Assembly of the mitochondrial membrane system: Partial sequence of a mitochondrial ATPase gene in Saccharomyces cerevisiae*

(restriction map/DNA sequence/ATPase proteolipid gene)

GIUSEPPE MACINO AND ALEXANDER TZAGOLOFF

Department of Biological Sciences, Columbia University, New York, New York 10027

Communicated by David E. Green, October 10, 1978

ABSTRACT The nucleotide sequence of mitochondrial DNA of a cytoplasmic "petite" mutant of Saccharomyces cerevisiae is reported. The DNA has a repeat length of 1000 base pairs and contains a genetic marker (oli-I) for the ATPase proteolipid. The nucleotide sequence reveals the presence of part of the structural gene of the subunit-9 proteolipid of the ATPase complex and an extended A+T-rich region adjacent to the carboxyl-terminal end of the gene. The structural gene sequence agrees with the primary structure of the protein. These studies point out the feasibility of using the DNA of appropriately marked "petite" mutants to obtain the sequence of mitochondrial genes.

Mitochondrial DNA of Saccharomyces cerevisiae codes for a limited number of proteins that are components of respiratory complexes and of the oligomycin-sensitive ATPase (1-6). Recent evidence suggests that the mitochondrial gene products correspond to the set of proteins translated within the organelle (6-8). In view of its relatively small size and the large number of currently known genetic markers, the mitochondrial genome of this yeast offers many advantages for DNA sequence studies.

In this communication, we report the nucleotide sequence of a segment of mitochondrial DNA obtained from a ρ− clone containing the oli-1 resistance locus (9). In other studies, the oli-1 marker was localized in the structural gene of a subunit polypeptide of the mitochondrial ATPase complex (7). Our data indicate that the DNA of the ρ− clone includes part of the ATPase gene adjoined by an extended A+T-rich region.

MATERIALS AND METHODS


Isolation of DS 400/A4. The ρ− strains were isolated from the respiratory competent haploid strain of S. cerevisiae D273-10B/A2I (α, met, ρ−, Cbl+ [p ph be]) (10). A fresh overnight culture was suspended in 0.5 M potassium phosphate buffer (pH 6.8) at a cell density of 106/ml. The suspension was treated with 10 μg of ethidium bromide per ml for 1 hr and immediately spread for single colonies on YPD plates (2% glucose/2% peptone/1% yeast extract/2% agar). After 3 days of incubation at 30°C, the ρ− mutants, which grew as small colonies, were picked and templated on fresh YPD plates in 50 position grids. Approximately 200 colonies were collected. This collection of primary ρ− clones was subjected to the following genetic tests. To verify the presence of the various antibiotic resistance markers in the mitochondrial DNA of the ρ− mutants, we crossreplicated the master plates on lawns of the antibiotic-sensitive haploid strain of S. carlsbergensis CB11 (a, ade-1, ρ−) spread on Wickerham’s minimal medium supplemented with 2% glucose. After incubation for 2 days, the protoplastically selective diploids were replicated on rutamycin, erythromycin, and paramomycin media (for exact compositions of the antibiotic media see ref. 2). Clones that gave rise to antibiotic-resistant diploids in the cross to CB11 were considered to have retained the resistance marker in their mitochondrial DNA. To test for the presence of the Mit loci (olf-1, xol-2, xol-3, cob-1, cob-2, pho-1, and pho-2) the master plates were cross-replicated on lawns of Mit− mutants carrying lesions in each of these loci (2, 4, 5, 11). The diploids issued from these crosses were replicated on YEFP medium (3% glycerol/2% ethanol/2% peptone/1% yeast extract/2% agar). The Mit marker was assumed to be present in the ρ− clones that grew on glycerol when crossed to the appropriate Mit− tester.

Out of the 200 colonies examined, 17 retained the olf-1, pho-2, cob-1, and cob-2 markers, but lost all the other markers tested. One of these clones, DS 400, was purified and treated with ethidium bromide under the same conditions used for the first mutagenesis. After the second ethidium bromide mutagenesis, 200 of the new ρ− clones were tested for the presence of the olf-1, cob-1, and cob-2 markers. Of these, 7 contained only the olf-1 resistance marker. The ρ− clone designated DS400/A4 was ascertained to have retained a segment of mitochondrial DNA with a repeat length of about 1 kilobase and was used for the sequence studies.

Preparation of Mitochondrial DNA. Cells grown on 2% glucose/1% peptone/1% yeast extract were converted to protoplasts by digestion with gluculase (Endo Lab, Garden City, NY), and mitochondria were prepared by a published procedure (12). For the preparation of DNA the mitochondria were briefly treated with DNase I (Worthington), washed, lysed with 2% Sarkosyl, and extracted with phenol. The DNA was purified by isopycnic centrifugation in CsCl (13).

Restriction Endonuclease Digestion of DNA. All digestions were carried out at 37°C with the exception of Taq I and Bcl I, which were done at 65°C. The reaction mixture contained 6 mM Tris-HCl (pH 7.5), 6 mM MgCl₂, and 6 mM 2-mercaptopropanethiol.

Abbreviations: ρ− refers to cytoplasmic "petite" mutants of S. cerevisiae. ρ− mutants have extended deletions in mitochondrial DNA and are characterized by pleiotropic deficiencies in respiratory functions. Mit− mutations are mutations in mitochondrial DNA that are expressed in a deficiency of mitochondrial ATPase or respiratory activity. oli-1, xol-2, and xol-3 are genetic loci that code for structural components of cytochrome oxidase; cob-1 and cob-2 for cytochrome b; and pho-1 and pho-2 for the ATPase.

* This paper is no. 24 of a series. The preceding paper in the series is ref. 5.
to ethanol (50 mM NaCl was added when EcoRI and Sal I were used). The EcoRI star digestion was carried out in the presence of 2 mM MnCl₂ instead of MgCl₂. The completeness of digestion was checked by gel electrophoresis in agarose (14).

**5'-End Labeling and Strand Separation of Restriction Fragments.** The restriction fragments were treated with alkaline phosphatase and 5'-end labeled with [γ-³²P]ATP in the presence of polynucleotide kinase (15). For strand separation, the labeled DNA was dissolved in 50 mM NaOH/10 mM EDTA and subjected to electrophoresis on 6–8% polyacrylamide gels. The gels were 1.5 mm thick and 40 cm long and were usually run overnight at 4°C. Labeled bands corresponding to single-stranded DNA were cut out and extracted from the gel (15).

Sequence of DNA. The sequence of the 5'-end-labeled single-stranded DNA was determined by the method of Maxam and Gilbert (15) on 0.4-mm thick polyacrylamide/urea gels (16).

**RESULTS**

Properties of ρ⁻ Clones with oli-I Resistance Marker. The seven ρ⁻ strains selected on the basis of their retention of the oli-I resistance marker exhibited a high degree of genetic stability. This was established by testing subclones of each strain for the presence of oli-I. In each case, out of 100–200 subclones tested, more than 95% retained the marker.

The repeat length of mitochondrial DNA was estimated from the cumulative molecular weights of the fragments generated by different restriction endonucleases. Hpa II digests of the DNA in most strains yielded a rather complex pattern of fragments whose total molecular weights were approximately 5000 base pairs (data not shown). The DNA of one strain (DS400/A4), however, was cleaved into two fragments by Hpa II and into a single fragment by a number of other endonucleases. Even though the unit length of the retained segment of mitochondrial DNA in DS400/A4 was estimated to be only 1060 base pairs, the overall size of the genome is much larger. Molecular species corresponding to multiples of 1060 base pairs were observed in partial digests with endonucleases that introduced only one cut. In undigested samples, most of the DNA migrated with a molecular weight greater than 10⁷. These results agree with previous findings that, in ρ⁻ mutants of *Saccharomyces cerevisiae*, the retained segment of mitochondrial DNA is reiterated and may attain the full length of the wild-type genome (18, 19).

Since mutations in the oli-I and pho-2 loci have been localized in the structural gene of subunit 9 of the ATPase complex (5, 7), the selected clones were checked for the presence of a pho-2 marker. The clones were crossed to the ATPase-deficient Mit⁻ tester, M302-34 (5), and the diploid progeny were scored for growth on glycerol. All the ρ⁻ clones, with the exception of DS400/A4, produced glycerol-positive diploids, indicating the presence of the pho-2 locus. Despite the fact that DS400/A4 had lost the pho-2 marker, it was chosen for sequence analysis because of the short repeat length of its DNA.

**Restriction Map of DS400/A4 Mitochondrial DNA.** The mitochondrial DNA of DS400/A4 was subjected to digestions with the restriction endonucleases listed in *Materials and Methods*. Only one fragment was detected on agarose gels when the DNA was digested with Hph I, Alu I, Hha I, Hae III, or Sst II (Fig. 1). The length of the fragments corresponded to approximately 1060 base pairs. (The fragment produced by Sst II, however, appeared to be slightly shorter. Of the 32 endonucleases tried, the only other enzymes capable of cleaving the DNA were Hpa II and EcoRI. The EcoRI cuts depended upon the substitution of Mn²⁺ for Mg²⁺, conditions under which the specificity of the endonuclease is reduced to a tetranucleotide (20). Both Hpa II and EcoRI star digests contained two fragments. The fragments obtained with Hpa II were 840 and 220 base pairs long; those with EcoRI star, 540 and 520 base pairs long.

Double and triple digestions with different combinations of enzymes indicated a clustering of the Hha I, Hph I, and Sst II sites near one of the two Hpa II sites. Similarly, the Hae III site was found to be close to the other Hpa II site. Alu I produced a new fragment when used in combination with Hpa II and cut the DNA approximately 200 base pairs distal to the short sequence that contains the Hpa II, Hph I, Hha I, and Sst II sites. Based on these data, it was possible to construct an unequivocal physical map of the DNA (Fig. 2). The two EcoRI star sites were 140 nucleotides away from the Hae III site in fragment A and 50 nucleotides away from the Alu I site in fragment C (Fig. 2).

![Fig. 1. Agarose gel analysis of DS400/A4 DNA. The following DNA samples were examined by electrophoresis in 1.5% agarose gels. Lanes a and j, Hae III digest of φX174 RF. DS400/A4 DNA was digested with: lanes b, Hph I; c, Alu I; d, Hha I; e, Sst II; f, Hae III; g, Hpa II; h, Hpa II + Alu I; i, Hpa II + Sst II; k, Hpa II + Hae III; l, Hpa II + Hha I; m, Hpa II + Hph I; n, Hae III + Alu I; o, Hae III + Sst II; p, Hae III + Hph I; and q, Alu I + Hha I + Hae III. Lengths in base pairs of φX174 RF fragments, indicated in the margin, are based on published data (17).](image-url)
Nucleotide Sequence of DS400/A4 Mitochondrial DNA.
The sequence of DS400/A4 DNA was obtained from restriction fragments that were labeled at the 5'-termini and separated into single strands. The restriction fragments used for the sequence determinations are shown in the map of Fig. 3. The sequences of most of the regions of the DNA were determined twice, starting from different restriction sites. In some cases the sequences of complementary strands helped to confirm the validity of the results. For example, the 5'-end-labeled single strands of fragment B produced by Hpa II digestion (see Fig. 2) yielded nucleotide sequences that were perfectly complementary to each other. It was also possible to obtain complementary sequences for the region of DNA containing the structural gene.

The sequence derived from the various fragments analyzed is presented in Fig. 3. From the known primary structure of subunit 9 of the ATPase, it was possible to recognize in the DNA sequence part of the structural gene of this protein. The structural gene region starts with the 45th amino acid from the amino terminus and continues uninterrupted until the last amino acid at the carboxyl terminus. There is an ochre termination codon immediately next to the last amino acid. Another ochre termination codon occurs a few bases further down the sequence. The sequence that follows the structural gene consists almost entirely of A+T (95%), with the exception of two short stretches that are very rich in G+C. One of the G+C-rich sequences contains one Hae III and three Hpa II sites. The existence of the two extra Hpa II sites was discovered when the sequence of the DNA from the Hae III site was determined. The additional two fragments produced by Hpa II are only a few nucleotides long and therefore were not detected on agarose gels.

The second G+C-rich region contains one Hpa II, two Sst II, and one Hha I sites. The two Sst II sites span a sequence 20 nucleotides long. The fact that the Sst II fragment seen on agarose migrates slightly faster than the full length of the repeat unit is consistent with the loss of the shorter fragment. The Hph I site, which, based on the restriction map, should also be present in this region, is not seen in the sequence. Since the sequence of the region was determined from Hpa II fragments, it is possible that there are several closely spaced Hpa II sites and that the sequence presented may be missing a few nucleotides corresponding to the Hph I site.

The nucleotide sequence of the DS400/A4 DNA also revealed that the two sites recognized by EcoRI star are A-A-A- T-T-C and T-A-A-T-T-C. Although there are numerous A-A-T-T tetrasequences in the DNA, these apparently are not cleaved under our conditions of incubation.

Several additional points emerge from the sequence. The first is the excellent agreement of the structural gene sequence with the amino acid sequence reported by Wachter et al. (7). The one exception is the residue at position 46. According to the protein sequence data, this residue should be a threonine. The nucleotide sequence, however, shows that the triplet at this position corresponds to a leucine. We have no explanation for this discrepancy. Second, even though yeast mitochondria use the codons of the universal code, there appears to be little codon degeneracy. This is seen in Table 1, which lists the codons for eight amino acids. Leucine is an especially interesting case because it has six codons, of which only two (UUA and CUA) are used for the seven residues present in the carboxyl-terminal portion of the protein. A preferential utilization of UUC is also seen for phenylalanine. The UUU codon at position 53 is due to a mutation in the DNA of the D273-10B/A21 strain in which a T has probably been substituted for an A since the protein of the wild-type oligomycin-sensitive parental strain D273-10B/A1 has a leucine at this position.

The absence of restriction sites made it impossible to obtain the entire sequence of fragment A. The sequencing gels of this fragment indicated, however, that the missing sequence consists almost entirely of A+T.

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Codon</th>
<th>Observed no. of codons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>GCA</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>U</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>0</td>
</tr>
<tr>
<td>Gly</td>
<td>GGA</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>U</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>0</td>
</tr>
<tr>
<td>Leu</td>
<td>UUA</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>CUA</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>U</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>0</td>
</tr>
<tr>
<td>Met</td>
<td>AUG</td>
<td>2</td>
</tr>
<tr>
<td>Phe</td>
<td>UUU</td>
<td>1*</td>
</tr>
<tr>
<td></td>
<td>UUC</td>
<td>4</td>
</tr>
<tr>
<td>Ser</td>
<td>UCA</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>U</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>AGU</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Val</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>U</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>0</td>
</tr>
</tbody>
</table>

* Amino-acid substitution in the oli-1 mutation.
FIG. 3. Sequence of DS400/A4 mitochondrial DNA. The sequence was derived from the restriction fragments shown in the diagram. It starts with the 5' end of the structural gene and ends with an adenine that is fused to the first nucleotide of the structural gene. The restriction sites of the various endonucleases are underlined. Asterisk denotes the position in the protein in which a phenylalanine has been substituted for a leucine in the oligomycin-resistant strain D273-10B/A21. The protein sequence of the proteolipid shown above the structural gene is based on the data of Wachter et al. (7).

DISCUSSION
The recent discovery of new antibiotic resistance (9, 21) and Mit− markers (1–6) in the mitochondrial genome of S. cerevisiae has allowed the construction of a fairly detailed genetic map of this genome (22). In a number of instances, mitochondrial mutations conferring resistance to antibiotics or leading to a loss of respiratory functions have also been localized on physical maps derived by restriction endonuclease analysis (23, 24). These advances have made it feasible to begin determining the sequences of those regions of the DNA that contain genes coding for components of the protein synthetic machinery or for subunit polypeptides of cytochrome oxidase, cytochrome b, and the oligomycin-sensitive ATPase.

Two different strategies can be used to obtain mitochondrial DNA of S. cerevisiae in amounts that are large enough for the sequence to be determined. Restriction fragments of the wild-type DNA can be cloned in a heterologous system with either phage or plasmid vectors. Alternatively, high yields of DNA may be obtained from ρ− mutants due to the amplification of the retained segments of mitochondrial DNA, which can
have repeat lengths of a few thousand base pairs or less. Such strains can be isolated on the basis of the retention of any genetically testable marker and are therefore potentially capable of providing DNA sequences from any region of the genome. The important advantage of using ρ− mutants is that they are easy to isolate and require a minimum initial investment of time.

The gene of the ATPase proteolipid (subunit 9) was chosen as a starting point for sequence studies for two reasons. The complete primary structure of the yeast proteolipid has been elucidated (7) and two well-characterized mitochondrial markers, oli-1 and pho-2, have been conclusively shown to be in the structural gene of the protein (5, 7). Because our intention was to determine the sequence of only the proteolipid gene, we selected ρ− clones, which retained the oli-1 resistance marker and whose mitochondrial DNA had a repeat length sufficiently short that its sequence could be determined by currently available techniques. DS400/A4 satisfied both criteria. The repeat unit in this clone is 1060 base pairs, and genetic tests indicate the presence of the oli-1 marker. Although most of the restriction sites in the DNA are clustered in two regions, the sequences of the fragments could be determined in almost their entirety.

The sequence of the DS400/A4 mitochondrial DNA has revealed several interesting features. In confirmation of the genetic data, only part of the structural gene of the proteolipid has been retained in the DNA, the rest having been lost as the result of a deletion that starts with the 44th amino acid from the amino terminus. The nondeleted part of the gene includes the entire carboxyl-terminal sequence beginning with the 45th amino acid. The nucleotide sequence is completely consistent with the known primary structure, with the exception of one amino acid residue, and shows that mitochondria use the codons of the universal code. The sequence also confirms the previously reported substitution of a phenylalanine for leucine in the proteolipid of the oligomycin-resistant parental strain D273-10B/A21 used in this study (7). Although the sequence data available at present are limited, they nonetheless suggest a highly nondegenerate code. For example, six out of seven leucine residues present in the carboxyl-terminal half of the proteolipid use the UUA codon. A preferential use of one codon is also seen in phenylalanine. Whether this is generally true of the rest of the genome will have to await further sequence data.

In addition to the structural gene, the retained segment of mitochondrial DNA of DS400/A4 contains a 1000-base-pair sequence proceeding from the carboxyl terminus. This sequence shows the presence of two ochre termination codons, one of which is immediately adjacent to the carboxyl-terminal amino acid. The rest of the sequence is extremely rich in A+T, with the exception of two short regions, which have a high content of G+C and contain most of the restriction sites of the DNA. The function of these nonstructural sequences cannot be specified at present, but one of the G+C-rich regions is highly palindromic, a property usually associated with promoters. This region, however, cannot be a promoter for the proteolipid gene because it is located on the carboxyl-terminal side of the structural gene sequence.

The deletion of the amino-terminal portion of the proteolipid is consistent with the loss of the pho-2 marker in the DS400/A4 clone. We have recently started to determine the sequence of the DNA of a ρ− mutant that contains the pho-2 marker and has a repeat length some 700 base pairs longer than DS400/A4. The restriction map of the DNA from this mutant indicates that the additional 700-base-pair sequence probably includes the rest of the proteolipid gene and perhaps the promoter region.

Preliminary sequencing data of the DNA from the new mutant also confirm that the end points of the deletion in DS400/A4 occur at the beginning and end of the sequence, as depicted in Fig. 3.

This research was supported by Grant HL 22174 from the National Institutes of Health, U.S. Public Health Service. We thank Drs. Catherine and Craig Squires for their advice and help in the sequence work.