Assembly of the Mitochondrial Membrane System

STRUCTURE AND NUCLEOTIDE SEQUENCE OF THE GENE CODING FOR SUBUNIT 1 OF YEAST CYTOCHROME OXIDASE

Susan G. Bonitz, Gloria Coruzzi, Barbara E. Thalenfeld, and Alexander Tzagoloff
From the Department of Biological Sciences, Columbia University, New York, New York 10027
Giuseppe Macino
From the Istituto di Fisiologia Generale, Universita di Roma, Rome, Italy

The oxi3 locus of yeast mitochondrial DNA has been sequenced in Saccharomyces cerevisiae D273-10B. The sequence was obtained from the mitochondrial genomes of a series of cytoplasmic "petite" mutants selected for the retention of genetic markers in the oxi3 locus. The oxi3 locus has been ascertainment to code for Subunit 1 of cytochrome oxidase. The Subunit 1 gene is 9,979 nucleotides long, consisting of seven to eight exons that account for only 16% of the gene sequence. The coding sequences have been identified on the basis of protein sequence homology with Subunit 1 of human cytochrome oxidase. The yeast Subunit 1 is 510 amino acid residues long and has a molecular weight of 58,000. In addition to the exon sequences, the Subunit 1 gene contains six to seven introns. The first four introns have long reading frames that are continuous with the exon coding sequences. These reading frames are potentially capable of coding for basic proteins with molecular weights ranging from 30,000 to 80,000. The first two introns of the gene have a sequence homology of 50%, while the reading frame of the fourth intron is 70% homologous with an intron of the apocytochrome b gene.

At least five stable transcripts have been found by Northern blot hybridizations with single-stranded DNA probes containing either exon or intron sequences. A 1.9-kilobase transcript hybridizes only with probes from the exon regions of the gene. This RNA species has been tentatively identified as the fully processed messenger of Subunit 1. Other transcripts are detected with intron probes. Three transcripts with sizes of 2.5, 2.4, and 0.85 kilobases appear to be stable excision products from the first, second, and fifth introns.

In the companion paper, (1) we reported the properties of a series of p− clones of Saccharomyces cerevisiae selected for the retention of markers in the oxi3 locus. The mitochondrial DNA (mtDNA) segments of the clones spanned the wild type sequence from 42 to 58 map units. This region has been shown in other studies to contain the oxi3 locus that codes for Subunit 1 of cytochrome oxidase (2, 3).

This research was supported by Research Grant GM25250 and a National Research Service Award (to B. E. T.) 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.

*Abbreviations used: p−, cytoplasmic "petite" mutants with long deletions in mitochondrial DNA; SDS, sodium dodecyl sulfate; kb, kilobase.

The mitochondrial genomes of the p− mutants have been used in the present study to sequence the oxi3 locus. A continuous nucleotide sequence of 10 kb has been obtained from approximately 43.5 to 58 map units. The sequence contains the entire gene of Subunit 1. The gene is 9,979 nucleotides long and is composed of seven to eight exons and six to seven introns. In addition to the exon-coding regions, the Subunit 1 gene has four other long reading frames. These are located in the introns and are potentially capable of coding for proteins with molecular weights of up to 80,000. One of the oxi3 introns has extensive sequence homology with the first intron of the apocytochrome b gene (4). The two genes also exhibit sequence homologies in some of the exon-intron boundaries, suggesting a common mechanism of splicing of the transcripts. The transcripts originating from the oxi3 region have been studied by Northern blot hybridization of total mitochondrial RNA to 5′ end-labeled DNA probes containing either exon or intron sequences. The results of these hybridizations indicate the messenger of Subunit 1 to be 1.9 kb long. The 1.9-kb transcript is the smallest and most abundant RNA detected by exon probes.

MATERIALS AND METHODS

Strains—The mtDNAs used for the sequence analysis were purified from the p− clones DS6, DS6/A400, DS6/A401, DS6/A402, DS6/A407, DS6/A422, and DS6/A462. The isolation of the clones and physical properties of their genomes have been reported in the accompanying paper (1).

Purification of mtDNA—Yeast was grown at 30°C to early stationary phase in 2% glucose, 1% yeast extract, 1% peptone. The procedures for the preparation of yeast mitochondria and purification of mtDNA have been described (1).

DNA Sequencing—Restriction fragments were labeled at the 5′ ends with [γ-32P]ATP (2000 to 3000 Ci/mmol, New England Nuclear Corp., Boston, MA) in a T4 polynucleotide kinase-catalyzed reaction (5). The labeled fragments were denatured in 90% formamide and strand-separated on 4%, 6%, or 8% polyacrylamide gels at 4°C (6). The labeled single strands were eluted from the gels and sequenced by the chemical derivatization method of Maxam and Gilbert (5).

Preparation of Mitochondrial RNA—The wild type strain of S. cerevisiae D273-10B/A1 (6) was grown to early stationary phase in 1 liter of a medium containing 2% galactose, 1% yeast extract, and 1% peptone. Approximately 10 g of wet weight cells were obtained from 1 liter of medium. The cells were washed in 1.2 M sorbitol and suspended in 30 ml of a solution containing 1.2 M sorbitol, 0.08 M KPO4, pH 7.5, 0.16 M 2-mercaptoethanol, 1 mM EDTA, and 1 mg/ml of Zymolase (Kurin Brewery Ltd., Japan). Virtually complete conversion of the cells to protoplasts occurred after 45 to 60 min of incubation at 32°C. The protoplasts were collected by centrifugation at 6,000 × g for 10 min and washed two times with 150 ml of 1.2 M sorbitol. The washed protoplasts were suspended in 75 ml of 0.7 M sorbitol, 10 mM Tris/acetate, pH 7.5, 1 mM EDTA, 0.1 mg/ml of bovine serum albumin, and 5 mM vanadium adenosine. The vanadyl adenosine was used as a ribonuclease inhibitor (7). The suspension was homog-
enized for 30 s in a Waring Blender and centrifuged two times at 1,000 \( \times g \). The supernatant was then centrifuged at 30,000 \( \times g \) for 10 min to pellet the mitochondria. The mitochondria were washed three times with 30 ml of 0.5 mM sorbitol, 0.05 mM Tris/acetate, pH 7.5, 1 mM EDTA. The washed mitochondria were lysed in 2 ml of 2% sodium dodecyl sulfate and immediately mixed with an equal volume of water-saturated phenol. The two phases were separated by centrifugation at 5,000 \( \times g \) for 10 min, and the water phase was dialyzed for 17 h against 2 liters of cold distilled water. Little degradation was observed under these conditions even after 1 week of storage.  

Northern Blots—Yeast mitochondrial RNA was denatured for 5 min at 60°C in 10 mM methylmercuric hydroxide, and 1 to 2 \( \mu g \) were loaded on 0.6-cm-wide wells of a 1% or 1.2% agarose gel containing 10 mM methylmercuric hydroxide. Electrophoresis was carried out at 5 V/cm for 4 to 5 h. The gel was treated with 14.3 mM 2-mercaptoethanol, stained with 5 \( \mu g \)/ml of ethidium bromide, and photographed. Prior to blotting, the gel was soaked in 7 mM iodoacetate buffer, pH 4.0.

The preparation of diazobenzyloxymethyl paper and the conditions for the transfer of RNA from the agarose gel to the paper were the same as described by Alwine et al. (8). Single-stranded DNA fragments labeled with \( ^{32}P \) at the 5' ends were hybridized to the diazobenzyloxymethyl strips (9). The diazobenzyloxymethyl strips were exposed to Kodak XR-1 film with an intensifying screen for 1 to 3 days at \(-80^\circC \).  

RESULTS  

**Sequencing Strategy**—The \( \rho^- \) clone DS6 has a mitochondrial genome with a unit length of 16.5 kb (1). Even though this clone was shown to contain the entire \( oxi3 \) gene, the large genomic size precluded its use for DNA sequencing. Most of the DNA sequences were therefore determined on the smaller genomes of the derivative clones which ranged from 2.3 to 6.1 kb in size, spanning the sequence from 42 to 58 map units. The regions of DNA (expressed in map units) sequenced in each clone are indicated in Table I. Since DS6/A442 had been shown to have a small internal deletion and inversion in its DNA, the sequence of the region retained in this clone was also obtained from a preparative Bgl II-Bam HI fragment from DS6.  

All the sequence data were obtained from single-stranded fragments labeled at the 5' ends. By choosing appropriate combinations of restriction endonucleases, it was possible to generate a sufficiently discrete number of fragments for subsequent resolution of the labeled mixture on the strand separation gels. All the restriction sites except those for \( Alu \) I and \( Rsa \) I had been mapped in each genome, thus reducing the number of different labelings necessary to complete the sequence.  

**Nucleotide Sequence of the \( oxi3 \) Locus**—The restriction map of the \( oxi3 \) locus is shown in Fig. 1. The sites used for labeling and the directions and extents to which the labeled fragments were read on the sequencing gels are indicated by the arrows. All the restriction sites used for labeling were crossed except for the \( Alu \) I site at 51.8 units. We therefore cannot be sure that the sequence of a small \( Alu \) I fragment may not have been missed. Most of the sequences were confirmed either from the complementary strands or from identical sequences obtained from different genomic DNAs.

**Table I**  

<table>
<thead>
<tr>
<th>Clone</th>
<th>Region sequenced</th>
<th>map units</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS6/A401</td>
<td>43.3-45.4, 47.2-50.7</td>
<td></td>
</tr>
<tr>
<td>DS6/A400</td>
<td>44.1-47.6</td>
<td></td>
</tr>
<tr>
<td>DS6/A402</td>
<td>47.1-51.7</td>
<td></td>
</tr>
<tr>
<td><strong>Bgl II-Bam HI</strong></td>
<td>50.6-53.2</td>
<td></td>
</tr>
<tr>
<td>DS6/A422</td>
<td>50.8-51.9, 52.8-58.2</td>
<td></td>
</tr>
<tr>
<td>DS6/A407</td>
<td>53.7-57.4</td>
<td></td>
</tr>
<tr>
<td>DS6/A462</td>
<td>50.2-51.9, 52.1-53.7</td>
<td></td>
</tr>
</tbody>
</table>

**FIG. 1. Physical map of the \( oxi3 \) locus.** The restriction fragments used for DNA sequencing are indicated by the arrows. The extent to which the sequences were read is represented by the lengths of the arrows. The map units are shown on the inner circle. The following symbols are used for the restriction sites: \( \Delta, \) \( HindII, \) \( \Delta, \) \( HpaII, \) \( \Delta, \) \( HaeII, \) \( \bullet, \) \( TaqI, \) \( \bullet, \) \( MboI, \) \( \bullet, \) \( MboII, \) \( \bullet, \) \( AluI, \) \( \circ, \) \( PvuII, \) \( \circ, \) \( HincII, \) \( \circ, \) \( HindIII, \) \( \circ, \) \( EcoRI, \) \( \circ, \) \( EcoRII, \) \( \circ, \) \( HhaI, \) \( \circ, \) \( RsaI, \) \( \bullet, \) \( HphI, \) \( \bullet, \) \( BglII, \) \( \circ, \) \( BamHI, \) \( \bullet, \) \( HpaI. \)  

As indicated in the previous paper, the restriction maps of the different \( \rho^- \) mtDNA segments were consistent with each other except for the DS6/A462 clone. This was confirmed by the DNA sequences which were identical in clones with inclusive or overlapping regions.  

A continuous sequence from 43.5 to 58 units is presented in Fig. 2. The sequence includes the region of DS6 where all the known \( oxi3 \) markers have been mapped (1). The sequence starts with an adenine + thymine-rich region. Although not shown, the adenine + thymine-rich sequence proceeds for at least 1 kb further upstream to the deletion end point of the DS6/A401 genome at approximately 42 map units. The ATG initiation at nucleotide +1 starts an open reading frame that continues for 2503 nucleotides before reaching a TAA termination codon. The amino acid sequence encoded in this reading frame is shown above the DNA sequence in Fig. 2. The assignment of codons is based on the universal code except for the termination codon UGA, which has been translated as tryptophan (10–12), and the CUN codons of leucine that have been shown to code for threonine in yeast mitochondria (13, 14). Another long reading frame occurs some 100 nucleotides downstream from the terminator of the first coding sequence. This reading frame is also very long, comprising 2399 nucleotides. There are four additional regions of variable lengths potentially capable of coding for proteins. These are seen between nucleotides +5073 and +6211, +6697 and +8148, +8169 and +8642, and +9422 and +9979. The last reading frame ending with an ochre terminator at nucleotide +9979 is followed by a long adenine + thymine-rich sequence only part of which is shown in Fig. 2. The six reading frames account for almost the entire sequence of the \( oxi3 \) locus.

**Identification of the Subunit 1 Exons**—Subunit 1 of yeast cytochrome oxidase migrates with an apparent molecular weight of 40,000 when electrophoresed on SDS-polyacrylamide gels (15, 16). A co-linear gene coding for a protein of
Sequence of Gene Coding for Subunit 1 of Cytochrome Oxidase

The nucleotide sequence of the intron reading frames is indicated in lowercase letters. The amino acid sequences of the intron reading frames are shown in lowercase letters.

Fig. 2. Nucleotide sequence of the oxi3 gene. The sequence of the nontranscribed strand is shown. The exon reading frames have been translated into the protein sequences shown in capital letters. The amino acid sequences of the intron reading frames are indicated in lowercase case. All the restriction sites are marked. The sequence starts at approximately 43.5 and ends at 58.2 map units.
FIG. 2—Continued
Sequence of Gene Coding for Subunit I of Cytochrome Oxidase

FIG. 2—Continued
**Sequence of Gene Coding for Subunit 1 of Cytochrome Oxidase**

**lys** asp asp tyr leu glu gly arg met ile lys ile asn arg lys glu ile thr ile lys thr lys his phe
AAA GAT GAT TAT TTA GTA GGA AGT ATG ARG ATA AAT AGA AAA CAA ATT ACT TCT TGT AAA ACA ACG TAT CAT TTT

lys val his gln gly lys tyr asn gly pro gly leu och
+4900
AAA GTT CAT CAA GGT AAA TAT AAT CCT GCC GGT TTA TAA TAA
TAA TAT TAT TCT CCC GGC GCC GCC GCC GCC GCC GCC GCC GCC GCC GCC GCC GCC GCC GCC GCC GCC GCC GCC GCC GCC

met glu ser arg met ile trp lys tyr his val arg phe gly glu gly ser phe ile trp
+5060
TTATATGCGGTTA ATG GAG AGC GCT ATG ATA TGA AAG TAT CAC GTA CGG TCC GTA GAG GGC TCT TTT ATG TGA

lys arg tyr glu ser asn asn asn asn gln val ile glu lys glu tyr asn leu leu leu tyr asp
+5210
AAA GGA TAT GAA AGT AAT AGT CAA GTA ATA AAG AAA GTA TTA AAA TIA AAT TAT GAT

lys leu gly pro tyr leu ala gly leu gly gly gly asp gly ser ile thr val gln asn ser ser ile lys
+5285
AAG TGG GGA CCT TAT TTA GGA GAT GGA TTT AAT GAA TGT ATC CTA ATT CAA TGT CTC TAT ACA AAA

lys ser lys tyr arg pro leu ile val val phe lys leu glu asp ile gln lys lys met ile lys ser lys tyr
+5360
AAA TCT AAA TAT AAG CCG TTA ATT ATT CTA GAA GAT TTA GAA GAA GGT TCT ATT CTA TAT TAT ATT

lys trh lys cys gly lys lys ile lys arg asn tyr val leu trp ile lys ile lys lys
+5435
TAA ACT AAA TGT GAA AAA GGT TAT AAA AAT ATT CGT AAT TAT GTA TTA CCT ATT CAT TGT AAA AAT

lys tyr trh leu asp asp ile lys trh ile lys cys gln leu lys ser lys lys lys lys
+5510
AAA AGA TAT GAA GGT AAT TAT AGA CAA TTA TAT ATT AAA TAA AAA TIA AAT TAT GAT

lys phe ile lys asp ile asp ile asp ile asp ile asp ile asp ile
+5985
ATT AAT TAT AAT ATT ATT AAT ATT CAA CTA AAT AAA AAT ATT ATT AAA ATT AAT AAA

lys ile ile asp ile lys arg ser ile asp ile ile ile ile ile ile ile ile ile ile ile
+6425
AAT TAA ATA GAT GGT AAA AAT ATT ATT TCT TTA GAT TTA AAT TAT ATT

Sequence of Gene Coding for Subunit 1 of Cytochrome Oxidase

leu lys asp tyr asn lys met asn tyr leu leu pro leu ile ile gly ala thr asp
+6675 tga gagatatgtcacaatatata ttt aac gat tat aat aac atg aac tac tat tta tta cca tta ata att gga gct aca gat

ath

thr ala phe pro arg ile asn asn ile ala phe thr trp val leu pro met gly leu val cys leu leu val thr ser thr
+6754 aca gca ttt cca aga att aat aac atg cct ttt tga tta cct atg ggg tta tga tgt gtt gtt aca tca act

leu val glu ser gly ala gly thr gly thr gly trp thr val tyr pro pro leu ser ser ile gln ala his ser gly pro
+6829 tta gta gaa tca ggt gct get aca ggg tga act gct tta gtt gta cca tta tca tca tca att cag tca cat tca gca cct

his

ath

ser val asp leu ala ile phe leu ala his leu thr ser ile ser leu leu gly ala ile asn phe ile val
+6904 gat gga gat tgt gtt tgt cca tta cat tta aca tca tca tca tta tgt gac gct att tgt ttc att gta

thr thr leu asn met arg thr asn gly met thr met his lys leu pro leu phe val trp ser ile phe ile thr
+6979 caa tca tta att aga aga aat ggt atg aga atg cat aat cca cta tta tga tca tca att ttc att aca

ath

ala phe leu leu leu leu ser pro val leu ser pro ile leu thr met leu met leu asp arg asn phe asn
+7054 ggg ttc tta tta tca tta cct gt tca ttc gct get aca tca tca tca tgt att aca tgc act tac act

thr ser phe phe glu val ala gly asp pro ile leu tyr gly his leu phe trp phe phe gly gln thr
+7129 act tca tca tca gaa gta gca gga ggt gct gcc cca ttc tta cca gaa gat cat tca ttc tgt ttt tca cca cct

his

ath

val ala thr ile ile ile leu ile ile thr asp asp met his phe ser lys cys trp lys leu leu lys trp
+7204 tgt gcc ctt att att att tca tta aca tca cta tgt gat ttt tca tta aaa taa tca

he11

ile thr asn ile ile ser thr leu phe phe ala leu phe val lys ile phe ile ser tyr asn asp gln gln asp
+7279 att aca ata att ata att cta tta ttt tca cca tca tta gtt gta cca tca tgt att tgt gtt gat gta
tga

lys ile ile asn ile thr ile ile leu lys asp asn ile lys arg ser ser gly thr thr arg lys ile leu
+7354 cag ata ata aat cat tta tta aar aag gat att att aag aca tcc tca gac act aca aag aca att aat

ath

asn ser ile ile lys phe asp glu trp ile ala gly leu ile asp gly asp gly tyr phe phe ile val ser
+7429 att tca ata aat aaa ttt aat gag tca tta ggt gca ttt ttt ggt att gat ggt ggt gtt ttt gtt gtt att gta

ath

lys lys tyr val ser leu glu ile thr val leu glu asp ile ala leu lys ile ile gln ile gln lys phe
+7504 aag aaa tat gta tca tca gta gaa att gta gca tta gaa gat gaa ata gct tta aar gaa at aat ttt

ath

gly gly ser ile leu arg ser gly val lys ala ile arg tyr arg ile ile ile lys
+7579 ggt gct tct att aca tta aga tca ggt gaa ata gtt att aga att aga tta ctt aat aac act gtt gta att aca

ath

leu ile asn ala val asn gly asn ile ile arg ser thr lys arg leu val gln phe asp lys lys val cys ile ile leu
+7654 tta att aat gca gtt aat ggt aat att aga aat act aca aag tta tta gtt cca tta att aat aaa taa gtt tgt att tta

ath

gly ile asp phe ile pro ile lys leu thr lys asp asn ser thr phe phe val gly phe phe asp asp gly
+7729 ggt atg tat att tat cca att aat tta act aar act gat aag taa gta gtt gat gtt gat gtt gat gtt ggt

ath

thr ile thr ser phe lys asp his pro glu leu thr lys thr val thr thr lys tyr leu gluu asp val
+7804 gca att att tca ttt aar aat aat cat cct caa tca aca aca act gta act aat aar aca tta gaa gat gta

gly

ath

gln glu tyr lys asn ile leu gly glu ile asp phe phe lys phe lys asp gly phe lys asp trp ser ile
+7879 caa gaa taa aat att tta gtt ggt gtt aat att att ttt gtt gaa tca caa aat ggt tat taa tca tca ctt att

ath

gln ser lys ile asp leu asn ile ile asp asp ile thr ile lys ile asp ile asp thr ile lys ile asp
+7954 cca tca aaa gat atn gta tta aat att gat att aat aca at aat aca tca aga aca ctt aca aar aar at ttt

ath

leu leu tyr leu ser lys glu phe tyr asp leu asp leu lys glu leu lys ala tyr asp ser asp ser ile gln
+8029 aar tta ttt taa att gta gaa ttt taa aat gaa taa aar act aar cct aat gtt aar tta cca cct cct cct gtt
cat ctt gtt
cat cta cca
ttt

his

ath

tyr lys ile thr leu phe asp phe lys ile lys trp lys asp lys asp lys ser asp asp ile lys
+8104 tat aar gca tta att ttt gaa aar taa taa gaa aar aar aar gta tat aar tat aar ttt taa gaa cca gtt gtt att att cct gta

ath

tyr asp ile thr asp lys his pro glu val tyr ile leu ile ile pro gly phe gly ile ile ser his
+8179 att at tat tat aar aac cct gaa gta tat att att tat att ctt gaa gtt gtt att att cct gta ctc ctc ctt

eco

ath

val val ser thr tyr ser lys lys pro val phe gly glu ile ser met val tyr ala met ala ser ile gly leu
+8259 gta gta tca aca ttt ttt aar cct gtt gtt ttt gtt gtt ttt gtt gtt gtt gtt gaa gta

ath

Fig. 2—Continued
Sequence of Gene Coding for Subunit 1 of Cytochrome Oxidase

**FIG. 2** Continued
The nucleotide sequence of the gene. The eight exons are indicated by the arrows above the one-letter amino acid sequence. The underlined sequences correspond to those regions of the protein where the homology with the human protein was 50% or greater.

The first exon (A1) starts with the ATG initiator at nucleotide +188 or +171. This exon codes for the first 55 to 56 amino acids from the NH2-terminal end. The homology with the human sequence averages 45%. The second exon (A2) is a short sequence (+2620 to +2655) coding for the next 10 to 12 amino acids and exhibiting 50% homology with the human protein. The third exon (A3) also codes for a short segment of the protein (+5168 to +5206). The 12 to 13 amino acids encoded in A3 are 60% homologous with the human Subunit 1. Exons A4 and A5 (+6718 to +7197 and +8203 to +8618) code for a total of 296 amino acids. This is the most conserved region, being almost 70% homologous with human Subunit 1. The sixth exon (A6) starts at nucleotide +9506. It is difficult to ascertain where this exon ends. There are 43 identities in the 83 amino acids encoded between nucleotides +9506 and +9754. We have somewhat arbitrarily indicated this exon to end at nucleotide +9781 since this gives the best fit with the human sequence. The last 18 amino acids of exon A6 have only two identities. The substitutions in the other 16 residues, however, involve acidic for acidic or neutral for neutral amino acids.

The last exon (A8) starts at nucleotide +9879 and ends with the COOH-terminal serine at nucleotide +9976. The human and yeast protein sequences in this region are only 25% homologous. The gap between what we have defined as the end of exon A6 and the beginning of the last exon is a 69-nucleotide-long sequence that can code for an additional 23 amino acids. Since there are only 10 amino acid residues in the comparable region of the human protein, it is possible that there is an additional small exon and two short introns in this sequence. We have not been able to find any significant homology in this region even by shifting the registers in the yeast sequence. Alternatively, the yeast Subunit 1 could have an insertion of 13 residues. It should be noted that the 69-nucleotide sequence contains three lysine (AAA) and seven asparagine (AAU) codons. Since these codons have been found to occur at high frequency in the intron coding frames of the apocytochrome b gene (4) and of the Subunit 1 gene reported here, we tend to favor the former possibility, namely that there is still a seventh short exon in the sequence. We have tentatively designated the seventh exon to extend from nucleotides +9830 to +9859. The two amino acid identities in the sequence include a proline which is usually found to be conserved. Until the exon-intron boundaries are precisely defined at the level of the messenger sequence, it is still not clear whether there are 7 or 8 exons.

The primary structure of the yeast subunit 1 encoded in the oxi3 locus based on the best fit with the human subunit is shown in Fig. 3. The reported sequence assumes that there are no major insertions or deletions between the two proteins. Although this cannot be ruled out at present, it is significant that the sequence homology of 68% in the fourth and fifth exons extends for 296 amino acids with only a single amino acid deletion in the yeast protein. This fact tends to support the idea that the two proteins are highly conserved both in terms of size and primary structure.

Based on our best approximation of the Subunit 1 sequence, the protein consists of 510 residues. The amino acid composition of the protein is presented in Table II. The composition derived from the gene sequence agrees reasonably well with two independently reported amino acid compositions of the \( \text{b} \) subunit. For the yeast Subunit 1 (21, 22), the molecular weight estimated from the amino acid content is 55,933. This value is slightly lower than measured by SDS-gel electrophoresis (15, 16). Similar discrepancies in size have been found for Subunit 3 of cytochrome oxidase (23) and cytochrome b (4). Although these differences are most likely due to under-

---

2 B. Barrell and F. Sanger, personal communication.
estimation of the true size of this class of hydrophobic proteins on SDS gels, it cannot be excluded that the mature proteins are smaller than indicated by their gene sequences due to post-translational processing.

Transcripts from the oxi3 Region of mtDNA—The restriction map of the oxi3 locus, together with the physical locations of the Subunit 1 coding sequences, has made it possible to examine the stable RNA species transcribed from this region. We were particularly interested in seeing whether DNA probes with exon or intron sequences could be used to identify the messenger of Subunit 1.

To study the oxi3 transcripts, the mitochondrial genomes of the p clones used for the sequence analysis were digested with different restriction endonucleases and the fragments labeled at the 5' termini with [32P]ATP. The labeled mixtures were separated into the single strands on polyacrylamide gels, eluted, and used as hybridization probes against Northern blots of total mitochondrial RNA. Prior to use, the DNA probes were sequenced for identification of strand (sense or antisense) and origin on the physical map. Since preliminary experiments indicated that the oxi3 transcripts were copied from the sense strand, only single-stranded probes complementary to this strand were used for the hybridizations.

The relation of the probes to the exon and intron regions of the gene is shown in Fig. 4. Four of the probes contained sequences confined to either exons A1, A4, A5, or A8. Because of the distribution of restriction sites, it was not possible to isolate probes containing pure exons A2 or A3. A number of probes were also obtained from each of the introns except those present in the COOH-terminal end of the gene where there are still uncertainties about the exon-intron boundaries.

The results of the hybridizations are summarized in Table III, and some representative autoradiographs of Northern blots are shown in Fig. 5. The most abundant transcript detected with each of the pure exon probes has a size corresponding to 1.9 kb. Although the exon probes also hybridize to larger transcripts, these are present in considerably lower concentrations. The higher molecular weight transcripts have not been studied in detail but could represent steady state levels of partially processed messenger. The fact that the 1.9-kb transcript is the smallest RNA species detected exclusively by the exon probes leads us to tentatively identify it as the subunit 1 messenger.

FIG. 4. DNA probes used for the Northern blot hybridizations. The eight exons and seven introns are depicted by the previously mapped on wild type mtDNA have been aligned with the gene. The probes used for the hybridizations are shown by the open bars in the lower part of the figure. Each probe has been numbered.

FIG. 5. Autoradiographs of Northern blots hybridized to oxi probes. The ethidium bromide-stained gels of total mitochondrial RNA (negative photographs) have been aligned with the autoradiographs. The sizes of the two mitochondrial rRNAs (3.2 and 1.6 kb) and of a 0.9-kb transcript are marked on the left-hand margins. The exon and intron probes are numbered as in Fig. 4.
The nucleotides have been numbered according to the convention described in the legend to Fig. 2. The repeated sequences described in the legend to Fig. 2.

In contrast to the above results, the intron probes hybridized to at least four different transcripts. Probes originating from the first or second introns showed pronounced hybridization to two similar transcripts estimated to be 2.4 and 2.5 kb long. These transcripts are probably identical with the 19 S RNAs recently described by Grivell et al. (24). The two transcripts are not always resolved on agarose gels from the first or second introns to detect both the 2.4- and 2.5-kb transcripts can be explained by cross-hybridization to homologous sequences in the two introns (see below for discussion of intron homologies). A third transcript that appears to consist of a pure intron sequence is detected by a probe from the fifth intron. This transcript is approximately 0.85 kb long and probably corresponds to the 11 S circular RNA reported by Arnberg et al. (25). One additional abundant higher molecular weight transcript (3.2 kb) was detected with a probe from the third intron. Since this transcript also hybridizes to different exons but not to the first two intron probes, it is likely to be a processing intermediate in which the early introns have been spliced out.

Deletion End Points and Localization of oxi3 Mutations—In the previous paper (1), oxi3 mutations were mapped to physically defined spans of the oxi3 locus. Some of the mutations can now be assigned to either intron or exon regions of the Subunit 1 gene. The complete nucleotide sequences of most of the mtDNA segments used in this study have allowed their deletion end points to be precisely defined. These data are presented in Table IV. Based on the marker retention of the clones and the nucleotide sequences of their mtDNAs, three mutations have been mapped to the intron regions. The two markers, M15-190 and M15-98, are present in DS6/A400 but absent in DS6/A402. Both mutations must, therefore, be located between nucleotides +282 and +2136 (i.e. the sequence present exclusively in DS6/A400) which includes most of the first intron of the gene. A similar analysis indicates that the mutation M3-9 falls in the second intron. Although the other mutations studied can also be assigned within physically defined limits, these contain both intron and exon sequences.
the last codon of exon A1 and ends 2434 nucleotides further downstream with an ochre terminator. Excluding the exon sequence, this frame codes for an additional 784 amino acids. An almost equally long reading frame is present in the second intron. Although the other two intron sequences are shorter, they nonetheless exceed the length of fortuitous reading frames (Fig. 8). The protein sequences average 50% identical homology of the two protein sequences encoded by the reading frames (Fig. 8). The protein sequences average 50% identical

There are six arginine codons of the CGN series in the first 3 introns of the Subunit 1 gene. None of these codons have been found in established yeast mitochondrial genes (4, 23, 27–29). The codons tabulated in Table VI also point to some interesting differences in the distribution of codons within single amino acid families. For example, there are nearly the same number of AUA and AUU isoleucine codons in the intron reading frames. In contrast, mitochondrial genes, including the Subunit 1 gene of cytochrome oxidase, use almost exclusively the AUA codon. A similar situation holds true for the two codons of phenylalanine. Whereas equal use is made of UUU and UUC in all the known mitochondrial genes, the introns exhibit a strong bias for the UUU codon. All the features described here for the intron reading frames of the Subunit 1 gene have also been noted in a long reading frame in the first intron of the apocytochrome b gene (4).

Several remarkable facts have emerged from a comparison of the Subunit 1 introns with one another and with the intron of the apocytochrome b gene. A computer search for identical sequences has revealed extensive homology between the first two introns of the Subunit 1 gene. Equally significant is the homology of the two protein sequences encoded by the reading frames (Fig. 8). The protein sequences average 50% identical

### Table VI

<table>
<thead>
<tr>
<th>Codon frequencies in the Subunit 1 gene</th>
<th>Codon frequencies in the Subunit 1 gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subunit 1 intron</td>
<td>Subunit 1 intron</td>
</tr>
<tr>
<td>Amino acid</td>
<td>Codon</td>
</tr>
<tr>
<td>Ala</td>
<td>GCA</td>
</tr>
<tr>
<td></td>
<td>U</td>
</tr>
<tr>
<td>Arg</td>
<td>AGA</td>
</tr>
<tr>
<td></td>
<td>G</td>
</tr>
<tr>
<td>Asn</td>
<td>AAU</td>
</tr>
<tr>
<td></td>
<td>C</td>
</tr>
<tr>
<td>Thr</td>
<td>ACA</td>
</tr>
<tr>
<td></td>
<td>U</td>
</tr>
<tr>
<td>Ile</td>
<td>AUU</td>
</tr>
<tr>
<td></td>
<td>A</td>
</tr>
<tr>
<td>Leu</td>
<td>UUA</td>
</tr>
<tr>
<td></td>
<td>G</td>
</tr>
</tbody>
</table>

*These codons have been tabulated from the gene sequences of the ATPase Subunits 6 and 9 (27, 29), cytochrome oxidase Subunits 2 and 3 (23, 28), and apocytochrome b (4).
amino acids over almost the entire lengths of the two introns. The results of Table VII summarize the divergence of the two sequences in terms of changes in each of these positions of the codons. These data indicate a small but significant preference for nucleotide substitutions in the third position.

An even more striking homology is evident between the first intron of the apocytochrome b and the fourth intron of the Subunit 1 genes (Fig. 9). In this case, the apocytochrome b intron is approximately 300 nucleotides longer. The two reading frames, however, can be aligned over 700 nucleotides with an excellent match in the amino acid sequences. The protein sequences are 70% homologous with a strong bias for base substitutions in the third positions of the codons (Table VII). We have also compared the sequences at the intron-exon boundaries, both within the Subunit 1 and with the apocytochrome b gene. The only significant homology near the splice points is found between the first two introns of the Subunit 1 and the first and fourth introns of the Subunit 1 and apocytochrome b genes. These homologies are shown in Fig. 10. All the other boundaries appear to have unique sequences.

**DISCUSSION**

The oxi3 locus is a complex region of yeast mitochondrial DNA located between the oligomycin and paromomycin resistance loci (31). During the past several years, a substantial body of evidence has been accumulated indicating that this locus represents a single mosaic gene coding for Subunit 1 of cytochrome oxidase. This conclusion is supported by the following observations. 1) Physical mapping of oxi3 mutations places a lower limit of 10 kb on the size of the gene (3). Since the apparent molecular weight of Subunit 1 is 40,000 (15, 16), the locus is at least nine times longer than the predicted length of a co-linear gene. 2) Mutations in oxi3 form a single genetic complementation group suggestive of a single gene (32). 3) The oxi3 region gives rise to different sized transcripts, some of which have been proposed to be stable intron excision products (24, 33, 34).

The DNA sequence of the oxi3 locus reported here fully supports the earlier interpretations about the organization of the Subunit 1 gene. Our sequence data indicate the gene to be 10 kb long, spanning the region of wild type mtDNA from 43.7 to 53 map units. Based on its orientation relative to the wild type map, the NH2-terminal end of the gene lies proximal to the paromomycin marker (Fig. 11). The gene is therefore transcribed from the same strand of DNA as the other two subunits of cytochrome oxidase (23, 28) and other recently sequenced mitochondrial genes (4, 27, 29). The Subunit 1 coding sequences are distributed among 7 to 8 different exons with lengths ranging from 36 to 480 nucleotides. Together, the exons dictate a protein whose sequence averages 50% homology with Subunit 1 of human cytochrome oxidase. Even though the exact junctions at the exon-intron boundaries

**Table VII**

<table>
<thead>
<tr>
<th>Nucleotide substitutions in the codons of the homologous intron reading frames</th>
<th>Total codons scored</th>
<th>Amino acid substitutions</th>
<th>Codons changed</th>
<th>No. of changes in different positions of the codons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>First</td>
</tr>
</tbody>
</table>

First and second introns of the Subunit 1 gene

First intron of apocytochrome b and fourth intron of the Subunit 1 gene

Fig. 8. Nucleotide and amino acid sequence homologies between the first and second introns of the oxi3 gene. The nucleotide and amino acid residues in the two introns have been numbered on the coordinates. The program of Staden (30) was used to search for nucleotide and amino acid homologies. The program was set to detect a minimum of nine out of 10 consecutive identities in the nucleotide sequence and three out of four identities in the amino acid sequences.

Fig. 9. Nucleotide and amino acid sequence homologies between the fourth intron of the oxi3 and the first intron of the apocytochrome b (cob) genes. The homologies were detected with the Staden program (30) as described in the legend to Fig. 8.
Sequence of Gene Coding for Subunit 1 of Cytochrome Oxidase

...cannot be specified at present, the homology of the yeast and human proteins is sufficiently high to provide a reasonably accurate idea of the overall length of the coding sequence. According to our estimates, the Subunit 1 exons are comprised of 1,533 nucleotides and represent only 16% of the total length of the gene. The yeast Subunit 1 consists of 510 amino acid residues with a molecular weight of 56,000. This value is 30% higher than measured on SDS-polyacrylamide gels (15, 16). Similar discrepancies in molecular weights have been found for Subunit 3 of cytochrome oxidase (23) and cytochrome b (4). We believe that the differences are most likely due to anomalous binding of SDS by these extremely hydrophobic proteins.

The identification of four of the exon sequences is supported by Northern blot hybridization studies. Single-stranded probes from exons A1, A4, A5, and A8 hybridize to a 1.9-kb transcript which we tentatively identify to be the fully processed messenger. The migration of this RNA on agarose gels corresponds to that of the 18S transcript described by Van Ommen et al. (33). Several other abundant transcripts are detected when the Northern blots are challenged with intron probes. Two stable transcripts hybridize to probes from the first or second introns. The sizes of these RNAs (2.4 and 2.5 kb) suggest that they may be excision products of the first two introns. A smaller transcript of 0.85 kb is also detected specifically with a probe from the fifth intron.

The apocytochrome b and Subunit 1 genes of yeast mitochondria are now well documented examples of mosaic genes. All the current evidence suggests that the messengers for both genes arise from processing events involving a stepwise removal of the intron sequences from large primary transcripts. The yeast mitochondrial genes offer a unique opportunity for studying the details of RNA splicing mechanisms for several reasons. First is the availability of a substantial number of mutants in which DNA sequence modifications, particularly in the intron regions, prevent normal splicing. In this context, it may be significant that some mutations in apocytochrome b introns have been shown to prevent the synthesis of cytochrome oxidase Subunit 1 (38–40). Some preliminary studies suggest a functional link in the excision of the first intron of apocytochrome b and the fourth intron of Subunit 1. An examination of the transcripts present in a mutant carrying a lesion in the first apocytochrome b intron has shown that the processing of the Subunit 1 messenger is arrested at the level of the fourth intron. 3

Several laboratories have proposed that the intron reading frames code for a specialized set of proteins required for RNA splicing (35, 36). Such a function is strengthened by the observation that certain intron mutations in apocytochrome b lead to an accumulation of precursor transcripts which are translated to high molecular weight proteins antigenically related to cytochrome b (40). These new proteins have been postulated to contain both exon- and intron-specific amino acid sequences. Although there are a number of attractive features in processing models requiring the synthesis of intron-encoded splicing enzymes (35, 36), not all of the evidence supports this interpretation. Morimoto et al. (34) have reported proper splicing of the 21S rRNA in r27 mutants defective in mitochondrial protein synthesis and therefore incapable of transcribing mitochondrial messengers. Along the same line, we have found that mitochondrial tRNA mutants are capable of processing the apocytochrome b messenger, albeit at a less efficient rate than wild type cells. 3 It should also be noted that reading frames coding for proteins with a high lysine and asparagine content are not restricted to the intron regions of mitochondrial genes. For example, in S. cerevisiae D273-10B, there is a 1500-nucleotide-long uninterrupted reading frame between the Subunit 2 and 3 genes of cytochrome oxidase. 3 A somewhat shorter reading frame of 500 nucleotides has also been found between the genes of Subunit 6 of the ATPase and apocytochrome b (29). Whether these sequences, as well as those of the introns, are expressed as translation products remains obscure at present.

Acknowledgments—We wish to express our appreciation to Dr. Bart Barrell and Dr. Frederick Sanger for providing us with the sequence of the human Subunit 1 of cytochrome oxidase prior to publication.

1 S. Bonitz, unpublished studies.
3 G. Coruzzi, unpublished studies.
REFERENCES