The ribosomal RNA genes of *Drosophila* mitochondrial DNA

Douglas O.Clary* and David R.Wolstenholme

Department of Biology, University of Utah, Salt Lake City, UT 84112, USA

Received 20 February 1985; Revised and Accepted 2 May 1985

ABSTRACT

The nucleotide sequence of a segment of the mtDNA molecule of *Drosophila yakuba* which contains the A+T-rich region and the small and large rRNA genes separated by the tRNA^{Val} gene has been determined. The 5' end of the small rRNA gene was located by S1 protection analysis. In contrast to mammalian mtDNA, a tRNA gene was not found at the 5' end of the *D. yakuba* small rRNA gene. The small and large rRNA genes are 20.7% and 16.7% G+C and contain only 789 and 1326 nucleotides. The 5' regions of the small rRNA gene (371 nucleotides) and of the large rRNA gene (643 nucleotides) are extremely low in G+C (14.6% and 9.5%, respectively) and convincing sequence homologies between these regions and the corresponding regions of mouse mt-rRNA genes were found only for a few short segments. Nevertheless, the entire lengths of both of the *D. yakuba* mt-rRNA genes can be folded into secondary structures which are remarkably similar to secondary structures proposed for the rRNAs of mouse mtDNA. The replication origin-containing, A+T-rich region (1077 nucleotides; 92.8% A+T), which lies between the tRNA^{Ile} gene and the small rRNA gene, lacks open reading frames greater than 123 nucleotides.

INTRODUCTION

We have shown that the mitochondrial DNA (mtDNA) molecule of *Drosophila yakuba* contains genes for two rRNAs, 22 tRNAs and 13 polypeptides (cytochrome c oxidase (CO) subunits I, II and III, ATPase subunit 6, Unidentified Reading Frames (URF) 1-6, 4L and A6L (ATPase 8)) (1-6). The genes found in the *D. yakuba* mtDNA molecule are the same as those found in mammalian mtDNAs (7-12), but their relative order differs. We have reported the nucleotide sequences of all of the polypeptide and tRNA genes and portions of the rRNA genes.

In this paper we present a sequence of the *D. yakuba* mtDNA molecule that includes the two complete rRNA genes together with the replication origin-containing A+T-rich region. We show that in spite of their unusually small size and high A+T content the secondary structure potential of both of the *D. yakuba* mt-rRNA genes is remarkably similar to the secondary structures which have been proposed for mouse mt-rRNAs (13-14).

MATERIALS AND METHODS

Details regarding preparation and identification of pBR322 and pBR325 clones of mtDNA from D. yakuba (Stock 2371.6, Ivory Coast), restriction enzyme digestions, electrophoresis, cloning of fragments into M13mp8 or M13mp9, purification of single-stranded and double-stranded M13 DNAs and preparation of viral DNAs containing partial deletions of cloned restriction fragments of mtDNA are given or referred to in (1,3).

All DNA sequences were obtained from M13mp8- or M13mp9-cloned fragments by the extension-dideoxyribonucleotide termination procedure (15) using [α -³²P]dATP (600 Ci/mmol; New England Nuclear) as described (1). The sequencing strategies used are given in Fig. 1.

Computer analyses of DNA sequences were carried out using the programs referred to in (3).

Isolation of D. yakuba RNA

Frozen ovaries, obtained by dissection of approximately 300 flies (16) were thawed into 3 ml of 4 M guanidinium isothiocyanate, 5mM sodium citrate (pH 7.0) 0.1 M β -mercaptoethanol and 0.5% sarkosyl and homogenized, and RNA was extracted using the hot phenol method described in (17,18). The resulting DNA was stored in sterile TE (10 mM Tris (pH 8.0) 1 mM EDTA) at -20°C until used.

S1 hybridization-protection

We used S1 hybridization-protection analysis (18,19) to determine the location of the 5' end of the small rRNA gene. By deletion of the D. yakuba mtDNA HindIII-C fragment contained in M13mp9, we produced a cloned fragment (X45-H2) which begins a short distance 3' to the FnuDII site in the small rRNA gene (Fig. 2) and extends through the A+T-rich region to the HindIII site 166 nucleotides from the 5' end of the URF2 gene (1). A radioactively (³²P)-labeled complement to the X45-H2 mtDNA insert was made by extension synthesis using Klenow fragment and the M13 sequencing primer. The reaction product was cleaved with FnuDII (37°C for 20 min), FnuDII was inactivated by incubation at 65°C for 15 min and an extension chase reaction using 0.5 mM dATP was carried out at 30°C for 15 min.

Hybridization. To 10 μ l of the final extension reaction product was added 100 μ l of D. yakuba total RNA, 100 μ l of yeast tRNA (100 mg), 2 μ l of sterile 3 M sodium acetate and 500 μ l of ethanol. The mixture was cooled to -70°C, centrifuged, and the precipitate lyophilized. A control was prepared by an identical procedure except that 100 μ l of sterile TE was substituted for the D. yakuba RNA solution. Fifty μ l of 40 mM PIPES (pH 6.4), 1 mM EDTA (pH

8.0), 0.4 M sodium chloride, 80% deionized formamide were added to each precipitate. The solution was heated to 65°C for 15 min and then incubated at 30°C for 4.5 hrs.

S1 digestion. To each of the products of hybridization was added 0.5 ml of 0.28 M NaCl, 0.05 M sodium acetate (pH 4.6), 4.5 mM ZnSO₄, 20 mg/ml of single-stranded salmon sperm DNA, and 1 µl of S1 nuclease (1000 U), and the mixture was incubated at 22°C for 15 min. The digestion mixture was extracted with phenol, and the nucleic acid component precipitated with 5 M ammonium acetate and isopropanol. The precipitate was washed with ethanol, dried, resuspended in 100 µl of water, ethanol precipitated, dried and resuspended in 3 µl of sequence loading buffer.

To generate a sequence ladder extending from the *Fnu*DII site, X45-H2 was sequenced using the Sanger method. To the reaction products was added 1 µl of *Fnu*DII (1:2 dilution in water), and the mixtures were incubated at 37°C for 15 min. The products of all reactions were separated by electrophoresis using a 6% acrylamide gel run at 1600 volts for 4 hrs.

Figure 1. A map of the *D. yakuba* mtDNA molecule showing the relative locations of the A+T-rich region (cross-hatched), the two rRNA genes (dotted), the origin (O) and direction (R) of replication and *Eco*RI and *Hind*III sites and fragments (A-E and A-F, respectively), (see (1) for details and references). The bar under the map indicates the segment sequenced. This segment is expanded below and the arrows beneath it show the extent and direction of individual sequences obtained, either from the ends of various restriction fragments or from deletion subfragments of restriction fragments. The 5'-3' direction of the sequence of this segment shown in Fig. 2 is from left to right.

←tRNA ^{ile} A+T rich region <u>GCAATTCATTA</u> ATAATAAAAACTCATCTTTTTTTTTTTTTTTTATTATTTACTATTTTTTTTTTAAAGAT	75
TACAATTTTTAAAAATTAATTGATTATATTAGATAACACATTTTAGGAATTCATAAAATATTTTTATAAAAAATT	150
ATTAAAAATTTTTATTTTCTATATAGATATATATGTAATATAAATCTTCAACTATATATAAATATATAAATAT	225
ATAATAATTAATTAATTTATTATATTATTTATATAAATCCAAGAATGTTATATAAAATTTTGAATATAGATTTTT	300
TTTTACAAATTATGTTACCATTTTTGGATTGTGAAATTTTTTATTAAATTGTTAACTATATTATAATATATTAA	375
ACATTTATATATATAGATTATCTATTAATTTAGAACTTAGTATACAAAAATTTTTTTTTTAAAAAAAAAAAAA	450
GAATAGGTTTTTTATTATTTTATAAATAATTATTAATTAGAAAAATTTATTTAATTAATAATTTTAAATTTAA	525
TATTTTAAAAATTATTATAATTTATTAATTAATTTATAAAATTTTTTAATTTATTTAATTAACAAATATTGTTTT	600
ATAATTTTTATTATAAAATAATTATTTTGTTAATAAATATAGTAAATTTTATTAATTTTAATTTTATTATTTT	675
AATGTATAAAAAAATAATTTTATTGTATAAATATTAATAAATAAATATGTAATAAGGTATTTAATTTAAGAA	750
ATTATTATATTTTATTATAGGGAATAAGTTATCAGATTTAGTGATTAATAAATAATTATTTTAAATAGTTTATTAT	825
AAGTTACAAATTTTATTATTTTTTTAAAAATTTTACTAATTTTAATTATTAATTTTATGATAAAAAATAATTT	900
GAATTTTATTTAATATACTAATATATATATATATATAAATAAATTAATTTTGTTTATAAGTTTTTTATTAT	975
TAATTATTATTTAAATTATCTATTATTTTTATTTTATTATTATTATAAAAAATTATTTTGATAAATTTTATTA	1050
AATTTATATTATAAAATAAATTTTTTTATTGTATATTTTAAAGTTTTATTTTGGCTTAAAAATTTGTTATTAGTT	1125
small rRNA→	
TGATTTATATGTAAATTTTGTGTGAATTTATATTTATTAAAAATAAATATATTATAAATTATTTATTCGCAG	1200
TAATTAATATTATTAATTAAGAAATTTAGAAATAGCAATATTAATAAGTATTGACCAAATGGTGCCAGCAGTC	1275
GCGGTTATACCAATAATACAAATAAATTTTTTTAGTATGAATTAAATTGATTTATTTAAAAATAAAATTAATAT	1350
PhxDII	
ATTAAGTGAAATTTTATATTTAAATTATTTTATAAAATAAATAATGAAGTTAAAAAATTTTAATAAAAACTAGC	1425
ATTAGATACCCCTATTATTTAAAAATGTAATAAATTGCTAAAGTAGTAATAGTTATGTTCTTGAAACTTAAAAAT	1500
TTGGCGGTATTTTAGTCTATCCAGAGGAACCTGTTTGTAAATCGATAATCCACGATGGACCTTACTTAAATTTGT	1575
AATCAGTTTATATACCGTCGTTATCAGAATATTTATAAGAATAAATATTCAATAATTTTAATAAAAAATTAT	1650
ATCAGATCAAGGTGTAGCTTATATTTAAGTAATAATGGGTTACAATAAATTTATTTAAACGGATAAAATTTATGAA	1725
AAAAATTTTGAAGGTGGATTGCTAGTAAAATTATAAGATTAATAATTTGATTTTAGCTCTAAAAATATGTACAC	1800
ATCGCCCGTCGCTCTTATTATTAAGTAAGATAAGTCGTAAACATAGTAGATGTACTCGAAAGTCTATCTAGAATC	1875
tRNA ^{val} →	
<u>ACAATTTAAAGCTTATTAAGTAAAGCATTTTCATTACATTGAAAAATTTTTGTGCAATCAATATAAATTCAT</u>	1950

```

large rRNA →
ATATTTTATTTATTTTAAATTATTTTATATAAAAAATTAGAAATAACTATAAAATTTAAAGTTTCTAGTATT 2025
.
.
.
GTTTAAAGAAAAATAATTTTAATAAAGTAGTATATTAGTATTGTAAAGAAAAATTGAAATAATTTGAAAAAATTTT 2100
.
.
.
ATTTTAAAGAAAAATTAATTTATTGTACCTGTGTATCAGCGTTTATTAATAAAAAATAAATATTTATTTTTC 2175
.
.
.
TCGATTTTAAAGAGTTAATATAATTTTAAGTTAATGTGACAAATTTATTTATAATATTATATTAGAAATGAAA 2250
.
.
.
TGTTATTCGTTTTTAAAGGTATCTAGTTTTTAAAGAAATAAATTTAATTTAGAAATTATAAATTTATTTAATTAT 2325
.
.
.
TTATTTAATTAATAATTATAATTTTAATATTTTATGGGATAAGCTATAAAAAATAATTTTAAAAATAATAAAT 2400
.
.
.
AAATTTAATAAATATATGCTTAGAATTAGCAATTATTAATAAATGTGTTATAATTTTATAAATTAATTTAT 2475
.
.
.
TTATTAATTTTAATTTATTTTAAAAATTAATTTTAAATTTTAAATTAAGTAATAATGATAGAATTAGTATA 2550
.
.
.
TAATATTGTTAAATAAATTTTATGAAAAGTTTAAATAAAGAATTCGGCAAAAAATGTTTCGCCTGTTTACA 2625
.
.
.
AAAAATGTCTTTTTGAATTATATATAAAGTCTAACCTGCCCACTGAAAAATTTTAAATGGCCGAGTATTTTGAC 2700
.
.
.
TGTGCAAGGTAGCATAATCATTAGTCTTTTAAATGAAGGCTGGAATGAATGTTGGACGAAATATTAACGTGTT 2775
.
.
.
CATTTAAATTTAAAAAGAAATTTTATTTTCTAGTCAAAAAGCTAAAATTAATTTAAAGACGAGAAGACCTATA 2850
.
.
.
AATCTTTATATTTTATTTTATTTTAAATTAAGATTAATTTTAAATTTTAAATTAATTTTATTTGCGGTGA 2925
.
.
.
TATTAATAATTTAAAAAATTTTAAATTTTAAAAAACATTAATTTATGAATAATTGATCCATTAATAATGATTAAAA 3000
.
.
.
AATTAAGTTACTTTAGGGATAACAGCGTAATTTTTTGGAGAGTTCATATCGATAAAAAAGATTGCGACCTCGAT 3075
.
.
.
GTTGGATTAAGATATAATTTTGGGTGTAGCGGTTCAAATTTTAAAGTCTGTTCCGACTTTTAAATCTTACATGATC 3150
.
.
.
TGAGTTCAAACCGGTGTAAGCCAGGTGGTTTCTATCTTTAAAAAATTATAATTTTATAGTACGAAAGGACCAAA 3225
.
.
.
TATTAATAAATTATATTTTATATAAGAAATTTATTTAATATATAAATAAAGTATTTTGG
trNAleuCUN →

```

Figure 2. Nucleotide sequence of the segment of the *D. yakuba* mtDNA molecule identified in Fig. 1. The sequence contains the small and large rRNA genes, separated by the tRNA^{val} gene (anticodon underlined). The A+T-rich region, previously mapped by electron microscope and restriction enzyme studies (16,20,21), is now defined as the 1077 nucleotide sequence lying between the 5' end of the tRNA^{ile} gene and the 5' end of the small rRNA gene (see text). Transfer RNA genes are boxed and the terminal regions of the two rRNA genes are boxed with broken lines. The sequence shown is the sense strand (5'-3') of the small and large rRNA genes and the tRNA^{val} gene, and the arrows indicate the direction of transcription of each gene. The location of a *Fnu*DII site used to determine the 5' terminal nucleotide of the small rRNA gene, and an *Eco*RI site in the large rRNA gene are shown. The entire tRNA^{ile} gene, the first 416 nucleotides of the A+T-rich region (with nucleotides 376, 412 and 415 omitted), the 3' terminal 222 nucleotides of the small rRNA gene, and the 5' terminal 138 nucleotides of the large rRNA genes are given in (1). The 3' terminal 296 nucleotides of the large rRNA gene and the entire tRNA^{leu}_{CUN} gene are given in (6).

RESULTS AND DISCUSSION

The mt-rRNA genes

The sequence of a continuous 3284 nucleotide section of the D. yakuba mtDNA molecule, determined using the strategies summarized in Fig. 1 is shown in Fig. 2. It has been determined that this D. yakuba sequence contains the entire genes for the mitochondrion's small and large rRNA subunits, separated by the tRNA^{val} gene.

The sequence of 418 nucleotides (26.1% G+C) upstream from the D. yakuba mt-tRNA^{val} gene (i.e., the 3' end of the small rRNA gene, Fig. 2) is 60.1% homologous to the corresponding region of the mouse small mt-rRNA gene (36.9% G+C; (8), see legend to Fig. 2). Comparisons to the 3' terminal 20 nucleotides of the small mt-rRNA of mosquito (Aedes albopictus) (22) indicated that the 3' terminal nucleotide of the D. yakuba small mt-rRNA gene is adjacent to the 5' terminal nucleotide of the tRNA^{val} gene (1). Upstream from the above mentioned 418 nucleotide region of the D. yakuba molecule, the sequence is very low in G+C nucleotides (14.6% over 371 nucleotides) compared to the 481 nucleotides (34.7% G+C) of the 5' end of the mouse small mt-rRNA gene (8) and only short segments of homology to the mouse gene have been found (Fig. 4a and b). Sequence homologies which would help determine the location of the 5' end of the D. yakuba small rRNA gene were not detected. Further, contrary to the situation in mammals where the tRNA^{phe} gene is located at the 5' end of the small rRNA gene, a sequence with the characteristics of a tRNA gene is not located anywhere in the sequence between the tRNA^{val} and tRNA^{ile} genes (Fig. 2).

In order to define the 5' end of the D. yakuba small rRNA gene, we have carried out S1 hybridization-protection analysis. Consistent with the low G+C content, very few restriction sites occur in the general region of the D. yakuba mtDNA molecule expected to contain the 5' end of the small rRNA gene. We used an M13 clone (X45-H2) which contains a deletion fragment of the HindIII-C fragment of D. yakuba mtDNA, one end of which is located a short distance downstream from a FnuDII site which in turn lies 600 nucleotides upstream from the 3' end of the small rRNA gene (Fig. 2). The other end of the X45-H2 insert is a HindIII site in the URF2 gene (1). For unknown reasons, we were unable to label the 5' end of the cleaved FnuDII site of the FnuDII-HindIII fragment using polynucleotide kinase and [α -³²P]dATP, even though the 5' ends of cleaved FnuDII sites of other fragments in the reaction labeled. We therefore continuously labeled a FnuDII-HindIII fragment by extension synthesis.

Figure 3. S1 protection analysis to determine the location of the 5' end of the small rRNA gene. Two autoradiographs (I, II) are shown, each resulting from electrophoresis through a 6% acrylamide gel. Lanes A, C, G and T of autoradiograph I (left) are part of a sequence ladder of the *Fnu*DII digestion products of a Sanger sequencing reaction of clone X45-H2. The + and - lanes in autoradiograph I contain the S1 digestion products of the continuously labeled X45-H2 clone cleaved at the *Fnu*DII site and hybridized with whole *D. yakuba* RNA and yeast tRNA (+ lane), or yeast tRNA alone (- lane). In the + lane the faint band (marked with an arrow head) lying above the four dark bands contains the largest protected fragment and corresponds to the band marked with an arrow head in the A lane. Reading of the sequence ladder above the latter band is ambiguous due to the presence of a band at the same position in each of the A, C, G and T lanes. This band probably represents an M13 fragment resulting from *Fnu*DII digestion, as it does not appear in the sequence ladder of the same region of the X45-H2 clone when *Fnu*DII digestion is omitted (autoradiograph II, right). All of the sequence shown in this latter autoradiograph can be read unambiguously. The A band marked with an arrow head is the same as that marked in autoradiograph I. A sequence which includes the bands corresponding to the protected fragments in autoradiograph I is given on the far right (3'-5', top to bottom) together with the reverse complement which corresponds to nucleotides 1082-1108 of the 5'-3' sequence shown in Fig. 2. The T* (nucleotide 1088 in Fig. 2) is interpreted as the 5' terminal nucleotide of the small rRNA gene (see text).

The results of the S1 hybridization-protection analysis are shown in Fig. 3. In the lane containing the S1 digestion products of *D. yakuba* RNA-protected, FnuDII-HindIII mtDNA fragments, four dark bands are evident above which is a single lighter band, and these five bands represent fragments differing by one nucleotide in length. The positions of the five bands in the gel relative to the accompanying sequence ladder indicate that the RNA molecules used in the analysis included some which were complementary to a region of the mtDNA molecule between the FnuDII site and nucleotides 183-187 upstream from the FnuDII site (nucleotides 1088-1092 in Fig. 2). This suggests that any of the nucleotides 1088-1092 may correspond to the 5' end of the small rRNA. Alternatively, the multiple protected fragments may have resulted from digestion by S1 nuclease of terminal nucleotides in some RNA-DNA hybrids due to the low G+C content of this region of the component molecules. We designate nucleotide 1088 (T), which corresponds in position to the band representing the largest protected fragment (Fig. 3), as the 5' terminal nucleotide of the *D. yakuba* small rRNA gene (Fig. 2). By this definition this gene contains 789 nucleotides.

The 1326 nucleotide sequence lying between the tRNA^{val} gene and the tRNA^{leu}_{CUN} gene (Fig. 2) is interpreted as the *D. yakuba* large rRNA gene. Close to the center of this sequence occurs an EcoRI site (nucleotides 2592-2597; Fig. 2). The sequence of 683 nucleotides between the beginning of this EcoRI site and the tRNA^{leu}_{CUN} gene is 23.4% G+C and includes two separate regions of 267 and 299 nucleotides which are 67.7% and 67.9% homologous, respectively, to corresponding sequences in the mouse large mt-rRNA gene (see legend to Fig. 5). These two regions of high homology are separated in the *D. yakuba* large rRNA gene by 117 nucleotides which include only one short segment of convincing homology (nine nucleotides, 100% homologous; Fig. 5d) to the corresponding 153 nucleotides of the mouse large mt-rRNA gene. From secondary structure considerations, we have previously suggested that the nucleotide which is adjacent to the 5' end of the tRNA^{leu}_{CUN} gene is the 3' terminal nucleotide of the large rRNA gene (6). The 643 nucleotides between the tRNA^{val} gene and the EcoRI site are only 9.5% G+C, compared to 32.4% G+C for the corresponding region of mouse mtDNA (8). Convincing homologies between this region of the *D. yakuba* and mouse mtDNA sequences have been found for only four short segments (10 to 24 nucleotides; Fig. 5a and b).

The nucleotide sequence of the large mt-rRNA gene of mosquito has recently been reported (23) and is 82.5% homologous to the *D. yakuba* large mt-rRNA gene. By comparison to a 44 nucleotide sequence reported to represent

the 5' terminus of the large mt-rRNA of mosquito, it was argued that the 5' terminus of the mosquito large mt-rRNA gene was the second nucleotide 3' to the tRNA^{val} gene. As the first 30 nucleotides (100% A+T) 3' to the mt-tRNA^{val} genes in *D. yakuba* and mosquito are 93% homologous it seems extremely likely that the 5' end of the *D. yakuba* mt-rRNA gene is adjacent to or within a small number of nucleotides from the end of the tRNA^{val} gene. Sequence data

Figure 4

Figures 4 and 5. Secondary structure models for the small rRNA gene (Fig. 4) and the large rRNA gene (Fig. 5) of *D. yakuba* mtDNA. The line drawings in Fig. 4a-c and Fig. 5a-d represent sections of the mouse small (12S) mt-rRNA gene and large (16S) mt-rRNA gene, respectively (8) drawn to scale and folded into the exact secondary structures proposed in (13,14). The sequences shown are the regions of the *D. yakuba* small rRNA gene and the large rRNA gene (Fig. 2) folded into the closest approximate fit to the corresponding sections of the respective mouse gene. The *D. yakuba* sequences and the mouse line drawings are numbered in each case from the 5' terminal nucleotide of the respective rRNA gene. All three sections of the small rRNA genes (Fig. 4a-c) and sections 1 and 2 of the large rRNA genes (Fig. 5a and 5b) are the same as those given for the mouse and human genes in (13,14). Sections 3 (Fig. 5c) and 2 (Fig. 5d) of the large rRNA genes are the combined sections 3 and 4, and 5 and 6, respectively, given for the mouse and human genes in (14).

Figure 4. The small rRNA gene. The corresponding accompanying solid lines marked a-c between nucleotides 1 and 371 in the *D. yakuba* sequence, and nucleotides 1 and 481 in the mouse sequence indicate regions of primary sequence homology (given as a percentage next to each *D. yakuba* region) between the two genes. A block of 13 nucleotides (337-349) in the homologous region c (nucleotides 333-351) of the *D. yakuba* gene are also 100% homologous to nucleotides 422-434 of the mouse gene. In Fig. 4b (lower right) is shown an alternative possible configuration of nucleotides 305-372 of the *D. yakuba* gene in which nucleotides 333-351 (broken line) are placed in a secondary structural location corresponding to that suggested for nucleotides 422-434 (broken line) of the mouse gene. The asterisks indicate corresponding nucleotides (305 and 372) in the main and inset *D. yakuba* sequences. The

sequences of the D. yakuba gene and of the corresponding region of the mouse gene which lie 3' to the directional arrow (Fig. 4b; nucleotides 371-372, D. yakuba; nucleotides 481-482, mouse) are 60.1% homologous. This latter value includes 13 single and 1 quadruple insertion/deletions postulated between the two genes, but not the single section of 47 nucleotides of the mouse gene which is missing from the D. yakuba gene (Fig. 4c; open triangle between nucleotides 728 and 729 of the D. yakuba sequence). The broad arrows between nucleotides 40 and 41 (Fig. 4a), and between nucleotides 286 and 287 (Fig. 4b) of the D. yakuba sequence indicate where in the mouse diagram a loop of 37 nucleotides, and a loop of 20 nucleotides, respectively, are inserted.

concerning the 3' end of the mosquito large mt-rRNA (23) are also consistent with the interpretation that in the D. yakuba mtDNA molecule the nucleotide adjacent to the tRNA^{leu}_{CUN} gene is the 3' terminus of the large rRNA gene.

The overall G+C contents of the D. yakuba small mt-rRNA gene (20.7%) and large mt-rRNA gene (16.7%) are very close to the values of 20.9% and 17.6% G+C determined by base composition analysis of the small and large mt-rRNAs respectively, of D. melanogaster (24).

Both of the D. yakuba mt-rRNA genes are considerably smaller than mammalian mt-rRNA genes (955 and 1582 in mouse; 789 and 1326 in D. yakuba).

Secondary structure of the small and large mt-rRNA genes

Secondary structure models for the small (12S) and large (16S) rRNAs encoded by mouse mtDNA (and human mtDNA) have been proposed (13,14). These models are derived from primary sequence and secondary structure comparisons to small and large rRNA genes from other sources (Zea mays chloroplast, Xenopus laevis and Saccharomyces cerevisiae nucleus), using as a basis experimentally established secondary structures of Escherichia coli 16S and 23S rRNAs. In view of the considerable differences in size and base composition between the 5' regions of both mt-rRNA genes of D. yakuba and mouse we considered it more worthwhile to determine the extent to which the D. yakuba genes could be folded into secondary structures similar to those proposed for mouse mt-rRNA genes, rather than to attempt to construct secondary structure models based on primary sequence comparisons. The results are shown in Figs. 4 and 5. Segments throughout the entire lengths of both of the D. yakuba mt-rRNA genes can be folded into secondary structures which show remarkable resemblance to the secondary structures of the corresponding segments of mouse mt-rRNA genes. Also the few convincingly homologous sequences which occur in the 5' regions of each of the corresponding D. yakuba and mouse genes occur at similar structural locations (Fig. 4a and b; Fig. 5a and b). Within the D. yakuba small rRNA gene component of one of these pairs of homologous sequences (c, Fig. 4b) occur 13 contiguous nucleotides which are 100% homologous to 13 nucleotides of the mouse small rRNA gene located

Figure 5. The large rRNA gene. The corresponding accompanying solid lines marked a-d between nucleotides 1 and the EcoRI site (Fig. 5c; nucleotides 644 through 649) of the D. yakuba sequence and nucleotides 1 to 813 of the mouse sequence indicate regions of primary sequence homology (given as a percentage next to each D. yakuba region) between the two genes. The sequences marked by directional arrows on the D. yakuba gene (Fig. 5c and d; nucleotides 644 to 910, and 1028 to the 3' end) are 67.8% and 67.9% homologous respectively, to the corresponding sequences marked by directional arrows (nucleotides 814-1084, and 1238-1559) on the mouse gene. The value for the former sequence includes four single and one double insertion/deletions postulated between the two sequences. The value for the latter sequence includes three single, one double and one quadruple insertion/deletions postulated between the two sequences, but not the sections of 17, 7 and 23 nucleotides of the mouse gene which are missing from the D. yakuba gene (Fig. 5d; open triangles between nucleotides 1292 and 1293, 1314 and 1315, and adjacent to the 3' terminus of

the D. yakuba sequence). The lines marked e which accompany the D. yakuba and mouse sequences (Fig. 5d) indicate a segment of nine nucleotides in each sequence which are 100% homologous and lie within a region (nucleotides 911-1027, D. yakuba; nucleotides 1085-1237, mouse) which shows little apparent primary sequence homology. In Fig. 5b, the broad arrow between nucleotides 433 and 434 of the D. yakuba sequence indicates where in the mouse diagram a distinct complex 146 nucleotide stem and loop structure is inserted.

elle and non-organelle rRNAs which indicate that variation in size is accounted for by deletion/insertion of blocks of contiguous nucleotides (25-30).

Data obtained from primary sequence and secondary structure comparisons of rRNAs from a wide variety of organisms have been used to argue for the existence in vivo and for functional significance of different helices, loops and interhelical regions in models of rRNA structure (25,28,29). It is clear from the present comparisons that most of the secondary structural features of the D. yakuba and mouse mt-rRNA genes have been maintained in spite of extensive divergence in base composition of these genes, particularly in the 5' regions, in the lines which led from a common ancestor to the present day mouse and D. yakuba. This strongly supports an argument in favor of there being a distinct functional significance for each of the components which are conserved in the secondary structure models of the mouse and D. yakuba mt-rRNA genes.

A secondary structure model was proposed for only the 3' half of the mosquito large rRNA gene, based mainly on comparisons to E. coli 23S RNA, as well as to mouse 16S mt-rRNA (23). This model exhibits most of the major features of the corresponding portion of our secondary structure model of the D. yakuba large mt-rRNA gene, but differs in details of stem and loop structures, most notably those involving nucleotides 1120 to 1214 (Fig. 5d).

The A+T-rich region

The sequence of 1077 nucleotides which separates the tRNA^{ile} gene and the small rRNA gene (Fig. 2) has an A+T content of 92.8% and lacks open reading frames larger than 123 nucleotides. This sequence constitutes the major portion of the A+T-rich region of the D. yakuba mtDNA molecule which was originally identified by electron microscope studies (16,20).

Replication of the molecule has been shown to originate within the A+T-rich region, approximately 250 to 450 nucleotides from the left end (21). The left half of this segment (nucleotides 261-330; Fig. 2) can be folded into a hairpin structure beginning with a 12 nucleotide pair stem (nucleotides 261-272; 319-330; Fig. 2) which includes three G-C pairs. This is of particular interest as replication of a number of circular DNA molecules including mammalian mtDNAs initiates within or close to regions with secondary structure

potential (31-37). A variety of secondary structures can be formed within the remainder of the A+T-rich region, mainly as the result of runs of As (up to 13, nucleotides 438-450; Fig. 2), runs of Ts (up to 19, nucleotides 29-47; Fig. 2) and runs of alternating As and Ts (up to 18, nucleotides 921-938; Fig. 2).

ACKNOWLEDGEMENTS

We wish to thank Felix Machuca for excellent technical assistance. This work was supported by National Institutes of Health Grant No. GM 18375.

*Present address: Department of Biochemistry, Stanford School of Medicine, Stanford, CA 94305, U.S.A.

REFERENCES

1. Clary,D.O., Goddard,J.M., Martin,S.C., Fauron,C.M.-R. and Wolstenholme,D.R. (1982) Nucl. Acids Res. 10, 6619-6637.
2. Clary,D.O., Wahleithner,J.A. and Wolstenholme,D.R. (1983) Nucl. Acids Res. 11, 2411-2425.
3. Clary,D.O. and Wolstenholme,D.R. (1983a) Nucl. Acids Res. 11, 4211-4227.
4. Clary,D.O. and Wolstenholme,D.R. (1983b) Nucl. Acids Res. 11, 6859-6872.
5. Clary,D.O. and Wolstenholme,D.R. (1984) Nucl. Acids Res. 12, 2367-2379.
6. Clary,D.O., Wahleithner,J.A. and Wolstenholme,D.R. (1984) Nucl. Acid Res. 12, 3747-3761.
7. 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.
8. Bibb,M.J., Van Etten,R.A., Wright,C.T., Walberg,M.W. and Clayton,D.A. (1981) Cell 26, 167-180.
9. Anderson,S., de Bruijn,M.H.L., Coulson,A.R., Eperon,I.C., Sanger,F. and Young,I.G. (1982) J. Mol. Biol. 156, 683-717.
10. Grosskopf,R. and Feldmann,H. (1981) Curr. Genet. 4, 151-158.
11. Koike,K., Kobayashi,M., Yaginuma,K., Taira,M., Yoshida,E. and Imai,M. (1982) Gene 20, 175-183.
12. Pepe,G., Holtrop,M., Gadaleta,G., Kroon,A.M., Cantatore,P., Gallerani,R., De Benedetto,D., Quagliariello,C., Sbisà,E. and Saccone,C. (1983) Biochem. Int. 6, 553-563.
13. Zweib,C., Glotz,C. and Brimacombe,R. (1981) Nucl. Acids Res. 9, 3621-3640.
14. Glotz,C., Zweib,C. and Brimacombe,R. (1981) Nucl. Acids Res. 9, 3287-3306.
15. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
16. Fauron,C.M.-R. and Wolstenholme,D.R. (1976) Nucl. Acids Res. 8, 2439-2452.
17. Feramisco,J.R., Smart,J.E., Burridge,K., Helfman,D.M. and Thomas,G.P. (1982) J. Biol. Chem. 257, 11024-11031.
18. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory.
19. Berk,A.J. and Sharp,P.A. (1977) Cell 12, 721-732.
20. Fauron,C.M.-R. and Wolstenholme,D.R. (1980) Proc. Natl. Acad. Sci. USA 73, 3623-3627.
21. Goddard,J.M. and Wolstenholme,D.R. (1980) Nucl. Acids Res. 8, 741-757.
22. Dubin,D.T., HsuChen,C.C., Timko,K.D., Azzolina,T.M., Pringle,D.L., and Ranzini,J.L. (1982) in Mitochondrial Genes, Slonimski,P., Borst,P. and

- Attardi, G., eds., pp. 89-98, Cold Spring Harbor Laboratory.
23. HsuChen, C.-C., Kotin, R.M. and Dubin, D.T. (1984) Nucl. Acids Res. 12, 7771-7785.
24. Klukas, C.K. and Dawid, I.B. (1976) Cell 9, 615-625.
25. Woese, C.R., Gutell, R., Gupta, R. and Noller, H. (1983) Microbiol. Rev. 47, 621-669.
26. Ware, V.C., Tague, B.W., Clark, C.G., Gourse, R.L., Brand, R.C. and Gerbi, S.A. (1983) Nucl. Acids Res. 11, 7795-7815.
27. Chan, Y.-L., Olvera, J. and Wool, I. (1984) Nucl. Acids Res. 12, 7819-7831.
28. Gray, M.W., Sankoff, D. and Cedergren, R.J. (1984) Nucl. Acids Res. 12, 5837-5852.
29. Noller, H.F. (1984) Ann. Rev. Biochem. 53, 119-162.
30. Clark, C.G., Tague, B.W., Ware, V.C. and Gerbi, S.A. (1984) Nucl. Acids Res. 12, 6197-6220.
31. Schaller, H. (1978) Cold Spring Harbor Symp. Quant. Biology 43, 401-408.
32. Hobom, G., Grosschedl, R., Lusky, M., Scherer, G., Schwarz, E. and Kossel, H. (1978) Cold Spring Harbor Symp. Quant. Biol. 43, 165-178.
33. Messer, W., Mejer, M., Bergmans, H.E.N., Hansen, F.G., von Meyenburg, K., Beck, E. and Schaller, H. (1978) Cold Spring Harbor Symp. Quant. Biol. 43, 139-145.
34. Tijan, R. (1978) Cold Spring Harbor Symp. Quant. Biol. 43, 655-662.
35. Attardi, G., Crews, S., Nishiguchi, J., Ojala, D. and Posakony, J. (1978) Cold Spring Harbor Symp. Quant. Biol. 43, 179-192.
36. Martens, P.A. and Clayton, D.A. (1979) J. Mol. Biol. 136, 327-351.
37. Bernardi, G., Baldacci, G., Bernardi, G., Faugeron-Fonty, G., Gaillard, C., Goursot, R., Huyard, A., Mangin, M., Marotta, R. and de Zamaroczy, M. (1980) in Organization and Expression of the Mitochondrial Genome, Developments in Genetics - pp. 241-250, Saccone, C. and Kroon, A.M. eds., Vol. 2, Elsevier/North Holland Biomedical Press, Amsterdam.