Mammalian Genome 10, 506–512 (1999).

Mammalian Genome
Incorporating Mouse Genome

The 5′ region of the COX4 gene contains a novel overlapping gene, NOC4

Nancy J. Bachman,[1,*] Wei Wu,[2] Timothy R. Schmidt,[2] Lawrence I. Grossman,[2] Margaret I. Lomax[1]

[1]Department of Anatomy and Cell Biology, The University of Michigan, 5724 Medical Science Building II, Box 0616, Ann Arbor, Michigan 48109-0616, USA
[2]Center for Molecular Medicine and Genetics, Wayne State University School of Medicine, Detroit, Michigan 48201, USA

Received: 17 August 1998 / Accepted: 15 December 1998

Abstract. We identified a novel human gene, NOC4 (Neighbor Of COX4), located 5′ to COX4, the gene for cytochrome *c* oxidase subunit IV, on Chr16q32-ter. Transcripts from this gene were identified among human expressed sequence tags. A full-length, 1.06-kb human retinal NOC4 cDNA encoded a 24-kDa, 210-amino acid hypothetical protein of unknown function. Northern hybridization analysis of human RNAs from various tissues detected NOC4 transcripts of 2.2 and 1.4 kb in all tissues examined, suggesting that NOC4 expression is ubiquitous. Transcription of both the COX4 and NOC4 genes initiates within a 250-bp intergenic promoter and occurs in opposite directions. The bidirectional promoter is G + C-rich, lacks TATA and CCAAT elements, and contains multiple potential binding sites for Sp1 and NRF-2/GABP. Two of the NRF-2/GABP sites are located within 14-bp direct repeats, a conserved feature of mammalian COX4 promoters. The NOC4 and COX4 genes are also linked in the rat, mouse, and bovine genomes. A NOC4-GFP fusion protein is located in both the nucleus and the cytoplasm, including the mitochondria.

Introduction

Cytochrome *c* oxidase (COX; EC 1.9.3.1) is a multisubunit complex that catalyzes the terminal redox reaction in the mitochondrial respiratory chain (Hatefi 1985). The 13-polypeptide subunits of the mammalian enzyme are encoded by both nuclear and mitochondrial genes. The three largest subunits (I–III) are encoded in mitochondrial DNA (mtDNA), whereas the ten smaller subunits are encoded in the nucleus. The mitochondrial subunits carry out the electron transfer and proton pumping activities of the complex (Capaldi 1990; Tsukihara et al. 1995, 1996). The exact function of the nuclear subunits is not known, but they are thought to be involved in the regulation and/or assembly of the enzyme complex (Kadenbach et al. 1987; Lomax and Grossman 1989; Poyton and McEwen 1996; Napiwotzki et al. 1997; Arnold and Kadenbach 1997; Grossman and Lomax 1997). Recent X-ray crystallographic analysis of bovine heart COX confirmed the presence of all 13 subunits in the enzyme complex and demonstrated that many nuclear-encoded subunits are intimately associated with the mitochondrial subunits (Tsukihara et al. 1995, 1996).

During analysis of the promoter region of the human COX4 gene, we identified a novel gene, NOC4 (Neighbor Of COX4), that encodes a 24-kDa protein of unknown function. The NOC4 gene is positioned in a head-to-head orientation with the COX4 gene and is thus transcribed from the non-coding strand of the COX4 genomic region. The transcription start sites of the two genes are within 250 bp of each other, suggesting that the intergenic region comprises a bidirectional promoter.

* *Present address:* Department of Biology, Hamilton College, Clinton, NY 13323, USA.

Correspondence to: M.I. Lomax

The nucleotide sequence data reported in this paper have been submitted to GenBank and assigned the accession numbers AF005889 (human COX4 gene), AF005888 (human NOC4 cDNA), and AF052621 (mouse NOC4 cDNA).

Materials and methods

Oligonucleotide primers. Oligonucleotide primers were synthesized in the University of Michigan DNA Synthesis Facility or the Macromolecular Core Facility of Wayne State University. Primers for amplification of intron 4 of COX4 (GenBank AF005889) were:
hCOX4-G 5′ CTCGTTATCATGTGGCAGAAGCAC 3′(nt 7244 to 7267)
PschIVL 5′ CAGCATCTCTCACTTCTTCCACTC 3′ (nt 8266 to 8283)
hCOX4-int4L 5′ ATCATGTGGCAGAAGCACTATGGT 3′ (nt 7250 to 7233) hCOX4-int4R 5′ GGGAGGGGGCCGTACACT 3′(nt 8142 to 8159)

DNA probes and PCR. The human COX4 cDNA insert was isolated from plasmid pCOX4.111 (Zeviani et al. 1987). Full-length NOC4 cDNAs from either adult human retina (I.M.A.G.E. Clone ID 219691; GenBank H80016) or mouse kidney (I.M.A.G.E. Clone ID 656449, GenBank AA244899) were from the I.M.A.G.E. Consortium (LLNL) (Lennon et al. 1996). DNA fragments used as hybridization probes were isolated by electrophoresis and labeled with [α-^{32}P]dCTP (>3000 Ci/mmole) by the random primer labeling method (Feinberg and Vogelstein 1983). We generated a probe to intron 4 of the human COX4 gene by two rounds of PCR amplification of human genomic DNA (500 ng; cell line 293), using nested primers. PCR was carried out with primers (3.5 pmoles each) corresponding to the ends of exon 4 (hCOX4-G) and exon 5 (PschIVL) in a 100-μl reaction containing 10 mM Tris-HCl (pH 8.3), 10 mM KCl, 4 mM $MgCl_2$, 0.2 mM each dNTP, and 10 units of AmpliTaq DNA polymerase, Stoffel fragment. Reactions were incubated at 94°C, 40 s; 63°C, 30 s, and 72°C, 5 min, for 5 cycles, then were annealed at 56°C for an additional 35 cycles. The gel-purified 1-kb PCR product was subjected to an additional round of PCR under similar conditions with nested internal primers that anchored to positions in exon/intron junctions (hCOX4-int4L and hCOX4-int4R).

DNA sequencing. The DNA sequence of exons 3, 4, and 5, and introns 3 and 4 of the COX4 gene was determined manually (CircumVent Thermal Cycle sequencing kit; New England Biolabs, Beverly, MA). The DNA sequence of the remaining regions of the COX4 gene and the NOC4 cDNAs was determined on duplex plasmid DNA by the University of Michigan DNA Sequencing Facility on an ABI model 373A automated DNA sequencer. DNA sequences were assembled with the ASSEMGEL program of PC/Gene (Intelligenetics, Campbell, Calif.). Database searches were carried out with the FASTA program of Genetics Computer Group (GCG) software (Madison, Wis.) on a VAX III computer under VMS or with the BLAST search algorithm at NCBI.

Genomic library screening. A human genomic library (Clontech, Palo Alto, Calif.) was screened under stringent hybridization and wash condi-

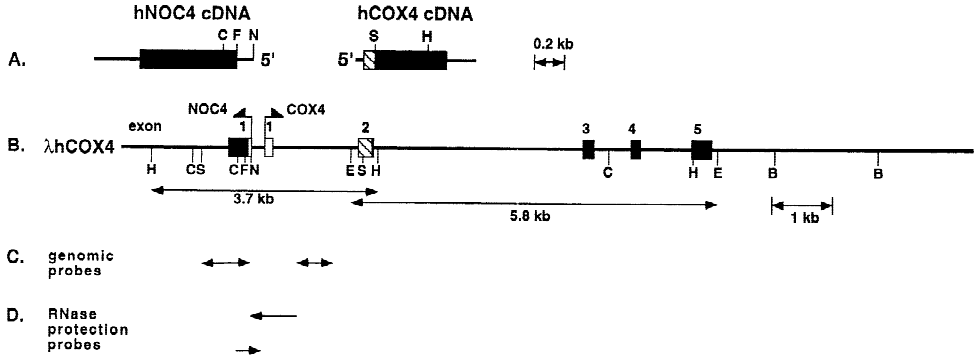

Fig. 1. The human COX4/NOC4 genomic region encoding the functional COX4 gene and the 5′ end of the linked gene NOC4. **A.** Restriction maps of the human COX4 and NOC4 cDNAs. Boxes represent coding regions; lines represent 5′ and 3′ UTRs. Hatched box denotes the presequence of COX IV. Filled boxes represent coding regions for the NOC4 hypothetical protein or the mature COX IV protein. Restriction enzyme sites are abbreviated as follows: *Bam*HI (B), *Eco*RI (E), *Hin*cII (C), *Hin*dIII (H), *Not*I (N), *Sac*I (S), *Sfi*I (F). **B.** Restriction map of the human COX4 and NOC4 genes. The genomic library contained partial *Sau*3A-digested human placental DNA cloned into the *Bam*HI site of λEMBL3 Sp6/T7. Exons are shown as box 1 (NOC4) or boxes 1–5 (COX4). Open boxes denote regions encoding the 5′ untranslated region (UTR); hatched boxes the presequence of COX IV; closed boxes the NOC4 polypeptide or the mature COX IV protein. There may be additional cleavage sites for *Hin*cII and *Sac*I 3′ to the COX4 gene that have not been mapped. Regions of λhCOX4 subcloned into plasmid vectors (3.7-kb *Hin*dIII and 5.8-kb *Eco*RI fragments) are indicated below the restriction map. **C.** Probes used in genomic Southern blots are indicated. **D.** Regions transcribed into antisense RNA probes for 5′ end mapping by RNase protections. The direction of the arrow indicates the 5′ to 3′ direction of the probe.

tions (Sambrook et al. 1989) with two probes, the entire 694-bp human COX4 cDNA and the intron 4 probe. Phage DNA was prepared (Ausubel et al. 1990) from two plaque-purified, positive clones. COX4 genomic regions were localized by restriction mapping and Southern hybridization with COX4 cDNA probes, subcloned into pGEM3Zf(+) (Promega, Madison, Wis.), and sequenced on both strands.

Northern analysis of NOC4 and COX4 mRNAs. A multi-tissue Northern blot (Clontech) containing various human tissue poly(A)$^+$ RNAs (2 μg) was probed with the 694-bp human liver COX4 cDNA, pCOX4.111, and the 1.06-kb *Eco*RI/*Hin*dIII insert from human retinal NOC4 cDNA. Labeled cDNA probes were separated from unincorporated nucleotides on a Bio-Spin 6 column (BioRad, Hercules, Calif.). The membrane was prehybridized for 30 min at 68°C in 5 ml ExpressHyb solution (Clontech). Equal molar amounts of COX4 (4 ng at 9.8×10^5 cpm/ng) and NOC4 (7 ng at 1.9×10^6 cpm/ng) cDNA probes were denatured with 100 μg sheared denatured salmon sperm DNA by heating to 100°C for 5 min and were added to 5 ml fresh ExpressHyb solution. The membrane was hybridized in the ExpressHyb solution containing probes for 90 min at 68°C, then washed according to the supplier's instructions to a final stringency of 0.1 × SSC, 0.1% SDS at 50°C. The blot was exposed for 5 h at −80°C with intensifying screens. Relative levels of COX4 and NOC4 transcripts were determined by volume image analysis of a 40-min exposure with a BioRad Model GS-250 Phosphorimaging System with Molecular Analyst v1.4 software.

RNase protection assays. A *Hin*dIII/*Pst*I genomic fragment spanning the 5′ end of COX4 and terminating at the *Pst*I site in intron 1 (nt 1481; Fig. 2) was the template for synthesis of ^{32}P-labeled antisense RNA. A *Hin*cII/*Hin*dIII fragment (nt 625–2937; GenBank AF005889) spanning the 5′ end of NOC4 was the template for synthesis of an antisense NOC4 RNA. Transcription reactions (800 ng of template) were incubated for 1 h at 37°C in a 20-μl reaction consisting of 40 mM Tris-HCl (pH 7.9), 6 mM $MgCl_2$, 2 mM spermidine-HCl, 10 mM DTT, 1 mM each of ATP, GTP, and UTP, 2.5 μM CTP, 50 μCi [α-^{32}P]CTP (~800 Ci/mmole), 20 units of RNasin (Promega), and 25 units of T7 RNA polymerase (Life Technologies, Gaithersburg, MD), followed by 15 min at 37°C with 1 unit of RNase-free DNase (Promega) to remove DNA. RNA probes were heat denatured and purified by gel electrophoresis.

RNase protection assays were carried out with an RPA II kit (Ambion, Austin, Tex.). Total HeLa RNA (30 μg) was ethanol precipitated with approximately 1×10^6 cpm of labeled antisense RNA, resuspended in 20 μl of hybridization buffer (80% formamide/100 mM sodium citrate, pH 6.4/300 mM sodium acetate, pH 6.4/1 mM EDTA), denatured 4 min at 90°C, then annealed overnight at 45°C. Control reactions contained 10 μg of yeast RNA in place of the HeLa RNA. RNA duplexes were digested with a mixture of RNases A and T1 (1:1000 dilution), incubated 30 min at 37°C, then precipitated at −20°C for at least 2 h. Pellets were resuspended in 95% formamide/0.025% each xylene cyanol and bromophenol blue/0.5 M EDTA/0.0025% SDS, heated 4 min at 90°C, and subjected to electrophoresis on a 6% acrylamide–7 M urea sequencing gel run 2 h at 60W.

Construction and localization of a NOC4-green fluorescent protein fusion. A T3 primer and a *Bam*HI adaptor primer based on the 3′ end of the human retinal NOC4 cDNA (5′GGTGGATCCTTTTGCACAAGTGTAGGACAGC) were used to amplify the 5′ flanking and protein coding regions by PCR. Reactions were carried out as recommended by the supplier (Gibco-BRL, Gaithersburg, Md.) with 20 ng template DNA, 10 μM each primer, 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 1.5 mM $MgCl_2$ for 30 cycles at 94°C 45 s, 55°C 45 s, 72°C 1 min 30 s. The PCR product was gel purified (Millipore Ultrafree MC filters), subcloned into pGEM-T (Promega) and sequenced. The 600-bp *Hin*dIII/*Bam*HI fragment containing the 5′ flanking and coding regions of NOC4 was then subcloned into the corresponding sites in the green fluorescent protein (GFP) vector derived from pcDNA3 (Invitrogen, Carlsbad, CA) and optimized for human codon usage (kindly provided by Brian Seed).

HeLa cells were seeded on glass cover slips and cultured at 37°C overnight prior to transfection. Transient transfection was performed with FuGENE 6 Transfection Reagent (Boehringer Mannheim, Indianapolis, IN) following the manufacturer's protocol. For each 35-mm petri dish, 1 μg of NOC4-GFP DNA in 3 μl FuGene 6 reagent was used. Cells were cultured for 48 h, followed by incubation in medium containing 1 mM MitoTracker Red CM-H2XRos (Molecular Probes, Eugene, OR) for 1 h. Cells on cover slips were washed subsequently in PBS, fixed with 2% paraformaldehyde in PBS for 5 min, washed again, and permeabilized in 0.1 M Tris-HCl (pH 7.7), 0.15 M NaCl, 1% BSA, 0.1% TX-100 (KB) at RT for 30 min. Nuclei were stained with 0.5 mg/ml DAPI for 10 min, washed in KB solution without TX-100, and mounted in 50% glycerol. Fluorescence was observed with a Zeiss laser scanning confocal microscope 310 equipped with a 63× oil lens.

Results

Isolation and characterization of the human COX4 gene. We isolated two clones containing the human COX4 gene by screening a genomic library with an intron 4 probe. Restriction mapping and Southern analysis with the human COX4 cDNA demonstrated that clone λhCOX4 contained two restriction fragments characteristic of the human COX4 gene: a 5.8-kb *Eco*RI fragment previously mapped to Chr 16q22-qter (Lomax et al. 1990) and a 3.7-kb *Hin*dIII fragment containing the 5′ end of the COX4 gene (Fig. 1). The 5.8-kb *Eco*RI fragment contained the remainder of the COX4 gene: exons 2–5 plus the 3′ flanking region (Fig. 1). The COX4

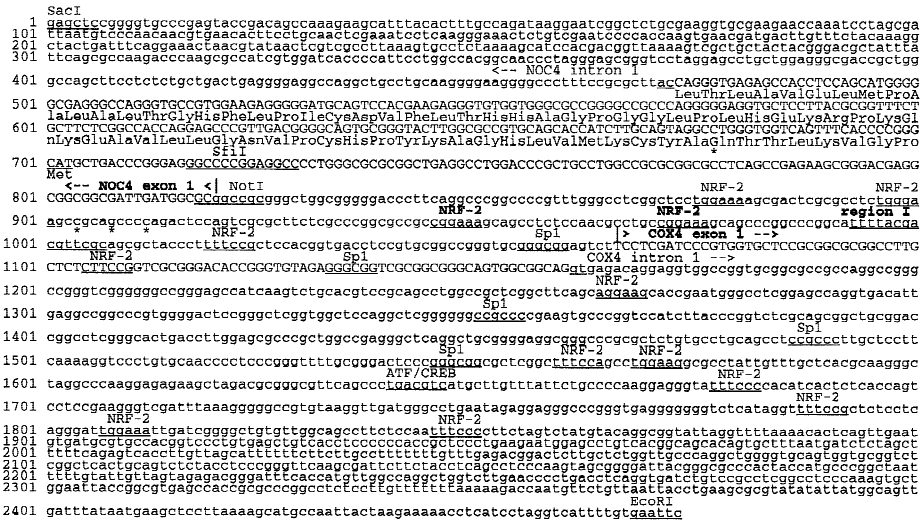

Fig. 2. Nucleotide sequence of 2.5 kb of genomic DNA comprising the bidirectional human COX4/NOC4 promoter. The sequence is numbered beginning with the *Sac*I site in intron 1 of the NOC4 gene, which is approximately 1 kb upstream of COX4 exon 1 (Fig. 1), and extends to the *Eco*RI site (nt 2473) in intron 1 (Fig. 1). The sequence of the NOC4 and COX4 5′ flanking regions plus COX4 exons 1–3 and part of exon 5 (GenBank AF005889) was determined from subclones of λhCOX4; the sequence of exons 3–5 was obtained from three independent subclones of two different PCR products of human genomic DNA. COX4 and NOC4 exons are in upper case, introns and flanking regions in lower case. The sequences of the amino terminus of the NOC4 protein (written in reverse orientation, from COOH → NH_2) is presented in the three-letter amino acid code below the nucleotide sequence. Restriction enzyme cleavage sites and consensus elements for binding by the transcription factor Sp1 (Kadonaga et al. 1986), NRF-2/GABP (Karim et al. 1990; Virbasius et al. 1993), and ATF/CREB (Montminy et al. 1986) are underlined. NRF-2 sites present in tandem 14-bp direct repeats in the promoter region and a 15-bp region (region I) conserved in human and bovine promoters are labeled in bold type. Leftward (⟨|) and rightward (|⟩) arrows indicate the 5′ ends of the longest available cDNAs for NOC4 and COX4, respectively. The sequence of the COX4 gene and cDNA were identical except for a G/A polymorphism at position 1150 in the 5′ UTR and a discrepancy of 11-bp at the 5′ end of the COX4 cDNA. Two adult brain ESTs (GenBank H49136 and H52746) showed nearly complete identity with the COX4 genomic region beginning at position 1066. Transcription start sites mapped by RNase protections are marked with asterisks (*).

gene thus spans approximately 7.4 kb. The location and size of the exons are similar to those of other mammalian COX4 genes from cow (Bachman 1995), rat (Yamada et al. 1990; Virbasius and Scarpulla 1990) and mouse (Carter and Avadhani 1991).

Identification of a novel overlapping gene, NOC4. FASTA searches of the GenBank/EMBL database with the presumptive COX4 promoter sequence showed high sequence identity to the 5′ flanking regions of all COX4 genes; however, this search also identified several ESTs (expressed sequence tags; Adams et al. 1991, 1995) derived from a different gene, named NOC4. There was sequence identity between the non-coding strand of the human COX4 genomic region and one EST from a human retina library, which began approximately 250 bp upstream of the human COX4 gene at nt 819 (Fig. 2) and extended for 350 bp to a 5′ splice consensus (Ohshima and Gotoh 1987) at nt 472, consistent with the presence of an intron in an expressed gene. On the basis of Southern hybridization under moderately stringent conditions, NOC4 is a single-copy gene in humans (N.J. Bachman, M.I. Lomax, data not shown).

NOC4 encodes a novel 24-kDa protein. The human retinal NOC4 cDNA was 1060 bp long with a 633-bp ORF (Fig. 3) encoding a 210-amino acid protein (molecular mass 24 kDa; pI = 5.91). The predicted protein sequence contains a potential membrane-spanning region (amino acids 52–78); comparison with a database of membrane-spanning proteins (Hofmann and Stoffel 1993) suggests that the N-terminal region (1-51) lies outside this hypothetical domain. A full-length 1025-bp NOC4 cDNA from mouse kidney (not shown) encodes a 207-amino acid protein (Fig. 3). The size of the predicted protein (23 kDa; pI = 5.60) differs from the human owing to a 9-bp deletion in the 5′ coding region of the mouse sequence. Database searching yielded no significant homology of the derived NOC4 proteins to known proteins.

NOC4 is widely expressed. Northern hybridization analysis with the human retinal NOC4 cDNA detected NOC4 transcripts in poly(A)$^+$ RNAs from all adult human tissues tested (Fig. 4). Two transcripts, 2.2 kb and 1.4 kb, were seen in all tissues (Fig. 4A). The size of the smaller transcript is consistent with the size of the NOC4 retinal cDNA (1.06 kb) if we assume a 300-residue poly(A) tail. However, the structural basis for the two different transcripts is unknown. Database searches with the human NOC4 cDNA sequence identified a large family of human ESTs derived from a variety of tissues and cell lines. Thus, NOC4, like COX4, appears to be widely expressed.

COX4 and NOC4 expression varies concordantly in different tissues. To assess the relative abundance of COX4 transcripts compared with those for NOC4, we probed a Northern blot of human tissues with a mixture of COX4 and NOC4 cDNAs. COX4 displayed 4- to 12-fold higher mRNA levels in each tissue than the levels of both NOC4 transcripts combined when the probes were corrected to equal specific activities (Fig. 4B). Although the level of COX4 mRNA was always higher than that of NOC4, the relative levels varied in a concordant manner. Levels of COX4 and of both NOC4 transcripts were highest in pancreas, showed moderate levels in heart, placenta, and skeletal muscle, and were lowest in liver (Fig. 4B).

The COX4/NOC4 intergenic region is a bidirectional promoter. The 5′ ends of the human COX4 and NOC4 genes were mapped by RNase protection (Figs. 2 and 5). Three protected bands of 157, 153 (major band), and 149 nt were identified for COX4, and a single protected band of 151 nt was identified for NOC4 (Fig. 5). These results placed the major transcription start site for COX4 at approximately 153 nt upstream of the exon 1/intron 1 junction, and for NOC4 at about 74 nt upstream of the ATG start codon (Fig. 2, *). On the basis of these results, transcripts of

```
human  gcgccatcaatcgccgccgcctcgtcccgcttctcggctgaggcgccgcgcggccaggcagcgggtccaggcctcagccgcgcgcccaggggcctccggg   100

human  gccctcccgggtcagcATGCCCGGGGTGAAACTGACCACCCAGGCCTACTGCAAGATGGTGCTGCACGGCGCCAAGTACCCGCACTGCGCCGTCAACGGG   200
human                MetProGlyValLysLeuThrThrGlnAlaTyrCysLysMetValLeuHisGlyAlaLysTyrProHisCysAlaValAsnGly    28
mouse                ************************************************************************************   28

human  CTCCTGGTGGCCGAGAAGCAGAAGCCGCGTAAGGAGCACCTCCCCCTGGGCGGCCCCGGCGCCCACCACACCCTCTTCGTGGACTGCATCCCCCTCTTCC   300
human  LeuLeuValAlaGluLysGlnLysProArgLysGluHisLeuProLeuGlyGlyProGlyAlaHisHisThrLeuPheValAspCysIleProLeuPheH   61
mouse  ***************Arg***Arg***************Pro***GlyAla***---------Ser*********************************   58

human  ACGGCACCCTGGCCCTCGCCCCCATGCTGGAGGTGGCTCTCACCCTGATTGATTCATGGTGCAAAGATCATAGCTACGTGATTGCTGGTTATTATCAAGC   400
human  isGlyThrLeuAlaLeuAlaProMetLeuGluValAlaLeuThrLeuIleAspSerTrpCysLysAspHisSerTyrValIleAlaGlyTyrTyrGlnAl   95
mouse  *****************Thr*************************************************Asn*****************************   92

human  TAATGAGCGAGTAAAGGATGCCAGTCCAAACCAGGTTGCAGAGAAGGTGGCCTCCAGAATCGCCGAGGGCTTCAGCGACACTGCGCTCATCATGGTAGAC   500
human  aAsnGluArgValLysAspAlaSerProAsnGlnValAlaGluLysValAlaSerArgIleAlaGluGlyPheSerAspThrAlaLeuIleMetValAsp   128
mouse  ************************************************************************Gly***Ala*****************   125

human  AACACCAAGTTTACGATGGACTGCGTAGCGCCTACGATCCACGTGTACGAGCACCATGAGAACAGATGGCGGTGCAGAGACCCACACCATGACTACTGTG   600
human  AsnThrLysPheThrMetAspCysValAlaProThrIleHisValTyrGluHisHisGluAsnArgTrpArgCysArgAspProHisHisAspTyrCysG   161
mouse  ***Ala*****************Ala************************Gln*********************************************   158

human  AAGACTGGCCAGAGGCACAGAGGATCTCAGCCTCGCTCCTGGACAGCCGGTCCTACGAGACGCTCGTGGATTTCGATAACCACCTGGATGACATTCGGAA   700
human  luAspTrpProGluAlaGlnArgIleSerAlaSerLeuLeuAspSerArgSerTyrGluThrLeuValAspPheAspAsnHisLeuAspAspIleArgAs   195
mouse  ************************************************************************************************Se   192

human  TGACTGGACAAACCCAGAGATCAATAAAGCTGTCCTACACTTGTGCTAGgcaggcaccgctgtgactgggctccgggccttttcccactacgttgaagaag   800
human  nAspTrpThrAsnProGluIleAsnLysAlaValLeuHisLeuCys---                                                     210
mouse  r*****************************************                                                          207

human  aaaacctatttttaaatgtaaataaaatatctggtagcctgtgtggaaagctgaccgttttaagaagtggcatgtgccttgaaaggggcagaatgttca   900

human  gtcggtcgtgtttttaacacagagtctctagaagaggtgcagacatcccgtctgactgtccctgtggactctctcagttgtatgttgctataatcctcca  1000

human  aatcaaagctcttttctgcttgtgcaagattgttcctattaaacagttttaactaaccttt                                          1060
```

Fig. 3. Sequence of the NOC4 cDNA and theoretical protein. The DNA sequence of the 1060-bp adult human retina cDNA (I.M.A.G.E cloneID 219691) is presented. The derived amino acid sequence for the human retina is shown below the nucleotide sequence in the three-letter code. Differences in the derived amino acid sequence for mouse kidney NOC4 (I.M.A.G.E. cloneID 656449; GenBank AF052621) are shown below the human protein sequence, with conserved residues indicated by an asterisk (*) and a short deletion indicated by dashes (–). 5′ and 3′ UTR sequences are shown in lower case. Poly(A) tails have been removed from the end of the cDNA sequence.

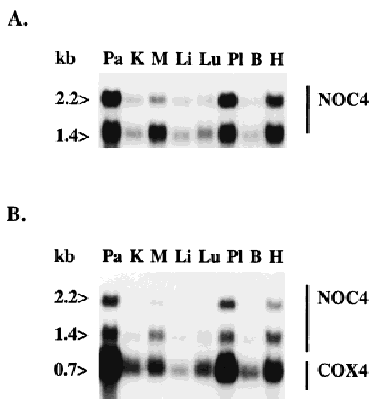

Fig. 4. Northern blot analysis of COX4 and NOC4 transcripts in human tissues. Tissue abbreviations are: Pa, pancreas; K, kidney; M, skeletal muscle; Li, liver; Lu, lung; Pl, placenta; B, brain; H, heart. Sizes of NOC4 transcripts were interpolated from RNA size markers (Clontech). **A.** Hybridization to the human retinal NOC4 cDNA. The 2.2- and 1.4-kb transcripts are indicated. **B.** Hybridization to both human retinal NOC4 and liver COX4 cDNAs. The two largest transcripts correspond to the 2.2- and 1.4-kb transcripts of NOC4; the 0.7-kb transcript (Zeviani et al. 1987) is from COX4.

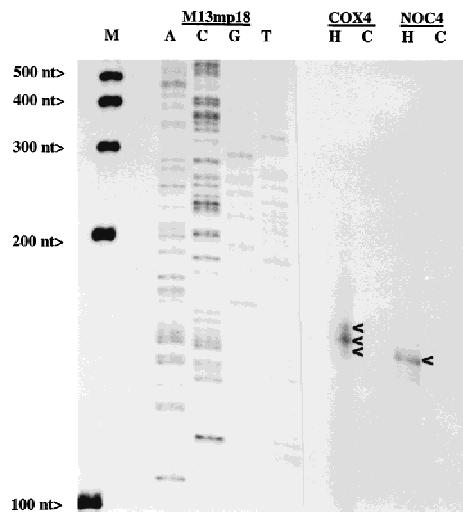

Fig. 5. Mapping the 5′ ends of COX4 and NOC4 genes by RNase protection analysis. ^{32}P-labeled antisense RNA probes (8×10^5 cpm of COX4 or 1×10^6 cpm of NOC4) overlapping the 5′ ends of each gene (see Fig. 1D) were prepared as described in Materials and methods and annealed to 30 μg HeLa RNA (H) or 10 μg control yeast RNA (C), digested with RNases A + T1 and fractionated on a 6% acrylamide–7 M urea sequencing gel. The antisense transcript for COX4 was 706 nt and for NOC4 was 425 nt (Fig. 1). Lanes 1 and 2, COX4 probe; lanes 3 and 4, NOC4 probe. Size markers: A, C, G, T, dideoxy sequencing reactions of M13mp18; M, RNA Century Markers (Ambion).

the two genes do not overlap. Each gene is transcribed from a separate strand of DNA and in the opposite orientation. They likely share promoter elements owing to the proximity of their transcription start sites (within 240 bp).

The 5′ flanking region of the human COX4 gene, like the analogous region in the previously characterized mammalian COX4 genes, is G + C rich (72% G + C in the COX4/NOC4 intergenic region, nt 819–1066), lacks TATA and CCAAT consensus elements, and contains multiple putative binding sites for the *ets*-domain transcription factor NRF-2/GABP (Karim et al. 1990; Virbasius and Scarpulla 1991; Carter et al. 1992) and Sp1 (Kadonaga et al. 1986; Fig. 2), but not NRF-1 (Evans and Scarpulla 1991).

Despite strong intragenic similarity among coding regions of mammalian COX4 and NOC4 genes, the sequence of the ~250 bp region separating the genes has evolved significantly. Nevertheless, a few short sequence elements have been evolutionarily conserved (Fig. 6). All mammalian COX4 promoters contain direct repeats with binding sites for NRF-2/GABP (Fig. 6). Two sites for NRF-2 are located within two 14-bp direct repeats at −74 to −61 and −48 to −35 upstream of the start site for transcription in the human gene. Human and bovine (Bachman 1995) COX4 promoters contain longer direct repeats (14 and 13 bp, respectively) than do rat (Virbasius and Scarpulla 1991) and mouse (Carter et al. 1992) promoters (11 and 9 bp). Functional analysis of rat and mouse COX4 genes has shown that the NRF-2 sites in the direct repeats are essential for maximal expression. A single consensus ATF/CREB site (Montminy et al. 1986) was located in intron 1 of the human COX4 gene. Consensus ATF/CREB binding sites are

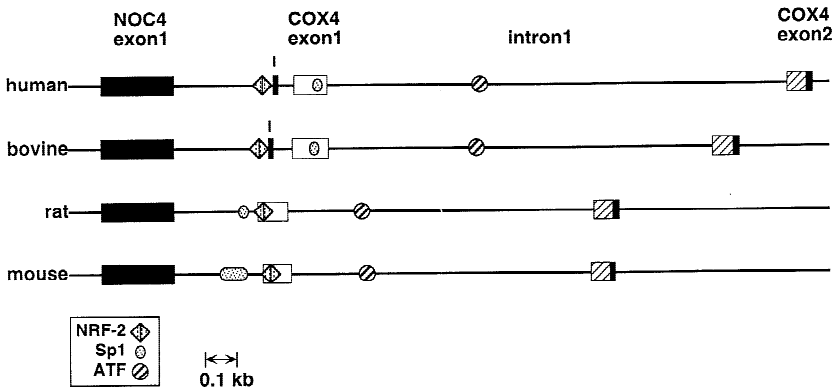

Fig. 6. Schematic diagram of mammalian COX4 promoters. The region surrounding the 5′ ends of the COX4 and NOC4 genes from human, bovine, rat, and mouse are displayed. The NOC4 and COX4 transcription units proceed to the left and right, respectively. Boxed regions indicate exons of each gene. Small triangles indicate direct repeats containing binding sites for NRF-2/GABP. Filled boxes (I) in bovine and human genes indicate a conserved 15-bp element. Binding sites for Sp1 (stippled circles) and ATF/CREB (hatched circles) are indicated.

also found in intron 1 of bovine (Bachman 1995) and rodent (Gopalakrishnan and Scarpulla 1994) COX4 genes (Fig. 6). In addition, a 15-bp element (TTTTACGACGTTCGC) identified from the alignment of the human and bovine COX4 promoters is precisely conserved just downstream of the direct repeats (Fig. 6; nt 993–1007 in the human sequence).

Subcellular localization of NOC4 protein. GFP fluorescence was observed in cells transfected with a NOC4-GFP fusion construct (Fig. 7, left panel, upper), whereas no signal is seen in the nontransfected cells visible in the same panel (lower). Both transfected and nontransfected cells showed mitochondrial and nuclear staining (Fig. 7, center panel). The NOC4-GFP signal (Fig. 7, right panel) is concentrated in the nucleus and in the cytosol; the latter includes mitochondria, as shown by the co-occurrence of blue and green (as sky blue) in the nucleus and the co-occurrence of red and green (as yellow and orange) in the mitochondria. In addition, the NOC4-GFP signal is present in the non-mitochondrial cytoplasm.

Discussion

COX4 in humans is adjacent to and probably shares a promoter with an expressed gene termed NOC4. These genes are also closely linked in the cow, mouse, and rat. The human and bovine genes show 93% DNA sequence identity in the 181-bp region corresponding to exon 1 of NOC4 and are also similar for an additional 100 nt upstream of the ORF. This 100-nt region is little conserved in a comparison of the rodent genes with either the human or bovine gene. NOC4 genes in all the species examined appear to be expressed. We identified several mouse ESTs with high sequence identity to human NOC4, including one from an uninduced rat PC-12 cell line (GenBank H31096; Lee et al. 1995).

The fact that NOC4 and COX4 are both highly expressed in pancreas, are expressed at moderate levels in heart, skeletal muscle, and human placenta, and are found at low levels in liver suggests that the regulation of these two genes is related. Previous studies of COX4 gene expression in rat (Virbasius and Scarpulla 1991) and mouse (Carter et al. 1992; Carter and Avadhani 1994) did not take into account the existence of a second functional gene in close proximity to COX4. Deletion analysis of the mouse COX4 promoter indicates that removal of an element between −600 and −142 increases COX4 expression twofold (Carter et al. 1992). This result is consistent with removal of a divergently transcribed gene that competes with COX4 for regulatory factors and/or transcription machinery. Deletion analysis of the rat COX4 promoter (Virbasius and Scarpulla 1991) showed that a small region (116 bp) of the COX4 gene containing tandem NRF-2 binding sites was sufficient to drive CAT expression from a basal promoter. It will be interesting to examine whether expression of the divergently transcribed NOC4 gene is similarly controlled by a small region or whether some of its controlling elements extend to within the COX4 gene itself.

Several other examples of bidirectional promoters are known in higher eukaryotes (Farnham et al. 1985; Lavia et al. 1987; Rizzo et al. 1990; Huang et al. 1990; Heikkila et al. 1993; Sun and Kitchingman 1994; Gaston and Fried 1994). Many of these examples are found in CpG islands adjacent to "housekeeping" genes and lack consensus TATA elements. A number of the bidirectional promoters in higher eukaryotes belong to coordinately regulated genes for related gene products, such as those for human Wilms tumor 1 and 2 (Huang et al. 1990), human α1 (IV) and α2 (IV) collagen (Heikkila et al. 1993), mouse Surf-3 (Rpl17a) and Surf-5 (Colombo et al. 1992), mouse and human Surf-1 and Surf-2 (Gaston and Fried 1994), human histidyl-tRNA synthetase and HO3 (O'Hanlon et al. 1995), and mitochondrial chaperonins 10 and 60 (Ryan et al. 1997).

Many genes that share bidirectional promoters encode proteins of related function. Consequently, we determined whether NOC4, like COX4, encodes a mitochondrial protein. We found that NOC4 occurs in the nucleus, cytosol, and mitochondria (Fig. 7). Examples of proteins encoded by one gene that target to more than one cellular compartment are known and have been termed sorting isozymes (Gillman et al. 1991; Martin and Hopper 1994). CCA1 in *Saccharomyces cerevisiae* is present in mitochondria, the nucleus, and the cytoplasm (Wolfe et al. 1994, 1996). Although involved in tRNA processing, the nuclear and cytoplasmic pools serve different cellular functions. Another example is the yeast *MOD5* gene, which is also involved in tRNA processing (Boguta et al. 1994; Zoladek et al. 1995).

The signals that mediate the targeting of sorting isozymes to various compartments have been elucidated in several cases. In the case of *MOD5,* two translation initiation sites are present (at aa 1 and 12); peptides starting at aa 1 target to mitochondria and cytosol, whereas those starting at aa 12 target to nuclei and cytosol. The region between 1 and 12, although having some features of mitochondrial targeting presequences, is not sufficient to direct a passenger protein to mitochondria without additional residues. Nuclear targeting of MOD5 requires a bipartite nuclear localization signal (NLS) located near the C-terminus (Tolerico et al. 1999). NOC4 shows several parallels to MOD5: it contains two potential translation initiation sites, at aa 1 and 14, and the region between 1 and 14 contains features of a mitochondrial targeting presequence.

Most nuclear-encoded, mitochondrially targeted proteins, including COX IV, contain amino-terminal targeting sequences that form amphiphilic α-helices (von Heijne 1986). These targeting sequences have basic amino acids on one face and hydrophobic residues on the other. NOC4 shares some key features of mitochondrial proteins, notably a predicted amphiphilic helix for amino acids 6–23 containing five basic residues (Lys and His) and no acidic residues. However, other properties of established mito-

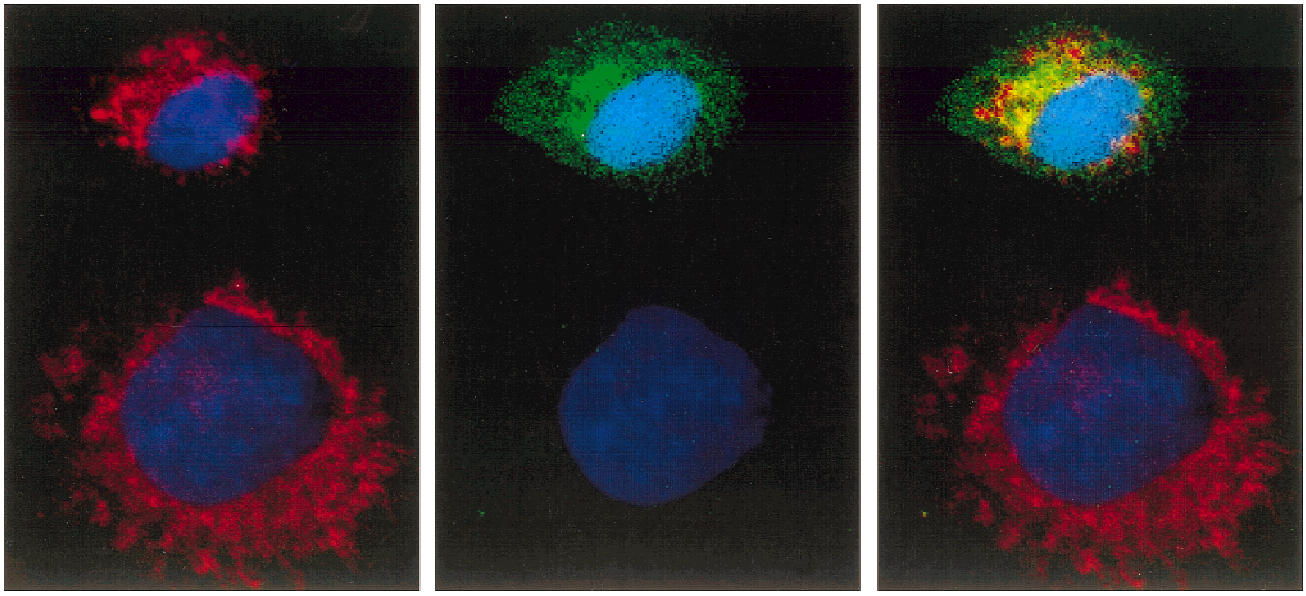

Fig. 7. Fluorescence microscopy of the NOC4-GFP fusion protein. HeLa cells transfected with NOC4-GFP in pcDNA3 were observed for GFP fluorescence and also stained to visualize nuclei and mitochondria. **Left panel:** *Red and blue channels:* Nuclei (DAPI; blue) and mitochondria (MitoTracker; red) are visible in a field showing a transfected (upper) and an untransfected (lower) cell. **Center panel:** *Green and blue channels:* The transfected cell nucleus is light blue and the cytoplasm green, showing NOC4-GFP is present in both the nucleus and the cytosol. **Right panel.** *Red, blue, green:* The yellow and orange present in the transformed cell cytoplasm indicate that NOC4-GFP is co-occurring with mitochondria.

chondrial import signals match weakly, mostly owing to low predicted values for hydrophobicity (0.11 vs. 4.5 or greater for most mitochondrial precursors), pI (9.75 vs. 12.78 for typical mitochondrially targeted proteins), and hydrophobic moment (0.27 vs. 5.0 or greater for most mitochondrial precursors). A similar case of weakly matching properties of mitochondrial import signals is seen for MOD5 (pI = 10.28; Boguta et al. 1994). Their results suggest that one mechanism utilized by a targeting isoform to distribute to several compartments is to have suboptimal localization signals for any one compartment.

Although the function of COX IV is not known, there is evidence that adenine nucleotides bind to COX IV and play a central role in regulation of COX activity (Arnold and Kadenbach 1997; Napiwotzki et al. 1997). We have here shown that a gene apparently co-regulated with COX4 is present in both the nucleus and the cytosol, including mitochondria. Such a protein could serve a regulatory role via nucleus-mitochondria cross-talk.

Acknowledgments. This work was supported by National Institutes of Health Grant GM48800 to M.I. Lomax. We thank several talented and dedicated undergraduate students, Camille Frasier, Calvin Lee, John Austin, and Nazema Siddiqui, for assistance with the library screening and restriction mapping of COX4 clones, and Dr. John Tomkiel for help with imaging. Support for computing facilities was provided in part by the General Clinical Research Center at the University of Michigan, funded by grant (MOIRR00042) from the National Center for Research Resources, National Institutes of Health, USPHS. The Confocal Imaging Core Facility at Wayne State University is supported in part by grants P30ES06639 from the National Institutes of Environmental Health Sciences and P30CA22453 from the National Cancer Institute.

References

Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH et al. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252, 1651–1656

Adams MD, Kerlavage AR, Fleischmann RD, Fuldner RA, Bult CJ et al. (1995) Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. Nature 377 (Suppl), 3–17

Arnold S, Kadenbach B (1997) Cell respiration is controlled by ATP, an allosteric inhibitor of cytochrome-*c* oxidase. Eur J Biochem 249, 350–354

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG et al. (1990) Escherichia Colt, Plasmids and Bacteriorphagy In *Current Protocols in Molecular Biology*, (New York: John Wiley and Sons) pp. 1.9.1–1.13.7

Bachman NJ (1995) Isolation and characterization of the functional gene encoding bovine cytochrome *c* oxidase subunit IV. Gene 162, 313–318

Boguta M, Hunter LA, Shen WC, Gillman EC, Martin NC et al. (1994) Subcellular locations of MOD5 proteins—mapping of sequences sufficient for targeting to mitochondria and demonstration that mitochondrial and nuclear isoforms commingle in the cytosol. Mol Cell Biol 14, 2298–2306

Capaldi RA (1990) Structure and function of cytochrome *c* oxidase. Annu Rev Biochem 59, 569–596

Carter RS, Avadhani NG (1991) Cloning and characterization of the mouse cytochrome *c* oxidase subunit IV gene. Arch Biochem Biophys 288, 97–106

Carter RS, Avadhani NG (1994) Cooperative binding of GA-binding protein transcription factors to duplicated transcription initiation region repeats of the cytochrome *c* oxidase subunit IV gene. J Biol Chem 269, 4381–4387

Carter RS, Bhat NK, Basu A, Avadhani NG (1992) The basal promoter elements of murine cytochrome *c* oxidase subunit IV gene consist of tandemly duplicated ets motifs that bind to GABP-related transcription factors. J Biol Chem 267, 23418–23426

Colombo P, Yon J, Garson K, Fried M (1992) Conservation of the organization of five tightly clustered genes over 600 million years of divergent evolution. Proc Natl Acad Sci USA 89, 6358–6362

Evans MJ, Scarpulla RC (1991) NRF-1: a *trans*-activator of nuclear-encoded respiratory genes in animal cells. Genes Dev 4, 1023–1034

Farnham PJ, Abrams JM, Schimke RT (1985) Opposite-strand RNAs from the 5′ flanking region of the mouse dihydrofolate reductase gene. Proc Natl Acad Sci USA 82, 3978–3982

Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132, 6–13

Gaston K, Fried M (1994) YY1 is involved in the regulation of the bidirectional promoter of the Surf-1 and Surf-2 genes. FEBS Lett 347, 289–294

Gillman EC, Slusher LB, Martin NC, Hopper AK (1991) MOD5 transla-

tion initiation sites determine N6-isopentenyladenosine modification of mitochondrial and cytoplasmic tRNA. Mol Cell Biol 11, 2382–2390

Gopalakrishnan L, Scarpulla RC (1994) Differential regulation of respiratory chain subunits by a CREB-dependent signal transduction pathway. J Biol Chem 269, 105–113

Grossman LI, Lomax MI (1997) Nuclear genes for cytochrome *c* oxidase. Biochim Biophys Acta 1352, 174–192

Hatefi Y (1985) The mitochondrial electron transport and oxidative phosphorylation system. Annu Rev Biochem 54, 1015–1069

Heikkila P, Soininen R, Tryggvason K (1993) Directional regulatory activity of *cis*-acting elements in the bidirectional α1(IV) and α2(IV) collagen gene promoter. J Biol Chem 268, 24677–24682

Hofmann K, Stoffel W (1993) TMBASE—a database of membrane-spanning protein segments. Biol Chem Hoppe-Seyler 374, 166

Huang A, Campbell CE, Bonetta L, McAndrews-Hill MS, Chilton-MacNeill S et al. (1990) Tissue, developmental and tumor-specific expression of divergent transcripts in Wilms tumor. Science 250, 991–994

Kadenbach B, Kuhn-Nentwig L, Buge U (1987) Evolution of a regulatory enzyme: cytochrome-*c* oxidase (Complex IV). Curr Top Bioenerg 15, 113–161

Kadonaga JT, Jones KA, Tjian R (1986) Promoter-specific activation of RNA polymerase II transcription by Sp1. Trends Biochem Sci 11, 20–23

Karim RD, Urness LD, Thummel CS, Klemz MJ, McKercher SR et al. (1990) The ETS-domain: a new DNA-binding motif that recognizes a purine-rich core DNA sequence. Genes Dev 4, 1451–1453

Lavia P, Macleod D, Bird A (1987) Coincident start sites for divergent transcripts at a randomly selected CpG-rich island of mouse. EMBO J 6, 2773–2779

Lee NH, Weinstock KG, Kirkness EF, Earle-Hughes JA, Fuldner RA et al. (1995) Comparative expressed-sequence tag analysis of differential gene expression profiles in PC-12 cells before and after nerve growth factor treatment. Proc Natl Acad Sci USA 92, 8303–8307

Lennon G, Auffray C, Polymeropoulos M, Soares MB (1996) The I.M.A.G.E. consortium: An integrated molecular analysis of genomes and their expression. Genomics 33, 151–152

Lomax MI, Grossman LI (1989) Tissue-specific genes for respiratory proteins. Trends Biochem Sci 14, 501–503

Lomax MI, Welch MD, Darras BT, Franke U, Grossman LI (1990) Novel use of a chimpanzee pseudogene for chromosomal mapping of human cytochrome *c* oxidase subunit IV. Gene 86, 209–216

Martin NC, Hopper AK (1994) How single genes provide tRNA processing enzymes to mitochondria, nuclei and the cytosol. Biochimie 76, 1161–1167

Montminy MR, Sevarino KA, Wagner JA, Mandel G, Goodman RH (1986) Identification of a cyclic-AMP responsive element within the rat somatostatin gene. Proc Natl Acad Sci USA 83, 6682–6686

Napiwotzki J, ShinzawaItoh K, Yoshikawa S, Kadenbach B (1997) ATP and ADP bind to cytochrome *c* oxidase and regulate its activity. Biol Chem 378, 1013–1021

O'Hanlon TP, Raben N, Miller FW (1995) A novel gene orientated in a head to head configuration with the human histidyl-tRNA synthetase (HRS) gene encodes an mRNA that predicts a polypeptide homologous to HRS. Biochem Biophys Res Commun 210, 556–566

Ohshima Y, Gotoh Y (1987) Signals for the selection of a splice site in pre-mRNA: Computer analysis of splice junction sequences and like sequences. J Mol Biol 195, 247–259

Poyton RO, McEwen J (1996) Crosstalk between nuclear and mitochondrial genomes. Annu Rev Biochem 65, 563–607

Rizzo MG, Ottavio L, Travali S, Chang C-D, Kaminska B et al. (1990) The promoter of the human proliferating cell nuclear antigen (PCNA) gene is bidirectional. Exp Cell Res 188, 286–293

Ryan MT, Herd SM, Sberna G, Samuel MM, Hoogenraad NJ et al. (1997) The genes encoding mammalian chaperonin 60 and chaperonin 10 are linked head-to-head and share a bidirectional promoter. Gene 196, 9–17

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press)

Sun Z, Kitchingman GR (1994) Bidirectional transcription from the human immunoglobulin V_H6 gene promoter. Nucleic Acids Res 22, 861–868

Tolerico H, Benko AL, Aris JP, Stanford DR, Martin NC et al. (1999) *S. cerevisiae* Mod5p-II contains sequences antagonistic for nuclear and cytosolic locations. Genetics 151, 57–75

Tsukihara T, Aoyama H, Yamashita E, Tomizaki T, Yamaguchi H et al. (1995) Structures of metal sites of oxidized bovine heart cytochrome *c* oxidase at 2.8 A. Science 269, 1069–1074

Tsukihara T, Aoyama H, Yamashita E, Tomizaki T, Yamaguchi H et al. (1996) The whole structure of the 13-subunit oxidized cytochrome *c* oxidase at 2.8 angstroms. Science 272, 1136–1144

Virbasius JV, Scarpulla RC (1990) The rat cytochrome *c* oxidase subunit IV gene family: tissue-specific and hormonal differences in subunit IV and cytochrome *c* mRNA expression. Nucleic Acids Res 18, 6581–6586

Virbasius JV, Scarpulla RC (1991) Transcriptional activation through ETS domain binding sites in the cytochrome *c* oxidase subunit IV gene. Mol Cell Biol 11, 5631–5638

Virbasius JV, Virbasius CA, Scarpulla RC (1993) Identity of GABP with NRF-2, a multisubunit activator of cytochrome oxidase expression, reveals a cellular role for an ETS domain activator of viral promoters. Genes Dev 7, 380–392

von Heijne G (1986) Mitochondrial targeting sequences may form amphiphilic helices. EMBO J 5, 1335–1342

Wolfe CL, Lou YC, Hopper AK, Martin NC (1994) Interplay of heterogeneous transcriptional start sites and translational selection of AUGs dictate the production of mitochondrial and cytosolic nuclear tRNA nucleotidyltransferase from the same gene in yeast. J Biol Chem 269, 13361–13366

Wolfe CL, Hopper AK, Martin NC (1996) Mechanisms leading to and the consequences of altering the normal distribution of ATP(CTP):tRNA nucleotidyltransferase in yeast. J Biol Chem 271, 4679–4686

Yamada M, Amuro N, Goto Y, Okazaki T (1990) Structural organization of the rat cytochrome *c* oxidase subunit IV gene. J Biol Chem 265, 7687–7692

Zeviani M, Nakagawa M, Herbert J, Lomax MI, Grossman LI et al. (1987) Isolation of a cDNA clone encoding subunit IV of human cytochrome *c* oxidase. Gene 55, 205–217

Zoladek T, Vaduva G, Hunter LA, Boguta M, Go BD, Martin NC, Hopper AK (1995) Mutations altering the mitochondrial-cytoplasmic distribution of Mod5p implicate the actin cytoskeleton and mRNA 3′ ends and/or protein synthesis in mitochondrial delivery. Mol Cell Biol 15, 6884–6894