Biochimica et Biophysica Acta, 1174 (1993) 63–71

BBAEXP 92512

Structural organization of the bovine gene for the heart/muscle isoform of cytochrome *c* oxidase subunit VIa

Ellen O. Smith [1] and Margaret I. Lomax

Department of Anatomy and Cell Biology, University of Michigan, Ann Arbor, MI (USA)

(Received 20 October 1992)

Key words: Mitochondrion; Tissue-specific expression; Promoter; MyoD1; Respiration; Transcription regulation

The bovine gene for the nuclear-encoded heart/muscle isoform of cytochrome *c* oxidase subunit VIa (*COX6A1*) was isolated from a library of bovine genomic DNA in lambda EMBL3 and sequenced. The gene spans 760 bp and comprises three exons and two small introns. Exon 1 encodes a 193 bp 5′ untranslated region, a 12 amino acid presequence, and the first 12 amino acids of the mature COX VIa protein. Exon 2 encodes amino acids 13 to 58, and exon 3 amino acids 59 to 85 plus the 35 bp 3′ untranslated region. Exons 2 and 3 are separated by a small intron of only 96 bp. All exon-intron boundaries matched the consensus splice junction sequences. *COX6A1* transcripts are present in RNA from bovine heart but not brain. Primer extension and ribonuclease protection assays were used to map the 5′ ends of *COX6A1* transcripts in heart; both methods identified several clusters of transcription initiation sites, indicating that *COX6A1* mRNA is heterogeneous at the 5′ end. The proximal 5′ flanking region is AT-rich and contains potential basal promoter elements, such as TATA and CCAAT boxes, associated with tissue-specific genes. A single consensus binding site for the muscle-specific transcription factor, MyoD1, was also located within this AT-rich region. The distal promoter region contained a perfect AP4 site plus potential binding sites for enhancer elements (NRF-1, Mt1, Mt3, and Mt4) proposed to regulate expression of genes for mitochondrial proteins.

Introduction

Cytochrome *c* oxidase (COX[1]; EC 1.9.3.1), the terminal enzyme complex of the mitochondrial electron transport chain, catalyzes the transfer of electrons from reduced cytochrome *c* to molecular oxygen [1]. In addition, the enzyme is involved in proton translocation across the mitochondrial inner membrane. In mammals, cytochrome *c* oxidase comprises thirteen polypeptide subunits. The three large subunits (I-III) are mitochondrial gene products that perform the catalytic functions of cytochrome *c* oxidase [2]. The ten smaller subunits are nuclear gene products [3,4] that may modulate cytochrome oxidase activity in response to different physiological signals or metabolic environments [5–7]. Amino acid sequence analysis of bovine heart and liver oxidases [8–11], as well as molecular genetic studies [12], have demonstrated that three nuclear subunits – COX VIa, VIIa and VIII – exist as tissue-specific isoforms, designated the heart (H) and liver (L) forms. This designation is based on the tissue from which these subunits were first isolated and does not imply restriction to that tissue. COX subunit VIa, a small, 84 amino acid integral membrane protein [13], was the first of these nuclear subunits shown to have isoforms [14,15]. Using isoform-specific cDNAs, Schlerf et al. [15] demonstrated that the gene for the H isoform is expressed only in heart and skeletal muscle and is thus restricted to contractile muscle, whereas the gene for the L (or non-muscle) isoform is expressed in all tissues, albeit at low levels in contractile muscle. The presence of two COX VIa isoforms in all mammals examined, i.e., cow [10], rat [14] and human [16,17], suggests that the genes for these isoforms probably arose by gene duplication before the mammalian radiation [12].

As part of our studies on tissue-specific expression of COX nuclear-encoded subunits, we recently isolated two cDNAs for COX VIa from a bovine heart cDNA library and demonstrated that this subunit, like most other COX nuclear subunits, is synthesized as a larger precursor containing a 12 amino acid N-terminal extension (presequence) [18]. To understand the transcrip-

Correspondence to: M.I. Lomax, Department of Anatomy and Cell Biology, University of Michigan, Ann Arbor, MI 48109-0616, USA.
[1] On leave from the Department of Biology, Madonna University, Livonia, MI 48150, USA.

Abbreviations: COX, cytochrome *c* oxidase; COX VIa, COX subunit VIa; COX VIa-H, heart/muscle isoform of COX VIa; COX VIa-L, liver isoform of COX VIa; *COX6A1*, genomic locus for bovine COX VIA-H; bp, base pair(s); kb, kilobase pair(s).

tional mechanisms controlling both basal and tissue-specific expression of COX nuclear subunits, we have begun to characterize representative COX genes. In this report, we present the isolation and structural analysis of the bovine gene for the heart/muscle isoform of COX VIa (*COX6A1*) and identify enhancer-like elements found in the promoters of genes for other respiratory proteins.

Materials and Methods

Southern blot analysis of bovine genomic DNA. Aliquots (10 μg) of high molecular weight calf thymus DNA (Sigma) were digested with restriction enzymes and subjected to electrophoresis on 15 cm 0.8% agarose gels at 30 v for 12 h. DNA was visualized with ethidium bromide, photographed, denatured and transferred to Nytran membranes by standard procedures [19]. Filters were hybridized in 40% formamide, 6 × SSPE, 1% SDS at 42°C with $5 \cdot 10^5$ cpm/ml ^{32}P-labeled cDNA probe labeled to a specific activity of $1\text{–}2 \cdot 10^8$ cpm/mg by the random-primer method [20]. After hybridization, filters were washed at 65°C in 0.1% SDS, 0.1 × SSC for 30 min and exposed to Kodak XAR-5 film with Dupont Cronex intensifying screens at −70°C.

Northern blot analysis of bovine RNAs. Total RNA was isolated from bovine heart and brain by the acid-phenol extraction method [21], from rat heart by the method of Chirgwin [22] followed by centrifugation over a CsCl step gradient, and from rat skeletal muscle by the method of Palmiter [23], as described previously [24]. Poly(A)$^+$ RNA isolated on an oligo(dT) spun column (Pharmacia) was separated by electrophoresis on denaturing formaldehyde-agarose gels [25] with RNA length standards (BRL 9.5 kb–0.24 kb RNA ladder). Transfer to Nytran membranes and hybridization with labeled cDNA was performed as recommended by the manufacturer.

Source of library. The bovine genomic library (Clontech BL1015j) had been constructed from *Sau*3A partial digest fragments of adult bovine genomic liver DNA size-fractionated to yield 8–22 kb fragments and cloned into the *Bam*HI site of EMBL-3 SP6/T7. The library was titered, plated on host strain LE392, and screened for COX VIa-H genomic clones by plaque hybridization [26] with a cDNA probe, as described in detail previously [27]. One clone, designated λCOX 6A1, was plaque-purified for further analysis. Phage DNA was isolated and digested with restriction enzymes for Southern blot analysis with the bovine heart COX VIa-H cDNA [18].

DNA Sequence analysis. Two *Eco*RI fragments, a 1 kb fragment containing the 5′ flanking region and a 2 kb fragment containing the entire *COX6A1* gene, were subcloned into plasmid vectors for DNA sequencing. The DNA sequence was determined by the dideoxy chain termination method [28] on alkali-denatured double-stranded plasmid DNA with the Sequenase version 2.0 kit (US Biochemicals) and [α-^{35}S]dATP (Amersham). Both commercial sequencing primers (USB) and synthetic oligonucleotide primers were used. DNA sequencing reactions were electrophoresed on 6% acrylamide-7 M urea gels. The sequence was confirmed by the University of Michigan DNA Sequencing Facility on a Model 373A automated DNA sequencer (Applied Biosystems, Foster City, CA) using synthetic primers, fluorescently-labeled dideoxynucleotides and Taq DNA polymerase. This method routinely yields at least 300 bp of sequence with 100% accuracy. DNA sequences were aligned using the ASSEMGEL program of PC/GENE (Intelligenetics, Mountain View, CA). The EUKPROM program, a 4-parameter weighted matrix algorithm [29], was used to identify

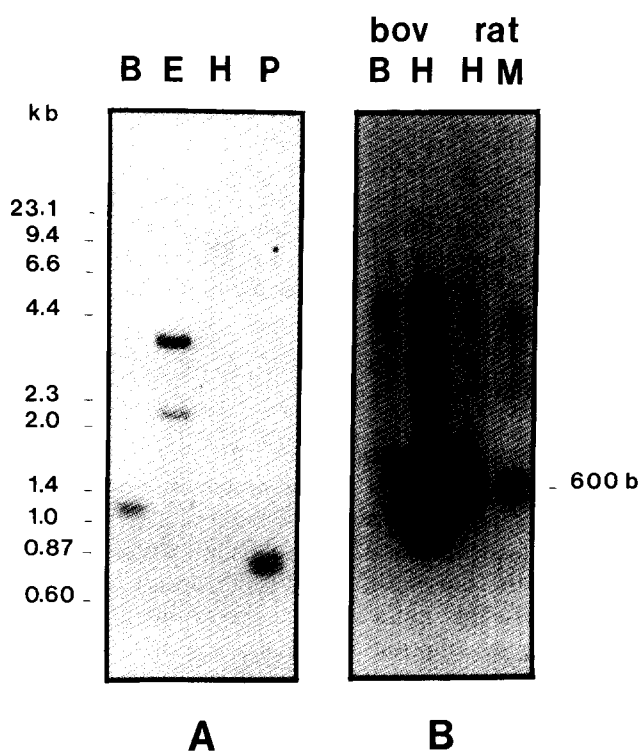

Fig. 1. Hybridization analysis of the *COX6A1* gene and transcripts. A. Genomic Southern blot analysis of bovine DNA. Bovine genomic DNA (10 μg) digested with *Bam*HI (B), *Eco*RI (E), *Hin*dIII (H), or *Pst*I (P) was separated by electrophoresis on an 0.8% agarose gel. DNA was transferred to Nytran membranes and hybridized with the BH-1 cDNA [18] under stringent hybridization conditions. Autoradiograms were exposed for 5 days with intensifying screens. B. Tissue-specific expression of the *COX6A1* gene. Northern blot analysis was performed on both bovine and rat RNAs. Poly(A)+ RNA (0.5 μg) isolated from bovine heart and brain and total RNA (20 μg) from rat heart and skeletal muscle was separated on denaturing formaldehyde-agarose gels and transferred to Nytran membranes. The membrane was hybridized with the BH-1 cDNA probe [18] and exposed to X-ray film with intensifying screens for 4 days. The autoradiogram was overexposed in an attempt to detect low level expression of *COX6A1* in brain and to visualize the rat mRNAs, which were not visible after overnight exposure.

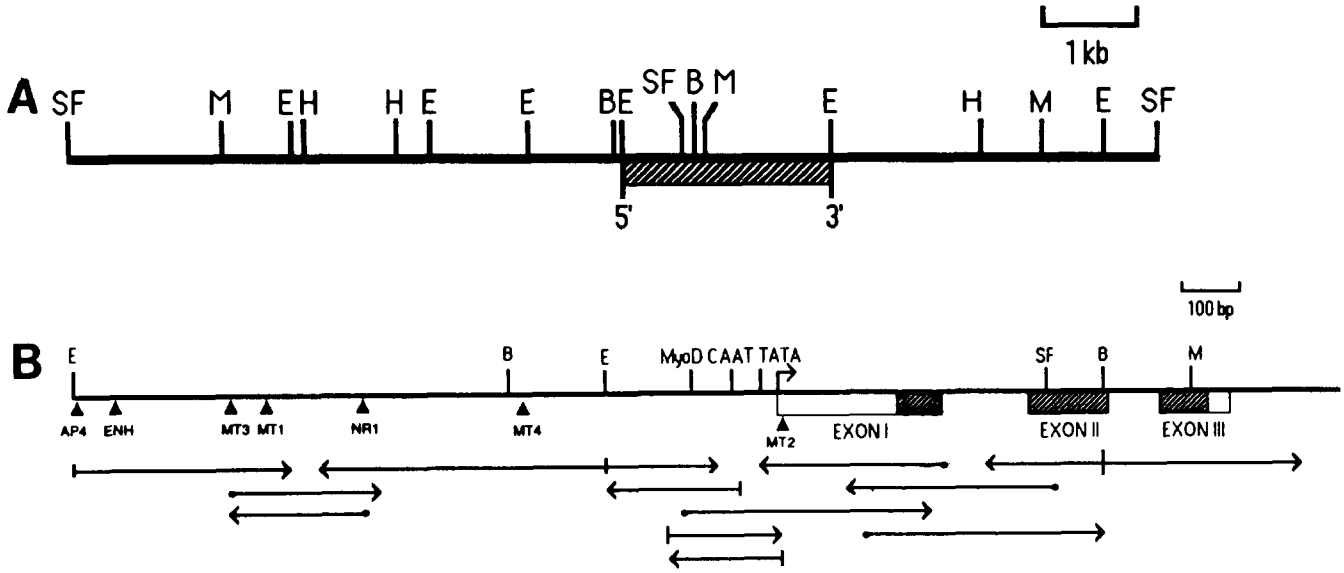

Fig. 2. Organization of the bovine *COX6A1* gene. A. Restriction map of genomic clone λCOX6AH-1 containing the *COX6A1* gene. The 2 kb *Eco*RI fragment (indicated by diagonal stripes) hybridized with both the BH-1 and BH-2 cDNAs [18], contained restriction sites for *Sfi*I, *Bam*HI, and *Sma*I, enzymes know to cleave the cDNA; therefore, this fragment was subcloned, mapped, and sequenced (see panel B). The 3 kb *Eco*RI fragment located 3′ to the gene hybridized with the BH-1 (3′), but not the BH-2 (5′), cDNA. Neither did it contain sites for *Sfi*I nor *Bam*HI. It was therefore assumed to contain a truncated pseudogene and was not characterized further. B. Organization and restriction map of the *COX6A1* gene. Open boxes denote 5′ and 3′ untranslated regions; striped boxes, coding sequence. B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; SF, *Sfi*I; M, *Sma*I. Arrows indicate the direction and extent of DNA sequencing. The location of synthetic sequencing primers is indicated by open circles. The furthermost transcription start site defined by primer extension studies is indicated by the arrow. Potential binding sites for transcription factors in the 5′ flanking region and promoter region are indicated.

potential eukaryotic promoters. We analyzed the gene sequence for recognition sites for eukaryotic sequence-specific transcription factors using the Wisconsin Genetics Computer Group program FIND and the data file provided by Dr. David Ghosh. Since this data file does not contain recognition sites for respiratory transcription factors, these sites were identified using QGSEARCH in PC/GENE.

Primer extension analysis. A synthetic 20-base anti-sense oligonucleotide primer (5′TCTCCCTTGGC-CGCACTGGC 3′) complementary to the DNA sequence-encoding amino acids 1 to 7 of mature COX VIa-H protein (bp 218–237, Fig. 3) was end-labeled with T4 polynucleotide kinase and [γ-^{32}P]ATP. Labeled primer was annealed with 0.5 μg poly(A)$^+$ RNA from bovine heart. Primer extension reactions were performed with AMV reverse transcriptase (BRL) at 43°C for 1 h. Extension products were analyzed by autoradiography after fractionation on a denaturing 6% acrylamide – 7 M urea gel with DNA sequencing reactions generated on M13mp18 DNA template and primed with the −40 primer as size markers.

RNAase protection analysis to map transcription start sites. An 800 bp *Eco*RI-*Bam*HI genomic fragment containing the 5′ flanking sequence, exon 1, intron 1, and part of exon 2 to the *Bam*HI site (Fig. 2B; codons 15 to 55, Fig. 3) was subcloned into BlueScript KS (Stratagene) for generation of an anti-sense riboprobe. DNA template preparation and generation of the anti-sense RNA by in vitro transcription with [α-^{32}P]UTP was essentially as described by Stratagene. The plasmid was linearized with *Eco*RI, treated with proteinase K, phenol extracted. Linearized plasmid DNA (1 μg) was transcribed in vitro with T7 RNA polymerase and [α-^{32}P]UTP. The ^{32}P-labeled RNA was treated with RNAase-free DNAaseI to remove the plasmid DNA template, phenol-extracted, and ethanol precipitated. RNAase-protection experiments were performed as previously described [19]. For each annealing condition, 5×10^5 cpm of anti-sense RNA probe was hybridized overnight with 5 μg poly(A)$^+$ RNA from bovine heart. Hybridization temperatures used were 55°, 60° and 65°C. The [^{32}P]RNA-mRNA hybrids were treated with a mixture of RNAase T1 and RNAase A to digest unprotected RNA, treated with proteinase K and phenol-chloroform extracted to remove protein, ethanol precipitated, then analyzed on a 6% polyacrylamide/7 M urea gel with M13mp18 sequencing reactions primed with the −40 sequencing primer (USB) as molecular length standards. An aliquot of ^{32}P-labeled RNA was annealed to 1 μg yeast tRNA as a negative control.

Results

Southern blot analysis of bovine genomic DNA. We previously reported the isolation and sequence of two bovine cDNAs for the heart/muscle isoform of COX

Fig. 3. Complete sequence of the *COX6A1* gene and its 5′ and 3′ flanking regions. The sequence of the gene is presented from near the *Eco*RI site in the promoter region. Promoter, exonic, and 3′ flanking sequences are presented in uppercase letters and intronic sequences in lowercase. The deduced protein sequence is presented and numbered above the DNA sequence. Intron-exon consensus splice junctions are overlined. The putative poly(A) addition signal is double-underlined. Potential TATA, CCAAT, MyoD1, AP4 and respiratory enhancer elements are underlined and labeled above the sequence. Restriction sites indicated in Fig. 2 or the text are underlined and labeled below the sequence. The gene is numbered from the furthest transcription start site (indicated by an * above the sequence) deduced from the longest primer-extended product (Fig. 4A). The location of the anti-sense primer used for the primer extension experiments is overlined (nts 217 to 236, codons 1 to 7 of the mature protein sequence).

VIa, designated COX VIa-H [18]. The overlapping cDNA sequences contained the entire coding region for the putative COX VIa-H precursor protein. To determine whether this bovine heart/muscle subunit is encoded by a single gene, or whether there are additional genomic sequences related to the COX VIa-H cDNAs as seen in the rat [15], we analyzed bovine genomic DNA by Southern blot hybridization analysis with the BH-1 cDNA-encoding the 3′ half of COX VIa-H mRNA [18]. This cDNA hybridized to a single 7 kb *Hin*dIII fragment, two *Eco*RI fragments (2 kb and 3 kb), and an 0.9 kb *Bam*HI fragment (Fig. 1A). Each of these fragments except the 3 kb-*Eco*RI fragment also hybridized with the BH-2 cDNA-encoding the 5′ half of COX VIa-H mRNA. These hybridization studies suggested that bovine COX VIa-H is encoded by a single copy gene, which we designated *COX6A1*. This result is in contrast to that of Schlerf et al. [15], who detected several fragments in the rat genome that hybridized to rat COX VIa-H cDNA.

Tissue-specific expression of the bovine COX6A1 gene. To demonstrate that transcription of the *COX6A1* gene, like the rat gene for COX VIa-H [15], is restricted to contractile muscle, we probed Northern blots of poly(A)$^+$ RNA from bovine heart and brain and total RNA from rat heart and skeletal muscle with the heart/muscle-specific cDNA probe, BH-1 (Fig. 1B). We overexposed this autoradiogram in order to determine whether low level transcription of this gene occurred in brain, but found a single, broad, major transcript of approx. 600 bp only in bovine heart, and not in bovine brain, RNA. The broad band could represent

```
                50            55         Intron 2-->
    oAlaPheIleProTyrHisHisLeuArgIleArgThrLys_
497 GGCATTCATCCCTTACCATCACCTCCGGATCCGCACCAAGgtacaccgacctcttgcggacgcagcaggcgggtgagccagggaatgggggacccccccat
                              BamHI
                                     Exon 3-->
                                        60            65            70            75
                                    __ProPheSerTrpGlyAspGlyAsnHisThrPheHisAsnProArgValAsnProLeuProT
597 actgacatctcgctccccttccctctctctccacagCCCTTCTCATGGGGAGACGGCAACCACACTTTTTTCCACAATCCCCGGGTCAACCCTCTGCCCA
                                                                                  SmaI
    80            85
    hrGlyTyrGluLysPro***
697 CGGGCTACGAAAAGCCTTGAGGCTTGGCTGATGCCCCCAGACACAATAAAATTGTGTGGAAGCTGTGTGTCCGAGGCTCCTTGGGGACCACGGGGAGGGG

797 CTGGTTCGAGCTGGGCGCGTGCGCACAGTGCAGTGGGTGCTCCACGTGCTCTTCCGGAGATCCTGACTCCCCAGGGACTCCAGGGAGCATCCTTGATGAA

897 CAGGTAGGAGATAGCCCCTCCCCGTTCACCTATTCATGATTAAGGCTCTAGGTTGTAAAAGGCTTTCACATACACTAATTTCTCATGATAACCCTGAGAG

997 GAAGGTATAATTTCGGTTCTCTTTTCTCCAGAGGAGGAATTTGAGATGCTCAGAGATTAAATGACTGAAATTCAGGGAGCTCACAGAGGCGGGTGGGAAC
                                                                                  SacI
1097 TACAAGGCTGCAGGGCCTGAAGTACCACGCGTTGCCTGACACCGACTGAGGGCCTCTCTGCGTTCCGGCTGTGTGGAGACAACAGTGATTTCAAACAGAT
         PstI
1197 CTACAGACCCATGTGAGAGGCTTCTTGAGCTAATTCTGGATTTGGAATATTTCAGTCATGCTCTACACCATCGTTCACGGGTCTCTGTATTATTTGGGGG

1297 AAGTGAGATGCTTCTTTTCTGGAAGTCTGATTCTCCTAGCTTCCAGGGACCCCGTTCAATGGGTCAAGCTCTGGATATTTCATTCACCTTCTAGGACAGT

1397 CAGCTCACTTGTCATTACCAGGGAGCCTGAGCGGAAACCCATGACCAACACACACACACACACAGAACTACTTTTGGCCTCCTCTCCTAACACTGCTTAC

1497 AGTGCACTTATGCTCACAGGTCCCAGTCAACTGAGAGAGTAAGAAGCATCTCTACCCACGAGCAGCCCCACCCTCCTAGAACAGCGCTGAGGTTCCTGGA

1597 CTCGCGCTCTTCCTTACAGCATAGTTATCTTATAAGCATCCTCATTTAATGTGACAGTTCTTATACATAAGAGAATTC
                                                                                  EcoRI
```

Fig. 3 (continued)

either autoradiographic spreading due to the long exposure and the high concentration of mRNA present in heart, or to size heterogeneity of the transcripts. Even on shorter (18 h) exposures, however, the *COX6A1* transcript appeared as a broad band, suggesting that the probe detected a population of molecules of slightly different length. The BH-1 cDNA probe also detected a similar sized transcript in rat heart and skeletal muscle. We conclude that *COX6A1* transcripts are the predominant species in bovine heart, and that expression of the *COX6A1* gene, like the rat COX VIa-H gene, is restricted to heart and skeletal muscle. Thus the *COX6A1* gene represents yet another example of a heart/muscle-specific gene.

Isolation and restriction mapping of the bovine gene for cytochrome c oxidase subunit VIa-H. Genomic clones were isolated by screening 10^6 phage from a bovine genomic library with the BH-2 cDNA [18]. Restriction mapping followed by Southern blot analysis of DNA from one clone, designated λCOX6AH-1, identified several restriction fragments that hybridized with the BH-1 cDNA (Fig. 2A): a 7 kb *Hin*dIII fragment, two *Eco*RI fragments (2 kb and 3 kb), and two *Bam*HI fragments (6 kb and 0.9 kb). The 7 kb *Hin*dIII and 2 kb *Eco*RI fragments hybridized with subregion probes from both the 5′ and 3′ ends of the cDNA. Therefore, the smaller 2 kb *Eco*RI fragment was subcloned into pUC13 and sequenced by the strategy indicated in Fig. 2B. An adjacent 1 kb *Eco*RI fragment containing 5′ flanking region was similarly subcloned and sequenced.

The sequence of the genomic region containing the *COX6A1* gene, the deduced COX VIa-H protein sequence, and the intron-exon junctions are presented in Fig. 3. The coding region comprises three exons and two small introns of 146 bp and 96 bp, respectively. Exon 1 encodes the 5′ UTR, a 12 amino acid presequence, and amino acids 1–12 representing the hydrophilic N-terminal domain of the mature protein. Exon 2 encodes amino acids 13–58, which include the 20 hydrophobic amino acids (amino acids 18 to 37) forming the transmembrane domain. Exon 3 encodes the hydrophilic C-terminal domain (amino acids 59–85) plus the 35 bases of 3′ UTR sequence present in the BH-1 cDNA. The cDNA sequence ends 2 bases after a putative polyadenylation signal (AATAAA); however, the gene probably extends for approx. 10–15 bp beyond this polyadenylation signal. No additional polyadenylation signal has been seen beyond the gene. The sequence of each intron/exon border agrees with consensus splice site sequences [30].

The 3.0 kb fragment downstream from the gene also hybridized with the BH-1 cDNA (Figs. 1A, 2A), but not the BH-2 cDNA probe (not shown). Furthermore, it contained neither the *Sfi*I nor *Bam*HI sites found in exons 1 and 2, respectively. This 3.0 kb fragment may contain either a duplicated exon or a processed pseu-

dogene. Because the 3 kb *Eco*RI fragment hybridized only with the BH-1 cDNA (3′ end), we assume it contains a truncated pseudogene and did not characterize it further. Thus, the restriction fragments seen on genomic Southern blots (Fig. 1A) could be accounted for by the structure of the *COX6A1* gene (Fig. 2A) and indicate that the bovine COX VIa-H isoform is encoded by a single gene.

Identification of transcription initiation sites. We used primer extension analysis of bovine heart mRNA to define the sites of transcription initiation. The primer used was a synthetic 20-base oligonucleotide complementary to the DNA sequence encoding amino acids 1–7 of mature COX VIa-H in exon 1 (Fig. 3). The oligonucleotide was labeled at the 5′ end with ^{32}P, annealed to bovine heart poly(A)$^+$ RNA, and extended with AMV reverse transcriptase. This analysis generated a heterogeneous collection of primer extended products (cDNAs) (Fig. 4A). Several clusters of cDNAs differing in length by 1 base were identified, suggesting that transcription is initiated imprecisely in the *COX6A1* gene. These potential transcription start sites all fell between the putative TATA box and the translation initiation codon. We also performed RNAase-protection analysis, using an anti-sense RNA probe extending from the *Eco*RI site upstream of the gene to the *Bam*HI site in exon 2, to confirm the heterogeneity at the 5′ end of *COX6A1* transcripts. The fragments protected from RNAase degradation were also heterogeneous in size but somewhat shorter than the primer-extended product (Fig. 4B). Since both methods gave multiple bands, we conclude that there are multiple transcription initiation sites for this gene in heart. Initiation at the furthest upstream site (designated by an *, Fig. 3) would result in a 181 bp 5′ UTR. Multiple start sites are consistent with and would explain the relatively broad band detected in Northern blot analysis of bovine heart RNA (Fig. 1B). Similar cases of heterogeneous transcription start sites in tissue-specific genes have been observed for the rat testis-specific cytochrome *c* gene [31] and the mouse muscle-specific acetylcholine receptor α-subunit gene [32]. Other COX genes exhibiting multiple transcription start sites are the rat *COX4* gene [33] and the mouse *COX5B* gene [34], although these genes are ubiquitously expressed and not tissue-specific.

Structure of the 5′ flanking region of the COX6A1 gene .To determine which sequence elements might be involved in transcriptional regulation of the *COX6A1* gene, we analyzed 1.3 kb of the 5′ flanking sequence of the gene (Fig. 3) for nucleotide sequence similarity to consensus binding sites for regulatory factors [35]. Although the coding region was GC-rich, averaging between 60–70% GC, the 300 bp region immediately preceding the transcription start site was AT-rich, averaging only 30% GC. Several potential basal promoter

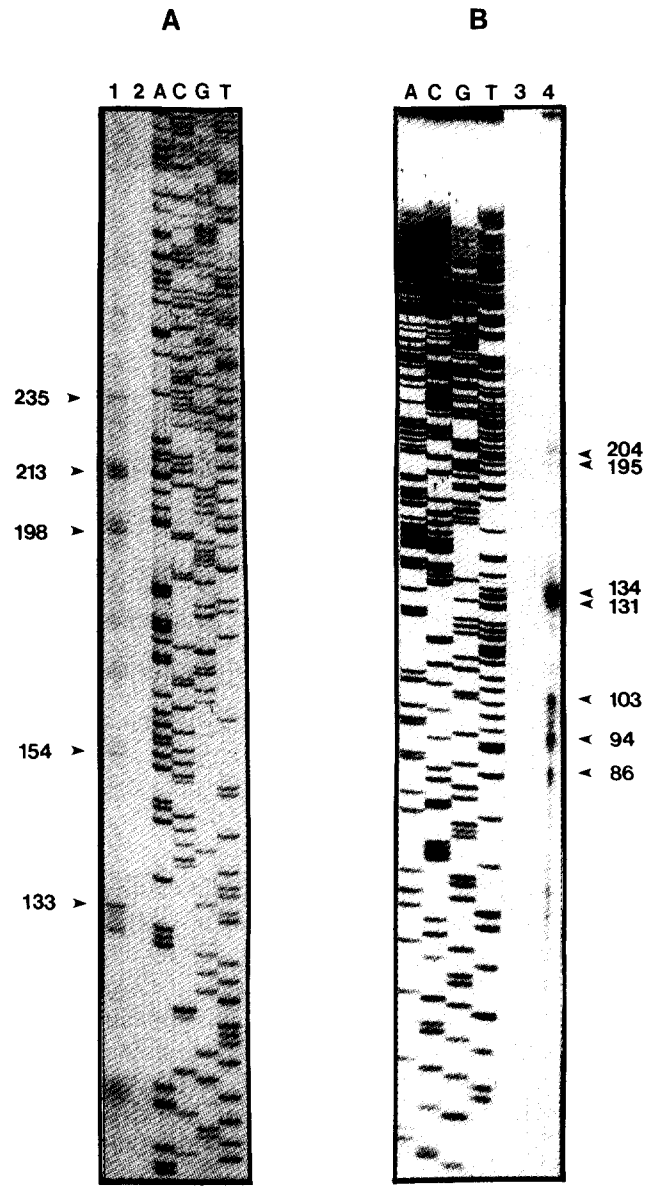

Fig. 4. Analysis of transcription start sites of the bovine heart *COX6A1* gene. A. Primer extension. Primer extension reactions were performed with 0.5 μg poly(A)+ RNA from bovine heart and an end-labeled 20-mer anti-sense primer that anneals to bp 217–236, Fig. 3. The molecular length standards were generated by performing DNA reactions on M13mp18 DNA template with the −40 sequencing primer (USB). Lane 1: The primer extension reaction with poly(A)+ RNA. Lane 2: Primer extension reaction with yeast tRNA. The sequencing reactions are indicated by A, C, G, T. The longest clearly visible primer extension product corresponds to transcription start site at a G residue on the coding strand. B. RNAase protection analysis. A uniformly labeled RNA probe complementary to exon 1, intron 1, and part of exon 2 was generated by in vitro transcription of a linearized plasmid containing this region. The anti-sense probe was annealed to 5 μg poly(A)+ RNA from bovine heart in hybridization buffer at 55°, 60° and 65°. A negative control containing 5 μg yeast tRNA was annealed at 55°. RNA-RNA hybrids were treated with RNAses A and T1 as described in Materials and Methods. A DNA sequencing ladder was generated with the −40 primer on M13mp18 template and Sequenase. Several clusters of transcription start sites were apparent, as had been seen with primer extension analysis (Fig. 4A).

elements, including both TATA and CCAAT boxes, were identified in this AT-rich proximal promoter region of the *COX6A1* gene (Fig. 3). A perfect TATA box [36] is located 64 bp 5′ to the transcription initiation site (−64 to −69). A consensus CCAAT box was found 45 bp upstream from the TATA box (−110 to −114). The sequence surrounding the CCAAT site matches the extended consensus sequence for the CP1 binding site, one of several transcription factors known to bind to the CCAAT-box [37]. The weighted matrix method for finding eukaryotic promoters [29] also identified this region as the most likely promoter region, with a score of −1.85. A single MyoD1 consensus binding site or E-box [38] is located just upstream from the CCAAT box in the AT-rich proximal promoter region (−183 to −188).

In contrast to the AT-rich proximal promoter region, the 1 kb region distal to it is slightly GC-rich and contains several sequences homologous to *cis*-regulatory elements or motifs found in several genes for mitochondrial proteins and believed to regulate mammalian respiratory genes [39]. A potential NRF-1 site (TGAGCATGTGC) [40,41] is located between −688 to −698 in the *COX6A1* gene matches the consensus NRF-1 sequence (TGCGCATGCGC) at 10 of 12 positions. Functional NRF-1 binding sites have been found in the promoters of the rat and human cytochrome *c* genes, the rat gene for cytochrome *c* oxidase subunit VIc, rat and human genes for MRP RNA, and human genes for cytochrome c_1 and ubiquinone binding protein [41]. One of the changes in the NRF-1 site in the *COX6A1* gene (underlined) is a C to A present also in the human cytochrome *c* gene.

A second enhancer element was first identified in the promoter of the human ATPase β subunit gene [42] and was shown to interact with nuclear proteins [42,43]. This element is also found in the human cytochrome c_1, and pyruvate dehydrogenase E1α subunit genes. A sequence (GACCACCAGG) found between −1208 and −1217 in the promoter of the *COX6A1* gene is related to the 3′ end of this element as defined by Kagawa and Ohta [43]. However, in human genes, this element occurs in an Alu repeat. Thus, the functional significance of this "enhancer" remains to be determined.

A third enhancer-like element for which a positive transcription factor has been identified by Wallace and co-workers as the OXBOX [44], a 13 base sequence found in the human genes for the F_1-ATPase β subunit and adenine nucleotide translocator. Although the OXBOX has been proposed to be a muscle-specific respiratory enhancer [44], no sequence similar to this enhancer was found in the 1.3 kb sequence upstream of the muscle-specific *COX6A1* gene. However, a sequence matching the OXBOX motif at 10 of 13 nucleotides has been identified in the promoter of the bovine gene for the heart/muscle isoform of COX subunit VIIa [45].

Several conserved sequence elements (Mt1, 3, and 4) that bind specific protein factors have been identified by Suzuki et al. [46,47] in the promoters of several genes for mitochondrial proteins. These sequences always occurred in the order Mt3, Mt1, Mt4. Interestingly, Mt3 and Mt4 sequences are also found in the D-loop region of animal mitochondrial DNAs [48]. Suzuki and co-workers have proposed that these sequences might play an important role in coordinate regulation of nuclear and mitochondrial genes for respiratory proteins. Sequences homologous to Mt3, Mt1 and Mt4 are present in the 5′ flanking regions of the human *COX6A1* gene, in the order proposed to be important for enhancer function [47]. However, since these consensus sequences are rather short, functional studies will be necessary to assess whether these regions are important in DNA-protein interaction. A 10 bp sequence (TCAGCTGTGG; −1293 to −1302) is a perfect match to the extended AP4 consensus binding site (YCAGCTGYGG) [35,49–51] and is also homologous to the Mt1 element (YCAGGTGTGG) identified in the promoters of the human cytochrome c_1 and F_1-ATPase subunit β genes [46].

Discussion

Analysis of bovine genomic DNA with isoform-specific probes identified only one genomic region encoding the heart/muscle isoform of COX VIa, suggesting that this isoform is encoded by a single-copy gene. This has been confirmed by isolation of a bovine genomic clone whose coding sequence is identical to the sequence of the bovine cDNAs for COX VIa-H [18]. The structure of the 5′ end of the *COX6A1* gene differs somewhat from that of other COX genes. Exon 1 encodes the 5′ untranslated region and the first 24 amino acids of the COX VI-H precursor protein comprising the short, 12 amino acid presequence plus the first 12 amino acids of the mature protein. By contrast, in the rat *COX4* [52] and human *COX5B* [53] genes, the longer presequence is encoded by a separate exon. This difference in the structure of COX genes may be due to the small size of the COX VIa-H presequence (12 amino acids) or may indicate that the N-terminal region of a longer protein has been recruited to function as a presequence. We note, however, that the presequence for the bovine gene for the heart/muscle isoform of COX subunit VIIa is encoded by two exons [45], suggesting that the COX VIIa-H presequence may be processed in two steps.

Our current knowledge of the tissue distribution of nuclear-coded COX subunits suggests that these subunits fall into two major classes: ubiquitous subunits present in all or most tissues, and tissue-specific iso-

forms found only in contractile muscle. One prediction arising from this tissue distribution is that genes for the ubiquitous subunits would be encoded by "housekeeping" genes. The promoters of housekeeping genes are GC-rich, contain many SP1 binding sites (GC-boxes), usually have no TATA box but may have CCAAT box homologies [54,55]. The genes for rat COX subunits IV [52] and VIc [56] and human subunit Vb [53] fall into this class. In contrast, the promoters of tissue-specific genes usually contain sequence elements that bind transcription factors TFIID (TATA box) and CBP (CCAAT box). As predicted, the promoter of the *COX6A1* gene has no features characteristic of "housekeeping" genes but instead contains TATA and CCAAT boxes, transcriptional control elements characteristic of tissue-specific genes. Since these two promoter elements are believed to lead to precise initiation of transcription, the multitude of transcription initiation sites is somewhat surprising. Imprecise transcription initiation may result in this case from the presence of many imperfect TATA and CCAAT elements in the AT-rich proximal promoter region, resulting in multiple RNA polymerase binding sites. Similarly, the promoter of the bovine gene for the heart/muscle isoform of COX subunit VIII is not GC-rich, but it has a unique transcription initiation site 25 bp downstream from a TATA element (M. Lomax, unpublished data). In contrast, the bovine gene for the heart/muscle isoform of COX VIIa [45] is expressed in a tissue-specific manner yet has structural features characteristic of a housekeeping gene: no TATA or CCAAT boxes but many GC boxes.

Mitochondrial biogenesis poses a unique problem: both mitochondrial and nuclear genes must be transcribed to produce the subunits of the four enzyme complexes of the electron transport chain. The recent characterization of several genes for mitochondrial proteins has led to identification of putative respiratory enhancer or activator elements and their transcription factors. The NRF-1 and Mt1, Mt3, and Mt4 elements identified in the 5′ flanking region of the bovine *COX6A1* gene may be important in regulating transcription of the gene. These elements are currently being analyzed for function.

Genes expressed only in striated muscle cells are regulated by the family of recently identified myogenic transcription factors such as MyoD1 [57,58]. A single MyoD1 binding site (E-box) is located in the proximal promoter region, just upstream from the TATA and CCAAT boxes. Whether this single E-box functions as a muscle-specific enhancer and which DNA sequences function as a cardiac-specific enhancers remain to be determined. No CArG box, a serum-response factor found in many muscle-specific genes [59] was found in the 5′ flanking region of the *COX6A1* gene. The sequence elements that regulate the expression of the *COX6A1* gene in skeletal and cardiac muscle are of great importance and are presently under investigation.

Acknowledgments

We thank Dr. N.J. Bachman, Northwestern University, for the bovine genomic library and Edna Gamliel for excellent technical assistance. This research was supported by grants NIH RO1 GM37086 and NSF DMB-90-00580 to M.I.L.

References

1 Hatefi, Y. (1985) Ann. Rev. Biochem. 54, 1015–1069.
2 Chomyn, A. and Attardi, G. (1987) in Current Topics in Bioenergetics (Lee, C.P., ed.), Vol. 15, pp. 295–329, Academic Press, New York.
3 Kadenbach, B., Kuhn-Nentwig, L. and Buge, U. (1987) in Current Topics in Bioenergetics (Lee, C.P., ed.), Vol. 15, pp. 114–162, Academic Press, New York.
4 Capaldi, R.A., Takamiya, S., Zhang, Y.Z., Gonzalez-Halphen, D. and Yanamura, W. (1987) in Current Topics in Bioenergetics (Lee, C.P., ed.), Vol. 15, pp. 91–119, Academic Press, New York.
5 Kadenbach, B. and Merle, P. (1981) FEBS Lett. 135, 1–11.
6 Poyton, R.O., Trueblood, C.E., Wright, R.M. and Farrell, L.E. (1988) Ann. Acad. Sci. NY 550, 289–307.
7 Capaldi, R.A. (1990) Ann. Rev. Biochem. 59, 569–596.
8 Merle, P. and Kadenbach, B. (1982) Eur. J. Biochem. 125, 239–244.
9 Kadenbach, B., Reimann, A., Stroh, A. and Huther, F.-J. (1988) in Oxidases and Related Redox Systems (King, T.E., Mason, H.E. and Morrison, M., eds.), pp. 653–668, Alan R. Liss, New York.
10 Yanamura, W., Zhang, Y.-Z., Takimiya, S. and Capaldi, R.A. (1988) Biochemistry 27, 4909–4914.
11 Capaldi, R.A. (1988) Trends Biochem. Sci. 13, 144–148.
12 Lomax, M.I. and Grossman, L.I. (1989) Trends Biochem. Sci. 14, 501–503.
13 Meinecke, L. and Buse, G. (1985) Biol. Chem. Hoppe-Seyler 366, 687–694.
14 Kadenbach, B., Hartmann, R., Glanville, R. and Buse, G. (1982) FEBS Lett. 138, 236–238.
15 Schlerf, A., Droste, M. and Kadenbach, B. (1988) EMBO J. 7, 2387–2391.
16 Fabrizi, G.M., Rizzuto, R., Nakase, H., Mita, S., Kadenbach, B. and Schon, E.A. (1989) Nucl. Acids Res. 17, 6409.
17 Fabrizi, G.M., Sadlock, J., Mita, S., Koga, Y., Rizzuto, R. and Schon, E.A. (1992) Gene 119, 307–312.
18 Smith, E.O., BeMent, D.M., Grossman, L.I. and Lomax, M.I. (1991) Biochim. Biophys. Acta 1089, 266–268.
19 Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual (Second Edition) Vol. 2, pp. 9.34–9.58, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
20 Feinberg, A.P. and Volgelstein, B. (1983) Anal. Biochem. 132, 6–13.
21 Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. 162, 156–159.
22 Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 14, 5294–5299.
23 Palmiter, R.D. (1974) Biochemistry 13, 3606–3615.
24 Lomax, M.I., Coucouvanis, E., Schon, E.A. and Barald, K.F. (1990) Muscle and Nerve 13, 330–337.
25 Lehrach, H., Diamond, D., Wozney, J.M. and Boedtker, H. (1977) Biochemistry 16, 4743–4751.

26 Benton, W.D. and Davis, R.W. (1977) Science 196, 180–182.
27 Lomax, M.I., Hsieh, C.-L., Darras, B.T. and Francke, U. (1991) Genomics 10, 1–9.
28 Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Nat. Acad. Sci. USA 74, 5463–5467.
29 Bucher, P. (1990) J. Mol. Biol. 212, 563–578.
30 Padgett, R.A., Grabowski, P.J., Konarski, M.M., Seiler, S. and Sharp, P.A. (1986) Ann. Rev. Biochem. 55, 1119–1150.
31 Virbasius, J.B. and Scarpulla, R.C. (1988) J. Biol. Chem. 263, 6791–6796.
32 Prody, C.A. and Merlie, J.P. (1991) J. Biol. Chem. 266, 22588–22596.
33 Virbasius, J.B. and Scarpulla, R.C. (1990) Nucl. Acids Res. 18, 6581–6586.
34 Basu, A. and Avadhani, N.G. (1991) J. Biol. Chem. 266, 15450–15456.
35 Locker, J. and Guzar, G. (1991) DNA Sequences 1, 3–11.
36 Bucher, P. and Trifonov, E.N. (1986) Nucl. Acids Res. 14, 10009–10026.
37 Chodosh, L.W., Baldwin, A.S., Carthew, R.W. and Sharp, P.A. (1988) Cell 53, 11–24.
38 Lassar, A.B., Paterson, B.M. and Weintraub, H. (1986) Cell 47, 649–656.
39 Nagley, P. (1991) Trends in Genetics 7, 1–4.
40 Evans, M.J. and Scarpulla, R.C. (1988) Mol. Cell. Biol. 8, 34–41.
41 Evans, M.J. and Scarpulla, R.C. (1990) Genes and Devel. 4, 1023–1034.
42 Tomura, H., Endo, H., Kagawa, Y. and Ohta, S. (1990) J. Biol. Chem. 265, 6525–6527.
43 Kagawa, Y. and Ohta, S. (1990) Int. J. Biochem. 22, 219–229.
44 Li, K., Hodge, J.A. and Wallace, D.C. (1990) J. Biol. Chem. 265, 20585–20588.
45 Seelan, R.S. and Grossman, L.I. (1992) Biochemistry 31, 4696–4704.
46 Suzuki, H., Hosokawa, Y., Nishikimi, M. and Ozawa, T. (1989) J. Biol. Chem. 264, 1366–1374.
47 Suzuki, H., Hosokawa, Y., Toda, H., Nishikimi, M. and Ozawa, T. (1990) J. Biol. Chem. 265, 8159–8163.
48 Suzuki, H., Hosokawa, Y., Nishikimi, M. and Ozawa, T. (1990) J. Biol. Chem. 266, 2333–2338.
49 Ghosh, D. (1990) Nucl. Acids Res. 18, 1749–1756.
50 Mermod, N., Williams, T.J. and Tjian, R. (1988) Nature 332, 557–561.
51 Chiu, R., Imagawa, M., Imbra, R.J., Bockovern, J.R. and Karin, M. (1987) Nature 329, 648–651.
52 Yamada, M., Amuro, N., Goto, Y. and Okazaki, T. (1990) J. Biol. Chem. 265, 7687–7692.
53 Lomax, M.I., Hsieh, C.-L., Darras, B.T. and Francke, U. (1991) Genomics 10, 1–9.
54 Dynan, W.S. and Tjian, R. (1985) Nature 316, 774–778.
55 Briggs, M.R., Kadonaga, J.T., Bell, S.P. and Tjian, R. (1986) Science 234, 47–52.
56 Suske, G., Enders, C., Schlerf, A. and Kadenbach, B. (1988) DNA 7, 163–171.
57 Lassar, A.B., Buskin, J.N., Lockshon, D., Davis, R.L., Apone, S., Hauschka, S.D. and Weintraub, H. (1986) Cell 47, 649–656.
58 Weintraub, H., Davis, R.L., Tapscott, S., Thayer, M., Krause, M., Benezra, R., Blackwell, T.K., Turner, D., Rupp, R., Hollenberg, S., Zhuang, Y. and Lassar, A.B. (1991) Science 251, 761–766.
59 Sartorelli, V., Webster, K.A. and Kedes, L. (1990) Genes and Develop. 5, 1811–1822.