Proc. Natl. Acad. Sci. USA
Vol. 76, No. 4, pp. 1663–1667, April 1979
Biochemistry

Nucleotide sequence of the mitochondrial structural gene for subunit 9 of yeast ATPase complex

(petite mutant DNA/dideoxy sequence method/A+T-rich sequences/codon usage/mtDNA)

LAMBERTUS A. M. HENSGENS, LESLIE A. GRIVELL, PIET BORST, AND JOHANNES L. BOS

Section for Medical Enzymology and Molecular Biology, Laboratory of Biochemistry, University of Amsterdam, Jan Swammerdam Institute, P. O. Box 60.000, 1005 GA Amsterdam, The Netherlands

Communicated by Severo Ochoa, January 22, 1979

ABSTRACT **We have determined the nucleotide sequence of a segment of *Saccharomyces* mtDNA that contains the structural gene for one of the subunits (the dicyclohexylcarbodiimide-binding protein) of the mitochondrial ATPase complex. The sequence fits the known amino acid sequence of this protein with the exception of one amino acid. Codon usage is biased in favor of A+T-rich codons. On both sides of the gene, the nucleotide sequence contains less than 4% (mol/mol) G+C for at least 180 nucleotides; these A+T sequences show no evidence of internal repetition. The gene and all the A+T-rich sequence preceding the gene are present in a 12S RNA that is the major transcript of this segment of mtDNA. The nature of the sequences responsible for binding ribosomes to mitochondrial mRNA and for termination of RNA synthesis is considered.**

The mtDNA of the yeast *Saccharomyces* is a 25-μm circle (1), which contains genes for rRNAs, tRNAs, and a limited number of proteins (see Fig. 1). Biochemical analysis of this DNA and its transcripts has uncovered several interesting features (see ref. 7 for review). The DNA contains only 18% (mol/mol) G+C (8), and about half of it is made up of segments that are nearly pure A+T (9). Intervening sequences have been found both in the gene for the large rRNA (10) and in the structural gene for apocytochrome *b* (7, 11, 12), and in both cases these intervening sequences have been shown to vary in size in closely related *Saccharomyces* strains (7).

To understand the biological significance of these features of yeast mtDNA, detailed information on strategic sequences in this DNA is indispensable. A suitable starting point for sequence analysis of protein-specifying genes is the *oli1* locus. This locus is known to code for subunit 9 (the so-called dicyclohexylcarbodiimide-binding protein) of the mitochondrial ATPase complex, and the amino acid sequence of this protein is known (13). Mutations at this locus may lead to oligomycin resistance of mitochondrial ATPase, and amino acid substitutions corresponding to these mutations have been determined (13, 14).

Slonimski and coworkers have isolated the petite mutant RP6, which can transmit oligomycin resistance in genetic crosses (15) and which contains a single amplified 1025-base pair fragment of wild-type mtDNA. This DNA has been used to map the *oli1* locus on the restriction fragment map of mtDNA (15), and we have now used RP6 as a convenient source of purified DNA to determine the nucleotide sequence of the structural gene for ATPase subunit 9 and the sequences surrounding this gene.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Yeast Strains and Isolation of mtDNA. The wild-type strain of *Saccharomyces cerevisiae* KL14-4A was obtained from R. J. Schweyen, Genetisches Institut, Munich University, West Germany. Petite mutant RP6 was obtained from P. P. Slonimski, Centre de Génétique Moléculaire du Centre National de la Recherche Scientifique (Gif-sur-Yvette, France). Retention of the *oli1* marker in this petite was checked by the replica-cross technique of Deutsch *et al.* (16) and found to be >99%

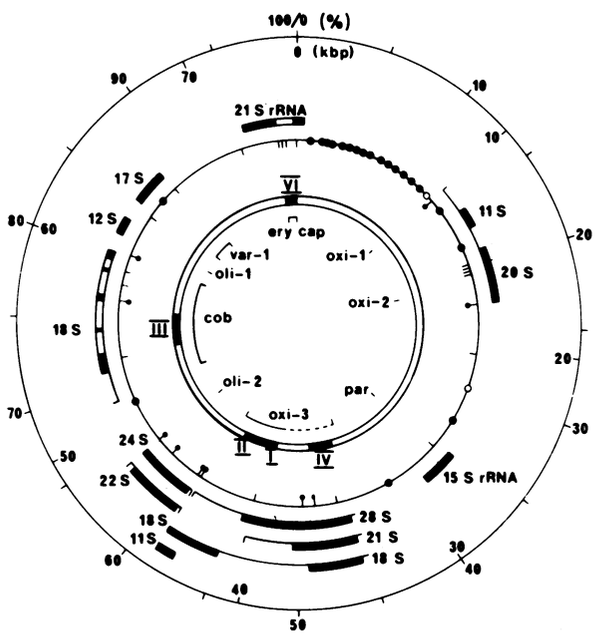

FIG. 1. Physical and genetic map of mtDNA from *Saccharomyces cerevisiae* strain KL14-4A. The genetic map in the inner ring is adapted from Schweyen *et al.* (2); *oxi1,2,3* are genes for subunits of cytochrome *c* oxidase, *cob* is the gene for apocytochrome *b*, *var1* is the gene for a ribosome-associated protein, and *ery, cap, par,* and *oli* are resistance determinants for erythromycin, chloramphenicol, paromomycin, and oligomycin, respectively. The black bars in the inner ring represent sequences present in KL14-4A mtDNA and absent in the mtDNA of *Saccharomyces carlsbergensis* at that position (3). The outer ring gives the position of recognition sites for endonucleases *Hin*dII + III (|) and *Eco*RI (⌈) and the approximate position of 4S RNA genes (4). The open circles are $tRNA_{met}$ genes (5). The approximate position of other transcripts is given outside the outer ring, the bars indicating the uncertainty in the exact position (ref. 6; unpublished experiments). The open sections in the 21S rRNA and in the 18S transcript of the *cob* region represent intervening sequences. Modified from ref. 7; kbp, kilobase pairs.

in the large-scale cultures used to isolate mtDNA by the procedure of Sanders *et al.* (17).

Isolation and Labeling of DNA Fragments. RP6 mtDNA was partially digested with endonuclease *Alu* I and end-labeled with [γ-^{32}P]ATP (The Radiochemical Centre, Amersham, England) and polynucleotide kinase as described by Birnstiel and coworkers (18). The fragments were purified by chromatography over Sephadex G-50, digested with endonuclease *Hap* II, and separated either in 1.5% agarose (19) or in 3.5% polyacrylamide slab gels (50 mM Tris citrate, pH 8.0/2.5 mM EDTA). Fragments were eluted from crushed gel segments by buffer extraction (20).

DNA Sequence Analysis. The sequences of DNA fragments labeled at one 5′ end were determined by the Maxam–Gilbert procedure (20) and by a modified chain-termination method (21) developed by Maat and Smith (22). In the latter method, the DNA fragments were incubated for 30 min at room temperature in 5 μl of a mixture containing 6.7 mM Tris·HCl (pH 7.6), 6.7 mM $MgCl_2$, 6.7 mM 2-mercaptoethanol, 50 mM NaCl, a specific dideoxynucleoside triphosphate (P-L Biochemicals) at 0.5 mM, the other three deoxynucleotide triphosphates at 0.02 mM, 25 pg of pancreatic DNase, and 1.5 unit of DNA polymerase I. Four incubations were run in parallel, and, in each, one of the dNTPs was replaced by the corresponding dideoxynucleoside triphosphate (in the T-terminated and G-terminated reactions) or a 500:1 mixture of dideoxynucleoside triphosphate and dNTP (A and C). The four reaction mixtures were analyzed on gels as in the Maxam–Gilbert procedure.

Enzymes. Restriction endonuclease *Alu* I was purchased from New England BioLabs; DNA polymerase I of *Escherichia coli* and pancreatic DNase (grade I) were from Boehringer Mannheim; and polynucleotide kinase from *E. coli* infected with phage T_4 was a gift from H. Van Ormondt. Endonuclease *Hap* II was prepared as described (23).

RESULTS

Colinearity of RP6 DNA with wild-type mtDNA

Although petite mutant mtDNAs consist of repeated segments of wild-type mtDNA, they often contain secondary rearrangements and deletions. In using such a mutant for sequence analysis, it is therefore essential to demonstrate that the secondary rearrangements are absent in the region of interest. This was done by three types of experiments: First, restriction enzyme analysis and electron microscopy of single-stranded DNA showed that the repeat units in RP6 mtDNA are in simple head-to-tail arrangement without inverted repeats. Second, Southern-type blotting experiments showed that the RP6 sequence hybridizes to the adjacent fragments of wild-type mtDNA (Fig. 2 *left*). By using the separated 340-base pair and 475-base pair *Hap* II and *Alu* I fragments (see Fig. 2 *right*) in these hybridization experiments, we also determined the orientation of RP6 on the wild-type mtDNA. Third, electron micrographs of hybrids between RP6 mtDNA and the 12S RNA transcribed from this region in wild-type strains (see Fig. 1) exhibited perfectly matched hybrids over most of the RP6 sequence, as indicated in Fig. 2 *left*. From these results, which will be presented in more detail elsewhere, we conclude that RP6 mtDNA is colinear with wild-type mtDNA in the region of interest.

Sequence analysis of the gene and its flanking regions

In preliminary experiments we found that the structural gene for subunit 9 of the ATPase complex lies over the three *Alu* I sites in the middle of the major *Hap* II fragment of RP6 mtDNA (see Fig. 2 *left*). To get end-labeled fragments suitable for sequence determination, RP6 mtDNA was partially digested with *Alu* I, and the fragments were labeled with polynucleotide kinase and [γ-^{32}P]ATP and then digested with *Hap* II. The resulting fragments labeled at one end (see Fig. 2 *right*) were separated by gel electrophoresis and used for sequence determinations. The sequence was determined from each of the three internal *Alu* sites proceeding towards the ends, and the overlapping sequences derived for the various *Alu* partial fragments were then aligned. Sequence analysis was done both by the Maxam–Gilbert method and by a modified chain-termination method using dideoxynucleoside triphosphates. The two methods often yielded complementary information and were used side by side.

Representative sequencing gels are shown in Fig. 3, and the

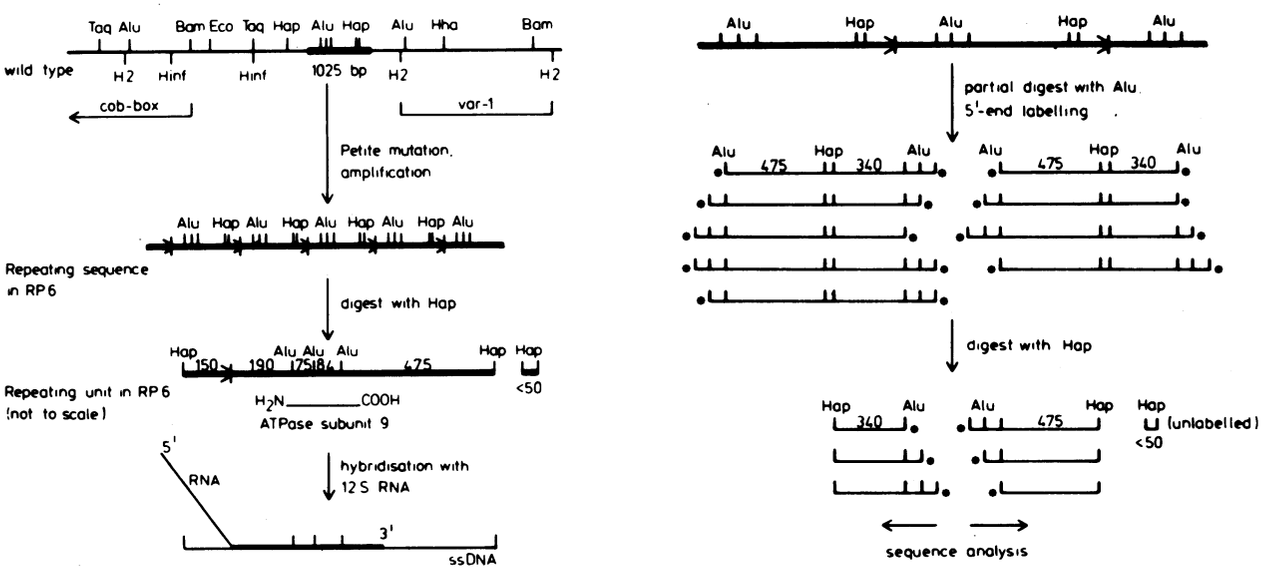

FIG. 2. Origin, structure, and transcription of petite RP6 mtDNA (*Left*) and partial digestion products used for nucleotide sequence analysis (*Right*). For reasons of clarity, only the *Hap* II sites in the immediate vicinity of the RP6 sequence on wild-type mtDNA are shown. The RP6 sequence itself is represented by a thickened bar. The limits of the repeating unit are indicated by >|. The lower part of *Left* shows a hybrid between RP6 mtDNA digested with *Hap* II and 12S RNA. See text for explanation. H2, recognition site for *Hin*dII. In *Right*, asterisks denote the presence of ^{32}P-labeled 5′ termini.

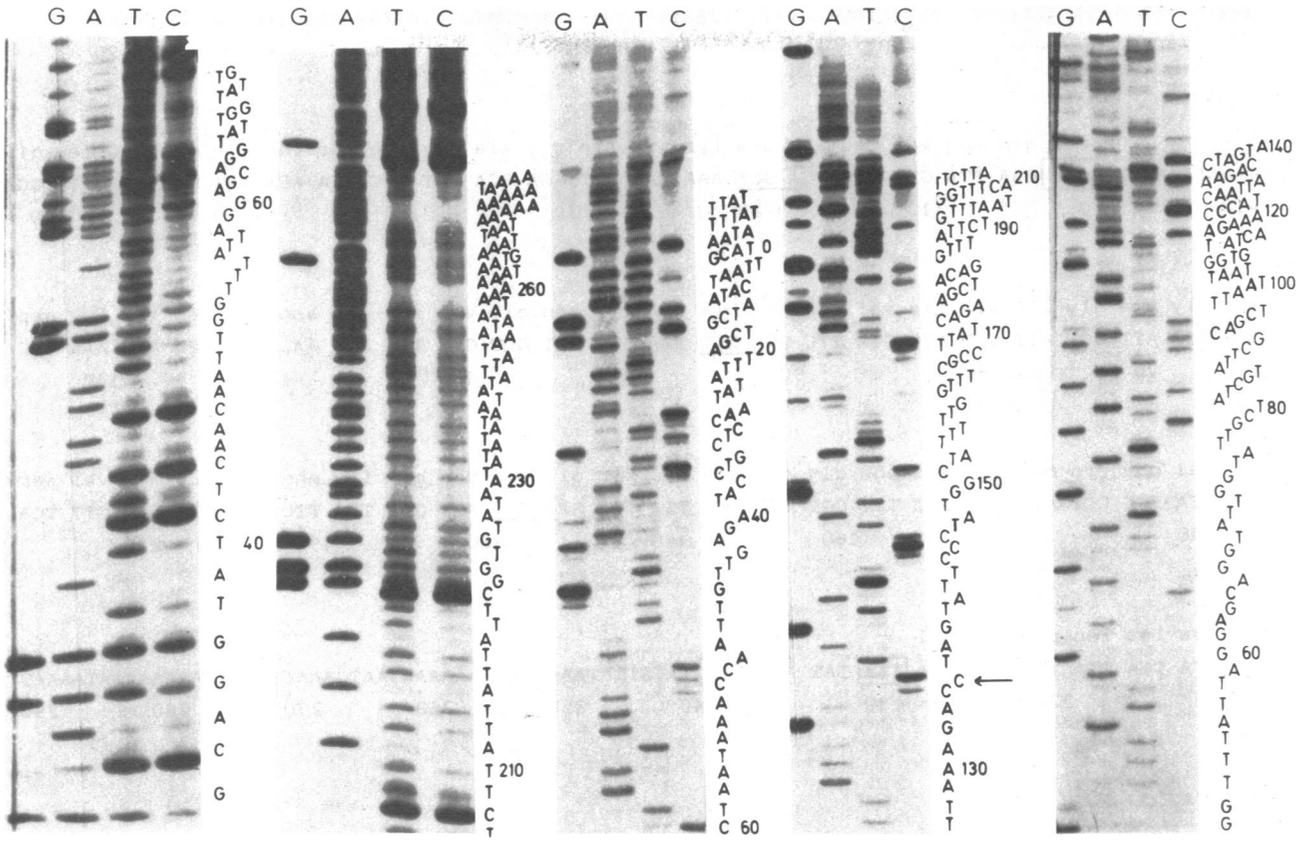

FIG. 3. Autoradiograms of DNA sequencing gels of fragments of RP6 mtDNA. The first two gels from left to right were obtained by using the Maxam–Gilbert technique (20), and the remainder were obtained by using the modified chain-termination method (ref. 21 and *Materials and Methods*). The arrow indicates the position where the derived amino acid sequence contains a Leu in contrast to the Thr in the published amino acid sequence (14). See text for further details.

sequence obtained is presented in Fig. 4. Comparison of our nucleotide sequence with the amino acid sequence determined by Sebald and Wachter (14) for the oligomycin-resistant version of the protein shows almost complete correspondence. The only discrepancy is at position 135–138, where Sebald and Wachter (14) find a threonine and we find a leucine. The DNA sequence at this position is unambiguous (see arrow in Fig. 3) and has been found in both strands. Since at least two point mutations are required to convert a threonine codon into a leucine codon, it is unlikely that we have analyzed a different variant of this gene. An error in the amino acid sequence is, therefore, the simplest explanation of the discrepancy, but other explanations are not excluded.

As a further check we have analyzed the sequence for recognition sites of known restriction enzymes. No additional *Alu* I site, besides the three known, is present, and the sequence contains no recognition site for any of the 23 enzymes that do not cut RP6 DNA.

The structural gene sequence is flanked by A+T-rich segments, which contain less than 4% G+C. On the 5′ side, the precise sequence is known up to nucleotide −108, but we have verified that Gs and Cs are rare up to nucleotide −250. At the 3′ end, the precise sequence is known down to nucleotide 300, but it remains A+T-rich down to at least nucleotide 400. We do not yet know exactly where the 12S RNA ends, but from the electron micrographs of DNA–RNA hybrids this is 70 nucleotides ± 100 from the end of the gene.

Codon usage

Codon usage is highly nonrandom, as shown in Fig. 5. Only 27 of the 61 amino acid-specifying codons are used, and the possibility that the mitochondrial translation system economizes on tRNA genes by not using all codons, although unlikely, can still not be excluded. Where a codon choice is possible, an A+T-rich codon is preferred. In the first codon position, this occurs 12 out of 13 times (expectation for random choice, 4 out of 13 times). In the third position, the preference for A or T is 63 out of 74 (expected, 39 out of 74).

If there is no selection for A+T-rich codons, the mole percent G+C of this gene would be 48. We find 33%, which is close to the minimum possible with this amino acid sequence (29%). This suggests that the low mole percent G+C of yeast mtDNA is not only due to the presence of noncoding A+T-rich stretches and the low G+C content of the rRNAs (23%; ref. 24) and tRNAs (35%; ref. 25), but that the selective pressure towards a low G+C content also operates on protein-specifying sequences.

Putative control elements and unusual structural features

We have scanned the region before the initiator ATG codon for a possible ribosome binding site. The most unusual structure present is the 14 nucleotide palindrome following the two lonely Gs at −31 and −26. Such palindromes are not rare, however, in very A+T-rich sequences, and its significance can only be assessed when the sequences flanking other protein-specifying genes in yeast mtDNA are known. Since the only transcript from this region is the 12S RNA that extends at least 400 nucleotides upstream of the gene (see Fig. 2 *left*), it is also possible that ribosomes bind to the 5′ end of this RNA and slide down to the first AUG.

There are no ribosome binding sites for bacterial ribosomes

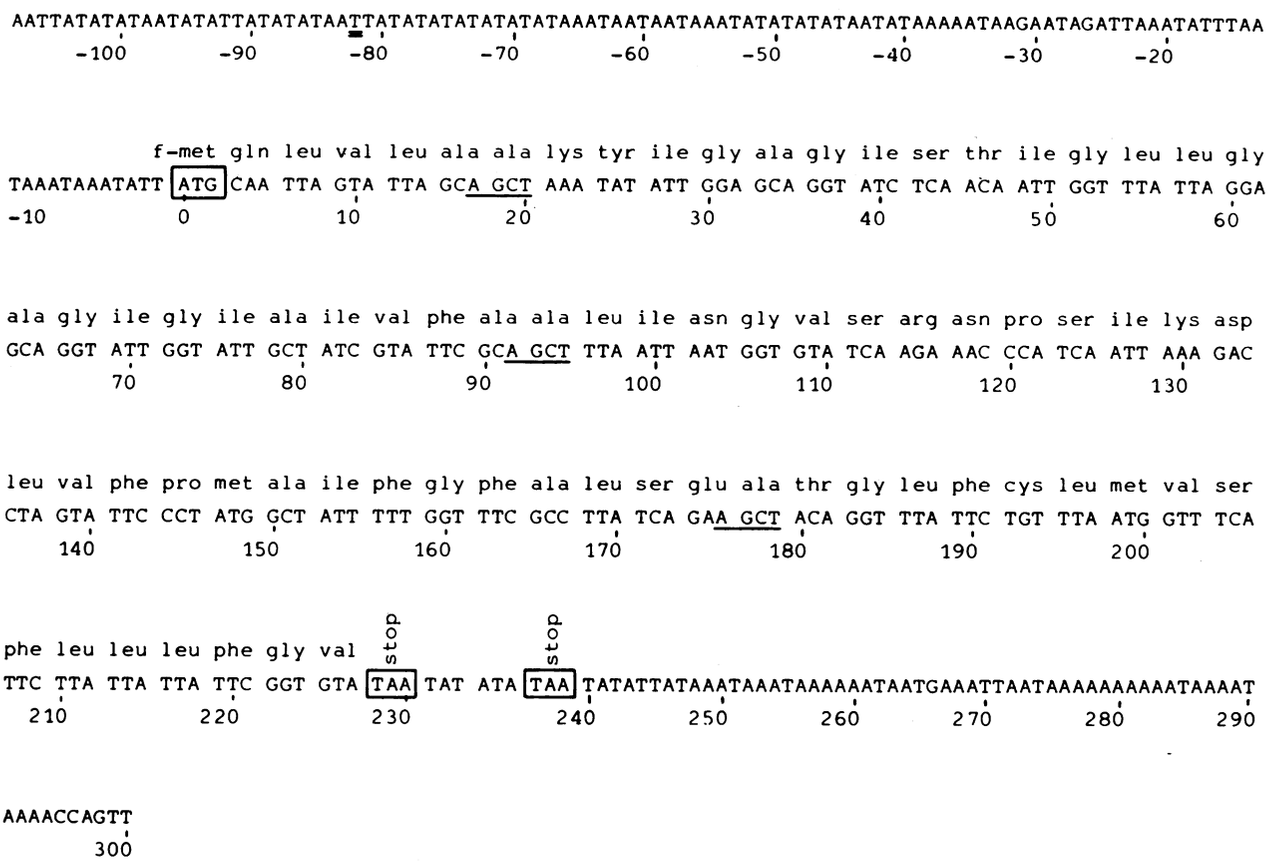

FIG. 4. Nucleotide sequence of one strand of the gene for subunit 9 of the ATPase complex and the derived amino acid sequence for this protein. Singly underlined sequences are the recognition sites for *Alu* I. The doubly underlined nucleotide is uncertain.

within 100 nucleotides upstream of the gene, although at several internal positions sequences occur that are complementary to a large portion of the Shine–Dalgarno sequence (ACCUCCUUA-OH) found at the end of bacterial 16S rRNA (26). Examples are at positions 29–33 (TGGAG) and 60–63 (GGAG). Such sites could account for the inability of bacterial ribosomes to effect the complete and correct synthesis of mitochondrial proteins (27).

The region following the gene contains several long $(dA)_n$ tracts up to $(dA)_{10}$, and the complementary strand used for 12S RNA synthesis, therefore, contains the same type of $(dT)_n$ tracts that appear to act as the termination signal for RNA synthesis, both in prokaryotic and eukaryotic systems (see refs. 28 and 29). The 12S RNA may terminate in this region (see Fig. 2 *left*), but we do not know exactly where.

1 \ 2	T	C	A	G	3
T	Phe 1 6 / Leu 11 0	Ser 0 0 5 0	Tyr 1 0 / Term 1 0	Cys 1 0 / Term 0 / Trp 0	T C A G
C	Leu 0 0 1 0	Pro 1 0 1 0	His 0 0 / Gln 1 0	Arg 0 0 0 0	T C A G
A	Ile 7 2 0 / Met 1•2	Thr 0 0 2 0	Asn 1 1 / Lys 2 0	Ser 0 0 / Arg 1 0	T C A G
G	Val 1 0 5 0	Ala 5 1 4 0	Asp 0 1 / Glu 1 0	Gly 8 0 2 0	T C A G

FIG. 5. Codons used for the synthesis of ATPase subunit 9.

DISCUSSION

Although the sequences of small segments of animal mtDNAs (30) and yeast mtDNA (31) have been determined before, Fig. 4 presents the sequence of a complete mitochondrial gene with its flanking sequences. This structural gene lacks an intervening sequence, which contrasts with the situation recently found for two other genes in this DNA. We have shown by RNA–DNA hybridization and by electron microscopy of hybrids that a 1160-base pair intervening sequence is present in the gene for the large rRNA (10) and confirmed the suspected (7, 11, 12) multiple intervening sequences in the 13S RNA that contains structural information for cytochrome *b* (A. C. Arnberg, L. A. Grivell, and G. J. B. Van Ommen, unpublished results). Whatever the function of intervening sequences in mitochondrial genes, they are obviously not essential.

Fig. 4 confirms some of the predictions of Bernardi and coworkers about the sequence organization of yeast mtDNA (9). The sequence contains a moderately G+C-rich gene between two very A+T-rich segments (cf. ref. 9), and these segments contain both oligo(dA·dT) and oligo[(dA-dT)·(dA-dT)] stretches (cf. ref. 32). Other predictions of the Bernardi model (9) are not borne out by our sequence, however: the gene is flanked neither by a G+T-rich cluster nor by a cluster of recognition sites for endonucleases *Hap* II and *Hae* III; the A+T-rich segments are transcribed; and they are not truly repetitive, in agreement with the results of DNA–DNA renaturation studies (1, 33) and the analysis of the mtDNA from another petite mutant that contains an A+T-rich repeat of only 68 nucleotides (31). Some repetition is obviously unavoidable in a random A+T sequence, and this may be sufficient for the faulty recombination thought to be the origin of the cytoplasmic petite mutants (9). The comparison of petite mtDNA sequences with the wild-type mtDNA from

which they are derived will soon provide precise information on the origin of petite mtDNA.

Finally, the excellent correspondence between nucleotide and amino acid sequence for this gene shows that (some) petite mutant DNAs contain precise copies of wild-type mtDNA segments, prepurified and faithfully replicated by nature. Although this had been inferred already from less precise analyses (see ref. 7), this is proved here directly at the nucleotide level.

Note Added in Proof. The sequence of a fragment of this gene (nucleotides 132–245) had been independently determined (34).

We are indebted to Dr. A. C. Arnberg, State University, Groningen, The Netherlands, for allowing us to quote the unpublished electron microscopical analyses of DNA–RNA hybrids; to Mr. J. Maat and Dr. H. Van Ormondt, State University, Leiden, The Netherlands, for helpful advice on sequence determination, communication of unpublished methods, and a gift of polynucleotide kinase; to Dr. W. Fiers, State University, Ghent, Belgium, for introducing us to sequence analysis of DNA; to Drs. A. Maxam and W. Gilbert, Harvard University, for a preprint of their sequencing method; and to Mr. A. Sonnenberg for his help in constructing a restriction map of the *oli1* region. This work was supported in part by a grant to P.B. and G. S. P. Groot from The Netherlands Foundation for Chemical Research (SON) with financial aid from The Netherlands Organization for the Advancement of Pure Research (ZWO).

1. Hollenberg, C. P., Borst, P. & Van Bruggen, E. F. J. (1970) *Biochim. Biophys. Acta* **209,** 1–15.
2. Schweyen, R. J., Weiss-Brummer, B., Backhaus, B. & Kaudewitz, F. (1978) *Mol. Gen. Genet.* **159,** 151–160.
3. Sanders, J. P. M., Heyting, C., Verbeet, M. Ph., Meijlink, F. C. P. W. & Borst, P. (1977) *Mol. Gen. Genet.* **157,** 239–261.
4. Van Ommen, G. J. B., Groot, G. S. P. & Borst, P. (1977) *Mol. Gen. Genet.* **154,** 255–262.
5. Borst, P., Bos, J. L., Grivell, L. A., Groot, G. S. P., Heyting, C., Moorman, A. F. M., Sanders, J. P. M., Talen, J. L., Van Kreijl, C. F. & Van Ommen, G. J. B. (1977) in *Mitochondria 1977: Genetics and Biogenesis of Mitochondria*, eds. Bandlow, W., Schweyen, R. J., Wolf, K. & Kaudewitz, F. (de Gruyter, Berlin), pp. 213–254.
6. Van Ommen, G. J. B. & Groot, G. S. P. (1977) in *Mitochondria 1977: Genetics and Biogenesis of Mitochondria*, eds. Bandlow, W., Schweyen, R. J., Wolf, K. & Kaudewitz, F. (de Gruyter, Berlin), pp. 415–424.
7. Borst, P. & Grivell, L. A. (1978) *Cell* **15,** 705–723.
8. Bernardi, G., Piperno, G. & Fonty, G. (1972) *J. Mol. Biol.* **65,** 173–189.
9. Prunell, A. & Bernardi, G. (1977) *J. Mol. Biol.* **110,** 53–74.
10. Bos, J. L., Heyting, C., Borst, P., Arnberg, A. C. & Van Bruggen, E. F. J. (1978) *Nature (London)* **275,** 336–338.
11. Slonimski, P. P., Claisse, M. L., Foucher, M., Jacq, C., Kochko, A., Lamouroux, A., Pajot, P., Perrodin, G., Spyridakis, A. & Wambier-Kluppel, M. L. (1978) in *Biochemistry and Genetics of Yeast*, eds. Bacilla, M., Horecker, B. L. & Stoppani, A. O. M. (Academic, New York), pp. 391–401.
12. Mahler, H. R., Hanson, D., Miller, D., Lin, D. D., Alexander, N. J., Vincent, R. D. & Perlman, P. S. (1978) in *Biochemistry and Genetics of Yeast*, eds. Bacilla, M., Horecker, B. L. & Stoppani, A. O. M. (Academic, New York), pp. 513–547.
13. Wachter, E., Sebald, W. & Tzagoloff, A. (1977) in *Mitochondria 1977: Genetics and Biogenesis of Mitochondria*, eds. Bandlow, W., Schweyen, R. J., Wolf, K. & Kaudewitz, F. (de Gruyter, Berlin), pp. 441–449.
14. Sebald, W. & Wachter, E. (1978) in *29th Mosbacher Colloquium on Energy Conservation in Biological Membranes*, eds. Schäfer, G. & Klingenberg, M. (Springer, Berlin), pp. 228–236.
15. DiFranco, A., Sanders, J. P. M., Heyting, C., Borst, P. & Slonimski, P. P. (1976) in *The Genetic Function of Mitochondrial DNA*, eds. Saccone, C. & Kroon, A. M. (North-Holland, Amsterdam), pp. 291–304.
16. Deutsch, J., Dujon, B., Netter, P., Petrochilo, E., Slonimski, P. P., Bolotin-Fukuhara, M. & Coen, D. (1974) *Genetics* **76,** 195–219.
17. Sanders, J. P. M., Flavell, R. A., Borst, P. & Mol, J. N. M. (1973) *Biochim. Biophys. Acta* **312,** 441–457.
18. Boseley, P. G., Tuyns, A. & Birnstiel, M. L. (1978) *Nucleic Acids Res.* **5,** 1121–1137.
19. Sanders, J. P. M., Borst, P. & Weijers, P. J. (1975) *Mol. Gen. Genet.* **143,** 53–64.
20. Maxam, A. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
21. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
22. Maat, J. & Smith, A. J. H. (1978) *Nucleic Acids Res.* **5,** 4537–4545.
23. DeFilippes, F. M. (1974) *Biochem. Biophys. Res. Commun.* **58,** 586–596.
24. Reijnders, L., Kleisen, C. M., Grivell, L. A. & Borst, P. (1972) *Biochim. Biophys. Acta* **272,** 396–407.
25. Martin, R., Schneller, J. M., Stahl, A. J. C. & Dirheimer, G. (1976) *Biochem. Biophys. Res. Commun.* **70,** 997–1002.
26. Shine, J. & Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 1342–1346.
27. Moorman, A. F. M., Verkley, F. N., Asselbergs, F. A. M. & Grivell, L. A. (1977) in *Mitochondria 1977: Genetics and Biogenesis of Mitochondria*, eds. Bandlow, W., Schweyen, R. J., Wolf, K. & Kaudewitz, F. (de Gruyter, Berlin), pp. 385–400.
28. Gilbert, W. (1976) in *RNA Polymerase*, eds. Losick, R. & Chamberlin, M. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 193–206.
29. Valenzuela, P., Bell, G. I., Venegas, A., Sewell, E. T., Maziarz, F. R., De Gennaro, L. J., Weinberg, F. & Rutter, W. J. (1977) *J. Biol. Chem.* **252,** 8126–8135.
30. Gillum, A. M. & Clayton, D. A. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 677–681.
31. Van Kreijl, C. F. & Bos, J. L. (1977) *Nucleic Acids Res.* **4,** 2369–2388.
32. Bernardi, G. & Timasheff, S. N. (1970) *J. Mol. Biol.* **48,** 43–52.
33. Christiansen, C., Christiansen, G. & Leth Bak, A. (1974) *J. Mol. Biol.* **84,** 65–82.
34. Macino, G. & Tzagoloff, A. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 131–135.