GENOMICS 10, 1-9 (1991)

Structure of the Human Cytochrome c Oxidase Subunit Vb Gene and Chromosomal Mapping of the Coding Gene and of Seven Pseudogenes

MARGARET I. LOMAX,* CHIH-LIN HSIEH,† BASIL T. DARRAS,‡ AND UTA FRANCKE†

*Department of Anatomy and Cell Biology, University of Michigan, Ann Arbor, Michigan 48109; †Department of Genetics and Howard Hughes Medical Institute, Stanford University Medical Center, Stanford, California 94305; and ‡Department of Pediatrics, New England Medical Center, Boston, Massachusetts 02111

Received September 14, 1990; revised December 14, 1990

Subunit Vb of mammalian cytochrome *c* oxidase (COX; EC 1.9.3.1) is encoded by a nuclear gene and assembled with the other 12 COX subunits encoded in both mitochondrial and nuclear DNA. We have cloned the gene for human COX subunit Vb (*COX5B*) and determined the exon–intron structure by both hybridization analysis and DNA sequencing. The gene contains five exons and four introns; the four coding exons span a region of approximately 2.4 kb. The 5′ end of the *COX5B* gene is GC-rich and contains many *Hpa*II sites. Genomic Southern blot analysis of human DNA probed with the human COX Vb cDNA identified eight restriction fragments containing COX Vb-related sequences that were mapped to different chromosomes with panels of human × Chinese hamster somatic cell hybrids. Because only one of these fragments hybridized with a 210-bp probe from intron 4, we conclude that there is a single expressed gene for COX subunit Vb in the human genome. We have mapped this gene to chromosome 2, region cen–q13.

INTRODUCTION

Cytochrome *c* oxidase (COX; EC 1.9.3.1), the terminal enzyme of the electron transport chain, transfers electrons from reduced cytochrome *c* to oxygen, in the process generating an electrochemical gradient across the mitochondrial inner membrane (reviewed in Hatefi, 1985; Capaldi, 1988). The mammalian enzyme is composed of 13 polypeptide subunits—3 encoded in mtDNA and 10 in nuclear DNA (reviewed in Kadenbach *et al.*, 1987; Capaldi, 1988; Chomyn and Attardi, 1987). The nuclear-coded COX subunits can be divided into two groups: those with muscle-specific isoforms (reviewed in Capaldi *et al.*, 1988; Lomax and Grossman, 1989) and those that are identical in all tissues (Yanamura *et al.*, 1988). The nuclear genes for these latter ubiquitous subunits are expressed constitutively and probably represent examples of "housekeeping" genes (Bird, 1986).

It has been possible to map human COX genes by means of Southern blot analysis of DNA from panels of somatic cell hybrids. For example, the single-copy gene for COX subunit VIII (*COX8*) was mapped to chromosome 11, region q12–q13 (Rizzuto *et al.*, 1989), and the gene for COX subunit IV (*COX4*) to chromosome 16 region q22–qter (Darras *et al.*, 1987; Lomax *et al.*, 1990). The isolation of a human COX Vb cDNA (Zeviani *et al.*, 1988) provided a probe for cloning and mapping the *COX5B* gene, whose chromosomal location was obscured by the presence of several *COX5B*-related genes or pseudogenes. We report here the isolation and DNA sequence of the expressed gene for human COX subunit Vb (*COX5B*) and the chromosomal location of the expressed gene and seven pseudogenes, as determined by analysis of panels of somatic cell hybrids with cDNA, genomic, and intron probes.

MATERIALS AND METHODS

Materials

Restriction enzymes were from BRL, New England Biolabs, or Boehringer-Mannheim. The DNA labeling system for "random-primer" labeling and [α-^{35}S]thiodATP for DNA sequencing were from Amersham. [α-^{32}P]dCTP for random primer labeling was from NEN. The Sequenase DNA sequencing kit was from USB.

Subcloning and Restriction Mapping of the COX5B Gene

Plasmid pCOX5bR3.4 containing the human *COX5B* gene was constructed by cloning a 3.4-kb *Eco*RI fragment from the 5′ end of λCOX5B-1 into the

0888-7543/91 $3.00

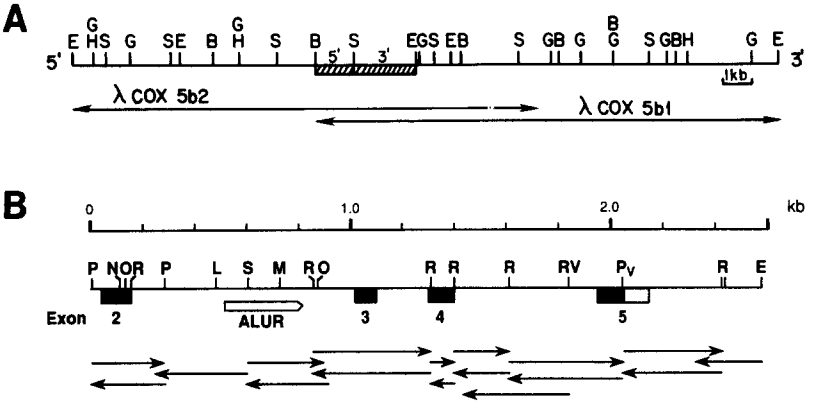

FIG. 1. Restriction maps of the *COX5B* gene. (**A**) Restriction map of the human *COX5B* locus, based on two overlapping genomic clones: λCOX5B-1 and λCOX5B-2. The hatched areas indicate the location and orientation of the *COX5B* gene, as defined by hybridization. Arrows define the genomic regions contained within each genomic clone. Restriction enzymes: B, *Bam*HI; E, *Eco*RI; G, *Bgl*II; H, *Hin*dIII; S, *Sst*I. (**B**) Restriction map of a 2.6-kb PSTI–*Eco*RI-fragment of a 3.4-kb *Eco*RI subclone of λCOX5B-1 (pCOX5bR3.4) encompassing the coding regions and 3′-untranslated region of the *COX5B* gene. Filled boxes represent coding (translated) exons, open boxes the 3′-untranslated region, and straight lines the introns. The open arrow (designated ALUR) denotes the position and orientation of the *Alu*I repeat in intron 2. Restriction enzymes: E, *Eco*RI; P, *Pst*I; L, *Sal*I; M, *Sma*I; N, *Not*I: O, *Nco*I: R, *Rsa*I; RV, *Eco*RV; S, *Sst*I; V, *Pvu*II. Arrows below the restriction map denote the sequencing strategy for the human *COX5B* gene.

*Eco*RI site of the plasmid vector BlueScriptKS$^-$ (Stratagene, La Jolla, CA). Sites for restriction enzymes that cleaved uniquely within the polylinker sequence were mapped within the genomic clone. Deletion subclones were generated by digesting the plasmid with enzymes that cleaved within both the genomic region and the polylinker, diluting the plasmid DNA to a concentration of 0.1–0.5 μg/ml, and religating the fragments. Deletions from internal *Pst*I, *Sal*I, *Sst*I, and *Eco*RV sites were generated in this manner and sequenced. Additional subclones of the 1.2-kb *Sst*I–*Eco*RV region were generated by cloning *Rsa*I fragments into the *Sma*I site of pUC13.

DNA Sequencing

Dideoxy DNA sequencing (Sanger *et al.*, 1977) was performed with Sequenase (USB) on duplex plasmid DNA (Chen and Seeberg, 1985) isolated by the alkaline lysis method (Birnboim and Doly, 1979) from mini-preps or from 250-ml cultures. Primers complementary to the T3 and T7 promoter sequences in BlueScriptKS$^-$ were used to sequence pCOX5bR3.4 and its deletion subclones. *Rsa*I subclones in pUC13 were sequenced using the M13 universal and reverse sequencing primers. Synthetic oligonucleotide primers (20-mers) were used to obtain the sequence of complementary strands in certain regions; they were synthesized in the University of Michigan DNA Synthesis Facility on an Applied Biosystems DNA synthesizer. DNA sequences were stored and analyzed using the ASSEMGEL program of PC/GENE (Intelligenetics, Mountain View, CA).

Somatic Cell Hybrids

A panel of 16 hybrid clones derived from seven independent fusion experiments between Chinese hamster and human cell lines (for summary see Yang-Feng *et al.*, 1986) was used for primary chromosomal assignment in human. Four rodent × human hybrids from series X, XVIII, and XIX containing different regions of human chromosome 2 were used for regional mapping (Alonso *et al.*, 1988; Barton *et al.*, 1989).

Probes

Two probes were used for primary mapping: Probe 1—a human pCOX5b.222 cDNA clone of 480 bp (Zeviani *et al.*, 1988); and Probe 2—a 1.2-kb *Sst*I–*Eco*RV fragment from genomic clone pCOX5bRS2.0 (the *Sst*I deletion clone of pCOX5bR3.4, which retained the 2.0-kb region from *Sst*I to *Eco*RI). Probe 2 contains exons 3 and 4, portions of introns 2 and 4, and all of intron 3 of human *COX5B*. For chromosomal assignment of the *COX5B* coding sequence, a 210-bp *Rsa*I fragment (1402–1612, Fig. 2) containing a portion of intron 4 of the *COX5B* gene (Probe 3) was used. The probes were isolated in 0.8% low-melting-point agarose and labeled with [α-^{32}P]dCTP by the random priming labeling method (Feinberg and Vogelstein, 1983).

Genomic Southern Blot Analysis

Human genomic DNA isolated from blood leukocytes was digested with the indicated restriction enzymes. DNA fragments (10 μg) from each digest were

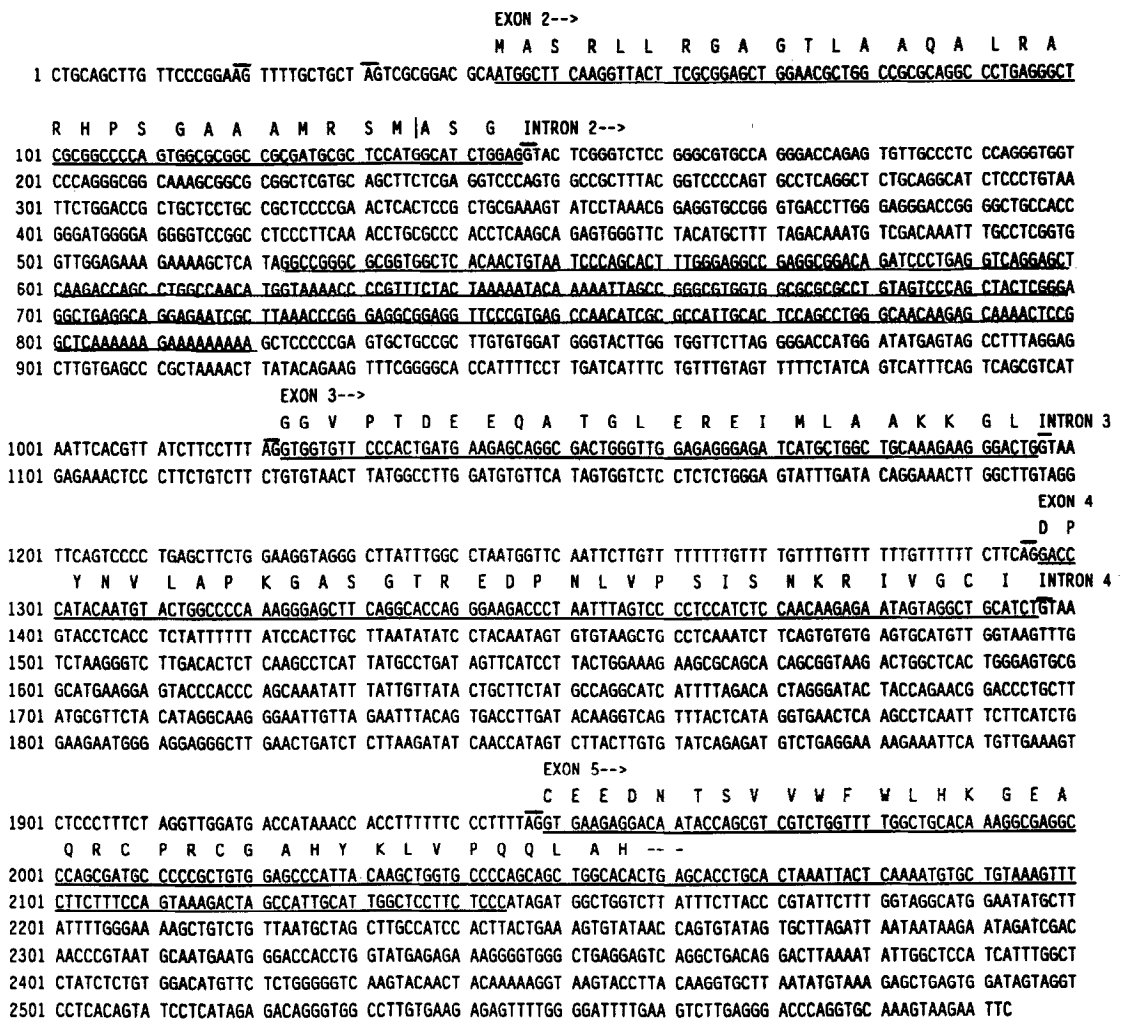

FIG. 2. DNA sequence of the human *COX5B* gene. The nucleotide sequence of 2593 bp of genomic DNA containing exons 2 through 5 and introns 2 through 4 is presented, starting in intron 1 at the *Pst*I site, which is designated nt 1, and continuing to the *Eco*RI site beyond the gene. Coding regions are underlined and the deduced amino acid sequence is presented above the DNA sequence. Consensus 5′-donor and 3′-acceptor splice sites, including the two potential 3′-acceptor sequences preceding the ATG initiation codon, are overlined. The *Alu* repeat sequence extending from base 521 to 820 is underlined.

separated by electrophoresis on 0.8% agarose gels, denatured, and transferred to Nytran nylon membranes (Schleicher & Schuell) by blotting. Southern blot hybridization analysis was performed at 42°C for 18–24 h in 50% formamide, 5× SSPE, 1% SDS, 50 μg/ml denatured salmon testis DNA with ^{32}P-labeled cDNA probe. After hybridization, membranes were washed at 60°C with decreasing concentrations of SSC containing 0.1% SDS. Restriction enzymes *Hin*dIII and *Pst*I were used for primary mapping, and *Hin*dIII was used for regional mapping. DNA from each of the hybrid and parental control cell lines was digested with enzyme, and then fragments were separated on 0.8% agarose gels and transferred to Hybond-N membranes (Amersham) by the method of Southern (1975). The conditions for prehybridization, hybridization, and washing of filters were described previously (Hsieh *et al.*, 1990). All filters were exposed to Kodak X-Omat AR films at −70°C for varying lengths of time.

RESULTS

Isolation of Genomic Clones for Human COX Vb

Southern blot analysis of human genomic DNA digested with *Pst*I revealed many fragments that hybridized with the pCOX5b cDNA (Zeviani *et al.*, 1988). We therefore anticipated that a genomic library would contain clones encompassing different genomic regions, including several pseudogenes. To isolate the human *COX5B* gene, we screened a human genomic library (Lawn *et al.*, 1978) containing *Alu*I–*Hae*III partial digest fragments of human DNA

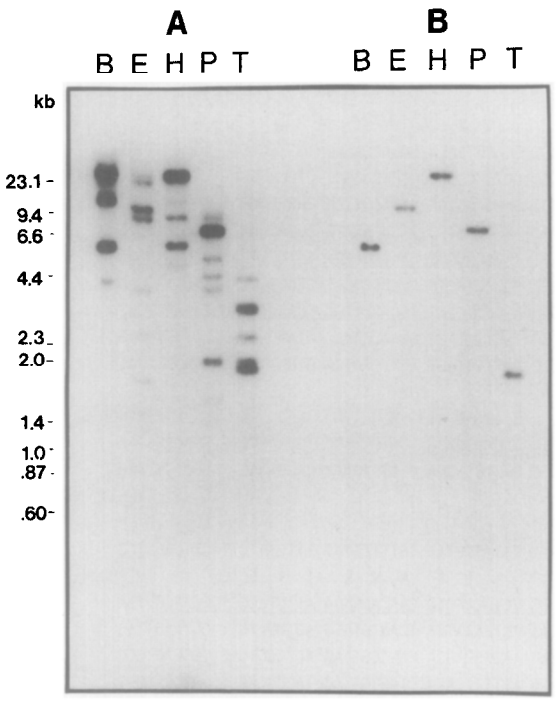

FIG. 3. *COX5B*-related sequences in the human genome. Genomic Southern blots of human DNA probed with (**A**) the pCOX5b cDNA (Probe 1) and (**B**) a 210-bp *Rsa*I fragment (Probe 3) from intron 4 of the *COX5B* gene. Restriction enzymes used are B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; T, *Taq*I.

cloned into the *Eco*RI site of bacteriophage λCharon4A (obtained from A. S. Lee with permission of Dr. T. Maniatis, Harvard University). Approximately 10^6 plaques were plated and screened by plaque hybridization techniques according to standard procedures (Benton and Davis, 1977) with Probe 1, the full-length human COX Vb cDNA insert from pCOX5b (Zeviani *et al.*, 1988), as hybridization probe. Phage DNA was isolated from small-scale lysates (Maniatis *et al.*, 1982) of the six plaques that gave duplicate hybridization signals and characterized by restriction mapping and Southern blot analysis with the pCOX5b cDNA (Probe 1). *Eco*RI digestion revealed only two types of clones, designated λCOX5B-1 and λCOX5B-2 (Fig. 1A). Southern blot analysis of phage DNA revealed that each clone contained an approximately 6-kb *Pst*I fragment that hybridized with the cDNA; therefore every clone isolated in this screen contained only one genomic region. Southern blot analysis with hybridization probes encoding either the 5′ end (a 115-bp *Eco*RI–*Nco*I fragment) or the 3′ end (a 100-bp *Pvu*II–*Eco*RI fragment) of the cDNA indicated that both λCOX5B-1 and λCOX5B-2 contained the entire *COX5B* coding region. Further restriction mapping confirmed that the two clones represented overlapping clones of the same genomic region (Fig. 1A). We therefore subcloned a 3.4-kb *Eco*RI fragment from the 5′ end of λCOX5B-1 for more detailed analysis.

The exon–intron structure of the *COX5B* gene and the sequencing strategy used are presented in Fig. 1B. This 3.4-kb *Eco*RI fragment contained the four coding exons, designated exons 2, 3, 4, and 5. The sequence of these four exons and that of the three introns are presented in Fig. 2. Exon 2 encodes the 31 amino acids of the presequence (amino acids −31 to −1) plus amino acids 1 to 3 of mature COX Vb; exon 3 encodes amino acids 3 to 28; exon 4 encodes amino acids 29 to 62; and exon 5 encodes amino acids 62 to 97 plus the 83-bp 3′-untranslated region. The DNA sequence of the coding exons and the 3′-untranslated region in this genomic region is essentially identical to the sequence of the cDNA, confirming that this gene encodes the pCOX5b cDNA. The only discrepancy between the genomic and the cDNA sequences occurred in exon 5. The sequence we determined for codons 77–78 of the human *COX5B* gene is GGC–GAC, which predicts the sequence Gly–Glu, whereas the published cDNA sequence is GGG–CAG or Gly–Gln. Since the bovine (Zeviani *et al.*, 1988), rat (Goto *et al.*, 1989), and mouse (Basu and Avadhani, 1990) cDNAs have GAC (Glu) at position 78, we assume that our genomic sequence (GGC–GAC) is correct and that the sequence (GGG–CAG) in the cDNA probably represents a sequencing error.

All intron–exon junctions of the *COX5B* gene (Fig. 2) conformed to the consensus splice junction sequences (Sharp, 1981; Padgett *et al.*, 1986), except for the 3′-splice junction preceding exon 2. The sequence preceding the ATG initiator Met codon in the pCOX5b cDNA is $(T)_{13}$CCA–ATG; only the CA (nucleotides 42–43, Fig. 2) preceding the ATG initiator Met codon is present in the genomic sequence at the end of intron 1. There is no consensus 3′-splice sequence preceding the CA, although there are at least two potential 3′-acceptor splice sequences just upstream at nucleotides 19–20 and 31–32. This observation and the poly(dT) tract in the 5′-untranslated region of this cDNA and other cDNAs isolated from this human endothelial cell library (e.g., see Fabrizi *et al.*, 1989) suggested that the cDNA contained a cloning artifact at the 5′ end. Additional support for this hypothesis is that the genomic sequence preceding the ATG initiation Met codon agrees well with the consensus sequence proposed by Kozak (1987) for initiation of translation of eukaryotic genes, suggesting that this sequence should be part of exon 2. This artifact in the cDNA precluded mapping exon 1 (encoding the 5′-untranslated region) by hybridization with the pCOX5b cDNA.

Chromosomal Mapping

Southern blots of human genomic DNA probed with the pCOX5b cDNA (Probe 1) revealed several

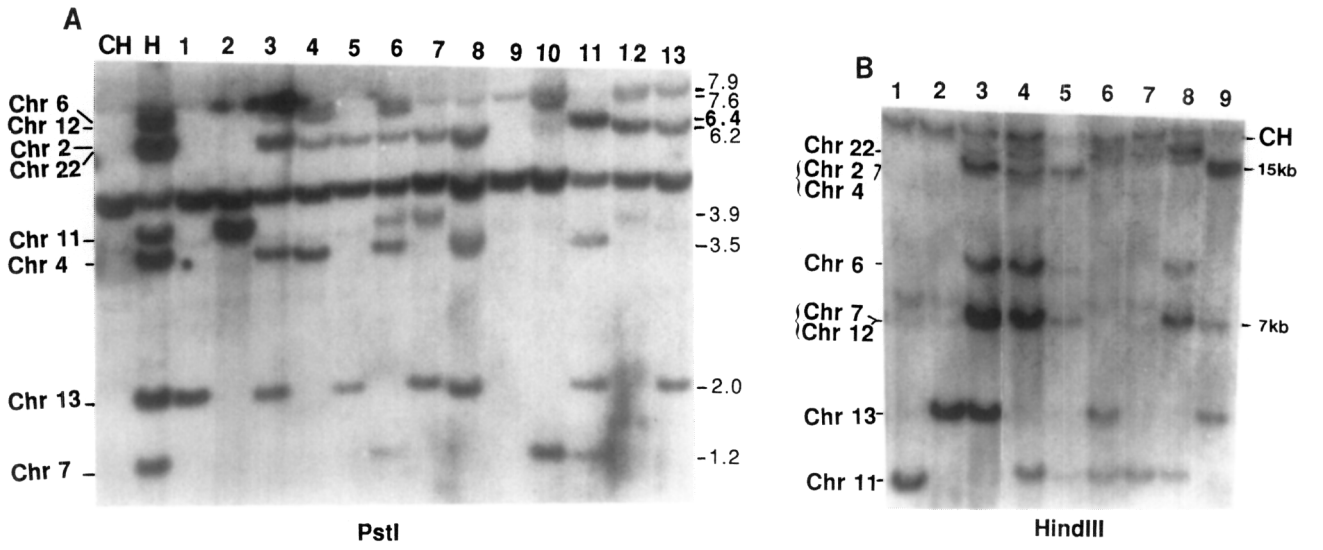

FIG. 4. Mapping of *COX5B* sequences. Hybridization of human *COX5B* cDNA (Probe 1) to *Pst*I-digested (**A**) and of the genomic exon/intron Probe 2 to *Hin*dIII-digested (**B**) Chinese hamster × human somatic cell hybrid and control DNA revealed discordant segregation of all human fragments. The chromosomal assignments for each band are marked on the left, and sizes in kb on the right. (**A**) The closely migrating 7.6- and 7.9-kb human fragments are separated in the hybrids: lanes **3, 4, 6, 10,** and **11** are positive for the 7.6-kb band. Similarly, the 6.4-kb human fragment containing the *COX5B* coding gene is separable from the 6.2-kb pseudogene fragment. Hybrids positive for the 6.4-kb fragment are in lanes **3, 8,** and **11.** (**B**) The heavy 15-kb *Hin*dIII fragment appeared to cosegregate with human chromosomes 2 and 4. After hybridization of the same filter with the intron Probe 3, the 15-kb fragment was detected in lanes **3, 4,** and **9** but not in lane **5.** These results led us to conclude that there are two overlapping 15.0-kb *Hin*dIII fragments that map to chromosomes 2 and 4, with that on chromosome 2 containing the COX5B coding gene. Based on comparison with results obtained with the *Pst*I-digested panel, two 7.0-kb *Hin*dIII fragments were postulated and assigned to chromosomes 7 and 12.

bands in each of five different restriction digests (Fig. 3A); however, Southern blots probed with a unique sequence probe (Probe 3), a 210-bp *Rsa*I fragment from intron 4, revealed a single hybridizing fragment in each digest (Fig. 3B). Initially, mapping of the human gene was attempted by hybridizing the ^{32}P-labeled cDNA probe to *Hin*dIII- and *Pst*I-digested genomic DNA from Chinese hamster × human somatic cell hybrids and controls. Chinese hamster bands of 4.7 and 17.5 kb were seen in *Pst*I- and *Hin*dIII-digested DNA, respectively (Fig. 4). The 4.7-kb *Pst*I fragment in human was not scored because it was indistinguishable from the Chinese hamster band. We were able to assign the two 15.0-kb *Hin*dIII fragments to different chromosomes by the intensity of hybridization and after assigning the coding gene with an intron probe. Sites on eight chromosomes were identified by the cDNA probe with both *Pst*I and *Hin*dIII hybrid panels (Fig. 4). The same fragments hybridized with the genomic probe 2. The presence or absence of the individual *Pst*I and *Hin*dIII fragments in the 16 hybrid cell lines is listed in Table 1 and compared with the human chromosome content of the hybrids. The *Pst*I and *Hin*dIII fragments that represent the same genetic locus could be identified by their pattern of segregation in the hybrid cell lines. The percentage of discordant hybrids is calculated for each fragment pair and each chromosome. It is apparent that each fragment is concordant only with a single chromosome and that they are all present on different chromosomes.

We then used the 210-bp human intron probe (Probe 3) to determine the localization of the coding gene. A single 15.0-kb *Hin*dIII fragment (Fig. 5, lane 1) and a single 6.4-kb *Pst*I fragment (data not shown) were detected in human control DNA, and two weakly hybridizing *Hin*dIII fragments of 4.4 and 3.2 kb were observed in rat DNA (Fig. 5, lane 3), while no cross-hybridization was detected with Chinese hamster DNA (Fig. 5, lane 2). The human 15.0-kb *Hin*dIII and 6.4-kb *Pst*I bands were present only in hybrid clones containing human chromosome 2. No human signal was detected in hybrids not containing human chromosome 2. The presence or absence of human chromosome 2 in hybrid cell lines was in perfect concordance with that of the human *COX5B* signal. All other human chromosomes were excluded by at least two discordant hybrids (Table 1). We have assigned *COX5B* to region cen–q13 of chromosome 2 (Fig. 6) by using four hybrids containing partially overlapping regions of human chromosome 2 in the absence of an intact chromosome 2. In Table 2, the chromosomal assignments of the *COX5B* coding gene and of seven pseudogenes (COX5BP1 to COX5BP7) are summarized. Hybrid cell lines that contained defined regions of chromosomes 4 and 11 were used to localize region-

TABLE 1

Correlation of COX5B-Specific Human Restriction Fragments and Human Chromosomes in Chinese Hamster × Human Somatic Cell Hybrids

	Signal																														
*Hin*dIII:	15.0	15.0	8.4	7.0	4.1	7.0	5.2	16.0																							
									Human chromosome																						
Hybrid *Pst*I:	6.4	3.5	7.9	1.2	3.9	7.6	2.0	6.2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X
XII-A	–	–	–	–	–	–	+	–	–	–	+	–	–	–	–	+	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–
XII-B	–	–	–	–	+	–	–	–	–	–	+	–	–	–	–	–	–	–	+	–	–	–	–	+	–	+	–	–	+	–	P
XII-C	–	–	–	–	–	–	+	–	–	–	+	–	–	–	–	–	–	–	–	–	+	+	+	P	–	P	–	–	–	–	P
XII-D	+	+	+	–	–	+	+	+	+	+	+	+	–	+	–	+	+	–	–	+	+	+	+	+	–	+	+	–	+	+	P
XII-E	–	+	+	–	–	+	–	+	+	–	+	+	–	+	–	+	–	–	–	+	–	+	+	+	–	+	+	–	+	+	R
XIII-A	–	–	–	–	–	–	+	+	–	–	–	–	L	–	L	+	+	–	–	–	+	P	+	–	–	+	–	+	–	+	P
XVII-A	–	+	+	+	+	+	–	+	–	–	+	P	–	+	+	–	–	+	+	+	–	–	+	–	–		–	+	+	+	+
XVII-B	+	–	+	–	+	+	–	–	–	+	P	–	–	+	–	–	–	–	+	+	–	+	–	–	–	+	–	+	+	–	+
XXI-A	–	–	+	–	+	–	+	+	–	–	–	–	+	+	–	+	–	–	+	–	+	–	+	–	+	+	+	+	+	+	–
XXI-B	–	–	–	–	+	–	–	+	+	–	–	–	–	–	P	+	–	–	P	–	–	+	–	–	+	–	+	–	P	+	+
31-A	–	–	+	–	+	+	–	+	+	–	–	P	–	+	–	–	P	–	P	+	R	–	+	+	–	+	+	+	+	+	P
31-B	+	+	–	+	–	–	+	–	–	P	–	+	–	–	+	–	+	–	–	–	+	+	+	–	–	+	+	–	–	–	+
31-C	–	–	+	–	–	–	+	+	–	–	+	P	+	+	–	–	–	+	–	–	R	+	–	+	+	–	–	+	–	+	P
XIII-B[a]	–	–	+	–	–	–	–	–	–	–	+	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
XVIII-A[a]	–	+	+	–	–	–	+	+	–	–	–	+	–	+	–	P	–	–	–	–	+	+	–	+	–	–	+	?	+	+	+
XXI-C[a]	–	–	+	+	–	+	–	+	+	–	+	–	+	+	+	+	–	+	P	+	–	–	+	–	L	L	+	–	+	+	+
Locus	% Discordant hybrids																														
COX5B									38	0	58	20	42	46	27	54	17	38	45	31	45	33	54	50	46	42	38	54	50	69	33
COX5BP1									31	25	42	0	50	38	18	46	25	31	55	23	55	42	31	42	54	50	31	62	42	46	33
COX5BP2									38	42	42	20	42	0	64	54	58	38	36	15	64	50	38	33	46	42	38	23	17	23	50
COX5BP3									46	25	58	20	33	54	0	62	25	15	36	38	55	58	46	58	38	67	46	46	58	62	33
COX5BP4									46	50	67	70	50	38	45	62	67	46	0	38	82	75	62	58	38	50	46	31	17	46	33
COX5BP5									23	25	42	20	58	15	45	54	42	38	36	0	73	50	38	33	54	42	38	38	17	38	33
COX5BP6									69	50	50	50	33	62	55	38	33	54	82	77	0	42	38	58	46	50	54	54	83	54	83
COX5BP7									31	58	58	40	42	23	55	31	50	46	55	38	55	50	31	42	38	50	31	31	33	0	50

Note. Data for chromosomes that are structurally rearranged (P, R, or ?) or present in fewer than 10% of cells (L) were excluded from calculation of percentage discordant hybrids.

[a] Hybrids used only in the *Pst*I panel were also excluded.

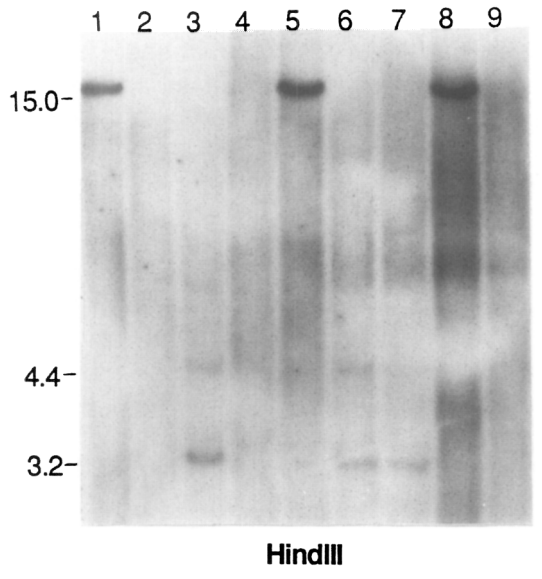

FIG. 5. Regional mapping of *COX5B* coding gene. Hybridization of ^{32}P-labeled human intron probe (Probe 3) to a Southern blot of *Hin*dIII-digested DNA from rodent × human hybrid cell lines and controls. Lane **1**, human diploid lymphoblastoid cells; lane **2**, Chinese hamster cells V79/380-6; lane **3**, rat hepatoma cell line; lanes **4–9**, rodent × human hybrid cell lines. Lanes **5** and **8** contain the 15.0-kb human *COX5B* fragment. Lanes **4** and **9** (Chinese hamster × human hybrids) have no detectable *COX5B* signal. Lanes **6** and **7** contain only the 4.4- and 3.2-kb rat signals. DNA in lanes **4–7** was extracted from hybrids that have retained different regions of human chromosome 2; lane **4**, hybrid C; lane **5**, hybrid B; lane **6**, hybrid A; lane **7**, hybrid D (Fig. 6). DNA in lane **8** was from hybrid that has retained an intact human chromosome 2, and that in lane **9** had no chromosome 2.

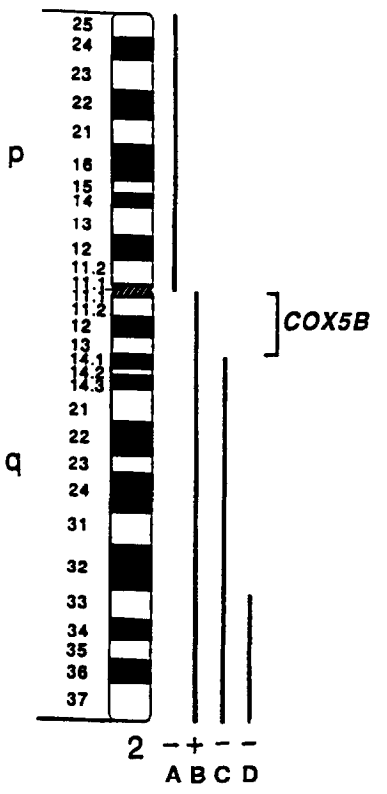

FIG. 6. Localization of *COX5B* coding gene on chromosome 2. Vertical bars illustrate the regions of chromosome 2 present in somatic cell hybrids A–D. A + indicates hybrid positive for human *COX5B* sequence, and − indicates hybrid negative for human *COX5B* sequence.

ally COX5BP1 and COX5BP4, respectively (data not shown).

DISCUSSION

Structural Features of the COX5B Gene

The *COX5B* gene is composed of five exons and four introns. Exon 1, which has not yet been localized, encodes only the 5′-untranslated region; exon 2 the presequence; and exons 3, 4, and 5 the remainder of the cDNA. It has been suggested that exons encode functional domains of proteins; thus it is sometimes possible to predict the exon–intron structure of a gene on the basis of functional domains of the protein. One prominent feature of most proteins destined for mitochondria is a short (20- to 40-amino-acid) basic N-terminal extension (presequence) that directs precursors to the mitochondrial outer membrane. Thus far, all genes for mitochondrial proteins with presequences have separate exons encoding the presequence. In the *COX5B* gene, this functional domain also is encoded by a separate exon, exon 2, as in the bovine *COX4* gene, the only other characterized COX gene encoding a subunit with a presequence (Bachman *et al.*, 1987). Goto *et al.* (1989) recently reported the sequence of a rat COX subunit Vb cDNA, which they assumed to be synthesized without a presequence, based on the fact that the rat cDNA, like the human pCOX5b cDNA, had an ATG Met codon preceding the sequence of the mature protein. However, the 54-bp rat sequence preceding this Met is an open

TABLE 2

Chromosomal Assignment of Different Fragments That Hybridized with the *COX5B* Probes 1 and 2

Gene name	Chromosome	Fragment size (kb)	
		*Pst*I	*Hin*dIII
COX5B	2cen–q13	6.4	15.0
COX5BP1	4cen–q31	3.5	15.0
COX5BP2	6	7.9	8.4
COX5BP3	7	1.2	7.0
COX5BP4	11q	3.9	4.1
COX5BP5	12	7.6	7.0
COX5BP6	13	2.0	5.2
COX5BP7	22	6.2	16.0

reading frame with a high degree of both nucleotide (82%) and amino acid sequence identity (84%) to the human sequence. A longer mouse cDNA (Basu and Avadhani, 1990) also contains an in-frame Met codon upstream, thus predicting a presequence. We conclude, therefore, that the rat clone is a partial cDNA lacking the 5′-untranslated region and the first 12 codons of the presequence.

The human cytochrome c_1 gene is the only gene in which two exons encode the presequence (Suzuki *et al.*, 1989); however, because the cytochrome c_1 precursor is processed in two separate steps, this two-exon structure still correlates with the two functional domains of the presequence. Thus, Gilbert's hypothesis for exon shuffling (Gilbert, 1985) appears to be valid in this case.

An additional feature common to most nuclear-coded COX subunits is a hydrophobic, putative membrane-spanning region flanked by hydrophilic N-terminal and C-terminal domains. COX subunits Va and Vb, however, are the two nuclear-coded COX subunits that have no such hydrophobic, membrane-spanning regions; it has also been determined experimentally that these two subunits do not span the mitochondrial inner membrane (Zhang and Capaldi, 1989). Thus, it was not possible a priori to predict the intron–exon structure of this gene.

None of the four introns of the *COX5B* gene sequenced thus far is larger than 1 kb. The largest, intron 2, contains an *Alu* repeat (nts 513–820, Fig. 2) with 92% sequence identity to the consensus genomic *Alu* repeat sequence (Schmid and Shen, 1985). This repeat is more similar to the consensus sequence for the older *Alu* repeat family than to the more recently evolved repeat family (Deininger and Slagel, 1988) and also contributes to the high G–C content of the intron. Intron 3 contained several interesting repeats, a direct repeat of CTTCTG near the beginning of the intron (nts 1111–1124, Fig. 2) and several direct repeats of TTGTTTT near the 3′ end (nts 1256–1288, Fig. 2) which could potentially be used to develop PCR markers to detect RFLPs.

The COX5B Gene May Be a Housekeeping Gene

Proteins that are essential components of intermediary metabolism and are found in most cell types are said to be encoded by "housekeeping" genes. Bird (1986) noted that housekeeping genes have several common structural features, most notably a high frequency of the rare dinucleotide CpG at their 5′ ends, resulting in a large number of *Hpa*II sites (CCGG). The promoter regions of housekeeping genes also lack certain consensus sequence elements such as TATA and CAAT boxes known to be associated with tissue-specific expression of genes (Maniatis *et al.*, 1987), but have multiple binding sites (G–C boxes) for the transcription factor SP1 (Dynan and Tjian, 1985). The *COX5B* gene also contains a large number of *Hpa*II sites near the 5′ end, i.e., in exon 2 and intron 1, and at least two consensus SP1 binding sites in the limited region of the first intron sequenced thus far (data not shown). These features suggest that the *COX5B* gene is also a housekeeping gene. Whether there are additional regulatory promoter or enhancer elements for respiratory proteins, such as NRF-1 binding sites found in the promoter of the cytochrome *c* gene (Evans and Scarpulla, 1989), awaits mapping and further analysis of the promoter region.

Chromosomal Localization of the Expressed COX5B Gene

We have identified eight different chromosomal sites of hybridization with probes containing both exon and intron sequences of *COX5B*. Since only the gene on human chromosome 2, region cen–q13, hybridized to an intron probe, we assigned the *COX5B* coding gene to this site. All the human fragments hybridized strongly under high stringency hybridization and washing conditions. This suggests that these sequences are pseudogenes of *COX5B* instead of partially related genes. Two other genes, *INHBB* (inhibin-β_B) and *MAL* (T cell differentiation protein), have been assigned to the same region in the same set of hybrids (Barton *et al.*, 1989; Alonso *et al.*, 1988). Genes for other nuclear-encoded subunits of cytochrome *c* oxidase have been mapped to different chromosomes: subunit IV (*COX4*) to chromosome 16, region q22–qter (Darras *et al.*, 1987; Lomax *et al.*, 1990) and subunit VIII (*COX8*) to chromosome 11, region q12–q13 (Rizzuto *et al.*, 1989).

ACKNOWLEDGMENTS

We thank Dr. A. S. Lee, USC, for the human genomic library, Dr. E. A. Schon, Columbia University, for the pCOX5b cDNA, E. Ziaja for screening the human genomic library, and D. BeMent and E. Gamliel for excellent technical assistance. This research was supported by NIH Grants GM30866 (M.I.L.) and GM26105 (U.F.), by NSF Grant DMB-90-05580 (M.I.L.), and by the Howard Hughes Medical Institute.

REFERENCES

1. Alonso, M. S., Barton, D. E., and Francke, U. (1988). Assignment of the T-cell differentiation gene MAL to human chromosome 2, region cen–q13. *Immunogenetics* **27:** 91–95.
2. Bachman, N. J., Lomax, M. I., and Grossman, L. I. (1987). Two bovine genes for cytochrome *c* oxidase subunit IV: A processed pseudogene and an expressed gene. *Gene* **55:** 219–229.
3. Barton, D. E., Yang-Feng, T. L., Mason, A. J., Seeburg, P. H., and Francke, U. (1989). Mapping of genes for inhibin

subunits α, β_A, and β_B on human and mouse chromosomes and studies of *jsd* mice. *Genomics* **5:** 91–99.

4. BASU, A., AND AVADHANI, N. G. (1990) Nucleotide sequence of cDNA for nuclear encoded subunit Vb of mouse cytochrome-*c* oxidase. *Biochim. Biophys. Acta* **1087:** 98–100.
5. BENTON, W. D., AND DAVIS, R. W. (1977). Screening lambda gt recombinant clones by hybridization to single plaques in situ. *Science* **196:** 180–182.
6. BIRD, A. P. (1986). CpG-rich islands and the function of DNA methylation. *Nature* **321:** 209–213.
7. BIRNBOIM, H. C., AND DOLY, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7:** 1513.
8. CAPALDI, R. A. (1988). Mitochondrial myopathies and respiratory chain proteins. *Trends Biochem. Sci.* **13:** 144–148.
9. CAPALDI, R. A., HALPHEN, D. G., ZHANG, Y.-Z., AND YANAMURA, W. (1988). Complexity and tissue specificity of the mitochondrial respiratory chain. *J. Bioenerg. Biomemb.* **20:** 291–311.
10. CHEN, E. Y., AND SEEBERG, P. H. (1985). Supercoil sequencing: A fast and simple method for sequencing plasmid DNA. *DNA* **4:** 165.
11. CHOMYN, A., AND ATTARDI, G. (1987). Mitochondrial gene products. *In* "Current Topics in Bioenergetics" (C. P. Lee, Ed.), Vol. 15, pp. 295–329, Academic Press, New York.
12. DARRAS, B. T., ZEVIANI, M., SCHON, E. A., AND FRANCKE, U. (1987). Sequences homologous to cytochrome *c* oxidase subunit IV are located on human chromosomes 14q21–qter and 16q22–q24. *Cytogenet. Cell Genet.* **46:** 603 (abstract).
13. DEININGER, P. L., AND SLAGEL, V. K. (1988). Recently amplified Alu family members share a common parental Alu sequence. *Mol. Cell. Biol.* **8:** 4566–4569.
14. DYNAN, W. S., AND TJIAN, R. (1985). Control of eukaryontic mRNA synthesis by sequence-specific DNA binding proteins. *Nature* **316:** 774–778.
15. EVANS, M. J., AND SCARPULLA, R. C. (1989). Interaction of nuclear factors with multiple sites in the somatic cytochrome *c* promoter. *J. Biol. Chem.* **264:** 14361–14368.
16. FABRIZI, G. M., RIZZUTO, R., NAKASE, H., MITA, S., KADENBACH, B., AND SCHON, E. A. (1989). Sequence of a cDNA specifying subunit VIIa of human cytochrome *c* oxidase. *Nucleic Acids Res.* **17:** 6409.
17. FEINBERG, A., AND VOGELSTEIN, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132:** 6–13.
18. GILBERT, W. (1985). Genes in pieces revisited. *Science* **228:** 823–824.
19. GOTO, Y., AMURO, N., AND OKAZAKI, T. (1989). Nucleotide sequence of cDNA for rat liver and brain cytochrome c oxidase subunit VIa (Vb). *Nucleic Acids Res.* **17:** 6388.
20. HATEFI, Y. (1985). The mitochondrial electron transport and oxidative phosphorylation system. *Annu. Rev. Biochem.* **54:** 1015–1069.
21. HSIEH, C.-L., STURM, R., HERR, W., AND FRANCKE, U. (1990). The gene for the ubiquitous octamer-binding protein Oct-1 is on human chromosome 1, region cen–q12, and near Ly-22 and Ltw-4 on mouse chromosome 1. *Genomics* **6:** 666–672.
22. KADENBACH, B., KUHN-NENTWIG, L., AND BUGE, U. (1987). *In* "Current Topics in Bioenergetics" (C. P. Lee, Ed.), Vol. 15, pp. 114–162, Academic Press, New York.
23. KOZAK, M. (1987). An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15:** 8125–8148.
24. LAWN, R. M., FRITSCH, E. F., PARKER, R. C., BLAKE, G., AND MANIATIS, T. (1978). The isolation and characterization of a linked δ- and β-globin gene from a cloned library of human DNA. *Cell* **15:** 1157–1174.
25. LOMAX, M. I., AND GROSSMAN, L. I. (1989). Tissue-specific genes for respiratory proteins. *Trends Biochem. Sci.* **14:** 501–503.
26. LOMAX, M. I., WELCH, M. D., DARRAS, B. T., FRANCKE, U., AND GROSSMAN, L. I. (1990). Novel use of a chimpanzee pseudogene from chromosomal mapping of human COX IV. *Gene* **86.**
27. MANIATIS, T., FRITSCH, E. F., AND SAMBROOK, J. (1982) "Molecular Cloning, A Laboratory Manual," Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
28. MANIATIS, T., GOODBOURN, S., AND FISCHER, J. A. (1987). Regulation of inducible and tissue-specific gene expression. *Science* **236:** 1237–1245.
29. PADGETT, R. A., GRABOWSKI, P. J., KONARSKA, M. M., SEILER, S., AND SHARP, P. A. (1986). Splicing of messenger RNA precursors. *Annu. Rev. Biochem.* **55:** 1119–1150.
30. RIZZUTO, R., NAKASE, H., DARRAS, B., FRANCKE, U., FABRIZI, G. M., MENGEL, T., WALSH, F., KADENBACH, B., AND SCHON, E. A. (1989). Gene specifying subunit VIII of human cytochrome *c* oxidase is localized to chromosome 11 and is expressed in both muscle and non-muscle tissues. *J. Biol. Chem.* **264:** 10595–10600.
31. SANGER, F., NICKLEN, S., AND COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74:** 5463–5467.
32. SCHMID, C. W., AND SHEN, C.-K. J. (1985). *In* "Molecular Evolutionary Genetics" (R. J. McIntyre, Ed.), pp. 323–356, Plenum, New York.
33. SHARP, P. A. (1981). Speculations on RNA splicing. *Cell* **23:** 643–646.
34. SOUTHERN, E. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98:** 503–517.
35. SUZUKI, H., HOSOKAWA, Y., NISHIKIMI, M., AND OZAWA, T. (1989). Structural organization of the human mitochondrial cytochrome c_1 gene. *J. Biol. Chem.* **264:** 1368–1374.
36. WEINER, A. M., DEININGER, P. L., AND EFSTRATIADIS, A. (1986). *Annu. Rev. Biochem.* **55:** 631–661.
37. YAMADA, M., AMURO, N., GOTO, Y., AND OKAZAKI, T. (1990). Structure of the rat gene for cytochrome *c* oxidase subunit IV. *J. Biol. Chem.* **265:** 7687–7692.
38. YANAMURA, W., ZHANG, Y.-Z., TAKIMIYA, S., AND CAPALDI, R. A. (1988). Tissue specific differences between heart and liver cytochrome c oxidase. *Biochemistry* **27:** 4909–4914.
39. YANG-FENG, T. L., DEGENNARO, L. J., AND FRANCKE, U. (1986). *Proc. Natl. Acad. Sci. USA* **83:** 8679–8683.
40. ZEVIANI, M., SAKODA, S., SHERBANY, A. A., NAKASE, H., RIZZUTO, R., SAMITT, C. E., DIMAURO, S., AND SCHON, E. A. (1988). Sequence of cDNAs encoding subunit Vb of human and bovine cytochrome c oxidase. *Gene* **65:** 1–11.
41. ZHANG, Y.-Z., AND CAPALDI, R. A. (1989). Subunit composition of the transmembrane parts of beef heart cytochrome c oxidase. *Ann. N.Y. Acad. Sci.* **555:** 3–7.