Volume 11 Number 2 1983 Nucleic Acids Research

Complete DNA sequence coding for the large ribosomal RNA of yeast mitochondria

Frédéric Sor and Hiroshi Fukuhara*

Institut Curie, Section de Biologie, Bâtiment 110, Centre Universitaire, 91405 Orsay, France

Received 5 October 1982; Revised and Accepted 20 December 1982

ABSTRACT

The mitochondrial gene coding for the large ribosomal RNA (21S) has been isolated from a rho⁻ clone of Saccharomyces cerevisiae. A DNA segment of about 5500 base pairs has been sequenced which included the totality of the sequence coding for the mature ribosomal RNA and the intron. The mature RNA sequence corresponds to a length of 3273 nucleotides. Despite the very low guanine-cytosine content (20.5 %), many stretches of sequence are homologous to the corresponding Escherichia coli 23S ribosomal RNA. The sequence can be folded into a secondary structure according to the general models for prokaryotic and eukaryotic large ribosomal RNAs. Like the E.coli gene, the mitochondrial gene contains the sequences that look like the eukaryotic 5.8S and the chloroplastic 4.5S ribosomal RNAs. The 5' and 3' end regions show a complementarity over fourteen nucleotides.

INTRODUCTION

In Saccharomyces cerevisiae, many mutations are known which affect the mitochondrial ribosomes. Most of these mutations are mapped within or near the ribosomal RNA genes of the mitochondrial DNA (mtDNA). Mutations conferring resistance to antibacterial drugs (see ref. 1), as well as temperature-sensitive mutations (2) have been described, and extensive genetic studies have been made using these mutations. The knowledge of ribosomal gene sequences and of their flanking regions is therefore a considerable help to an understanding of the genetic properties of yeast mitochondria which are often very unusual (recombination, deletion, inversion, and so forth). Also, the structure of mitochondrial ribosomal RNA itself is of particular interest for the study of the ribosome, since the availability of the specific mutations can help to define the functional domains of the ribosomal architecture. These reasons prompted us to analyze the sequences of the rRNA genes. The primary structure of the small ribosomal RNA (15S) of yeast mitochondria has been described previously (3,4). The present report describes the gene for the large ribosomal RNA (21S rRNA).

0305-1048/83/1102-0339$2.00/0

MATERIALS AND METHODS

Source of the 21S rRNA gene.

The rho$^-$ mutant, F11, derived from the wild type strain IL8-8C (α trp$_1$ his$_1$; originally from the Centre de Génétique Moléculaire du CNRS, Gif) has been used for the isolation of the gene. IL8-8C carries a chloramphenicol resistance mutation, C^R_{321}, and an erythromycin resistance mutation, E^R_{514}. It is also a ω^+ strain, that is, it contains the intron ω in the 21S rRNA gene. The rho$^-$ strain F11 has retained all of these markers. Restriction analysis of F11 mtDNA and hybridization experiments with 21S rRNA (5) have shown that the mutant has a mtDNA sequence of 12000 base-pairs (bp) (repeated in tandem ; cf. ref. 6) and contains the totality of the 21S rRNA transcription unit, but deleted for 84 % of the wild type mtDNA (11). A few other rho$^-$ mutants (E1, E2 and E3) derived from the same wild type strain have been also used for confirmation of sequences. Their mtDNAs contained parts of the 21S rRNA gene (see Figure 1).

Isolation of mtDNA, restriction analysis and DNA sequencing.

Mitochondria were purified from protoplasts and treated with pancreatic DNase I to eliminate nuclear DNA contamination. MtDNA was purified by CsCl isopiknic centrifugation. These procedures have been described previously by Casey et al. (7). Restriction analysis of mtDNA was performed as detailed by Wésolowski et al. (8). DNA sequences were determined according to the chemical method of Maxam and Gilbert (9) using either 5' or 3' end-labelling. S1 nuclease mapping experiments were performed under the conditions described previously (10).

RESULTS AND DISCUSSION

Restriction map of the region coding for the 21S rRNA.

The mtDNA of the mutant F11 is a typical tandemly repeated rho$^-$ genome. No abnormal rearrangements of sequence was found with respect to the wild type map of this region (presence of all restriction fragments identical to the wild type, except for the junction-fragments of the repeating units). Single and double digestions using various combinations of restriction enzymes allowed us to establish the map shown in Figure 1A. All the restriction sites used for sequencing have been sequenced across from other sites (Figure 1B).

Sequence of the 21S rRNA gene.

Figure 2 shows the nucleotide sequence of the 21S rRNA gene and its flanking regions. The sequence corresponds to the area specified in Figure 1. The sequence corresponding to the mature RNA is 3273 bp long (1 to 2688 and 3832 to 4416), split by the intron ω (2689 to 3831). The guanine-cytosine content is 20.5 %. There are many long sequences of pure adenine-thymine bases, as in the case of the mitochondrial 15S rRNA gene.

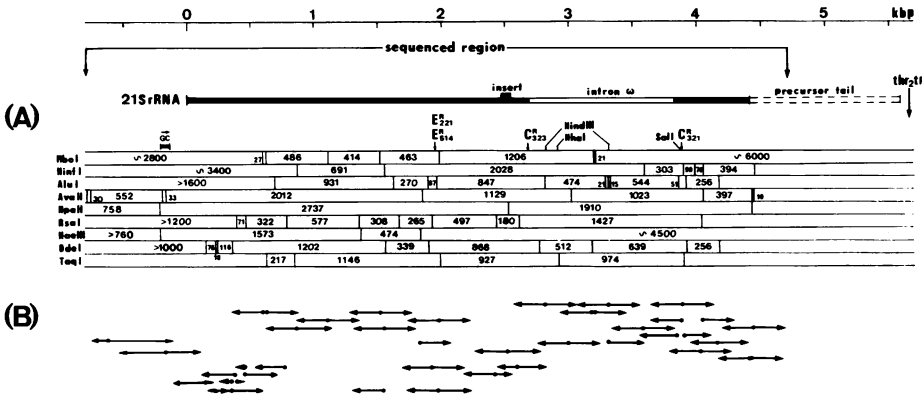


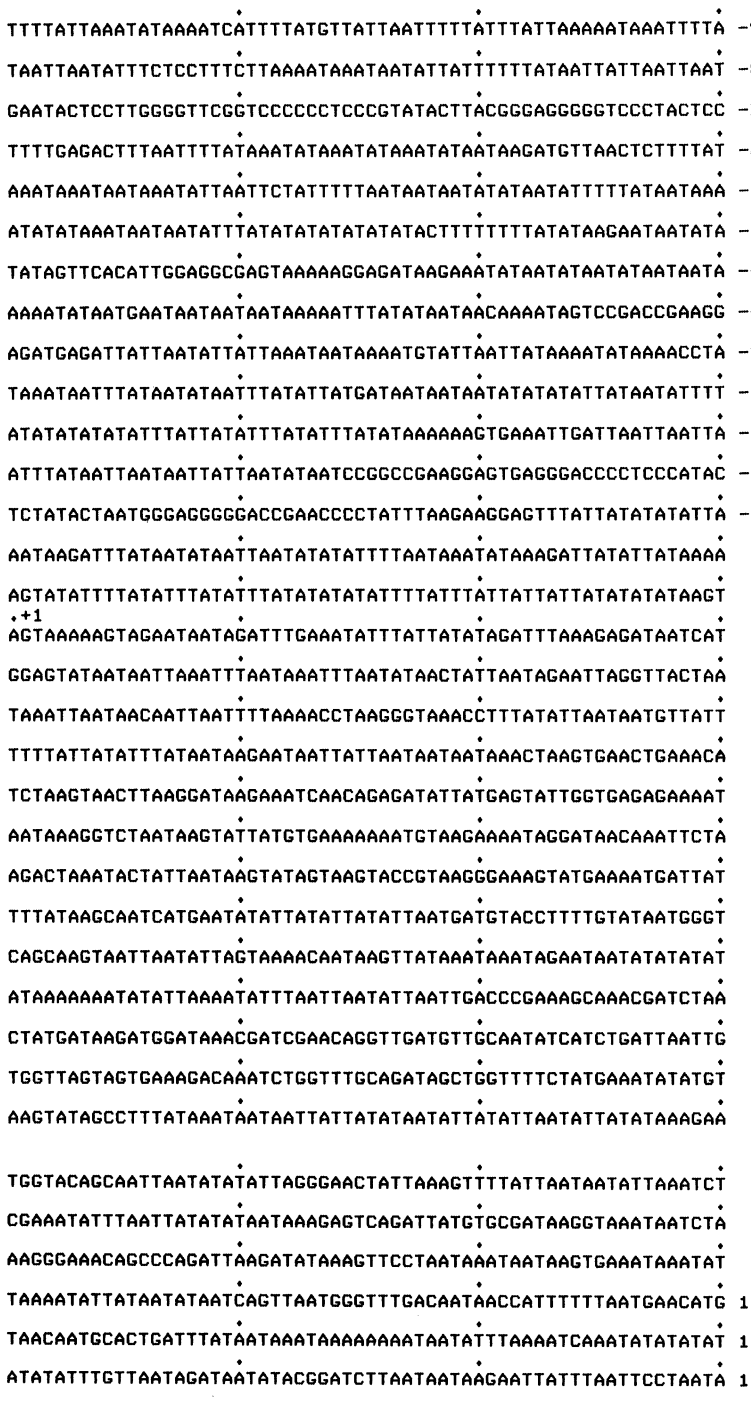
Figure 1 : (A) Restriction map of the sequenced region. The top line shows the mature 21S rRNA. Thick lines indicate the localization of the rho⁻ mtDNAs used for the rRNA gene sequencing. Lengths of fragments are in base-pairs. The wavy line indicates the intron. gc means a GC cluster with the indicated orientation (see text). E^R_{514}, E^R_{221}, C^R_{323} and C^R_{321} are drug resistance genetic markers. ω is the intron. (B) Sequencing scheme. Each arrow indicates the position of labelling (5' or 3') and the length of sequence read.

5' region of the gene.

In yeast mitochondria, the large and small rRNA belong to two separate transcription units (23,24). Levens et al. (11) have shown, by guanylyl transferase capping reaction, that the initiating nucleotide of the primary transcript of 21S rRNA is also the first nucleotide of the mature 21S rRNA. They have also determined the first nucleotides of the RNA : 5' AGU(UUUU)AGUAGAAUAAUAGA 3' (bases in parentheses are uncertain according to the authors). S1 mapping experiments of Osinga and Tabak (12) clearly identified this sequence on mtDNA as AGTAAAAAGTAGAATAATAGA. We have confirmed this sequence and its location. These authors also sequenced about 200 bp flanking the 5' end of the gene and suggested a possible role of the sequence -1 to -9 in transcription. Far upstream, we have found a "GC clusters" (at -215). Such a cluster is known to be present in many copies in the yeast mtDNA, but their role is not known (see ref. 10).

3' region of the gene.

Merten et al. (13) have shown that the primary transcript of the 21S rRNA gene is a large molecule of 5.1-5.4 kilobases (electron microscopy), which is subsequently processed by (i) excision of the 1.1 kb intron and (ii) excision of the 3' extension of about 1.2 kb. Also Bos et al. (14) have estimated by S1 mapping experiments that the 3' end of the mature RNA should lie about 570 bases downstream from the end of the intron. Our S1 mapping experiment (Figure 3) shows that the end of the mature RNA


TTTTATTAAATATAAAATCATTTTATGTTATTAATTTTTATTTATTAAAAATAAATTTTA -901
TAATTAATATTTCTCCTTTCTTAAAATAAATAATATTATTTTTTATAATTATTAATTAAT -841
GAATACTCCTTGGGGTTCGGTCCCCCCTCCCGTATACTTACGGGAGGGGGTCCCTACTCC -721
TTTTGAGACTTTAATTTTATAAATATAAATATAAATATAATAAGATGTTAACTCTTTTAT -661
AAATAAATAATAAATATTAATTCTATTTTTAATAATAATATATAATATTTTTATAATAAA -601
ATATATAAATAATAATATTTATATATATATATATACTTTTTTTTTATATAAGAATAATATA -541
TATAGTTCACATTGGAGGCGAGTAAAAAGGAGATAAGAAATATAATATAATATAATAATA -481
AAAATATAATGAATAATAATAATAAAAATTTATATAATAACAAAATAGTCCGACCGAAGG -421
AGATGAGATTATTAATATTATTAAATAATAAAATGTATTAATTATAAAATATAAAACCTA -361
TAAATAATTTATAATATAATTTATATTATGATAATAATAATATATATTATAATATTTT -301
ATATATATATATTTATTATATTTATATTTATATAAAAAAGTGAAATTGATTAATTAATTA -241
ATTTATAATTAATAATTATTAATATAATCCGGCCGAAGGAGTGAGGGACCCCTCCCATAC -181
TCTATACTAATGGGAGGGGGACCGAACCCCTATTTAAGAAGGAGTTTATTATATATATTA -121
AATAAGATTTATAATATAATTAATATATATTTTAATAAATATAAAGATTATATTATAAAA -61
AGTATATTTTATATTTATATTTATATATATATTTTATTTATTATTATTATATATATAAGT -1
.+1
AGTAAAAAGTAGAATAATAGATTTGAAATATTTATTATATAGATTTAAAGAGATAATCAT 60
GGAGTATAATAATTAAATTTAATAAATTTAATATAACTATTAATAGAATTAGGTTACTAA 120
TAAATTAATAACAATTAATTTTAAAACCTAAGGGTAAACCTTTATATTAATAATGTTATT 180
TTTTATTATATTTATAATAAGAATAATTATTAATAATAATAAACTAAGTGAACTGAAACA 240
TCTAAGTAACTTAAGGATAAGAAATCAACAGAGATATTATGAGTATTGGTGAGAGAAAAT 300
AATAAAGGTCTAATAAGTATTATGTGAAAAAAATGTAAGAAAATAGGATAACAAATTCTA 360
AGACTAAATACTATTAATAAGTATAGTAAGTACCGTAAGGGAAAGTATGAAAATGATTAT 420
TTTATAAGCAATCATGAATATATTATATTATATTAATGATGTACCTTTTGTATAATGGGT 480
CAGCAAGTAATTAATATTAGTAAAACAATAAGTTATAAATAAATAGAATAATATATATAT 540
ATAAAAAAATATATTAAAATATTTAATTAATATTAATTGACCCGAAAGCAAACGATCTAA 600
CTATGATAAGATGGATAAACGATCGAACAGGTTGATGTTGCAATATCATCTGATTAATTG 660
TGGTTAGTAGTGAAAGACAAATCTGGTTTGCAGATAGCTGGTTTTCTATGAAATATATGT 720
AAGTATAGCCTTTATAAATAATAATTATTATATAATATTATATTAATATTATATAAAGAA 780
TGGTACAGCAATTAATATATATTAGGGAACTATTAAAGTTTTATTAATAATATTAAATCT 840
CGAAATATTTAATTATATATAATAAAGAGTCAGATTATGTGCGATAAGGTAAATAATCTA 900
AAGGGAAACAGCCCAGATTAAGATATAAAGTTCCTAATAAATAATAAGTGAAATAAATAT 960
TAAAATATTATAATATAATCAGTTAATGGGTTTGACAATAACCATTTTTTAATGAACATG 1020
TAACAATGCACTGATTTATAATAAATAAAAAAAATAATATTTAAAATCAAATATATATAT 1080
ATATATTTGTTAATAGATAATATACGGATCTTAATAATAAGAATTATTTAATTCCTAATA 1140

TGGAATATTATATTTTTATAATAAAAATATAAATACTGAATATCTAAATATTATTATTAC 1200

TTTTTTTTTTAATAATAATAATATGGTAATAGAACATTTAATGATAATATATATTAGTTA 1260

TTAATTAATATATGTATTAATTAAATAGAGAATGCTGACATGAGTAACGAAAAAAAGGTA 1320

TAAACCTTTTCACCTAAAACATAAGGTTTAACTATAAAAGTACGGCCCCTAATTAAATTA 1380

ATAAGAATATAAATATATTTAAGATGGGATAATCTATATTAATAAAAATTTATCTTAAAA 1440

TATATATATTATTAATAATTATATTAATAATATATATAATTATATTATATATTTTTTTTT 1500

ATATAATATAAACTAATAAAGATCAGGAAATAATTAATGTATACCGTAATGTAGACCGAC 1560

TCAGGTATGTAAGTAGAGAATATGAAGGTGAATTAGATAATTAAAGGGAAGGAACTCGGC 1620

AAAGATAGCTCATAAGTTAGTCAATAAAGAGTAATAAGAACAAAGTTGTACAACTGTTTA 1680

CTAAAAACACCGCACTTTGCAGAAACGATAAGTTTAAGTATAAGGTGTGAACTCTGCTCC 1740

ATGCTTAATATATAAATAAAATTATTTAACGATAATTTAATTAAATTTAGGTAAATAGCA 1800

GCCTTATTATGAGGGTTATAATGTAGCGAAATTCCTTGGCCTATAATTGAGGTCCCGCAT 1860

GAATGACGTAATGATACAACAACTGTCTCCCCTTTAAGCTAAGTGAAATTGAAATCGTAG 1920

TGAAGATGCTATGTACCTTCAGCAAGACGGGAAGACCCTATGCAGCTTTACTGTAATTAG 1980

ATAGATCGAATTATTGTTTATTATATTCAGCATATTAAGTAATCCTATTATTAGGTAATC 2040

GTTTAGATATTAATGAGATACTTATTATAATATAATGATAATTCTAATCTTATAAATAAT 2100

TATTATTATTATTATTAATAATAATAATATGCTTTTCAAGCATAGTGATAAAACATATTT 2160

ATATGATAATCACTTTACTTAATAGATATAATTCTTAAGTAATATATAATATATATTTTA 2220

TATATATTATATATAATATAAGAGACAATCTCAATTGGTAGTTTTGATGGGGCGTCATTA 2280

TCAGCAAAAGTATCTGAATAAGTCCATAAATAAATATATAAAATTATTGAATAAAAAAAT 2340

AATATATATTATATATATTAATTATAAATTGAAATATGTTTATATAAATTTATATTTATT 2400

GAATATATTTTAGTAATAGATAAAAATATGTACAGTAAAATTGTAAGGAAAACAATAATA 2460

ACTTTCTCCTCTCTCGGTGGGGGTTCACACCTATTTTTAATAGGTGTGAACCCCTCTTCG 2520

GGGTTCCGGAACTTAAATAAAAATGGAAAGAATTAAATTAATATAATGGTATAACTGTGC 2580

GATAATTGTAACACAAACGAGTGAAACAAGTACGTAAGTATGGCATAATGAACAAATAAC 2640

ACTGATTGTAAAGGTTATTGATAACGAATAAAAGTTACGCTAGGGATAatttacccccctt 2700

gtcccattatattgaaaaatataattattcaattaattatttaattgaagtaaattgggt 2760

gaattgcttagatatccatatagataaaaataatggacaataagcagcgaagcttataac 2820

aactttcatatatgtatatatacggttataagaacgttcaacgactagatgatgaagtgga 2880

gttaacaataattcatccacgagcgcccaatgtcgaataaataaaatattaaataaatat 2940

caaaggatatataaagatttttaataaatcaaaaaataaaataaaatgaaaaatattaaa 3000

aaaaatcaagtaataaatttaggacctaattctaaattattaaaagaatataaatcataa 3060

ttaattgaattaaatattgaacaatttgaagcaggtattggtttaattttaggagatgct 3120

tatattcgtagtcgtgatgaaggtaaactatattgtatgcaatttgagtgaaaaaataag 3180

Figure 2 : Sequence of the 21S rRNA gene and its flanking regions. The non coding (RNA-like) strand is written from 5' to 3' end. Position +1 corresponds to the first nucleotide of the mature rRNA (which is also the first nucleotide of the primary transcript). Position 4416 is the last nucleotide of the mature rRNA, as determined by S1 mapping (see Figure 4), with $\pm$ 1 nucleotide uncertainty. In the precursor RNA the 3' end extends further 1.2 kb. The intron is from 2689 to 3831 (lower case letters). The non-excised 66 bp mini-insert is from 2466 to 2532. Two additional A's are present between 2534 and 2535 in ω^- strains which lack the intron and the mini-insert. E^R_{514}, C^R_{323} and C^R_{321} mutations are at 1951, 2685 and 3884 respectively (see Figure 5).

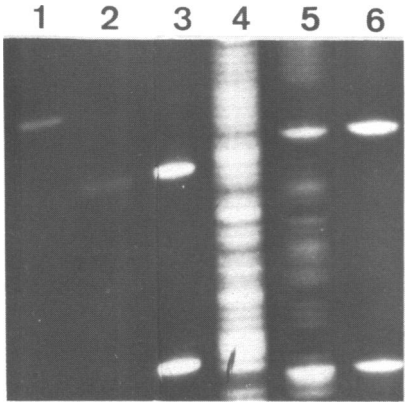


Figure 3 : Determination of the 3' end of the gene sequence coding for the mature 21S rRNA. The coding strand of an AvaII fragment (397 bp, see Figure 1A) was labelled at its 3' end with ^{32}P-α-dCTP, and annealed to the mature 21S rRNA. The hybrid was digested with S1 nuclease and electrophoresed on a denaturing 5 % polyacrylamide gel. Lane 1 : undigested AvaII fragment ; lane 2 : DNA from the digested hybrid ; lanes 3-6 : a sequencing gel used as reference ladders. Note that the AvaII fragment was shortened by 19 nucleotides after S1 nuclease treatment.

corresponds to the position 4416 (possibly ± 1 base) of the gene. This 3' end does not seem to have any detectable heterogeneity. Immediately before the 3' end, we find a 14 bp sequence (4397 to 4410) which is complementary to the beginning of the gene (10 to 23). Such a head-tail complementarity is known to exist in the large rRNA of bacteria.

The 3' end is followed by a sequence very rich in GC (4422 to 4478). This sequence diverges much from the typical GC clusters mentioned above. It shows a possible secondary structure shown in Figure 4, which may have a role in the splicing of

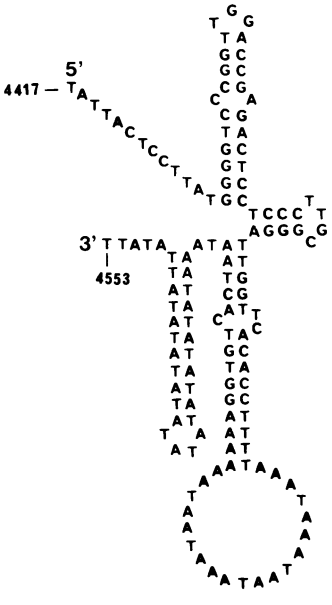
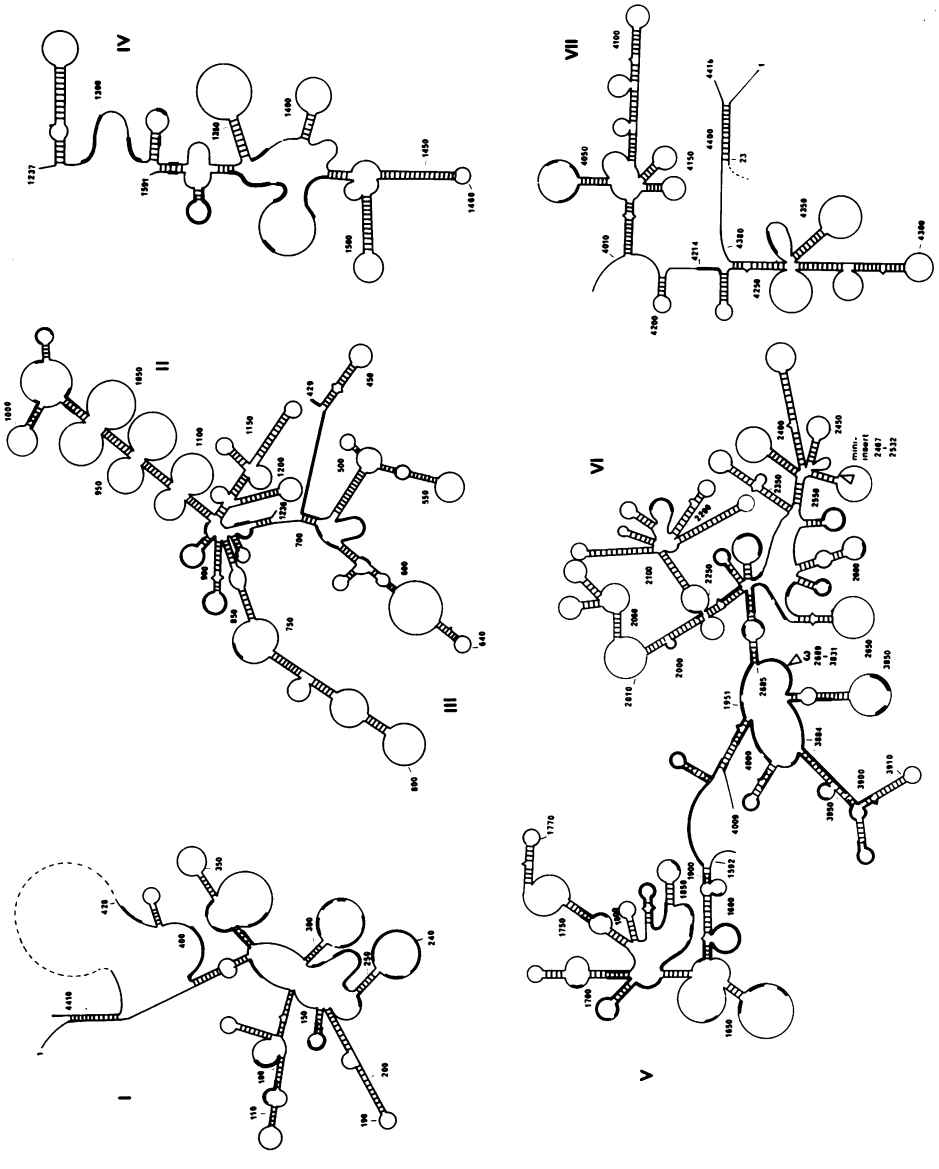


Figure 4 : A potential secondary structure of the sequence which follows the 3' end of the mature RNA sequence.


I
II
III
IV
V
VI
VII

the precursor RNA at this position.

Intron and insertion.

The region 2689 to 3831 corresponds to the intron ω. This segment is present only in the strains called ω^+. The intron and its flanking sequences have been previously studied by Dujon (15), who used a rho^- strain (IL8-8C/R53) derived from the same original strain as ours. The intron sequences obtained with these two isomitochondrial strains completely agree. About 160 bp ahead of the intron there is a 66 bp mini-insert (2466 to 2532) described by the same author. This insert is absent in ω^- strains (IL16-11D used by Dujon and FF1210-6C examined by us, data not shown). Both ω^- strains have however two additional A's between 2534 and 2535. The mini-insert is transcribed into the mature RNA without excision (S1 mapping experiments, not shown). Such a non-excised short insert has been also found in the mitochondrial 15S rRNA gene (10).

Secondary structure of the 21S rRNA.

The gene sequence can be folded into a secondary structure according to the general models (16,17,18) proposed for the large rRNAs of various organisms. Figure 5 shows a tentative folded structure, drawn in the style of Branlant et al. (17). Details of base pairing in this figure have many other possibilities, but long distance interactions first proposed for bacterial RNAs can be remarkably reproduced with the mitochondrial sequence, despite its unusually low guanine-cytosine content. Conservation of the paired regions by compensating variations seems convincing in this case. The mitochondrial large rRNA gene looks much like E.coli 23S rRNA gene, since the 5' region of the mitochondrial gene contains a sequence (from the 5' end to about 160) reminescent of the 5.8S rRNA of eukaryotic ribosomes (cf. 20). Also a sequence which may be related to the chloroplast 4.5S rRNA is found in the 3' region (4216 to the 3' end) (cf. 18,21,22). The mtDNA of Aspergillus nidulans, a filamentous fungus with a much smaller mitochondrial genome, appears to lack the 5.8S and 4.5S rRNA counterparts (19).

Recently, we have identified the erythromycin mutation E^R_{514} at the position 1951 (submitted for publication). At 734 and 790 bp downstream (excluding the intron) there are two chloramphenicol mutations, C^R_{323} and C^R_{321}, previously identified by Dujon (15). In the secondary structure model (Figure 5, domain V-VI) all these mutations are found together within a small loop, suggesting a specific role of this region in the peptide elongation cycle. This region is one of the best conserved sequences of the ribosomal RNA genes in those organisms that have been studied so far.

Figure 5 : A possible secondary structure of the mitochondrial large rRNA, as deduced from the gene sequence. The sequence was folded into a tentative secondary structure according to Branlant et al.'s model (17) ; it is divided into seven domains (I to VII). Details of pairing remain to be worked out in comparison with other rRNA structures. Thick lines indicate the sequences homologous with the E.coli 23S rRNA.

ACKNOWLEDGEMENTS

We received financial supports from the CNRS (ATP 3644) and from the DGRST (81.E.1206).

REFERENCES

1. Dujon, B., Colson, A.M. and Slonimski, P. (1977) in Mitochondria 1977 : Genetics and Biogenesis of Mitochondria, Bandlow, W. et al., Eds., pp. 579-669, W. de Gruyter, Berlin.
2. Bolotin-Fukuhara, M., Faye, G. and Fukuhara, H. (1977) Molec. Gen. Genet. 152, 295-305.
3. Sor, F. and Fukuhara, H. (1980) Compt. Rend. Acad. Sci. (Paris) 291, (série D), 933-936.
4. Sor, F. and Fukuhara, H. (1982) in Mitochondrial Genes, Slonimski, P.P. et al., Eds., pp. 255-262, Cold Spring Harbor Laboratory, Cold Spring Harbor.
5. Lewin, A., Morimoto, R., Rabinowitz, M. and Fukuhara, H. (1978) Molec. Gen. Genet. 163, 257-275.
6. Faye, G., Fukuhara, H., Grandchamp, C., Lazowska, J., Michel, F., Casey, J., Getz, G.S., Locker, J., Rabinowitz, M., Bolotin-Fukuhara, M., Coen, D., Deutsch, J., Dujon, B., Netter, P. and Slonimski, P.P. (1973) Biochimie 55, 779-792.
7. Casey, J., Hsu, H.J., Rabinowitz, M. and Fukuhara, H. (1974) J. Molec. Biol. 88, 717-733.
8. Wésolowski, M., Monnerot, M. and Fukuhara, H. (1980) Current Genetics 2, 121-129.
9. Maxam, A. and Gilbert, W. (1980) in Methods in Enzymology, Grossman, L. and Moldave, K., Eds., Vol. 65, pp. 499-559, Academic Press, New York.
10. Sor, F. and Fukuhara, H. (1982) Nuc. Acids Res. 10, 1625-1633.
11. Levens, D., Ticho, B., Ackerman, E. and Rabinowitz, M. (1981) J. Biol. Chem. 256, 5226-5232.
12. Osinga, K.A. and Tabak, H.F. (1982) Nuc. Acids Res. 10, 3617-3626.
13. Merten, S., Synenki, R.M., Locker, J., Christanson, T. and Rabinowitz, M. (1980) Proc. Natl. Acad. Sci. USA 77, 1417-1421.
14. Bos, J.L., Osinga, K.A., van der Horst, G., Hecht, N.B., Tabak, H.F., van Ommen, G.-J.B. and Borst, P. (1980) Cell 20, 207-214.
15. Dujon, B. (1980) Cell 20, 185-197.
16. Noller, H.F., Kop, J., Wheaton, V., Brosius, J., Gutell, R.R., Kopylov, A.M., Dohme, F., Herr, W., Stahl, D.A., Gupta, R. and Woese, C. (1981) Nuc. Acids Res. 9, 6167-6189.
17. Branlant, C., Krol, A., Machatt, M.A., Pouyet, J., Ebel, J.P., Edwards, K. and Kössel, H. (1981) Nuc. Acids Res. 9, 4303-4324.
18. Glotz, C., Zwieb, C., Brimacombe, R., Edwards, K. and Kössel, H. (1981) Nuc. Acids Res. 9, 3287-3306.
19. Köchel, H.G. and Küntzel, H. (1982) Nucl. Acids Res. 10, 4795-4801.
20. Nazar, R.N. (1980) FEBS Letters 119, 212-214.
21. Machatt, M.A., Ebel, J.P. and Branlant, C. (1981) Nuc. Acids Res. 9, 1533-1549.
22. MacKay, R.M. (1981) FEBS Letters 123, 17-18.
23. Faye, G., Kujawa, C. and Fukuhara, H. (1974) J. Molec. Biol. 88, 185-203.
24. Faye, G., Kujawa, C., Fukuhara, H. and Rabinowitz, M. (1976) Biochem. Biophys. Res. Commun. 68, 476-482.