Isolation and Sequence of the Structural Gene for Cytochrome c Oxidase Subunit VI from *Saccharomyces cerevisiae*

Richard M. Wright‡, Christine Ko, Michael G. Cumskey‡, and Robert O. Poyton§

From the Department of Molecular, Cellular and Developmental Biology, University of Colorado at Boulder, Boulder, Colorado 80309

Using synthetic oligodeoxyribonucleic acid probes we have identified and isolated COX6, the structural gene for subunit VI of cytochrome c oxidase from *Saccharomyces cerevisiae*. The nucleotide sequence of COX6 predicts an amino acid sequence, for the mature subunit VI polypeptide, which is in perfect agreement with that determined previously. The nucleotide sequence of COX6 also predicts that subunit VI is derived from a precursor with a highly basic 40-amino acid NH₂-terminal presequence. This precursor has been observed after in vitro translations programmed by yeast poly(A⁺) RNA. Northern blot analysis of poly(A⁺) RNA from strain D273-10B reveals that COX6 is homologous to three RNAs of 1800, 900, and 700 bases in length. By means of Southern blot analysis, the cloned gene was shown to be co-linear with yeast chromosomal DNA and to exist in a single copy in the yeast genome. An additional open reading frame, consisting of 82 codons, terminates 22 codons upstream from COX6. It is "in frame" with the COX6 coding region.

The assembly of a functional mitochondrion requires the coordinate interaction of both mitochondrial and nuclear genomes (1, 2). We have approached the study of this interaction through investigation of the yeast heterologomeric membrane protein, cytochrome c oxidase. This enzyme, which catalyzes the terminal events of electron transport, is located in the mitochondrial inner membrane. It is composed of three polypeptide subunits which are coded by mitochondrial DNA (I, II, III) and six polypeptide subunits coded by nuclear DNA (IV, V, VI, VII, VIIa, VIII) (cf. Refs. 3 and 4). Much of what is known concerning the structure and organization of the enzyme has been reviewed recently (2, 5). To facilitate study of the mechanisms coordinating the interaction between nuclear and mitochondrial genomes in the assembly of cytochrome c oxidase, we have begun to characterize the nuclear genes coding for subunits IV, V, VI, VII, VIIa, and VIII (6, 7).

In this communication we describe the molecular cloning and DNA sequence of the gene, COX6, for subunit VI. We also identify a 40-amino acid NH₂-terminal presequence on subunit VI and three RNAs that are homologous to COX6.

**MATERIALS AND METHODS**

**RESULTS**

Cloning the Subunit VI Gene—A clone, contained in the yeast YEpl3 clone bank of Nasmyth and Tatchell (29), which carries the COX6 gene was identified by hybridization to synthetic oligonucleotides. Fig. 1 illustrates the origin of these sequences with respect to the amino acid sequence of mature subunit VI (4, 8). The probes employed to screen the YEpl3 clone bank were selected from the protein sequence on the basis of predicted minimum codon ambiguity and without regard to codon use bias (30).

The 6-1 and 6-2 oligonucleotides shown in Fig. 1 were 5' end-labeled and used as hybridization probes to screen cells from the YEpl3 clone bank. Initially, hybridization conditions were optimized for the 6-2 probe. All colonies showing hybridization to 6-2 were then screened using the 6-1 probe. Plasmid DNA was prepared from single colonies showing hybridization to both probes. This DNA was cleaved with the restriction endonuclease PstI, electrophoresed through agarose, blotted to nitrocellulose (17), and hybridized separately to the 6-1 and 6-2 probes as indicated under "Materials and Methods." Clones which corresponded to the strongest hybridization signals obtained from both probes were retained. We selected one clone, YEpl3-3, for further characterization.

A restriction endonuclease cleavage map of YEpl3-3 was produced using Southern blot hybridization and co-digestion analyses (Fig. 2). The plasmid is 15.2 kb long and contains a 4.5-kb insert. No cleavage sites have been found within the insert for Clal, PstI, SalI, or XhoI. Single cleavage sites were found within the insert for BamHI, EcoRI, PvuII, and Smal; and neither BamHI site in YEpl3 was regenerated by cloning this Sau3A-terminated fragment.

**DNA Sequence of the Subunit VI Gene**—We identified the coding region for subunit VI by hybridizing the synthetic oligonucleotides to restriction endonuclease-cleaved, Southern-blotted YEpl3-3 plasmid DNA. Fortuitously, the 6-1 and 6-2 probes hybridized to different EcoRI fragments, thus

*This research was supported by Grants GM 29838 and GM 30228 from the National Institutes of Health. 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.

This paper is dedicated to the memory of Dr. Murray Rabinowitz.

§ Established investigator of the American Heart Association, 1978–1983. To whom correspondence should be addressed; the Department of Molecular, Cellular and Developmental Biology, Campus Box 347, University of Colorado, Boulder, CO 80309.

1 "Materials and Methods" are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are available from the Journal of Biological Chemistry, 9560 Rockville Pike, Bethesda, MD 20814. Request Document No. 84M-1491, cite the authors, and include a check or money order for $1.20 per set of photocopies. Full size photocopies are also included in the microfilm edition of the Journal that is available from Waverly Press.

2 The abbreviations used are: kb, kilobase pair; bp, base pairs; ORF, open reading frame; HPLC, high-performance liquid chromatography.

15401
precisely localizing the coding region and defining the polarity of the gene. *HindII*-BglII subfragments of YEpl3 were cloned into *HindII*-BamHI-cleaved M13mp8 and M13mp9 where they were subjected to dideoxy sequence analysis (12, 25). The sequencing strategy employed is illustrated in Fig. 2, and the DNA sequence of the subunit VI gene is shown in Fig. 3.

Several features of this sequence are of interest to us. The DNA sequence is in perfect agreement with the protein sequence of subunit VI as previously determined (4, 8). It is extended beyond the NH$_2$-terminal serine residue of the mature protein by 40 amino acids, yielding a predicted preprotein of 148 amino acids and 19,968 daltons. The predicted sequence encompasses nucleotides +1 through +120. Three tandem TAA termination codons immediately follow the coding region. The 200-bp 3' are very A:T-rich (75%) and follow by more random base composition. No AAUAAA sequence is found. Nor do we find the TTTTTATA sequence observed to be necessary for termination of transcription in some yeast genes (33).

No additional ATG codons have been found, in-frame, within 229 bp 5' of what is postulated to be the initiator ATG at +1 bp. However, in-frame ATG does occur between positions −309 and −307 and positions −231 and −229. The former defines the beginning of a 246-bp open reading frame which terminates in TAG-TAG 57 bp from the subunit VI initiator ATG. Only two potential, canonical “TATA-like” sequences can be identified in the DNA sequence. These lie between −381 and −373 (G TAT AAG TA) and −402 and −395 (A TAT ACA). Both of these sequences are upstream (62 or 85 bp) from the ORF. Hence, the upstream ORF appears to interpose between the subunit VI gene and potential promoter related sequences. The ORF could code for a polypeptide of 82 amino acids and 10,715 daltons. It bears no sequence relation to any of the other cytochrome c oxidase subunits (4).

We have examined the protein-coding region of the subunit VI gene for evidence of codon use bias (Table I). Of 15 leucines, 13 are coded by TTA or TTG; all 14 arginines are coded by AGA or AGG; 7 of 9 phenylalanines are coded by TTT. Similar codon use biases have been observed in other yeast genes (cf. Ref. 30). They are not maintained in the upstream ORF.

The amino acid composition of mature subunit VI (MS) differs in some marked ways from that of the presequence (PS). Most amino acids occur in equivalent proportion when the mature sequence is compared with the presequence. The striking exceptions are the following: Arg (MS 6.5%, PS 17.5%); Ile (MS 1.8%, PS 7.5%); Thr (MS 2.7%, PS 10%); Asp (MS 8.3%, PS 0%); Glu (MS 14.8%, PS 0%). The presequence contains 25% basic amino acids and lacks acidic amino acids. A trypsin-like or cathepsin β-like proteolytic cleavage site (Arg-Lys) precedes the mature protein by 1 amino acid (Tyr) residue.

**Copy Number and Genomic Environment of the Subunit VI Gene**—We have examined nuclear DNA from wild type *Saccharomyces cerevisiae*, strain D273-10B, in an effort to char-
endonuclease cleavage sites for the enzymes used in this study are indicated. We have marked the postulated precursor protein to begin at nucleotide +1 and the mature protein to begin at nucleotide +121 (arrow). Some of the elements common to yeast genes have been indicated, including the canonical “TATA” sequences and the CAAG box (47).

Fig. 3. Sequence of the subunit VI gene. The nucleotide sequence of the entire subunit VI coding region is shown. Restriction endonuclease cleavage sites for the enzymes used in this study are indicated. We have marked the postulated precursor protein to begin at nucleotide +1 and the mature protein to begin at nucleotide +121 (arrow). Some of the elements common to yeast genes have been indicated, including the canonical “TATA” sequences and the CAAG box (47).
characterize the genomic environment of the subunit VI gene. Figs. 4 and 5 show Southern blots of D273-10B nuclear DNA cleaved by various restriction endonucleases. The hybridization probe used to track the subunit VI gene was nick-translated BglII fragments 4 and 5 from pVIB4+5. This probe contains the entire coding sequence of the subunit VI gene as well as flanking DNA both 3′ and 5′ of the gene. The presence of a single BamHI site in the YEpVI-3 insert serves to orient the organization of the nuclear DNA. Cleavage of the DNA appears to be complete inasmuch as plasmid DNA included in the digestion is cleaved to completion.

We conclude that there is only a single copy of the subunit VI gene in nuclear DNA for the following reasons. 1) Cleavage of nuclear DNA with restriction endonucleases which cut outside of the insert in YEpVI-3 all produce only one hybridization signal (Fig. 4). This signal can be tracked as a single band upon double and triple cleavage, showing that hybridization occurred on a unique fragment of DNA. 2) A number of enzymes which cleave within the YEpVI-3 insert produce either the same hybridization pattern from nuclear DNA as cleavage of the plasmid alone or detect nuclear sequences which abut within the plasmid-derived probe (Fig. 5). These patterns depend upon the choice of probe employed. For example, the BglII fragments 4 and 5 probe shown in Fig. 5 hybridizes to two EcoRI fragments, which do not co-migrate with fragments from the YEpVI-3 plasmid, since a single EcoRI site occurs in the subunit VI gene. On the other hand, HindIII and SstI which cut twice in the subunit VI gene, each produce one fragment which co-migrates with the plasmid-derived fragment and one which does not. BglII cleavage produces only BglII fragments 4 and 5 from either nuclear or plasmid DNA. These fragments also co-migrate. Double cleavage with BglII and HindIII or BglII and SstI also produce precisely co-migrating pairs of bands. Similar experiments have been conducted with other restriction fragments (HindIII fragments 2 and 3 and SstI fragment 3) from within the subunit VI gene as probes. In all cases the data support explicit co-linearity of YEpVI-3 and the nuclear DNA. Since no other hybridization can be detected with any probe, a second copy of the gene is unlikely to exist.

It is interesting that the subunit VI gene lies very close to a region which is essentially devoid of restriction sites for greater than 17,000 base pairs of DNA (Fig. 6).

Identification of Subunit VI Transcripts in D273-10B—We have examined wild-type S. cerevisiae, strain D273-10B, for evidence of RNA homologous to the YEpVI-3 insert. Total RNA and poly(A+) RNA were prepared from cells grown on YPD medium, lactate medium, or shifted from high to low glucose. RNAs were then fractionated on CH3HgOH-agarose gels, blotted to nitrocellulose, and hybridized with nick-translated BglII fragments 4 and 5 from pVIB4+5 (Fig. 7). Two major RNAs and a weakly hybridized RNA were found. They range in size from approximately 1800 bases to 900 bases and 700 bases, any one of which could readily code for the subunit VI protein. When these experiments were done with probes derived from the protein-coding region only (data not shown), the same hybridization pattern was observed. The sizes of either of the two prominent transcripts shown in Fig. 7 are in excellent agreement with the size of the transcriptional unit, inferred by DNA sequence analysis to be 800 to 900 bp, and with sucrose gradient fractionation studies of subunit VI RNA (9). From the experiment shown in Fig. 7 it appears that the subunit VI gene is glucose-shift derepressible, as suggested by previous studies (9).

Presubunit VI Has an NH2-terminal Extension—The DNA sequence of COX6 (Fig. 3) predicts that subunit VI has an NH2-terminal presequence. To obtain direct evidence for a preprotein precursor to subunit VI, poly(A+) RNA from glucose-shifted D273-10B cells was translated in vitro. When these translation products were immunoprecipitated with anti-subunit VI antibody and electrophoresed on sodium decyl sulfate-polyacrylamide gels, a protein of 18,100 daltons is observed (Fig. 8). This protein is labeled with [35S]methionine, an amino acid which can be incorporated only at the NH2-terminal position according to the COX6 DNA sequence. This protein is also competed away by the addition of purified subunit VI to the immunobinding reaction. Together, these results indicate that this protein is a precursor which contains mature subunit VI and a 40-amino acid presequence. A precursor of similar size has been found previously by others (9, 37, 38).

DISCUSSION

In this study we have presented the complete nucleotide sequence of the structural gene, COX6, for subunit VI of yeast cytochrome c oxidase. It is located on a unique fragment of nuclear DNA and therefore appears to be present as a single copy in the yeast genome. The DNA sequence of the coding and flanking regions of COX6 have revealed a number of interesting features including: the presence of a 40-amino acid presequence at the NH2 terminus of mature subunit VI; the presence of an open reading frame which terminates 57 base pairs upstream from the initiation codon of the COX6 gene and which is in-frame with it; and the presence of sequences which have been proposed to be important for transcription initiation and termination in other yeast genes.
COX6 gene is co-linear with genomic DNA and therefore has the possibility that subunit VI is made from a larger polypeptide with an in-frame ATG initiation codon. This codon is the only in-frame ATG initiation codon. Since it is clear from Figs. 4 and 5 that the cloned COX6 gene is 40 codons away from an in-frame ATG initiation codon. This codon is the only in-frame ATG between the protein-coding sequence for subunit VI and the termination codon for the upstream open reading frame. Since it is clear from Figs. 4 and 5 that the cloned COX6 gene is co-linear with genomic DNA and therefore has not been rearranged during cloning, these results suggest that the subunit VI polypeptide is preceded by a presequence which is 40 amino acids long. A polypeptide which is immunoreactive with anti-subunit VI serum and which is larger than subunit VI is synthesized in vitro in reticulocyte lysates programmed with yeast poly(A*) RNA. We have found that this polypeptide is approximately 5200 daltons larger than subunit VI and can be labeled with [35S]methionine, which can only occur at a primary translation product of the COX6 gene. Thus, it is most likely a primary translation product of the COX6 gene.

Subunit VI Protein-coding Sequence—The polypeptide sequence deduced from the COX6 gene between nucleotides +120 and +444 (Fig. 3) shows perfect agreement with the amino acid sequence of the subunit VI polypeptide (4, 8). Since subunit VI has been proposed previously to be synthesized from a larger precursor (9, 37, 38), it was of interest to determine if the coding sequence of the polypeptide would indicate the presence of either a NH2- or COOH-terminal extension. The perfect alignment of the serine carboxyl group of the subunit VI polypeptide with a serine codon which precedes a triple termination codon clearly rules out the possibility that subunit VI is made from a larger polypeptide with a COOH extension. On the other hand, the nucleotide sequence between positions 1 and 120 clearly predicts an NH2-terminal presequence. Indeed, the NH2-terminal serine codon of the subunit VI polypeptide is 40 codons away from an in-frame ATG initiation codon. This codon is the only in-frame ATG between the protein-coding sequence for subunit VI and the termination codon for the upstream open reading frame. Since it is clear from Figs. 4 and 5 that the cloned COX6 gene is co-linear with genomic DNA and therefore has not been rearranged during cloning, these results suggest that the subunit VI polypeptide is preceded by a presequence which is 40 amino acids long. A polypeptide which is immunoreactive with anti-subunit VI serum and which is larger than subunit VI is synthesized in vitro in reticulocyte lysates programmed with yeast poly(A*) RNA. We have found that this polypeptide is approximately 5200 daltons larger than subunit VI and can be labeled with [35S]methionine, which can only occur at the initiation codon of the COX6 gene. Thus, it is most likely a primary translation product of the COX6 gene.

The presequence of subunit VI has a high content of basic amino acids which are randomly dispersed throughout, lacks acidic amino acids, and contains a trypsin-like or cathepsin β-like proteolytic cleavage site (Arg-Lys) near its carboxy-terminal junction with mature VI. There is no primary sequence homology between this presequence and the three other presequences (pre-CCP, pre-Tu, and preproteolipid) which have been characterized so far on other nuclear-coded mitochondrial proteins (i.e. yeast cytochrome c peroxidase (39); yeast mitochondrial elongation factor Tu (40); and the Neurospora crassa proteolipid subunit of ATP synthetase (41)). However, it is interesting that all four presequences are...
basic. In three of the four (pre-Tu, preproteolipid, presumunit VI), the basic amino acids are distributed throughout the sequence; making it unlikely that they contain a transmembranous segment like that proposed for pre-CCP (39). It is also interesting that within each of the four presquences there are clusters of dibasic amino acids which could be processing sites for trypsin or cathepsin β-like proteases. Moreover, there is homology in the positioning of one of these sites in the Neurospora preproteolipid and yeast pre-pre VI. In both, the dibasic amino acid cluster occurs at positions –2 and –3 with respect to the NH₂ terminus of each mature polypeptide. For preproteolipid the sequence of this site is –5'-Lys-Arg-X-NH₂ terminus and in pre-pre VI it is –5'-Arg-Lys-X-NH₂ terminus. This observation raises the possibility that this site may be important for the removal of the presquence from the mature protein and suggests that a similar enzyme may process both of these presquences.

5′-Flanking Sequences—Two features of the 5′-flanking sequence of COX6 are of interest. First, there is an 82-codon region upstream from the ORF which is reminiscent of a number of other yeast genes (Refs. 34-36). A second feature of the 5′-flanking sequence of COX6 which is of special interest is the finding that T-A-T-A sequences are normally found from COX6 is of special interest is the finding that T-A-T-A sequences are normally found from COX6 is reminiscent of a number of other yeast genes (e.g. HIS3, 42; LEU2, 35; and URA3, 34). At present, the role, if any, of the products of these ORFs in the expression of their downstream genes is unclear (cf. Refs. 34–36).

The predicted amino acid sequence from this reading frame may process both of these presequences. For preproteolipid the sequence of this site is –5'-Flanking

Acknowledgments—We would like to thank Tom Patterson and Cindy Trueblood for valuable advice and discussion throughout this work. In addition we thank Bob Thompson, Steve Cline, and Linda Breeden for their help in the synthesis of the oligonucleotides used in this study. Cathy Inouye’s careful and speedy attention to the preparation of this manuscript is gratefully acknowledged.

REFERENCES

Yeast Cytochrome Oxidase Subunit VI Structural Gene

Yeast cytochrome oxidase subunit VI from Saccharomyces cerevisiae was studied by Richard M. Wright, Christine Ko, Michael C. Cumsky, and Robert O. Poyton.

**Materials and Methods**

**Enzyme Extraction**

The yeast strain was grown in YPD broth at 30°C for 24 hours. The cells were harvested by centrifugation and washed twice with 50 mM sodium phosphate buffer, pH 7.2. The resulting cell pellets were resuspended in 50 mM sodium phosphate buffer, pH 7.2, and sonicated for 1 minute at 20% energy output. The supernatant was collected by centrifugation at 15,000 g for 10 minutes at 4°C.

**SDS-PAGE and Western Blotting**

The crude extract was fractionated by electrophoresis through a 12% polyacrylamide gel and transferred to a nitrocellulose membrane. The membrane was incubated with polyclonal antibodies raised against yeast cytochrome oxidase subunit VI. The membrane was then incubated with horseradish peroxidase-conjugated secondary antibodies and developed with chemiluminescent substrate.

**Mass Spectrometry**

The purified protein was analyzed by mass spectrometry to determine its amino acid sequence.

**Conclusion**

The results of this study provide insights into the structure and function of yeast cytochrome oxidase subunit VI, which may have implications for understanding oxidative phosphorylation in eukaryotic cells.