. (1980) Proc. Nat. Acad. Sci. USA 77, 201-204
38. Baer, R.J. and Dubin, D.T. (1981) Nucl. Acids Res. 9, 323-337
39. Dujon, B. (1980) Cell 20, 185-197
40. Allet, B. and Rochaix, J.-D. (1979) Cell 18, 55-60
41. Edwards, K. and Kossel, H. (1981) Nucl. Acids Res. 9, 2853-2869
42. Nomiyama, H., Sakaki, Y. and Tagaki, Y. (1981) Proc. Nat. Acad. Sci. USA 78, 1376-1380
43. Veldman, G.M., Klootwijk, T., deRegt, V.C.H.F., Planta, R.I., Branlant, C., Krol, A. and Ebel, J.-P. (1981) Nucl. Acids Res. 9, 6935-6952

44. Greenwell, P., Harris, R.J. and Symons, R.H. (1974) Eur. J. Biochem. 49, 539-554
45. Eckermann, D.J. and Symons, R.H. (1978) Eur. J. Biochem. 82, 225-234
46. Herr, W. and Noller, H.F. (1979) J. Mol. Biol. 130, 421-432
47. Kearsey, S.E. and Craig, I.W. (1980) Nature 290, 607-608
48. Bkan, H., Wright, C.T., Bibb, M.J., Wallace, D.C. and Clayton, D.A. (1981) Proc. Nat. Acad. Sci. USA 78, 3789-3793
49. Blanc, H., Adams, C.w. and Wallace, D.C. (1981) Nucl. Acids Res. 9, 5785-5795
50. Branlant, C., Krol, A., Machatt, A. and Ebel, J.-P. (1981b) Nucl. Acids Res. 9, 293-307
51. Gourse, R.L., Thurlow, D.L., Gerbi, S.A. and Zimmermann, R.A. (1981) Proc. Nat. Acad. Sci. USA 78, 2722-2726
52. Kazemie, M. (1975) Eur. J. Biochem. 58, 501-510
53. Kazemie, M. (1976) Eur. J. Biochem. 67, 373-378
54. Branlant, C., Krol, A., Sriwidada, J. and Brimacombe, R. (1976) Eur. J. Biochem. 70, 483-492
55. Zwieb, ., Glotz, C. and Brimacombe, R. (1981) Nucl. Acids Res. 9, 3621-3640
56. Noller, H.F. and Woese, C.R. (1981) Science 212, 403-411
57. Stiegler, P., Carbon, P., Zuker, M., Ebel, J.-P. and Ehresmann, C. (1981) Nucl. Acids Res. 9, 2153-2172
58. Sor, F. and Fukuhara, H. (1980) C.R. Acad. Sc. Paris, 291, 933-936
59. Salim, M. and Maden, B.E.H. (1981) Nature 291,205-208
60. Rubtsov, P.M., Musakhanov, M.M., Sakharyev, V.M., Krayev, A.S., Skryabin, K.G. and Bayev, A.A. (1980) Nucl. Acids Res. 8, 5779-5794
61. Ehresmann, B., Backendorf, C., Ehresmann, C. and Ebel, J.-P. (1977) FEBS Lett. 78, 261-266
62. Nomura, M., Yates, J.L., Dean, D. and Post, L.E. (1980) Proc. Nat. Acad. Sci. USA 77, 7084-7088
63. Woese, C.R., Magrum, L.J., Gupta, R., Siegel, B., Stahl, D.A., Kop, J., Crawford, N., Brosius, J., Gutell, R., Hogan, J.J. and Noller, H.F. (1980) Nucl. Acids Res. 8, 2275-2293
64. Glotz, C. and Brimacombe, R. (1980) Nucl. Acids Res. 8, 2377-2395
65. Carbon, P., Ebel, J.-P. and Ehresmann, C. (1981) Nucl. Acids Res. 9, 2325-2333
66. Woese, C.R., Fox., G.E., Zablen, L., Uchida, T., Bonen, L., Pechman, K., Lewis, B.J. and Stahl, D. (1975) Nature 254, 83-86
67. Herr, W., Chapman, N.M. and Noller, H.F. (1979) J. Mol. Biol. 130, 433-449
68. Prince, J.B., Hixson, S.S. and Zimmermann, R.A. (1979) J. Biol. Chem. 254, 4745-4749
69. Brown, W.M., George, M. and Silson, A.C. (1979) Proc. Nat. Acad. Sci. USA 76, 1967-1971
70. Cordingley, J.S. and Turner, M.J. (1980) Mol. Biochem. Parasitol. 1, 91-96
71. Gray, M.W. (1981) Mol. Cell. Biol. 1, 347-357
72. Leister, D.E. and Dawid, I.B. (1974) J. Biol. hem. 249, 5108-5118
73. O'Brien, T.W., Denslow, N.D., Harrille, T.O. and Hessler, R.A. (1980) in The Organization and Expression of the Mitochondrial Genome, Kroon, A.M. and Saccone, C., Eds., North-Holland, Amsterdam
74. Czempiel, W., Klose, J. and Bass, R. (1976) FEBS Lett. 62, 259-262
75. Tan, H.V., Gavrilovic, M. and Schapira, G. (1974) FEBS Lett. 45, 299-303
76. Garrett, R.A. (1979) in International Review of Biochemistry, Offord, R.E., Ed., Vol. 25, pp. 121-177, University Park Press, Baltimore
77. Faye, G. and Sor, F. (1977) Molec. Gen. Genet. 155, 27-34

78. Neupert, W., Miller, F., Michel, R. and Hallermayer, G. (1979) Methods Enzymol. 56, 79-92
79. Kleisen, .M. and Borst, P. (1975) Biochim. Biophys. Acta 390, 78-81
80. Hanas, J., Linden, G. and Stuart, K. (1975) J. Cell. Biol. 65, 103-111
81. Spithill, T.W., Shimer, S.P. and Hill, G.C. (1981) Mol. Biochemical Parasitol. 2, 235-255
82. Heyting, C., Talen, J.L., Weijers, P.J. and Borst, P. (1979) Molec. Gen. Genet. 168, 251-277
83. Hahn, U., Lazarus, C.M. Lunsdorf, H. and Kuntzel, H. (1979) Cell 17, 191-200
84. De Vries, H., De Jonge, J.C., Bakker, H., Meurs, H. and Kroon, A. (1979) Nucl. Acids Res. 6, 1791-1803
85. Heckman, J.E. and RajBhandary, U.L. (1979) Cell 17, 583-595
86. Brosius, J., Palmer, M.L., Kennedy, P.J. and Noller, H.F. (1978) Proc. Nat. Acad. Sci. USA 75, 4801-4805