Assembly of the Mitochondrial Membrane System

DNA SEQUENCE AND ORGANIZATION OF THE CYTOCHROME \( b \) GENE IN SACCHAROMYCES CEREVISIAE D273-10B*

(Received for publication, March 4, 1980)

Francisco G. Nobrega and Alexander Tzagoloff
From the Department of Biological Sciences, Columbia University, New York, New York 10027

The mitochondrial genomes of cytoplasmic "petite" (\( \rho^- \)) mutants of Saccharomyces cerevisiae have been used to sequence the cytochrome \( b \) gene. A continuous sequence of 6.2 kilobase pairs has been obtained from 71.4 to 80.2 units of the wild type map. This region contains all the cytochrome \( b \) genes previously assigned to the \( cob1 \) and \( cob2 \) genetic loci. Analysis of the DNA sequence has revealed that in the strain D273-10B, the cytochrome \( b \) gene is composed of three exons. The longest exon (b1) codes for the first 252 to 253 amino acids from the NH2-terminal end of the protein. The next two exons (b2 and b3) code for 16 to 18 and 115 to 116 amino acids, respectively. The complete cytochrome \( b \) polypeptide chain consists of 385 amino acids. Based on the amino acid composition, the yeast protein has a molecular weight of 44,000. The three exon regions of the cytochrome \( b \) gene are separated by two introns. The intron between b1 and b2 is 1414 nucleotides long and contains a reading frame that is potentially capable of coding for another protein of 384 amino acid residues. The second intron is 733 nucleotides long. This sequence is rich in A+T and includes a G+C cluster that may be involved in processing of the cytochrome \( b \) messenger. The organization of the cytochrome \( b \) region in S. cerevisiae D273-10B is somewhat less complex than has been reported for other yeast strains in which exon b1 appears to be further fragmented into three smaller exons.

The availability of \( \rho^- \) mutants with well characterized genomes has made it possible to sequence the cytochrome \( b \) region and deduce the primary structure of the protein. Our studies show that in the D273-10B strain of Saccharomyces cerevisiae, the cytochrome \( b \) gene is constituted of three exons. The coding and intervening sequences have been identified from the homology of the amino acid sequences of the yeast and mammalian cytochrome \( b \).

MATERIALS AND METHODS

Strains—The isolation and physical characterization of the \( \rho^- \) strains have been reported (1).

Purification of mtDNA—The \( \rho^- \) mutants were grown at 30°C in 10 liters of liquid YPD medium (2% glucose, 1% yeast extract, and 2% peptone). The cells were harvested in early stationary phase of growth. Mitochondria were isolated and mtDNA was purified as described previously (1).

DNA Sequencing—Restriction fragments were labeled at the 5' ends with [\( \gamma \)32P]ATP (2,000 to 3,000 Ci/mmol, New England Nuclear) in the presence of T4 polynucleotide kinase (2). In most instances, the labeled fragments were denatured either in 0.1 M NaOH or 80% formamide and separated into single strands on polyacrylamide gels (2). In some cases, the labeled fragments were subjected to a second endonucleolytic cleavage before separation of the double-stranded DNA fragments on polyacrylamide gels. The 5' end-labeled fragments were sequenced by the chemical derivatization method of Maxam and Gilbert (2). In most experiments, the single strands were clearly separated without evidence of significant cross-contamination. The absence of any appreciable background radioactivity on the strand separation gels indicated that there were few if any single-stranded nicks in the restriction fragments.

RESULTS

Sequencing Strategy—The \( \rho^- \) clones used for the sequencing of mtDNA originated from the respiratory-competent haploid S. cerevisiae D273-10B/A21 (3). The most complex clone DS400/A12 has a mitochondrial genome with a unit length of 7.6 kbp spanning the region of wild type mtDNA from 69.3 to 80.2 units (1). All the other \( \rho^- \) clones were derived from DS400/A12. The unit lengths of their mtDNAs ranged from 0.16 to 3.5 kbp (1).

Although partial sequence data was obtained from preparative restriction fragments of the DS400/A12 mtDNA, the genome size of this clone was much too large to be sequenced directly. Our strategy, therefore, was to sequence the mtDNA segments of the simpler \( \rho^- \) mutants. Due to the smaller genome sizes, there were fewer restriction sites and it was possible to find combinations of enzymes that generated a sufficiently discrete number of fragments for adequate resolution of 5' end-labeled mixtures by electrophoresis on polyacrylamide gels. In addition, the smaller genome sizes allowed for a greater extent of amplification of the retained segments of mtDNA.

Most of the sequence was determined from the mtDNAs of DS400/M8, DS400/M4, and DS400/M11. Some remaining

* This research was supported by Grant HL22174 from the National Institutes of Health, United States Public Health Service. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† Present address, Instituto de Quimica-Bioquimica, Cidade Universitaria, C.P. 20780, Sao Paulo, Brazil. Recipient of a FAPESP fellowship since November, 1979.

The abbreviations used are: \( \rho^- \) refers to cytoplasmic petite mutants resulting from long deletions in mitochondrial DNA. Mit loci are markers in mitochondrial genes that code for subunit polypeptides of cytochrome oxidase, cytochrome \( b \) (\( cob1 \) and \( cob2 \)), and the oligomycin-sensitive ATPase. The antibiotic resistance loci referred to in this paper are par (paromomycin), cap (chloramphenicol), and oli (oligomycin). bp, base pair.
gaps were filled by sequencing the genomes of the low complexity clones DS400/N1, N2, N24, N28, and N31. The data obtained from the latter clones were also used to locate mutational sites in defined regions of the DS400/A12 mtDNA. This information was important in identifying the coding elements in the cytochrome b region.

**Nucleotide Sequence of DS400/A12**—The spans of mtDNA represented by the *ρ* clones used for the sequence analysis are shown in Fig. 1. All the mtDNA segments have been aligned with the restriction map of DS400/A12. The fragments sequenced from each genome are indicated by the arrows. The sequence between 71.4 and 74.3 units was determined primarily from the genomes of DS400/M8 and M4 and the smaller mtDNA segments of DS400/N1 and N24. This region has numerous restriction sites and most of the sequences could be confirmed from overlapping fragments and complementary strands.

The middle region of DS400/A12 (74.3 to 75.5 units) was sequenced in the clones DS400/N23, N28, N31, and preparative HincII and HindIII fragments from DS400/A12 (Fig. 1). The sequence of the other half of the mtDNA (75.4 to 80.2 units) was obtained from DS400/M11 and in part from DS400/N2. The sequence of this region has a few short gaps around some of the *Hpa II* sites that could not be crossed from neighboring restriction sites. In addition to the small gaps, only partial sequences were obtained in the region from 69.3 to 71.4 units. Even though we do not have the complete sequence of this region, examination of the sequencing gels indicates that it is highly rich in A+T. The only potential coding sequence that could be detected straddles the *Tag I*/*Hinfl/Mbo II*/*Hph I* site cluster at 69.9 units. These sites occur in a relatively G+C-rich sequence of 70 to 80 nucleotides that has been identified as the glutamic tRNA gene previously mapped in the general vicinity of the cytochrome *b* region (4).

The nucleotide sequence from 71.4 to 80.2 units is presented in Fig. 2. A survey of the sequence reveals only three regions with a sufficiently high G+C content and uninterrupted reading frames to have possible coding functions. The first region starts with an AUG initiator at nucleotide +1 and continues in frame for 1917 nucleotides before ending with an ochre terminator. The average G+C content of this sequence is 25%. The other two registers of the same DNA strand or the three registers of the complementary strand contain numerous ochre and amber termination codons. The reading frame is preceded by 139 nucleotides consisting almost entirely of A+T (=139 to −1).

The second coding region occurs between nucleotides +2174 and +2227, being only some 50 nucleotides long. The G+C content of this short sequence is 35%. The third region of interest (+2883 to +3308) also has a unique reading frame capable of generating a continuous amino acid sequence of 141 residues. This reading frame terminates with an ochre codon and is followed by an A+T-rich sequence. The G+C content of the coding sequence is 26%.

The amino acid sequences encoded by the three reading frames are indicated above the DNA sequences in Fig. 2 and are based on the usual codon assignments of the universal code with two exceptions. Yeast and human mitochondria have recently been shown to use UGA as a tryptophan codon (5,6). Yeast mitochondria have also been found to translate CUA as threonine instead of leucine (7). Both deviations from the standard code have been taken into account in deriving the amino acid sequences.

The 2800-nucleotide sequence downstream from the ochre terminator in the third coding region consists of semirepetitive A+T-rich sequences interspersed with numerous G+C clusters. This is particularly evident in the span from 76.2 to 80.2 units where there are nine G+C clusters, each having one or more *Hpa II* and *Hae III* sites.

**Structural Gene Coding Elements**—The apocytochrome b gene of yeast mitochondria has been shown to have a mosaic structure (8–10). Although cytochrome b has been purified from *S. cerevisiae* (11, 12), *Neurospora crassa* (13), and bovine (14) mitochondria, none of the proteins have been sequenced. These two circumstances made it impossible to distinguish the exon and intron components of the gene from the DNA sequence alone.

A fortunate solution to this dilemma came from a comparison of the DNA sequences of the yeast and bovine mitochondrial DNAs. Drs. F. Sanger and B. Barrell at the MRC Laboratory in Cambridge, England, have obtained extensive sequence data on bovine mtDNA. Their analysis of the bovine DNA revealed a gene that codes for a protein with amino acid sequence homology to three separate amino acid sequences encoded in the cytochrome b region of the yeast genome. Since the homology of the two proteins is high (approximately 50%), it has permitted a fairly precise definition of the exon-intron interfaces in the yeast sequence. It should be stressed, however, that the identification of the coding elements in the yeast sequence is contingent on two assumptions. The first is that there are no common intervening sequences in the yeast and bovine genes coding for homologous amino acid sequences, and the second is that no major deletions or insertions of amino acids exist in the two proteins. The first homologous amino acid sequence is encoded by Fig. 1. *ρ* genomes used for DNA sequencing. The restriction map of the DS400/A12 mtDNA is shown in the second-innermost circle. The symbols used to denote the restriction sites are: Hinf I (▲), *Hpa II* (◇), *Hae III* (◇), *Hph I* (◆), *Mbo I* (◇), *Mbo II* (◇), *Alu I* (◇), *Tag I* (◇), *Hinfl* (◇), *Hha I* (◇), *Eco RI* (◇), *Eco HI* (◇), *Bcl I* (◇), *Bgl II* (◇), *Bam HI* (◆), *Sac II* (◇), *Rsa I* (△). Only the *Rsa I* sites used for sequencing are indicated. The wild type map units are shown in the inner circle. The entire DS400/A12 mtDNA spans the wild type map from 69.3 to 80.2 units. The physical limits of the low complexity genomes are depicted by the heavy-lined arcs. The restriction fragments and the approximate distances sequenced in the clones DS400/N7, N9, and N31 clones whose genomes were also sequenced are not shown on the map.
DNA Sequence and Organization of the Cyt b Gene

The juncture of the sequence corresponds to a continuous region of the wild type mtDNA. Only the sequence of the nontranscribed strand is shown. The amino acid sequences of the three cytochrome b exons are indicated in roman type. The amino acid sequence of the long reading frame that extends into the first intron has been typed in italics. All the restriction sites except RsaI are marked. The sequence from nucleotides +4841 to +4948 may have some errors.
nucleotides +1 to +759. This exon (b1) starts with the first AUG initiator downstream from the A+T-rich region and generates a polypeptide chain with 252 to 253 amino acid residues. The second exon (b2) is considerably shorter and is encoded by nucleotides +2174 to +2227. At most, b2 codes for 18 amino acids. The COOH-terminal portion of the protein is encoded by exon b3 whose limits extend from nucleotides +2958 to the ochre terminator at +3308. Exon b3 codes for 115 to 116 amino acids.

The primary structure of yeast cytochrome b encoded by DNA Sequence and Organization of the Cyt b Gene

FIG. 2—Continued
the three exons is shown in Fig. 3. The protein consists of 385 amino acid residues and has a molecular weight of 44,000. There are 2 amino acids at the exon-intron junctures that are not homologous to the bovine cytochrome b. One occurs between b1 and b2 and the other occurs between b2 and b3. If the splicing mechanism does not involve the generation of new codons at the junctures, the alternative amino acids at residue 253 are glutamine or histidine, and at residue 270 are aspartic acid or valine.

Localization of Cytochrome b Mutations—The DNA sequence analysis of the low complexity p-clones has made it possible to ascribe precise physical limits to their mtDNA segments and to map certain cytochrome b mutations in the exon and intron regions. The physical limits of the mtDNA segments indicated by the DNA sequence is in excellent agreement with their previous localization in the DS400/A12 mtDNA from restriction analysis (1). The genotypes of the clones and the nucleotide sequences retained in their genomes are summarized in Table I. Based on the results, the mutations have been assigned to the spans of the DS400/A12 mtDNA shown in Fig. 4. The four cob2 mutations M10-152, M17-71, and M33-119 fall entirely within exon b1. The fifth cob2 marker studied, M9-228, is present in DS400/N7 and N9. The N7 mtDNA contains part of exon b1 and in addition extends for 256 nucleotides into the first intron. Since the mi- mutant M9-228 synthesizes a truncated form of apocytochrome b (15), the mutation is likely to be in exon b1. The other five markers have previously been mapped in the cob1 locus (15). M8-181 is the only marker retained in the clones DS400/N23 and N28. The genomes of these clones contain the entire exon b2 plus some flanking sequences from the neighboring introns. At present, it is not certain whether M8-181 is an exon or intron mutation. The three mutations
DNA Sequence and Organization of the Cyt b Gene

fMet-Ala-Phe-Arg-Lys-Ser-Val-Tyr-Leu-Ser-Val-Val-Asn-Ser-Tyr-Ile-Ile-Asp-Ser-Pro-Gln-Pro-Asp-Ser
Ile-Asn-Tyr-Trp-Trp-Asn-Met-Gly-Ser-Leu-Gly-Leu-Cys-Leu-Val-Ile-Gln-Ile-Val-Thr-Gly-Ile-Phe-Met
Ala-Met-His-Tyr-Ser-Asn-Ile-Glu-Leu-Ala-Phe-Ser-Ser-Val-Glu-His-Ile-Arg-Asp-Val-His-Asn-Gly
Tyr-Ile-Leu-Arg-Tyr-Leu-His-Ala-Asn-Gly-Ala-Ser-Phe-Phe-Met-Val-Met-Phe-Met-His-Met-Ala-Lys-Gly
Leu-Tyr-Tyr-Gly-Ser-Tyr-Arg-Ser-Pro-Arg-Val-Thr-Leu-Trp-Asn-Gly-Va I L -Ile-Phe-Leu-Thr-Ile
Ala-Thr-Ala-Phe-Leu-Gly-Tyr-Cys-Cys-Val-Tyr-Gly-Gln-Met-Ser-His-Trp-Gly-Ala-Thr-Ile-Thr-Asn-Leu
Phe-Sera-Ile-Pro-Phe-Val-Gly-Asn-Asp-Ile-Val-Ser-Trp-Leu-Trp-Gly-Phe-Ser-Ser-Pro-Thr
Ile-Gln-Arg-Phe-Ala-Leu-His-Tyr-Leu-Val-Pro-Phe-Ile-Ile-Ala-Met-Val-Ile-Met-His-Leu-Ala
Leu-His-Ile-Gly-Ser-Ser-Pro-Leu-Gly-Ile-Thr-Glu-Asp-Arg-Ile-Pro-Met-Ser-Ser-Phe-Lys-Phe-Leu-Thr
Val-Leu-Phe-Tyr-Val-Phe-Leu-Phe-Met-Leu-Ile-Leu-Ile-Val-Phe-Val-Thr-Ser-Asn-Thr
Ile-Phe-Asp-Leu-Va I L -Val-Phe-Leu-Phe-Met-Leu-Ile-Leu-Ile-Val-Phe-Val-Thr-Ser-Asn-Thr
Leu-Asp-Pro-Thr-Pro-Glu-Val-Trp-Tyr-Leu-Leu-Pro-Phe-Tyr-Ala-Ile-Leu
Leu-Pro-Tyr-Ala-Ile-Leu-Arg-Ser-Ile-Pro-Asp-Lys-Leu-Leu-Val-Ile-Thr-Met-Phe-Ala-Ile-Leu
Val-Leu-Leu-Leu-Leu-Leu-Pro-Phe-Arg-Ser-Val-Gly-Trp-Tyr-Leu-Leu-Ser-Lys-Phe-Phe
Phe-Ala-Phe-Asp-Val-Leu-Gly-Gln-Ile-Gly-Ala-Cys-His-Val-Glu-Pro-Tyr-Va I L -Val
Met-Gln-Ile-Ala-Thr-Phe-Ile-Ile-Phe-Val-Ile-Leu-Val-Pro-Val-Leu-Thr-Ile-Glu-Asn
Val-Leu-Phe-Tyr-Ile-Ile-Gly-Arg-Val-Asn-Lys

Fig. 3. Amino acid sequence of yeast apocytochrome b. The primary structure has been deduced from the DNA sequences of the three exons. The uncertain amino acid residues at the two exon-intron interfaces are indicated by asterisks. If the splicing events do not generate new codons, the two possible choices are glutamine or histidine at residue 253 and aspartic acid or valine at residue 270.

Table I
Sequences retained in low complexity ρ− clones
The numbering of the nucleotides is the same as in Fig. 2. Three of the low complexity clones have small internal deletions. Deletions: in DS400/N23, from +2125 to +2174; in DS400/N2, from +3376 to +3397; and in DS400/N7, from −20 to +503.

| Clone    | Genetic marker | Retained sequence |
|----------|----------------|-------------------|
| DS400/N24 | None           | −104 to +67       |
| DS400/N1  | M10-152        | +68 to +365       |
| DS400/N2  | M7-40          | +3105 to +3375, +3398 to +3718 |
| DS400/N7  | M9-228         | −109 to −21, +504 to +1130 |
| DS400/N9  | M9-228, M33-119| +74 to +983       |
| DS400/N23 | M8-181         | −2075 to +2124, +2175 to +2446 |
| DS400/N28 | M8-181         | +1853 to +2451    |
| DS400/N31 | M6-200         | −740 to −390, +2539 to 2800 |
| DS400/M8  | M9-228, M33-119| −380 to +1137     |
| DS400/M4  | M21-71, M17-162|                  |
| DS400/M11 | M7-40, M8-53   | +2765 to +6125    |

M7-40, M8-53, and M13-101 map in the region of exon b3. The M7-40 marker is retained in DS400/N2 whose genome contains most of exon b3 and a 410-nucleotide sequence downstream from the gene. M7-40 has also been shown to synthesize a prematurely terminated polypeptide (10) and is probably a mutation in exon b3. The M8-53 and M13-101 markers map in a region that includes part of exon b3 and part of the intron separating b2 from b3. These two mutations could be either in exon b3 or in the adjoining intron.

Fig. 4. Location of cytochrome b mutations in relation to the exon and intron regions of the cytochrome b gene. The restriction sites of the DS400/A12 genome from 71 to 77 units are indicated on the lower line. The symbols used for the restriction sites are the same as in Fig. 1. The cytochrome b exons are depicted by the solid bars. The hatched bar represents the reading frame in the fmt intron. The heavy lines show the physical limits of the low complexity genomes. The physical limits of the cytochrome b mutations are indicated by the open boxes.

The only mutation that clearly lies outside of the exon regions if M6-200. This marker is present in the low complexity clone DS400/N31. The restriction map (1) and DNA sequence of the mtDNA segment in DS400/N31 suggest that it arose from a deletion in the internal region of the parental DS400/M4 mtDNA from 71.3 to 75.2 units. On the linear map, the genome of this clone, therefore, represents two separate spans of the DS400/A12 mtDNA. One lies between nucleo-
tides +2539 and +2800. The other sequence of DS400/N31 originates from a region upstream from the gene (70.8 to 71.3 units). Since the latter sequence is outside of the gene, the M6-200 mutation must be in the second intron between nucleotides +2539 and +2800.

**Nucleotide Sequences of the Cytochrome b Introns**—The three cytochrome b exons define two intron regions. The first intron (b1/b2) is 1414 nucleotides long. The beginning 1158 nucleotides of this intron contain a reading frame that is continuous with the reading frame of exon b1 (Fig. 2). There are several interesting features about this sequence. The intron reading frame has an AUG initiation codon 1 or 2 amino acids downstream from the end of exon b2. The reading frame starting with this initiator is 1152 nucleotides long and is capable of generating a protein with 384 amino acids (Fig. 2). The hypothetical protein has an unusual composition with a very high content of lysine, tyrosine, and asparagine. It is also significant that the intron sequence has a G+C content of only 23%. This is somewhat lower than the usual range of 27 to 31% G+C found for yeast mitochondrial genes (16–20).

The second intron (b2/b3) is 733 nucleotides long. Intron b2/b3 contains A+T-rich sequences as well as short stretches of nucleotides with a moderate G+C content. There are no reading frames of sufficient length in this intron to code for a protein. In addition to the semirepetitive A+T sequences, there is a G+C cluster with three closely spaced Hpa II sites and a single Hae III site (+2769 to +2800). It is of interest that the terminal 14 nucleotides of the Hpa II/Hae III site cluster (+2787 to +2800) has the sequence 3'-GAGGAA-AGCCCTCCA-5'. An inversely homologous sequence, 5'-CTCTTCTGGGTTT-3', occurs some 500 nucleotides upstream in the intron (+2233 to +2335). These sequences have previously been found to be common to many G+C clusters in the yeast genome and have been proposed to serve as recognition sites for processing of mitochondrial transcripts (21). A possible secondary structure of the intron region assuming base pairing between the two inverted repeats is shown in Fig. 5.

Although we have not completed our analysis of the sequences adjacent to the intron-exon interfaces, it should be noted that exons b2 and b3 have a common pentanucleotide sequence, CCTGA, close to their 5' termini. These sequences encode Pro-Asp and Pro-Glu in exons b2 and b3, respectively. To determine whether this short sequence is important in splicing will require additional information about the exact intron-exon juncptures and a more detailed analysis of the exon and intron sequences.

**Codon Utilization**—The cytochrome b gene reflects the same bias in codon utilization as do other yeast mitochondrial genes (16–20). Most of the codons terminate in an A or U and certain codons are not used (Table II). For example, the CGN family for arginine has not been detected in any of the genes despite the fact that the cognate tRNA gene is present in the yeast genome. Certain codons that were not found in the genes sequenced earlier appeared in the cytochrome b gene. These are UGC for cysteine and GUG for valine. Table II also lists the codons found in the intron coding frame. This frame contains six codons that are absent in the bona fide genes.

The amino acid composition of yeast cytochrome b derived from the DNA sequence is shown in Table III. The experimentally determined compositions of N. crassa and bovine cytochromes b are also listed for comparison. With the exception of a few amino acids, the compositions of the three different proteins are quite similar.

**Junction of DS400/A12**—The DS400/A12 genome is a circular DNA with a tandemly repeated 7.6-kbp segment of mtDNA (1). The restriction map of DS400/A12 indicated that it partially overlaps with the mtDNA segment of another ρ mutant DS401 (1). Both clones have a common 920-bp Hpa II fragment with an internal HinfI site. Since DS401 was selected for the retention of the oliI and loss of the cytochrome b markers, its DNA segment lies on the cap side of the cytochrome b gene (21).

In order to determine the juncture nucleotides of the DS400/A12 genome, the sequence of the Hpa II-HinfI fragments from each of the strains were compared (see Fig. 6 for fragments sequenced). The sequences of the fragments were found to be identical up to 11 nucleotides away from the Hpa II site (Fig. 7). The sequence divergence in DS400/A12 starts from nucleotide +6125, indicating that the 11 nucleotides preceding the Hap II site are part of a sequence on the other side of the juncture of the DS400/A12 genome. The Hpa II site in DS400/A12, therefore, is located at 68.3 map units and is distinct from the Hpa II site of the DS401 mtDNA at 60.2 units.

The orientation of the cytochrome b exons relative to the wild type physical map confirms previous conclusions about the direction of transcription of the cytochrome b gene (10, 22). The gene is transcribed from the same DNA strand as the genes for subunits 6 and 9 of the ATPase (16, 17, 19) and subunits 2 and 3 of cytochrome oxidase (18, 20). The positions of these genes on the physical map of D273-10B and the direction in which they are transcribed are illustrated in Fig. 8.

**Organization of the Cytochrome b Gene in D273-10B and Other Strains of Yeast**—Several lines of evidence point to substantial differences in the organization of the cytochrome b gene among laboratory strains of S. cerevisiae. The D273-10B strain used in the present study has a somewhat shorter mitochondrial DNA (70 kbp) than do most other strains examined (75 kbp) (23). The smaller genome size in D273-10B is due in part to a 3.4-kbp deletion in the region of cytochrome b (23). The restriction maps shown in Fig. 9 compare the maps of D273-10B and KJ14-4A, a strain whose mitochondrial genome is akin to a number of other well-characterized yeast strains (24). The two restriction maps are identical from 72.4 to 77.8 units. The region between the Eco RI site at 72.4 units and the HincII/Hha I sites at 71 units, however, is approximately 3.4 kbp shorter in D273-10B.

2 S. Bonitz, Columbia University, unpublished studies.
TABLE II
Codon utilization in the cytochrome b and other genes of yeast mtDNA

| Codon | Cyt b | LG.* | Other genes* |
|-------|-------|------|--------------|
| Ala   | GCA   | 11   | 2            |
|       | U     | 10   | 7            |
|       | C     | 0    | 3            |
|       | G     | 1    | 0            |
| Arg   | AGA   | 11   | 10           |
|       | G     | 0    | 0            |
|       | CGA   | 0    | 0            |
|       | U     | 0    | 2            |
|       | C     | 0    | 1            |
|       | G     | 0    | 0            |
| Asn   | AAU   | 17   | 46           |
|       | C     | 3    | 1            |
| Asp   | GAU   | 8-9  | 15           |
|       | C     | 0    | 1            |
| Cys   | UGU   | 3    | 5            |
|       | C     | 1    | 1            |
| Gln   | CAA   | 4-5  | 11           |
|       | G     | 2    | 3            |
| Glu   | GAA   | 5    | 11           |
|       | G     | 0    | 1            |
| Gly   | GAA   | 5    | 1            |
|       | U     | 19   | 13           |
|       | C     | 0    | 0            |
|       | G     | 1    | 1            |
| His   | CAU   | 12-13| 4            |
|       | C     | 0    | 0            |
| Ile   | AUA   | 1    | 6            |
|       | U     | 33   | 30           |
|       | C     | 7    | 2            |
| Leu   | UUA   | 43   | 35           |
|       | G     | 0    | 1            |

| Codon | Cyt b | LG.* | Other genes* |
|-------|-------|------|--------------|
| Lys   | AAA   | 7    | 44           |
|       | G     | 0    | 3            |
| Met   | AUG   | 16   | 2            |
| Phe   | UUU   | 21   | 21           |
|       | C     | 15   | 2            |
| Pro   | CCA   | 9    | 0            |
|       | U     | 10   | 4            |
| Ser   | UCA   | 16   | 2            |
|       | U     | 9    | 8            |
|       | C     | 0    | 0            |
|       | G     | 0    | 0            |
|       | AGU   | 3    | 6            |
| Thr   | ACA   | 5    | 6            |
|       | U     | 7    | 3            |
|       | C     | 0    | 1            |
|       | G     | 0    | 0            |
|       | CUA   | 4    | 2            |
|       | U     | 0    | 3            |
|       | C     | 0    | 0            |
|       | G     | 0    | 0            |
| Trp   | UGA   | 7    | 7            |
|       | G     | 0    | 1            |
| Val   | GUA   | 20-21| 12           |
|       | U     | 12   | 4            |
|       | C     | 1    | 0            |
|       | G     | 3    | 0            |
| Tyr   | UAU   | 18   | 28           |
|       | C     | 3    | 1            |

* Codons found in the reading frame from +763 to +1914.
* These codons were tabulated from the gene sequences of two subunits of the ATPase (16, 19) and two subunits of cytochrome oxidase (18, 20).

TABLE III
Amino acid composition of yeast, N. crassa, and bovine cytochromes b

| Amino acid | Yeast residues | mol % | mol % | mol % | N. crassa residues | mol % | mol % | mol % | Bovine residues | mol % | mol % | mol % |
|------------|----------------|-------|-------|-------|---------------------|-------|-------|-------|----------------|-------|-------|-------|
| Ala        | 22             | 5.7   | 7.2   | 7.2   | 11                  | 2.8   | 4.1   | 2.2   | 46             | 3.0   | 3.8   | 2.7   |
| Arg        | 2              | 7.3-7.5| 8.6   | 7.5   | 10                  | 1.0   | 1.2   | 0.9   | 30             | 1.8   | 3.8   | 2.7   |
| Asx        | 4              | 2.5-3.1| 7.4   | 3.5   | 12                  | 6.5   | 7.1   | 6.6   | 75             | 3.0   | 2.7   | 3.0   |
| Cys        | 8              | 3.1-3.4| 2.4   | 3.1   | 13                  | 1.8   | 1.6   | 1.8   | 75             | 3.0   | 2.7   | 3.0   |
| Glu        | 11             | 10.6  | 8.0   | 9.4   | 41                  | 11.2  | 12.0  | 15.7  | 30             | 9.3   | 6.6   | 4.7   |
| Gly        | 7              | 3.0   | 2.7   | 3.0   | 21                  | 5.4   | 4.0   | 4.0   | 20             | 9.3   | 6.6   | 4.7   |
| His        | 16             | 4.1   | 4.7   | 5.9   | 16                  | 4.1   | 4.7   | 5.9   | 16             | 4.1   | 4.7   | 5.9   |
| Ile        | 28             | 7.3   | 7.9   | 5.9   | 28                  | 7.3   | 7.9   | 5.9   | 28             | 7.3   | 7.9   | 5.9   |
| Leu        | 16             | 18    | 2.7   | 3.0   | 16                  | 18    | 2.7   | 3.0   | 16             | 18    | 2.7   | 3.0   |
| Lys        | 36             | 9.3   | 6.6   | 4.7   | 36                  | 9.3   | 6.6   | 4.7   | 36             | 9.3   | 6.6   | 4.7   |

* Values taken from Ref. 13.
* Values taken from Ref. 14.

MAP UNITS

Fig. 6. Restriction fragments of DS401 and DS400/A12 sequenced to determine the juncture of the DS400/A12 mtDNA. Only parts of the restriction maps of the two DNAs are shown. The arrows show the fragments sequenced. The symbols for the restriction sites are the same as in Fig. 1.

In addition to the differences in the restriction maps of the cytochrome b region, there is evidence from genetic studies and R-looping that strains with the 3.4-kbp insert have a more complex organization of the gene. Several laboratories have concluded that in some yeasts, the cytochrome b gene consists of five exons and four introns. Genetic and restriction mapping of the five exons (box 5/4, 8, 1/9, 2, and 6) and four introns
In the accompanying article (1), we described the properties of a set of ρ− clones with different genetic markers in the cytochrome b gene. The clones were characterized with respect to the physical properties of their mtDNAs and were used to construct a detailed restriction map of the cytochrome b region in S. cerevisiae D273-10B. In the present study, the clones have been further exploited for the purpose of sequencing the cytochrome b gene.

The nucleotide sequence of yeast mtDNA has been obtained from 71.4 to 80.2 units. This span of the genome contains all the cytochrome b markers represented by our collection of mit− mutants. The sequence includes three coding regions that are proposed to be the exons of the yeast cytochrome b gene. The three exons generate a protein that is 50% homologous to bovine cytochrome b based on the first exon codes for 252 to 253 residues from the NH2-terminal end of the protein. The second exon consists of a very short reading frame that codes for a second protein with 384 amino acids. Whether this sequence in fact corresponds to a separate gene is not known at present. There are several features of the sequence that tend to argue against a coding function. Most yeast mitochondrial genes sequenced to date contain 27 to 31% G+C (16–20). In contrast, the G+C content of the intron sequence is only 23%. The amino acid composition of the intron product is also unusual, being very rich in asparagine, lysine, tyrosine, and other amino acids whose codons do not have a G or C. Finally, the reading frame makes use of certain codons such as the CGN family for arginine and UGG for tryptophan that have not appeared in any of the identified yeast mitochondrial genes. Despite these reservations, we cannot exclude the possibility that the sequence might serve some real function, perhaps by coding for a protein necessary for processing of the cytochrome b messenger. This possibility is currently being investigated.

The sequence of the second intron has an A+T-rich character. This intron also has a short sequence with a high G+C content and several Hpa II and Hae III sites. Part of this sequence has been further examined for the purpose of sequencing the cytochrome b gene.

Fig. 7. Nucleotide sequence of DS401 and DS400/A12 in the region of the juncture. The sequence of DS400/A12 has been numbered according to the convention used in Fig. 2.

Fig. 8. Physical map of the mitochondrial genome of D273-10B. The cap (chloramphenicol) and par (paromomycin) resistance loci are provided for orientation. The direction of transcription is indicated by the arrow.

Fig. 9. Organization of the cytochrome b genes in S. cerevisiae D273-10B and KL14-4A. The restriction map in the cytochrome b region is based on the published data of Grivell et al. (24).

TABLE IV
Exon and intron regions of the cytochrome b gene in S. cerevisiae D273-10B and KL14-4A

| Exon | Intron | Size (bp) | Locus | Exon Size | Locus |
|------|--------|----------|-------|-----------|-------|
| b1   | b2     | 756-759  | b2    | 600       | box 5/4 |
|      | b3     | 345-348  | b3    | 600       | box 6  |

a Taken from Ref. 25.
b Taken from Ref. 8.

(box 3, 10, and 7) (8, 10) agree with measurements of the duplex and loop regions observed in electron micrographs of RNA-DNA hybrids (25). The most recent estimates of the sizes of the exons and introns are presented in Table IV. These data suggest that the b1 exon of D273-10B is split into three smaller exons in KL14-4A and related strains. The two extra introns in KL14-4A have been mapped between the Eco RI site (72.4 units) and HincII site (71 units), the region where the two strains differ by 3.4 kbp. Assuming that the general pattern of restriction sites is an index of overall sequence homology, the rest of the gene appears to have a similar organization in both strains. As shown in Fig. 9, the extra 3.4 kbp in KL14-4A can be accounted for by the two introns located between the Eco RI and HincII sites.

DISCUSSION

In the accompanying article (1), we described the properties of a set of ρ− clones with different genetic markers in the cytochrome b gene. The clones were characterized with respect to the physical properties of their mtDNAs and were used to construct a detailed restriction map of the cytochrome b region in S. cerevisiae D273-10B. In the present study, the clones have been further exploited for the purpose of sequencing the cytochrome b gene.

The nucleotide sequence of yeast mtDNA has been obtained from 71.4 to 80.2 units. This span of the genome contains all the cytochrome b markers represented by our collection of mit− mutants. The sequence includes three coding regions that are proposed to be the exons of the yeast cytochrome b gene. The three exons generate a protein that is 50% homologous to bovine cytochrome b based on the sequence of the bovine gene.3

The first exon codes for 252 to 253 residues from the NH2-terminal end of the protein. The second exon consists of a very short reading frame that codes for a second protein with 384 amino acids. Whether this sequence in fact corresponds to a separate gene is not known at present. There are several features of the sequence that tend to argue against a coding function. Most yeast mitochondrial genes sequenced to date contain 27 to 31% G+C (16–20). In contrast, the G+C content of the intron sequence is only 23%. The amino acid composition of the intron product is also unusual, being very rich in asparagine, lysine, tyrosine, and other amino acids whose codons do not have a G or C. Finally, the reading frame makes use of certain codons such as the CGN family for arginine and UGG for tryptophan that have not appeared in any of the identified yeast mitochondrial genes. Despite these reservations, we cannot exclude the possibility that the sequence might serve some real function, perhaps by coding for a protein necessary for processing of the cytochrome b messenger. This possibility is currently being investigated.

The sequence of the second intron has an A+T-rich character. This intron also has a short sequence with a high G+C content and several Hpa II and Hae III sites. Part of this

3 Dra. F. Sanger and B. Barrell, private communication.
cluster is inversely homologous to a 14-nucleotide sequence further upstream in the intron. These inverted repeated sequences may be involved in some initial processing event leading to the removal of part of the intron sequence.

Although the yeast cytochrome b gene has been found to have a mosaic structure in all the yeast strains examined, the number of coding elements appears to vary depending on the strain. *S. cerevisiae* D273-10B has a relatively short mitochondrial genome, lacking a 3.4-kbp insert in the cytochrome b region. This extra 3.4-kbp sequence is present in most other yeast strains (8–10, 25). Van Ommen et al. (25) have presented evidence that the 3.4-kbp insert occurs in a region of the cytochrome b gene where there are two long introns of approximately 2 and 1.4 kbp. The structure of the cytochrome b gene in D273-10B can be easily reconciled with the five-exon and three-intron structure proposed for other yeast strains if it is assumed that exon B1 is further fragmented into three smaller exons. In fact, the sizes of the first three cytochrome b exons in KL14-4A measured in RNA-DNA hybrids (25) agree reasonably well with the single b1 exon of D273-10B reported here.

The definition at the DNA sequence level of the exon and intron regions makes it possible to study the location of mitochondrial mutations that abolish the synthesis of cytochrome b. The DNA sequences of the low complexity p clones have already provided some information about the location of 10 different cytochrome b mutations from our collection of mit mutants. Based on the genotypes of the clones sequenced, all the cob2 mutations have been found to fall within the first cytochrome b exon. The cob1 mutants, however, represent a more heterogeneous group. While some cob1 mutants appear to be a longer exon (b3), other mutants map either in the short middle exon or in the flanking introns. At least one mutation (M6-200) falls completely inside of the intron sequence between b2 and b3. Such strains will be useful in future studies on the mechanism by which the exons are spliced in the messenger RNA.

Cytochrome b is an important electron transfer carrier of the mitochondrial respiratory chain. Spectral and kinetic evidence indicate the presence of two distinct b type cytochromes (b1 and b2) (28, 29). The question of whether these are physically different proteins has never been resolved. Most of the genetic data, however, are consistent with the existence of a single cytochrome b gene. This has been confirmed by the sequence of the gene reported here. Furthermore, the primary structure of the yeast cytochrome b now allows the chemical basis for the two spectral species to be tested experimentally.

REFERENCES
1. Nobrega, F. G., and Tzagoloff, A. (1980) *J. Biol. Chem.* 255, 9821–9827
2. Maxam, A. M., and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. U. S. A.* 74, 560–564
3. Tzagoloff, A., Akai, A., and Foury, F. (1976) *FEBS Lett.* 65, 391–395
4. Wesolowski, M., and Fukuhara, H. (1979) *Mol. Gen. Genet.* 170, 261–275
5. Macino, G., Coruzzi, G., Nobrega, F. G., Li, M., and Tzagoloff, A. (1979) *Proc. Natl. Acad. Sci. U. S. A.* 76, 3784–3786
6. Barrell, B. G., Bankier, A. T., and Drouin, J. (1979) *Nature* 282, 189–194
7. Li, M., and Tzagoloff, A. (1979) *Cell* 18, 47–53
8. Slonimski, P. P., Pajot, P., Jacq, C., Foucher, M., Perrodin, G., Kochko, A., and Lamouroux, A. (1978) in *Biochemistry and Genetics of Yeast: Pure and Applied Aspects* (Bacila, M., Horecker, B. L., and Stoppani, A. M., eds) pp. 399–368, Academic Press, New York
9. Alexander, N. J., Vincent, R. D., Perlman, P. S., Miller, D. H., Hanson, D. K., and Mahler, H. R. (1979) *J. Biol. Chem.* 254, 2471–2479
10. Haid, A., Schwyeney, R. J., Bechmann, H., Kaudewitz, F., Solioz, M., and Schatz, G. (1979) *Eur. J. Biochem.* 94, 451–464
11. Lin, L.-F., Clejan, L., and Beattie, D. S. (1978) *Eur. J. Biochem.* 87, 171–179
12. Katun, M. B., Pool, L., and Groot, G. S. P. (1976) *Eur. J. Biochem.* 68, 35–105
13. Weiss, H., and Ziganke, B. (1976) in *Genetics and Biogenesis of Chloroplasts and Mitochondria* (Bücher, Th., Neupert, W., Sebald, W., and Werner, S., eds) pp. 259–266, North-Holland, Amsterdam
14. Von Jagow, G., Schagger, H., Engel, W. D., Machleidt, W., Machleidt, I., and Kolb, H. J. (1978) *FEBS Lett.* 91, 121–125
15. Tzagoloff, A., Foury, F., and Akai, A. (1979) *Mol. Gen. Genet.* 149, 33–42
16. Macino, G., and Tzagoloff, A. (1979) *J. Biol. Chem.* 254, 4617–4623
17. Hensgens, L. A. M., Grivel, L. A., Borst, P., and Bos, J. L. (1979) *Proc. Natl. Acad. Sci. U. S. A.* 76, 1663–1667
18. Coruzzi, G., and Tzagoloff, A. (1979) *J. Biol. Chem.* 254, 9324–9330
19. Macino, G., and Tzagoloff, A. (1980) *Cell* 20, 507–511
20. Thalenfeld, B., and Tzagoloff, A. (1980) *J. Biol. Chem.* 255, 6173–6180
21. Tzagoloff, A., Macino, G., Nobrega, M., and Akai, A. (1980) *Curr. Genet.*, in press
22. Claissen, M. L., Spyrakakis, A., Wamhier-Kluppel, M. L., Pajot, P., and Slomiński, P. P. (1978) in *Biochemistry and Genetics of Yeast: Pure and Applied Aspects* (Bacila, M., Horecker, B. L., and Stoppani, A. M., eds) pp. 369–390, Academic Press, New York
23. Morimoto, R., and Rabinowitz, M. (1979) *Mol. Gen. Genet.* 170, 25–48
24. Grivell, L. A., Arnborg, A. C., Boer, P. H., Borst, P., Bos, J. L., Van Bruggen, E. F. J., Groot, G. S. P., Hecht, N. B., Hensgens, L. A. M., Van Ommen, G. J. B., and Tabak, H. F. (1979) *ICN-UCLA Symp. Mol. Cell. Biol.* 18, 305–324
25. Van Ommen, G. J. B., Groot, G. S. P., and Grivell, L. A. (1979) *Cell* 18, 511–523
26. Groot, G. S. P., Van Harten-Looibroek, N., and Kreike, J. (1978) *Biochim. Biophys. Acta* 517, 457–463
27. Beattie, D. S., Chen, U.-S., Clejan, L., and Lin, L.-F. H. (1979) *Biochemistry* 18, 2400–2406
28. Chance, B., Wilson, D. F., Dutton, P. L., and Erucinska, M. (1970) *Proc. Natl. Acad. Sci. U. S. A.* 66, 1175–1179
29. Rieske, J. S. (1976) *Biochim. Biophys. Acta* 456, 195–247