Structural Analysis of Two Genes Encoding Divergent Forms of Yeast Cytochrome *c* Oxidase Subunit V

MICHAEL G. CUMSKY,[†] CYNTHIA E. TRUEBLOOD, CHRISTINE KO, AND ROBERT O. POYTON*

Department of Molecular, Cellular and Developmental Biology, University of Colorado at Boulder,
Boulder, Colorado 80309-0347

Received 25 March 1987/Accepted 7 July 1987

In *Saccharomyces cerevisiae*, subunit V of the inner mitochondrial membrane protein complex cytochrome *c* oxidase is encoded by two nonidentical genes, *COX5a* and *COX5b*. Both genes are present as single copies in *S. cerevisiae* and in several other *Saccharomyces* species. Nucleotide sequencing studies with the *S. cerevisiae* *COX5* genes reveal that they encode proteins of 153 and 151 amino acids, respectively. Overall, the coding sequences of *COX5a* and *COX5b* have nucleotide and protein homologies of 67 and 66%, respectively. They are saturated for nucleotide substitutions that result in a synonymous codon, indicating a long divergence time between these two genes. Nucleotide sequences flanking the *COX5a* and *COX5b* coding regions exhibit no significant homology. The *COX5a* protein, pre-subunit V_a, contains a 20-amino-acid leader peptide, whereas the *COX5b* protein, pre-subunit V_b, contains a 17-amino-acid leader peptide. These two leader peptides exhibit only 45% homology in the primary sequence, but have similar predicted secondary structures. By analyzing the RNA transcripts from both genes we have found that *COX5a* is a contiguous gene but that *COX5b* contains an intron. Surprisingly, the *COX5b* intron interrupts the AUG codon that initiates translation of the pre-subunit V_b polypeptide and contains a 5' donor splice sequence that differs from that normally found in yeast introns.

Cytochrome *c* oxidase is a key enzyme in the regulation of cellular energy production in eucaryotes (18). As a member of the electron transport chain of the inner mitochondrial membrane, this enzyme catalyzes the reduction of oxygen in a reaction that is essentially irreversible (57, 59). Since this is the only irreversible reaction in the mitochondrial electron transport chain, it is equivalent to the committed step in a metabolic pathway. As such it is an important control point in the regulation of electron flow through the chain (18). At present, it is unclear how eucaryotic cells alter their cytochrome *c* oxidase activity levels in response to changes in energy demand. Is it by changing (increasing or decreasing) the number of assembled holoenzyme molecules in the inner membrane? Is it by modulating the activity of individual enzyme molecules? Or is it by the assembly of isologous forms of subunits, which have different effects on catalysis, into holoenzyme molecules (4, 12, 23, 24, 26)?

To study these and related questions, we are examining various aspects of the biogenesis of cytochrome *c* oxidase in *Saccharomyces cerevisiae*. This enzyme consists of nine different polypeptide subunits (46); subunits I, II, and III are encoded by mitochondrial DNA, whereas subunits IV, V, VI, VII, VIIa, and VIII are encoded by nuclear DNA. The structural genes encoding all but one of these subunits (subunit VII) have been identified and characterized (6, 11-13, 20, 25, 30, 33, 42, 51, 53, 60, 61). Subunits I, II, and III as well as subunits IV, VI, and VIIa are each encoded by a unique gene. In contrast, subunit V is encoded by a small gene family composed of two nonidentical genes, *COX5a* and *COX5b* (12).

Previously, we isolated and partially characterized both *COX5* genes (12, 13) and showed that the protein product of either can assemble into holocytochrome *c* oxidase. Subsequently, two other laboratories also isolated and sequenced

a *COX5* gene from *S. cerevisiae* (25, 51). However, because only partial nucleotide sequences for *COX5a* and *COX5b* have been available until now the relation between *COX5a* and *COX5b* and this *COX5* gene has been unclear. In studies described here we demonstrate that *COX5a* and this recently described *COX5* gene are identical. We also demonstrate that *COX5a* and *COX5b* exist in single copy and are markedly different. They encode proteins that differ in 55 of 154 amino acids, their 5'- and 3'-flanking regions lack homology, and one of them, *COX5b*, possesses an intron that interrupts its initiation codon. In the accompanying paper (54) we describe the effects of null mutations in *COX5a* and *COX5b* on in vivo cytochrome *c* oxidase activity, cell respiration, and growth on nonfermentable carbon sources. We also demonstrate that these two genes encode interchangeable mature protein subunits that are expressed at markedly different levels.

MATERIALS AND METHODS

Plasmids, strains, and growth media. The *S. cerevisiae* strains used for complementation studies and transcript analysis were D273-10B (*mata*, ATCC 24657), JM28 (*mata leu2-3 leu2-112 his4-580 ura3-52 trp1-289 ade2 cox5a-1*) (13), and JM43 (*mata leu2-3 leu2-112 his4-580 ura3-52 trp1-289*) (12). Other yeast strains used were *S. cerevisiae* C258, *S. kluyveri* 51-242, *S. diastaticus* 55-99, *S. carlsbergensis* 57-47, *S. chevalieri* 61-22, *S. bayanus* 77-104, *S. italicus* C-105, and *S. oviformis* C247 obtained from H. Phaff (University of California, Davis) and *S. cerevisiae* A364, *S. cerevisiae* S288C, and *S. capensis* 74-44 obtained from the Yeast Genetic Stock Center (University of California, Berkeley). All yeast strains were grown at 28°C on either YPD or semisynthetic galactose medium (34) for nonselective growth. Prototrophic selection for transformants carrying YEp13-552 or YEp13-511 was performed on SD medium (52) supplemented with 40 µg of histidine, adenine, tryptophan, or uracil per ml as required. Respiratory proficiency was

* Corresponding author.

[†] Present address: Department of Molecular Biology and Biochemistry, University of California, Irvine, CA 92717.

FIG. 1. DNA sequencing strategy. (A) Partial restriction map and sequencing strategy of the *COX5a* gene. Symbols: coding sequence; flanking region that was sequenced; individual M13 clone that was sequenced; direction and approximate position of *COX5a* transcripts; the probe (*EcoRI*-*ApaI*) used for S1 mapping and RNA and genomic blots. Restriction sites: E, *EcoRI*; Hc, *HincII*; H, *HindIII*; Bg, *BglII*; S, *Sau3A*; X, *XhoII*; A, *ApaI*; P, *PstI*; and Hp, *HpaI*. (B) Same information for the *COX5b* gene. The kinks in the arrows and within the gene represents the *COX5b* intron. The kinked arrow that is starred depicts primer extension-cDNA sequence analysis of the mRNA. Additional restriction sites: B, *BamHI*; D, *DraI*; and C, *ClaI*. The *BamHI*-*BglII* fragment was used as the probe for S1 mapping experiments, whereas the *BglII*-*BglII* fragment was used in RNA and Southern blot analysis. bp, Base pairs.

tested on YPEG medium (34). Solid media contained 2% Bacto-Agar (Difco Laboratories, Detroit, Mich.).

Escherichia coli strains used for propagating plasmids were either HB101 or RR1 (5). A *recA* derivative of strain JM103 (35) was used as a host for all phage cloning. *E. coli* strains were grown at 37°C in LB or YT medium (37). When grown selectively, transformed strains were propagated in either of the above media containing 100 µg of ampicillin per ml.

The shuttle plasmid YEp13 (7) was used in this study. The derivative plasmids YEp13-552 and YEp13-511, carrying the *COX5a* and *COX5b* genes, respectively, have been described (12, 13, 33). They were isolated from the yeast library constructed by Nasmyth and Tatchell (41) as part of 6.3- and 4.8-kilobase (kb) inserts, respectively, of yeast DNA in the vector YEp13. The position of the *COX5b* gene on the 4.8-kb yeast insert was determined by analyzing transcripts derived from it. By hybridizing poly(A)⁺ RNA to restriction fragments from within the insert we found that only one region, a 1.2-kb fragment between two *BglII* restriction sites (Fig. 1B), encodes a transcript that is readily detected. The position of the *COX5a* gene on the 6.3-kb yeast insert was determined by performing low-stringency hybridizations against a probe derived from *COX5b*. Restriction fragments derived from these plasmids or various sub-

clones were nick translated and used as hybridization probes (see Results).

DNA sequence analysis. DNA sequencing was performed after cloning restriction fragments from either gene into the M13 phage vectors mp8 and mp9 (36) or mp18 and mp19 (41). In all cases the dideoxy chain-termination method was used (49). Reactions were analyzed on 6% acrylamide buffer gradient gels (3) as well as with multiple loadings on 6% acrylamide sequencing gels. All gels were dried before exposure to SB5 film. At least two independent isolates from each subclone were sequenced.

Transcript analysis. RNA isolation, fractionation, blotting, and hybridization procedures were essentially as described previously (12). The 5' ends of the transcripts were analyzed with S1 and exonuclease VII protection procedures (2, 56). Briefly, a 5' end-labeled fragment extending well upstream of each *COX5* gene (an *EcoRI*-*ApaI* fragment for *COX5a* and a *BamHI*-*BglII* fragment for *COX5b*) (Fig. 1) was used as the probe. Yeast poly(A)⁺ RNA or total RNA prepared from transformants carrying either gene on multicopy plasmids, was ethanol precipitated with 5' end-labeled probe, and the mixture then suspended in a formamide-containing hybridization buffer (2). After being heated at 72°C for 15 min to denature the probe, the mixture was hybridized at 49°C for 4 to 6 h. After hybridization, samples were treated with S1 nuclease (2,000 to 5,000 U; Boehringer Mannheim Biochemicals, Indianapolis, Ind.) or exonuclease VII (1 to 2 U; Bethesda Research Laboratories, Inc., Gaithersburg, Md.) for 30 to 45 min at 37°C. Nucleic acids were then precipitated with ethanol, washed, suspended in formamide loading buffer, and resolved on sequencing gels. Dideoxy reactions from known templates were used as standards to position the precise ends of the transcripts.

RNA sequence analysis. The sequence of the mature *COX5b* mRNA through the splice junctions was determined by performing cDNA primer extension analysis of the mRNA, in the presence of dideoxynucleoside triphosphates, by a method described recently by McPheeters et al. (34a). Briefly, 10 to 20 µg of yeast poly(A)⁺ RNA was annealed for 3 min at 60°C with 0.2 to 0.3 pmol of an end-labeled, synthetic 23-base oligonucleotide primer (5'-TTTCTGTTCT AAGTTTGGCATAT-3'), which primes between positions +110 and +132 on *COX5b* (Fig. 2). After rapid cooling of the mixture in a dry ice-ethanol bath, the hybrids were extended with 1 to 10 U of avian myeloblastosis virus reverse transcriptase (Life Sciences, Inc.) in the presence of the appropriate ratios of deoxynucleoside triphosphates to dideoxynucleoside triphosphates. After 30 min at 37°C, the reactions were stopped and analyzed for DNA sequences as described above.

Computer analysis of DNA and protein sequences. Manipulation of DNA sequence data, including codon usage, nucleotide and amino acid sequence homology, and protein hydropathy, was aided by an IBM personal computer. DNA sequences were analyzed by using search algorithms devised by Schneider et al. (50), Pustell and Kafatos (47, 48), and Wilbur and Lipman (58). The Pustell and Kafatos algorithms were obtained as software provided by International Biotechnologies Inc. (New Haven, Conn.), and the Wilbur and Lipman algorithm was accessed on line as the IFIND program in Bionet (Intelligenetics, Palo Alto, Calif.). Protein primary sequence homologies were identified and evaluated by using the algorithms of Lipman and Pearson (29). Secondary structures were predicted from primary sequences as described by Chou and Fasman (10) by using a computer program written by David McPheeters (University of Colo-

<u>COX5b</u>	GTACTGTTTGTTCATTATTATATCGGCAATGACACTCACCAATTTCTTTTTTTTGATT	-317
<u>COX5a</u>		
	TGTATTGTTTCGATAAGGCAACGAACGAGTACTCGAAAAATGTCTAGGACCCCTTAAAATTACTGAGGGGTTGAGAAAA	-238
	AGATCTGCAGGCAGACCGCTCCCTACGCTTCTAAATAGA	
	TACCGTGCAAAAGACGAAAAAGACGAATTTTCATTGATTATATTTTATAAATGAAGTGTGCATTAAACAATAGACC	-159
	ATGTACTTAGCACTTCTGCCTACATAGTAATTATTCATCATTTGAAGTGCACATTGCAAAAATCGTCCAATAGACGTT	
	AATTATTTCAATTTAATATTCTTTGTCAGGAACTTTCAATGGAATAACGCCACATATTTCATTGTAAGAECATGTAT	-80
	TTCTCGTTTGTGATTGGCCCTACGTTTGTGCGCATACAAGCGGGGCGAAAAAGTTAAAAAGAACATCAATTTAAATA	
	AACAAACACTGATTTTGTGTTTGTAGTTTAAAGATATCCATTACTAACATTCGAGGTGTACAAGCACAGTTTGTGA	-1
	AGAACCATCAGTCCTGATTTTGTCTCTGTTTGAAGCAATGCACCAACACAAAGATCAACTAAGAACGCATCTACA	
	<div> <div>GTG TTG</div> <div>CGT ACT</div> <div>TCT</div> <div>CTT ACT</div> <div>AAA</div> <div>---</div> <div>---</div> <div>GGG</div> <div>---</div> <div>GCA CGG</div> <div>CTA</div> <div>ACT</div> <div>GGG</div> <div>ACG</div> <div>AGA</div> <div>TTT</div> </div> <div> <div>ATG TTA</div> <div>CGT AAC</div> <div>ACT</div> <div>TTT ACT</div> <div>AGA</div> <div>GCT</div> <div>GGT</div> <div>GGA</div> <div>CTA</div> <div>TCA</div> <div>CGT</div> <div>ATT</div> <div>ACA</div> <div>TCC</div> <div>GTA</div> <div>AGA</div> <div>TTC</div> </div>	51
	Met Leu Arg Thr Ser Leu Thr Lys --- --- Gly --- Ala Arg Leu Thr Gly Thr Arg Phe	60
	Asn Thr Phe Arg Ala Gly Leu Ser Arg Ile Thr Ser Val	
	<div> <div>GTT CAA</div> <div>ACA</div> <div>AGG</div> <div>GCC</div> <div>CTT TCG</div> <div>AAG</div> <div>GCA</div> <div>ACA</div> <div>TTG</div> <div>ACA</div> <div>GAT CTG</div> <div>CCC</div> <div>GAA</div> <div>AGA</div> <div>TGG</div> <div>GAA</div> <div>AAT</div> </div> <div> <div>GCT CAA</div> <div>ACA</div> <div>CGT</div> <div>GCT</div> <div>CTT TCC</div> <div>AAC</div> <div>GCT</div> <div>GCT</div> <div>GGA</div> <div>ATG</div> <div>GAT CTG</div> <div>CAA</div> <div>TCA</div> <div>CGA</div> <div>TGG</div> <div>GAG</div> <div>AAC</div> </div>	111
	Val Gln Thr Lys Ala Leu Ser Lys Ala Thr Leu Thr Asp Leu Pro Glu Arg Trp Glu Asn	120
	Ala His	
	<div> <div>ATG CCA</div> <div>AAA</div> <div>TTA</div> <div>GAA</div> <div>CAG</div> <div>AAA</div> <div>GAG</div> <div>ATT</div> <div>GCA</div> <div>GAT</div> <div>AAT</div> <div>TTG</div> <div>ACA</div> <div>GAA</div> <div>CGT</div> <div>CAA</div> <div>AGG</div> <div>CTT</div> <div>CCA</div> </div> <div> <div>ATG CCA</div> <div>TCC</div> <div>ACT</div> <div>GAG</div> <div>CAG</div> <div>CAG</div> <div>GAT</div> <div>ATT</div> <div>GTC</div> <div>AGT</div> <div>ATG</div> <div>TTG</div> <div>AGT</div> <div>GAA</div> <div>CGT</div> <div>CAA</div> <div>AAA</div> <div>TTA</div> <div>CCA</div> </div>	171
	Met Pro Asn Leu Glu Gln Lys Glu Ile Ala Asp Asn Leu Thr Glu Arg Gln Lys Leu Pro	180
	Ser Thr	
	<div> <div>TGG AAA</div> <div>ACT</div> <div>CTC</div> <div>AAA</div> <div>ATC</div> <div>GAG</div> <div>GAA</div> <div>ATC</div> <div>AAA</div> <div>GCA</div> <div>GCT</div> <div>TGG</div> <div>TAC</div> <div>ATA</div> <div>TCC</div> <div>TAC</div> <div>GGC</div> <div>GAG</div> <div>TGG</div> </div> <div> <div>TGG</div> <div>GCA</div> <div>CAG</div> <div>CTT</div> <div>ACT</div> <div>GAG</div> <div>CCT</div> <div>GAA</div> <div>AAG</div> <div>CAA</div> <div>GCT</div> <div>ATG</div> <div>TGG</div> <div>TAC</div> <div>ATT</div> <div>TCT</div> <div>TAC</div> <div>GGA</div> <div>GAA</div> <div>TGG</div> </div>	231
	Trp Lys Thr Leu Asn Asn Glu Glu Ile Lys Lys Ala Val Trp Tyr Ile Ser Tyr Gly Glu Trp	240
	Ala Gln	
	<div> <div>GGA CCT</div> <div>AGA</div> <div>AGA</div> <div>CCT</div> <div>GTA</div> <div>CAC</div> <div>GGA</div> <div>AAA</div> <div>GGC</div> <div>GAT</div> <div>GTT</div> <div>GCA</div> <div>TTT</div> <div>ATA</div> <div>ACT</div> <div>AAA</div> <div>GGA</div> <div>GTA</div> <div>TTT</div> </div> <div> <div>GGC</div> <div>CCA</div> <div>AGA</div> <div>AGA</div> <div>CCT</div> <div>GTA</div> <div>TTG</div> <div>AAT</div> <div>AAG</div> <div>GGT</div> <div>GAT</div> <div>TCC</div> <div>AGT</div> <div>TTT</div> <div>ATT</div> <div>GCC</div> <div>AAA</div> <div>GGT</div> <div>GTT</div> <div>GCT</div> </div>	291
	Gly Pro Arg Arg Pro Val His Gly Lys Gly Asp Val Ala Phe Ile Thr Lys Gly Val Phe	300
	Leu Asn	
	<div> <div>TTA GGG</div> <div>TTA</div> <div>GGA</div> <div>ATC</div> <div>TCA</div> <div>TTT</div> <div>GGG</div> <div>CTC</div> <div>TTT</div> <div>GCT</div> <div>TTA</div> <div>GTG</div> <div>AGA</div> <div>CTA</div> <div>TTA</div> <div>GCC</div> <div>AAT</div> <div>CCT</div> <div>GAA</div> </div> <div> <div>GCA</div> <div>GGC</div> <div>CTA</div> <div>CTA</div> <div>TTT</div> <div>TCA</div> <div>GTC</div> <div>GGA</div> <div>CTT</div> <div>TTT</div> <div>GCT</div> <div>GTC</div> <div>GTC</div> <div>AGG</div> <div>---</div> <div>ATG</div> <div>GCC</div> <div>GGT</div> <div>GGC</div> <div>CAA</div> </div>	351
	Leu Gly Leu Gly Ile Ser Phe Gly Leu Phe Gly Leu Val Arg Leu Leu Ala Asn Pro Glu	357
	Ala Phe	
	<div> <div>ACT CCA</div> <div>AAG</div> <div>ACT</div> <div>ATG</div> <div>AAC</div> <div>AGG</div> <div>GAA</div> <div>TGG</div> <div>CAG</div> <div>TTG</div> <div>AAA</div> <div>TCA</div> <div>GAC</div> <div>GAG</div> <div>TAT</div> <div>CTG</div> <div>AAG</div> <div>TCA</div> <div>AAA</div> </div> <div> <div>GAC</div> <div>CCA</div> <div>AAG</div> <div>ACC</div> <div>ATG</div> <div>AAT</div> <div>AAG</div> <div>GAG</div> <div>TGG</div> <div>CAG</div> <div>CTA</div> <div>AAG</div> <div>AGT</div> <div>GAC</div> <div>GAA</div> <div>TAT</div> <div>TTG</div> <div>AAG</div> <div>TGG</div> <div>AAG</div> </div>	411
	Thr Pro Lys Thr Met Asn Arg Glu Trp Gln Leu Lys Ser Asp Glu Tyr Leu Lys Ser Lys	417
	Asp Ala Lys	
	<div> <div>AAT GGC</div> <div>AAT</div> <div>CCT</div> <div>TGG</div> <div>GGA</div> <div>GGT</div> <div>TAT</div> <div>TCT</div> <div>CAA</div> <div>GTT</div> <div>CAA</div> <div>TCT</div> <div>AAA</div> <div>TAA</div> <div>GTAGACGAGGAAAAATAAAA</div> </div> <div> <div>AAT</div> <div>GCT</div> <div>AAT</div> <div>CCT</div> <div>TGG</div> <div>GGT</div> <div>GGT</div> <div>TAT</div> <div>TCT</div> <div>CAG</div> <div>GTC</div> <div>CAA</div> <div>TCT</div> <div>AAA</div> <div>TAA</div> <div>ACAGGCCAGAAATTTTGAA</div> </div>	475
	Asn Ala Asn Pro Trp Gly Gly Tyr Ser Gln Val Gln Ser Lys Ter	481
	TTGTTTCGTATATTCGGTGTGTTGGGGTATAAGTAGATTGTTTCATATATACGCATTGGTCTTAGTTTCAGTAGGTTGA	554
	AGGAAATCAATAGGGGTAAAGATGTGCATGTTTTTCAGCTAGTCTGTGACCTGTACGAAAGTGAATATCTTATT	558
	TTACTTAGTTCTCTGTACCTTCTCTGCAAATATCATTGTTGTTACTCGAAGAAGAAAAAATAATCATGGAAT	633
	TGGAAAAAAGTCCAATCTAAAAATAAAAAATGTTCAAGTCATCGATTCTATGGGAGGAAAA	700

FIG. 2. Sequences of the *COX5b* and *COX5a* genes. The DNA sequences and the amino acid sequences derived from them are shown. The boxed areas indicate homologous DNA sequences within the coding regions of *COX5a* and *COX5b*. Differences in the derived amino acid sequences of subunits V_b (top line) and V_a (bottom line) are as indicated. The DNA sequence of the *COX5a* gene shown here is in perfect agreement with DNA sequences for a *COX5* gene that have been published previously (25, 51). For the DNA sequence of *COX5b*, the position of the intron is bracketed, sequences homologous to consensus splicing sequences in other yeast genes are underlined, and a short open reading frame that lies between the 5' end of the transcript and the initiation codon is overlined. Large arrows denote the ends of the major transcripts; small arrows denote the ends of the minor transcripts. The vertical arrow designates the junction between mature subunit V_a and its presequence.

rado). The location of hydrophobic domains was determined by the SOAP algorithm (27) with a rolling average of nine amino acid residues.

Miscellaneous methods. Procedures for polyacrylamide gel electrophoresis, immunoblotting, yeast and *E. coli* transformations, restriction enzyme analysis, agarose gel electrophoresis, nucleic acid preparations, preparation of poly(A)⁺ RNA, and nucleic acid hybridizations have all been described (12). Other methods (end labeling, nick translations, fragment purification) were performed by established procedures (31).

RESULTS

Nucleotide sequences of *COX5a* and *COX5b*. The sequencing strategies and DNA sequences for *COX5a* and *COX5b* are presented in Fig. 1 and 2. The sequence of the coding region of *COX5a* shows a single uninterrupted open reading frame of 153 codons (Fig. 2), predicting a primary translation product of 17,139 daltons. Because the NH₂-terminal amino acid sequence of mature subunit V_a has been determined (45), and there are no additional open reading frames immediately upstream, both the initiation codon and the site of proteolytic processing of the subunit V_a precursor (pre-V_a) are readily deduced (Fig. 2). Pre-V_a contains a 20-amino-acid leader peptide. The complete nucleotide sequence of *COX5a* shown in Fig. 2 confirms and extends a partial sequence we have published previously for this gene (12). Moreover, the deduced amino acid sequence and composition of the *COX5a* gene product are in perfect agreement with the amino acid composition and partial amino acid sequence determined previously for the subunit V polypeptide isolated from holocytochrome *c* oxidase enzyme preparations (45, 46). Together, these results establish unequivocally that *COX5a* encodes the subunit V isolog, V_a, that has been previously referred to as subunit V (46). Since the nucleotide sequence of *COX5a* is identical with the nucleotide sequences reported recently for a subunit V gene isolated in other laboratories (25, 51), it is clear that these latter genes are independent isolates of *COX5a*.

The protein-coding region of *COX5b* contains 151 codons (Fig. 2) and predicts a primary translation product with a molecular weight of 17,195. Because the DNA sequence upstream of the mature protein failed to reveal an in-frame ATG that could initiate translation of the protein, the methionine that initiates translation of the *COX5b* gene product (and hence the length and composition of the V_b leader peptide) was initially unclear. However, because this region contains the sequence TACTAAC, one of the highly conserved sequences found in all yeast introns (19, 28, 44), we presumed that the *COX5b* mRNA was spliced and that splicing was involved in the generation of a translational initiation codon. The existence and boundaries of the intron were confirmed experimentally (described below). It lies between -88 and +2 base pairs (Fig. 2). As such, it splits the A and T of the initiation codon, thus making splicing obligatorily required for translation-initiation. In addition to the intron the *COX5b* 5'-flanking sequence contains a short open reading frame 30 bases upstream from the initiation codon (Fig. 2).

As yet, the NH₂-terminal sequence of mature subunit V_b has not been determined. However, it is clear from in vitro translation studies (13, 33) with hybrid-selected mRNA that the primary translation product of *COX5b* is larger than either mature subunit V_a or V_b on sodium dodecyl sulfate-polyacrylamide gels. Assuming that this difference in appar-

ent molecular weight is attributable to the presence of a leader peptide on the *COX5b* protein product, the most likely location for the NH₂ terminus of mature V_b can be deduced by aligning its amino acid sequence with that determined for mature V_a (45). This alignment suggests that pre-V_b has a 17-amino-acid leader peptide.

Of note is the fact that neither the sequence of pre-V_a nor pre-V_b agrees with the preliminary sequencing results reported previously by Cerletti et al. (9) for a putative precursor to subunit V. Since both pre-V_a and pre-V_b contain methionine only at their NH₂ termini, and not at positions 1, 9, and 10 as reported previously (9), it is clear that the polypeptide studied by Cerletti et al. (9) is something other than a precursor to subunit V_a or V_b.

In Fig. 2, the *COX5a* and *COX5b* coding sequences have been aligned for optimal homology by using the algorithm of Lipman and Pearson (29) after scoring for amino acid identity, the presence of insertions or deletions, and amino acid replacements based on the frequency with which similar replacements occur in protein families within the Dayhoff data base (14). Within the 452 nucleotides of coding sequence that are compared, there are 150 differences. Of these differences, 49 correspond to silent base substitutions, i.e., substitutions that do not alter the amino acids specified. These silent substitutions are distributed throughout the coding region but are enriched in the carboxyl-terminal 27 amino acids, a region where the overall homology is also higher than it is elsewhere in the coding region.

The frequency of codon usage in both *COX5a* and *COX5b* is typical of that found in yeast genes that are expressed at low levels. With the codon bias index of Bennetzen and Hall (1) as a measure of overall use of preferred codons, *COX5a* has a higher level of bias (0.19) than *COX5b* (0.05).

Both *COX5* genes are present in single copy and are detectable in several different *Saccharomyces* species. To determine the copy number of each *COX5* gene within the haploid yeast genome, we performed Southern blot analysis (hybridization) of D273-10B genomic DNA digested with restriction enzymes that cut either outside or within each gene. The patterns observed for hybridization to a *COX5a*- or *COX5b*-specific probe indicate that each gene is present in single copy (data not shown). We examined three additional *S. cerevisiae* strains as well as the closely related (*S. diastolicus*, *S. chevalieri*, *S. capensis*, *S. italicus*, and *S. oviformis*) and more distantly related (*S. kluyveri*, *S. carlsbergensis*, and *S. bayanus*) *Saccharomyces* species (32). Figure 3 shows the *EcoRI* fragments that carry *COX5a* and *COX5b* sequences. The hybridization patterns for all strains, except *S. kluyveri*, are identical. Thus, all but one of the *Saccharomyces* strains tested carry both *COX5a* and *COX5b*. Since the extent of DNA reassociation between *S. kluyveri* and *S. cerevisiae* is extremely low, compared with that between *S. cerevisiae* and the other species tested here (32), the *COX5* genes in *S. kluyveri* are probably too divergent, relative to the *COX5* genes in *S. cerevisiae*, to be detected under our hybridization conditions. A similar conclusion has been reached for the *SUC* genes in *S. cerevisiae* and *S. kluyveri* (8).

Transcript analysis. RNA blots of either whole cell RNA or poly(A)⁺ RNA reveal that *COX5a* and *COX5b* produce different-sized transcripts. The *COX5a* gene gives rise to a monodisperse transcript approximately 900 bases long. In contrast, *COX5b* hybridizes to two polydisperse transcripts that range in size from 800 and 700 bases (Fig. 4). The *COX5b* transcript pattern shown here differs slightly from that presented previously (12) insofar as the two poly-

FIG. 3. *COX5* genes in different *Saccharomyces* strains. Samples (10 μ g) of total DNA from each of the indicated *Saccharomyces* strains were prepared and digested to completion with *Eco*RI. The digests were divided in half; the duplicate samples were run on a 0.7% agarose gel, which was then subjected to Southern blot analysis with the *COX5a* (A) and *COX5b* (B) probes described in Fig. 1A and B, respectively. The last four strains on the right in A and B are different *S. cerevisiae* laboratory strains.

disperse transcripts are better resolved here, due to a different gel system, and we observe no hybridization with the third, larger transcript seen earlier. The probe used for the experiment shown in Fig. 4 was an oligonucleotide specific for the *COX5b* coding sequence, whereas the probe used for the previous study was a 1.2-kb *Bgl*II restriction fragment that included most of the *COX5b* coding sequence as well as sequences 3' to *COX5b*. Therefore, the larger transcript observed earlier is not a *COX5b* transcript but is from a gene 3' to *COX5b*. At present, the reason(s) for the polydispersity of the two *COX5* transcripts and their rela-

FIG. 4. Transcript analysis of the *COX5a* and *COX5b* genes. (A) A 20- μ g sample of poly(A)⁺ RNA prepared from the yeast strain D273-10B was fractionated on a 1.8% formaldehyde-agarose gel, blotted, and hybridized against a nick-translated, *COX5a*-specific probe (Fig. 1A). (B) A 5- μ g sample of total RNA from the transformant JM28-511 grown selectively on SD medium without leucine (52) was analyzed in a similar manner by using an end-labeled synthetic 23-base oligonucleotide that hybridizes to positions +110 to +132 in the *COX5b* gene (see Materials and Methods and legend to Fig. 5).

tionship(s) to one another are unclear. However, it should be noted that *COX5b* copy number has no effect on the transcript patterns; *COX5b* transcripts in both D273-10B (the wild type) and JM28-511 (the transformant carrying multiple copies of *COX5b* used for the Northern blot in Fig. 4) are polydisperse.

To determine whether either gene contains an intervening sequence and, at the same time, to localize the 5' end(s) of their transcripts, we performed protection experiments with nuclease S1 and exonuclease VII (2, 56). Figure 5A shows the results of a representative experiment for *COX5a*. The probe used was a 5' end-labeled *Eco*RI-*Apa*I restriction fragment (Fig. 1). The size of the protected fragment is essentially the same for either nuclease, indicating that no introns are present in the 5' half of the transcript. The same results are obtained with the wild-type strain, D273-10B, and with strain JM28-552, which overexpresses *COX5a*. This analysis also indicates that the *COX5a* transcript seen in Fig. 4 has four different 5' ends (two major and two minor), positioned 12 to 32 nucleotides upstream of the translational initiation codon (Fig. 2). Similar analysis of the 3' half of the gene (data not shown) also indicates the absence of an intron.

In contrast to the above results for *COX5a*, nuclease protection experiments with a 5' end-labeled *Bam*HI-*Bgl*II restriction fragment from *COX5b* (Fig. 1) reveal a clear difference in size of the fragments protected from digestion with S1 nuclease and exonuclease VII (Fig. 5B). Digestion with S1 gives a fragment of approximately 90 base pairs, whereas digestion with *Exo*VII produces a fragment of approximately 220 base pairs. This finding, obtained both with RNA from D273-10B and JM 28-511, a strain that overexpresses *COX5b*, indicates that there is a short intervening sequence in the NH₂-terminal portion of the coding

FIG. 5. Analysis of the 5' termini of transcripts from *COX5a* (A) and *COX5b* (B). S1 nuclease and exonuclease VII analyses were performed as described in Materials and Methods. The protected ends in each RNA preparation were sized by comparison with the adenine lane (ddA) or thymidine lane (ddT) of a sequencing ladder from a known template. (A) Poly(A)⁺ RNA (50 µg) from lactate-grown D273-10B cells (A⁺) or total RNA (50 µg) from the transformant JM28-552 (552) were analyzed. (B) Poly(A)⁺ RNA (50 µg) from galactose-grown D273-10B cells (A⁺) or total RNA from the transformant (511) were analyzed. The last three lanes on the right (511) contained 100, 50, and 100 µg of this RNA, respectively.

region of *COX5b*. To further characterize this intron the *COX5b* mRNA sequence was determined by primer extension analysis in the presence of dideoxy nucleoside triphosphates (34a). From this analysis (Fig. 6A), it is clear that the intron extends from positions -87 to +1 of the *COX5b* gene (Fig. 6B) and interrupts the ATG codon used to initiate translation of pre-*V_b*. Another interesting feature of this intron is the use of GCATGT, rather than GTAPyGT, as the 5' donor splice sequence. This is surprising because a common feature of all intervening sequences in eucaryotic mRNAs is the use of a 5'-GT sequence at the donor splice junction (39). It is not clear whether the presence of this nonconsensus sequence affects the efficiency of splicing, although this possibility seems unlikely because the RNA sequence determined in Fig. 6 gives no indication of unspliced mRNA.

Comparison of *COX5a* and *COX5b* gene products. Based on the optimized alignment shown in Fig. 7, it is clear that 66% of the amino acid residues in *V_a* and *V_b* are identical and that there are two long regions of perfect homology: residues 53 to 66 and residues 102 to 133 (*V_a* numbering). In those

regions where primary sequence homology is low, due to the presence of either conservative or nonconservative amino acid replacements, the conservation of predicted secondary structure is essentially equivalent to the extent of primary sequence conservation. Overall, 65% of the residues within these regions lie in domains of similar predicted secondary structure. Thus, it would appear that in these regions amino acids have not been replaced in a way that conserves secondary structure. If they were, one would expect a conservation of secondary structure that exceeds the conservation of primary sequence. In contrast, it appears that the local hydropathy along the *V_a* and *V_b* polypeptide chains is quite highly conserved, suggesting that amino acid replacements have been made so as to conserve hydrophobic and hydrophilic domains. Isologs *V_a* and *V_b* share two hydrophobic domains: a short domain that extends between amino acids 7 and 12 (*V_a* numbering) and a long domain that extends between amino acids 75 and 97. The latter domain is long enough to span a membrane, either as an α -helix, β -structure, or self-paired, antiparallel β -pleated sheet. However, its average hydropathy is less than expected for a transmembrane domain (27).

The primary sequence homology of the leader peptides in pre-*V_a* and pre-*V_b* is lower (45% identical residues) than that of the mature subunits. Despite this low level of homology, both leader peptides have nearly perfect homology at their junction with the mature subunits, both possess the same

FIG. 6. Primer extension and dideoxy sequencing of the *COX5b* mRNA. (A) Approximately 15 µg of Poly(A)⁺ RNA and 0.2 to 0.3 pmol of an end-labeled, 23-base synthetic oligonucleotide primer (5'-TTTCTGTTCTAAGTTTGGCATAT-3') were annealed and extended with avian myeloblastosis virus reverse transcriptase in the presence of dideoxynucleoside triphosphates as described in Materials and Methods. The reactions were then analyzed on a 6% buffer gradient sequencing gel. The figure shows the portion of the gel containing the sequence of the spliced transcript, which is written at the right. (B) Diagram showing the position of the intron within *COX5b*. bp, Base pairs.

FIG. 7. Comparison of the primary sequences, predicted secondary structures, and hydropathies of subunits V_a and V_b . The primary sequences of V_a (top line) and V_b (bottom line) were aligned by the algorithm of Lipman and Pearson (29). The optimized alignment uses a colon to denote identity and a dot for a conservative replacement. The amino acid residue number is given above the primary sequence of V_a and below the primary sequence of V_b . The junction of the leader peptide with each mature subunit is labeled with an arrow. The predicted secondary structures, denoted above the aligned primary sequences, were determined by the rules of Chou and Fasman (10) and are displayed with their symbols. The average local hydropathies of each polypeptide are displayed below the aligned primary sequences. Symbols: ●, V_a ; ○, V_b .

number of basic amino acid residues, and both have similar hydropathies and predicted secondary structures (Fig. 7). This latter observation is interesting in light of recent models that emphasize the importance of amphipathic secondary structures within leader peptides (17, 55). Although leader peptides from nucleus-encoded mitochondrial proteins have been assumed to have α -helical structures (55) the predicted

secondary structure for the leader peptides on both V_a and V_b suggests a hairpin model composed of a self-paired, amphipathic, antiparallel β -pleated sheet. In this model, the second basic amino acid within each subunit V leader (Arg-8 for V_a and Lys-8 for V_b) would fall inside of a predicted β -turn, and the first and third basic amino acids within each would lie on the same side of the β -pleated sheet.

DISCUSSION

In this study, we have demonstrated that the *COX5a* and *COX5b* genes encode proteins of divergent sequence and are present in single copy in several species of the budding yeast *Saccharomyces*. This latter finding, in conjunction with the finding that a strain carrying null alleles of both genes is totally deficient in cytochrome *c* oxidase activity and respiration (see accompanying paper [54]), strongly suggests that neither additional copies of either gene nor other forms of subunit V exist.

By aligning the nucleotide sequences of the coding regions of *COX5a* and *COX5b* as shown in Fig. 2 and omitting from consideration the three deletions in *COX5a* and one deletion in *COX5b*, we calculate that the nucleotide and protein homologies between both genes are 67 and 66%, respectively. This high degree of sequence homology leaves little doubt that *COX5a* and *COX5b*, like the genes for other isoproteins, arose by gene duplication followed by sequence divergence (16). To estimate divergence time from the number of silent changes (i.e., nucleotide substitutions that result in a synonymous codon) between *COX5a* and *COX5b* we used the method of Perler et al. (43). The percent corrected divergence value for the two genes is 130, a value which indicates that these genes are saturated for silent changes and which can, therefore, only be used to calculate a minimum divergence time. Assuming a rate of 7×10^{-9} silent substitutions per nucleotide per year (38, 43), we estimate that the *COX5* gene duplication occurred at least 130 million years ago. Additional evidence that *COX5a* and *COX5b* did not arise from a recent gene duplication event comes from the observation (Fig. 3) that both genes are present in several *Saccharomyces* species.

In contrast to the high degree of nucleotide sequence homology observed within the coding regions of *COX5a* and *COX5b*, the 5'- and 3'-flanking sequences of these two genes exhibit little, if any, significant homology. In addition to lack of overall sequence homology there are some differences in the 5'-flanking regions of these genes that have potential interest for their regulation. First, a sequence related to TATA $\uparrow$ T (i.e., TATA box) (22) is present in the *COX5b* 5'-flanking region (between positions -187 and -191) but is absent from the corresponding region of *COX5a* (cf 15). Second, *COX5b* contains an intron within its coding region. This intron is especially interesting because (i) it splits the initiation codon of pre-*V_b* and hence must be spliced before the initiation of translation, (ii) its 5' donor splice sequence, 5'-GCATGT, differs from the 5' junction consensus sequence, 5'-GTAPyGT, normally found in yeast introns, and (iii) the 5' end of *COX5b* mRNA, but not *COX5a* mRNA, contains a short open reading frame 30 nucleotides upstream from the subunit *V_b* initiation codon. This reading frame could play a role in the translational regulation of *COX5b* as has been observed for the upstream open reading frames of *GCN4* (21). Since it is clear from the accompanying paper (54) that *COX5a* and *COX5b* are expressed differentially, the importance of these differences is currently under study.

ACKNOWLEDGMENTS

We thank Tom Patterson for providing the sequence ladder used to map the 5' end of the *COX5a* gene, Richard Wright for DNA from strain D273-10B, and John Trawick for codon bias index calculations. David McPheeters' advice and assistance in the primer extension analysis is also gratefully acknowledged.

This work was supported by Public Health Service grants GM29838 and GM30228 from the National Institutes of Health.

M.G.C. was supported in part by a postdoctoral fellowship from the Damon Runyon-Walter Winchell Cancer Fund and in part by a fellowship from the National Institutes of Health.

LITERATURE CITED

1. Bennetzen, J. C., and B. D. Hall. 1982. Codon selection in yeast. *J. Biol. Chem.* **257**:3026-3031.
2. Berk, A. J., and P. A. Sharp. 1978. Spliced early mRNA's of simian virus 40. *Proc. Natl. Acad. Sci. USA* **75**:1274-1278.
3. Biggin, M. D., T. J. Gibson, and G. F. Hong. 1983. Buffer gradient gels and ³⁵S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. USA* **80**:3963-3965.
4. Bisson, R., and G. Schiavo. 1986. Two different forms of cytochrome *c* oxidase can be purified from the slime mold *Dictyostelium discoideum*. *J. Biol. Chem.* **261**:4373-4376.
5. Bolivar, F., R. L. Rodriguez, M. C. Betlach, and H. W. Boyer. 1977. Construction and characterization of new cloning vehicles. I. Ampicillin-resistant derivatives of the plasmid pMB9. *Gene* **2**:75-93.
6. Bonitz, S. G., G. Coruzzi, B. E. Thalenfeld, A. Tzagoloff, and G. Macino. 1980. Assembly of the mitochondrial membrane system. Structure and nucleotide sequence of the gene coding for subunit I of yeast cytochrome *c* oxidase. *J. Biol. Chem.* **255**:11927-11941.
7. Broach, J. R., J. N. Strathern, and J. B. Hicks. 1979. Transformation in yeast. Development of a hybrid cloning vector and isolation of the *CAN1* gene. *Gene* **8**:121-133.
8. Carlson, M., and D. Botstein. 1983. Organization of the *SUC* gene family in *Saccharomyces*. *Mol. Cell. Biol.* **3**:351-359.
9. Cerletti, N., P. C. Bohni, and K. Suda. 1983. Import of proteins into mitochondria. Isolated yeast mitochondria and a solubilized matrix protease correctly process cytochrome *c* oxidase subunit V precursor at the NH₂ terminus. *J. Biol. Chem.* **258**:4944-4949.
10. Chou, P. Y., and G. D. Fasman. 1978. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol. Relat. Areas Mol. Biol.* **47**:45-148.
11. Coruzzi, G., and A. Tzagoloff. 1979. Assembly of the mitochondrial membrane system. DNA sequence of subunit 2 of yeast cytochrome oxidase. *J. Biol. Chem.* **254**:9324-9330.
12. Cumsky, M. G., C. Ko, C. E. Trueblood, and R. O. Poyton. 1985. Two nonidentical forms of subunit V are functional in yeast cytochrome *c* oxidase. *Proc. Natl. Acad. Sci. USA* **82**:2235-2239.
13. Cumsky, M. G., J. E. McEwen, C. Ko, and R. O. Poyton. 1983. Nuclear genes for mitochondrial proteins. Identification and isolation of a structural gene for subunit V of yeast cytochrome *c* oxidase. *J. Biol. Chem.* **258**:13418-13421.
14. Dayhoff, M. 1978. Atlas of protein sequence and structure, vol. 5. National Biomedical Research Foundation, Silver Spring, Md.
15. DeHaan, M., A. P. G. M. Van Loon, J. Kriek, R. T. M. J. Vaessen, and L. A. Grivell. 1984. The biosynthesis of the ubiquinol-cytochrome *c* reductase complex in yeast. DNA sequence analysis of the nuclear gene coding for the 14 kDa subunit. *Eur. J. Biochem.* **138**:169-177.
16. Doolittle, R. F. 1981. Similar amino acid sequences: chance or common ancestry? *Science* **214**:149-159.
17. Douglas, M. G., M. T. McCammon, and A. Vassarotti. 1986. Targeting proteins into mitochondria. *Microbiol. Rev.* **50**:166-178.
18. Erecinska, M., and D. F. Wilson. 1982. Regulation of cellular energy metabolism. *J. Membr. Biol.* **70**:1-14.
19. Fouser, L. A., and J. D. Friesen. 1986. Mutations in a yeast intron demonstrate the importance of specific conserved nucleotides for the two stages of nuclear mRNA splicing. *Cell* **45**:81-93.
20. Fox, T. D. 1979. Five TGA "stop" codons occur within the translated sequence of the yeast mitochondrial gene for cytochrome *c* oxidase subunit II. *Proc. Natl. Acad. Sci. USA* **76**:6534-6538.
21. Hinnebusch, A. G. 1984. Evidence for translational regulation of the activation of general amino acid control in yeast. *Proc. Natl. Acad. Sci. USA* **81**:6442-6446.

22. Guarente, L. 1984. Yeast promoters: positive and negative elements. *Cell* 36:799–800.
23. Kadenbach, B. 1983. Structure and evolution of the "Atmungsferment" cytochrome *c* oxidase. *Angew. Chem. Int. Ed. Engl.* 22:275–282.
24. Kadenbach, B., and A. Stroh. 1984. Different reactivity of carboxylic groups of cytochrome *c* oxidase polypeptides from pig liver and heart. *FEBS Lett.* 173:374–380.
25. Koerner, T. J., J. Hill, and A. Tzagoloff. 1985. Cloning and characterization of the yeast nuclear gene for subunit 5 of cytochrome oxidase. *J. Biol. Chem.* 260:9513–9515.
26. Kuhn-Nentwig, L., and B. Kadenbach. 1985. Isolation and properties of cytochrome *c* oxidase from rat liver and quantification of immunological differences between isozymes from various rat tissues with subunit-specific antisera. *Eur. J. Biochem.* 149:147–158.
27. Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157:105–132.
28. Langford, C. J., F.-J. Klinz, C. Donath, and D. Gallwitz. 1984. Point mutations identify the conserved, intron-contained TACTAAC box as an essential splicing signal sequence in yeast. *Cell* 36:645–653.
29. Lipman, D. J., and W. R. Pearson. 1985. Rapid and sensitive protein similarity searches. *Science* 227:1435–1442.
30. Maarse, A., A. Van Loon, H. Reizman, I. Gregor, G. Schatz, and L. Grivell. 1984. Subunit IV of yeast cytochrome *c* oxidase: cloning and nucleotide sequence of the gene and partially amino acid sequencing of the mature protein. *EMBO J.* 3:2831–2837.
31. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
32. Martini, A. V., and C. P. Kurtzman. 1985. Deoxyribonucleic acid relatedness among species of the genus *Saccharomyces* Sensu Stricto. *Int. J. Syst. Bacteriol.* 35:508–511.
33. McEwen, J. E., M. G. Cumsky, C. Ko, S. D. Power, and R. O. Poyton. 1984. Mitochondrial membrane biogenesis: characterization and use of Pet mutants to clone the nuclear gene coding for subunit V of yeast cytochrome *c* oxidase. *J. Cell. Biochem.* 24:229–242.
34. McEwen, J. E., C. Ko, B. Kloeckener-Gruissem, and R. O. Poyton. 1986. Nuclear functions required for cytochrome *c* oxidase biogenesis in *Saccharomyces cerevisiae*. Characterization of mutants in 34 complementation groups. *J. Biol. Chem.* 261:11872–11879.
- 34a. McPheeters, D. S., A. Christiansen, E. T. Young, G. Stormo, and L. Gold. 1986. Translational regulation of expression of the bacteriophage T4 lysozyme gene. *Nucleic Acids Res.* 14:5813–5826.
35. Messing, J., R. Crea, and P. H. Seeburg. 1981. A system for shotgun DNA sequencing. *Nucleic Acids Res.* 9:309–321.
36. Messing, J., and J. Viera. 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. *Gene* 19:269–276.
37. Miller, J. H. 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
38. Miyata, T., and T. Yasunaga. 1981. Rapidly evolving mouse α -globin-related pseudogene and its evolutionary history. *Proc. Natl. Acad. Sci. USA* 78:450–453.
39. Mount, S. M. 1982. A catalog of splice junction sequences. *Nucleic Acids Res.* 10:459–472.
40. Nasmyth, K. A., and K. Tatchell. 1980. The structure of transposable yeast mating type loci. *Cell* 19:753–764.
41. Norrander, J., T. Kempe, and J. Messing. 1983. Construction of improved M13 vectors using oligonucleotide-directed mutagenesis. *Gene* 26:101–106.
42. Patterson, T. E., and R. O. Poyton. 1986. Cox8, the structural gene for yeast cytochrome *c* oxidase subunit VIII. *J. Biol. Chem.* 261:17192–17197.
43. Perler, F., A. Efstratiadis, P. Lomedico, W. Gilbert, R. Kolodner, and J. Dodson. 1980. The evolution of genes: the chicken preproinsulin gene. *Cell* 20:555–566.
44. Pikielny, C. W., D. L. Teem, and M. Rosbash. 1983. Evidence for the biochemical role of an internal sequence in yeast nuclear mRNA introns: implications for UI RNA and metazoan mRNA splicing. *Cell* 34:395–403.
45. Power, S. D., M. A. Lochrie, and R. O. Poyton. 1984. The nuclear-coded subunits of yeast cytochrome *c* oxidase. III. Identification of homologous subunits in yeast, bovine heart, and *Neurospora crassa* cytochrome *c* oxidase. *J. Biol. Chem.* 259:6575–6578.
46. Power, S. D., M. A. Lochrie, K. A. Sevarino, T. E. Patterson, and R. O. Poyton. 1984. The nuclear-coded subunits of yeast cytochrome *c* oxidase. I. Fractionation of the holoenzyme into chemically pure polypeptides and the identification of two new subunits using solvent extraction and reversed phase high performance liquid chromatography. *J. Biol. Chem.* 259:6564–6570.
47. Pustell, J., and F. C. Kafatos. 1982. A convenient and adaptable package of DNA sequence analysis programs. *Nucleic Acids Res.* 10:51–59.
48. Pustell, J., and F. C. Kafatos. 1984. A convenient and adaptable package of computer programs for DNA and protein sequence management, analysis and homology determination. *Nucleic Acids Res.* 12:643–655.
49. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463–5467.
50. Schneider, T. D., G. D. Stormo, M. A. Yarus, and L. Gold. 1984. Delila system tools. *Nucleic Acids Res.* 12:129–140.
51. Seraphin, B., M. Simon, and G. Faye. 1985. Primary structure of a gene for subunit V of the cytochrome *c* oxidase from *Saccharomyces cerevisiae*. *Curr. Genet.* 9:435–439.
52. Sherman, F., G. Fink, and C. W. Lawrence. 1979. Methods in yeast genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
53. Thalenfeld, B. E., and A. Tzagoloff. 1980. Assembly of the mitochondrial membrane system. Sequence of the *Oxi2* gene of yeast mitochondrial DNA. *J. Biol. Chem.* 255:6173–6180.
54. Trueblood, C. E., and R. O. Poyton. 1987. Differential effectiveness of yeast cytochrome *c* oxidase subunit V genes results from differences in expression not function. *Mol. Cell Biol.* 7:3520–3526.
55. von Heijne, G. 1986. Mitochondrial targeting sequences may form amphipathic helices. *EMBO J.* 5:1335–1342.
56. Weaver, R. F., and C. Weissman. 1979. Mapping of RNA by a modification of the Beck-Sharp procedure: the 5' termini of 15S β -globin mRNA precursor, and mature 10S β -globin mRNA have identical map coordinates. *Nucleic Acids Res.* 7:1175–1193.
57. Wikstrom, M. 1981. Cytochrome oxidase. A synthesis, p. 155. Academic Press, Inc., New York.
58. Wilbur, W. J., and D. J. Lipman. 1983. Rapid similarity searches of nucleic acid and protein data banks. *Proc. Natl. Acad. Sci. USA* 80:726–730.
59. Wilson, D. F., M. Stubbs, R. L. Veech, M. Erecinska, and H. A. Krebs. 1974. Equilibrium relations between the oxidation-reduction reactions and adenosine triphosphate synthesis in suspensions of isolated liver cells. *Biochem. J.* 140:57–64.
60. Wright, R. M., L. K. Dircks, and R. O. Poyton. 1986. Characterization of COX9, the nuclear gene encoding the yeast mitochondrial protein cytochrome *c* oxidase subunit VIIa. Subunit VIIa lacks a leader peptide and is an essential component of the holoenzyme. *J. Biol. Chem.* 261:17183–17191.
61. Wright, R. M., C. Ko, M. G. Cumsky, and R. O. Poyton. 1984. Isolation and sequence of the structural gene for cytochrome *c* oxidase subunit VI from *Saccharomyces cerevisiae*. *J. Biol. Chem.* 259:15401–15407.