Two bovine genes for cytochrome *c* oxidase subunit IV: a processed pseudogene and an expressed gene

Nancy J. Bachman **, Margaret I. Lomax ** and Lawrence I. Grossman *

Department of Cellular and Molecular Biology, Division of Biological Sciences, The University of Michigan, Ann Arbor, MI 48109 (U.S.A.)

SUMMARY

We have isolated and analyzed 17 clones from a bovine genomic library in phage λCharon28 probed with a bovine liver cDNA for cytochrome *c* oxidase subunit IV. Restriction enzyme mapping and Southern analysis indicated that these clones represent only two genomic regions. One region was shown by nucleotide sequencing to contain a subunit IV pseudogene of the processed type. The other class of clones contained the 5' region of a putative expressed gene; the region consists of two exons and two introns, with one exon encoding exclusively the domain representing the presequence present on newly synthesized subunit-IV polypeptides. Genomic Southern analysis indicated that these two clones probably represent the only sequences in the bovine nucleus that share nucleotide sequence identity with the liver subunit IV cDNA when utilizing moderately stringent hybridization conditions.

INTRODUCTION

Eukaryotic cytochrome *c* oxidases (EC 1.9.3.1) (reviewed in Capaldi et al., 1983; Azzi, 1980), proteins of central importance in oxidative respi-

ration, are well known to consist of polypeptides encoded by both the nucleus and the mitochondria. The three polypeptides encoded and expressed in the mitochondria (Anderson et al., 1981; Mason and Schatz, 1973) must interact with five to six nuclear gene products in yeast (Power et al., 1984; Gutweniger et al., 1981) or nine to ten in mammals (Kadenbach et al., 1983; Merle and Kadenbach, 1980a) to form a functional enzyme. The catalytic

---

* *Correspondence to*: Dr. L.I. Grossman at his present address, Department of Molecular Biology and Genetics, Wayne State University School of Medicine, Detroit, MI 48201 (U.S.A.) Tel. (313)577-5326.

** Present addresses: (N.J.B.) Department of Biology, Franklin and Marshall College, Lancaster, PA 17604 (U.S.A.) Tel. (717)291-3948; (M.I.L.) Department of Microbiology and Immunology, The University of Michigan Medical School, Ann Arbor, MI 48109 (U.S.A.) Tel. (313)936-0847.

Abbreviations: aa, amino acid(s); bp, base pair(s); COX IV, cytochrome *c* oxidase subunit IV; *coxIV*, gene or mRNA coding for COX IV; kb, 1000 bp; Myr, million years; nt, nucleotide(s); SSC, 0.015 M NaCl, 0.015 M $Na_3 \cdot$ citrate, pH 7.0.

and proton-pumping functions of cytochrome oxidase and the site of interaction with cytochrome c all reside on the mitochondrial subunits, whereas the function(s) of the nuclear-encoded subunits are unknown. Most of the nuclear-encoded subunits of cytochrome oxidase are synthesized as larger polypeptide precursors and imported into mitochondria, where they are processed into their mature forms (Lewin et al., 1980; Mihara and Blobel, 1980; Schmelzer and Heinrich, 1980; Schmelzer et al., 1982) and assembled into a functional complex.

We recently reported the isolation and characterization of a liver cDNA from the domestic cow *Bos taurus* coding for COX IV (Lomax et al., 1984). This cDNA encodes 32 bp of 5′-untranslated sequence, 22 aa of an N-terminal extension (presequence), and aa 1–105 of the mature protein. We have begun to analyze the organization of *coxIV* genes and describe here the isolation of two bovine *coxIV* genes, a processed pseudogene containing mutations that preclude synthesis of a normal protein product, and the 5′ end of a putative expressed gene that contains introns. These two genes coding for bovine COX IV comprise most, and probably all, of the bovine genomic sequences closely related to the *coxIV* cDNA.

MATERIALS AND METHODS

(a) Materials

T4 DNA ligase and *Escherichia coli* DNA polymerase I were obtained from Boehringer-Mannheim. *E. coli* DNA polymerase I Klenow fragment was purchased from New England Biolabs. Restriction endonucleases were from the above suppliers, BRL, or IBI. M13 DNAs and 15- or 17-bp primers were from P-L Biochemicals. Reagents for dideoxy sequencing were purchased as a kit from BRL. Nitrocellulose membranes were from Millipore (HATF, 0.45 μm) or Schleicher & Schuell (BA85, 0.45 μm). Deoxyadenosine 5′-(α-[^{35}S]thio) triphosphate (600 Ci/mmol, 8 mCi/ml) and deoxynucleotide 5′-[α-^{32}P] triphosphates (800 Ci/mmol or 3000 Ci/mmol, 10 mCi/ml) were from Amersham.

(b) Isolation of bovine genomic *coxIV* genes

A genomic library (Woychik et al., 1982) containing *Mbo*I partial digest fragments of bovine DNA cloned into the *Bam*HI site of λ Charon 28 was plated on the host strain, K802, both obtained from Fritz Rottman (Case Western Reserve University). Cloned DNA was isolated by standard methods (Lomax et al., 1984; Maniatis et al., 1982; Birnboim and Doly, 1979). High M_r genomic DNA was isolated from 10-g portions of frozen bovine liver (Blin and Stafford, 1976). The *coxIV* cDNA [pCOX4.419 (Lomax et al., 1984)] insert was separated from pBR322 on freeze-thaw 5 to 20% sucrose gradients (Baxter-Gabbard, 1972; El-Gewely and Helling, 1980) after digesting 100–250 μg of cDNA with *Pst*I.

Recombinant clones were screened with the nick-translated (Maniatis et al., 1982) *Pst*I fragment of pCOX4.419 or with single-stranded probes prepared by primer extension from M13mp11 subclones of the cDNA. Bacteriophage plaques or bacterial colonies on nitrocellulose filters (Maniatis et al., 1982; Benton and Davis, 1977) were hybridized at 42°C in 6 × SSC, 50% formamide with the nick-translated DNA probe for 16 to 24 h. The filters were washed 3 times for 1 h in 2 × SSC, 0.1% SDS at room temperature, and twice for 1 h in 1 × SSC, 0.1% SDS at 65 to 68°C.

(c) Analysis of genomic DNA for *coxIV*-related genes

Southern transfers were incubated with nick-translated or single-stranded probes for 16 to 24 h at 65 to 68°C in a hybridization solution containing 6 × SSC. For the cDNA (56% G + C), this is about 37°C below the hybrid melting temperature. Filters containing transferred fragments from cloned DNAs were hybridized and washed according to Maniatis et al. (1982). Genomic DNA transfers were prehybridized in 6 × SSC, 5 × Denhardt's solution, 250 μg/ml of denatured salmon sperm DNA and 0.1% SDS at 68°C for at least 6 h. Hybridization for 24 h with NaOH-denatured nick-translated probe was done in the prehybridization mixture containing only 1 × Denhardt's solution. The filters were washed twice in 2 × SSC for 30 min at room temperature, twice in 1 × SSC for 1 h at 68°C, and twice in 0.5 × SSC for 1 h at 68°C.

(d) Nucleotide sequence analysis

M13 subclones were sequenced by the dideoxy chain-termination method (Sanger et al., 1977) utilizing either [α-^{32}P]dATP or [α-^{35}S]dATP and the conditions specified in the BRL dideoxy sequencing kit. Chemical sequencing (Maxam and Gilbert, 1980) was performed on 3'-end-labeled fragments (Maniatis et al., 1982) with the modifications (Rubin and Schmid, 1980) previously noted (Hudspeth et al., 1982). Products of sequencing reactions were separated on 6% or 11% polyacrylamide gels containing 8 M urea or 50% formamide. Sequencing gels were run at up to 55°C to minimize effects of secondary structure due to the high G + C content of the DNA.

RESULTS AND DISCUSSION

(a) Identification of bovine genomic clones coding for COX IV

To identify genomic clones coding for bovine COX IV, we screened a bovine genomic library by hybridization with the previously isolated subunit IV cDNA (Lomax et al., 1984), designated pCOX4.419. Of approx. 750000 plaques screened, 28 purified plaques were characterized by rescreening with subregion probes and by constructing restriction endonuclease site maps of their DNAs (Fig. 1). Eleven of these 28 plaques failed to hybridize with subregion probes. The remaining 17 plaques represented multiple isolates of four different recombinant clones, denoted phages λBCO4.1, λBCO4.2, λBCO4.3 and λBCO4.5. Comparison of restriction maps indicated that three of the phage contain overlapping fragments from the same genomic region, whereas the fourth contains a different genomic region (Fig. 1). Rescreening with subregions of the
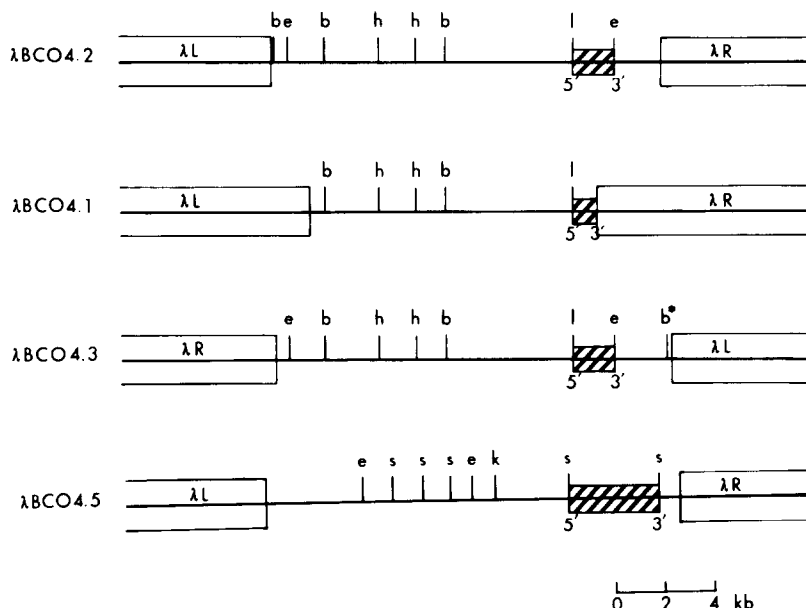

Fig. 1. Restriction maps of λ cytochrome oxidase subunit IV clones. Restriction enzymes: b, BglII; e, EcoRI; h, HindIII; k, KpnI; l, SalI; s, SstI. All EcoRI, SalI, KpnI and SstI sites are shown for λBCO4.5, and all BglII, EcoRI, HindIII and SalI sites are shown for the others. The hatched areas indicate the region in each clone that hybridized to the coxIV cDNA. The orientation of the coding sequences, determined by hybridization of probes derived from 5' (148 bp PstI-AvaI) and 3' (294 bp AvaI-PstI) regions of the coxIV cDNA or, for clone λBCO4.5, by nucleotide sequencing, is indicated below each map. A BglII site that was rarely cleaved in λBCO4.3 is indicated by an asterisk. The open boxes to the left (λL) and right (λR) indicate the vector arms of λ Charon 28. The insert sizes are: λBCO4.1, 11.6 kb; λBCO4.2, 16.2 kb; λBCO4.3, 16.0 kb; λBCO4.5, 16.9 kb.
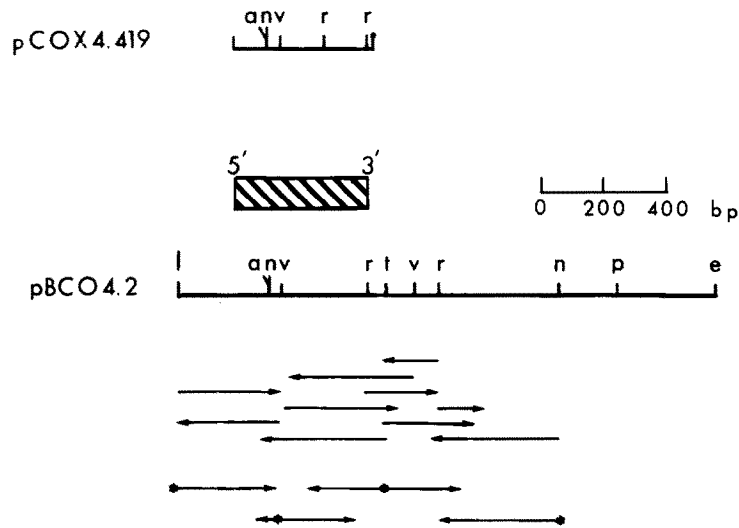
Fig. 2. Restriction map and sequencing strategy of pBCO4.2. Restriction enzymes: a, *Apa*I; e, *Eco*RI; l, *Sal*I; n, *Nco*I; p, *Pst*I; r, *Rsa*I; t, *Taq*I; v, *Ava*I. The region in pBCO4.2 homologous to the cDNA is indicated by the hatched box. Strategies used to determine the nucleotide sequence of pBCO4.2, which is a subclone of λBCO4.2 (Fig. 1), are indicated below the map. Ends labeled for Maxam and Gilbert (1980) sequencing are indicated by asterisks. Unmarked arrows indicate the regions sequenced using the dideoxy chain termination method.

cDNA indicated that λBCO4.5 hybridized only to a 5'-specific probe whereas the other three clones hybridized to both 5'-and 3'-specific probes.

(b) The prevalent class of genomic clones contains a processed pseudogene

Phages λBCO4.1, λBCO4.2, and λBCO4.3 each appear to contain the same region of the bovine genome, based on the similarity of their restriction site maps (Fig. 1); of these clones, only λBCO4.2 was analyzed further. A 1.7-kb *Sal*I-*Eco*RI fragment (Fig. 1) that contained sequences homologous to the subunit IV cDNA was subcloned into pUC8. The nucleotide sequence of 1033 bp of the resulting subclone (pBCO4.2) was determined from the *Sal*I site (Figs. 2 and 3). This fragment appears to contain a processed pseudogene with the following features. (i) The gene is colinear with the cDNA (except for two small deletions) and thus contains no introns; (ii) a poly(A)-rich sequence [A$_{15}$(GAAA)$_3$] at nt 840–866 in pBCO4.2 is approx. 15 nt downstream from a putative poly(A) addition signal [(T)AATAA(G)] (Proudfoot and Brownlee, 1976) and 116 bases downstream from the end of the

translation stop codon (Fig. 3); (iii) the sequence is flanked by two 10-bp direct repeats (AGAAAT-GAAT), generally assumed to be associated with the retroposition of a processed mRNA or a cDNA into the chromosome (Soares et al., 1985).

Although the sequence of pBCO4.2 is highly homologous to the subunit IV cDNA, several features of the processed pseudogene suggest that it could not encode a functional protein. Perhaps most significant is a single nucleotide substitution at nt 220, changing the initiation codon from ATG (Met) to ATA (Ile); the latter is known to function as a translation initiation codon only in mitochondria (Anderson et al., 1982). In addition, two small deletions have occurred, one of which removes a single nucleotide between nt 510 and 512 and produces a frameshift after aa 97; this frameshift puts a downstream termination codon (nt 656–658) in frame and would lead to the synthesis of an altered protein truncated by 22 aa (Fig. 3). There are several other substitutions in this region, suggesting that several different changes contributed to the net loss of one nucleotide. A second region (corresponding to nt positions 378–386 in the cDNA) contains a deletion of CTC, which maintains the reading frame created by the

(A)₁   GTCGACAAGAATGATGGCACAGATTTGC

```
  29   TCTGGGTTTGCTGTTTCCTCGTCATTTTGATTTTGTCATGGTGTGTGTCCTTACATGAATATC

  92   TTCTAATTTCATTCACAGCAGGGACAAAACCTGTAGAAAATGCTTTGTAATAAATTTAATGAG

 155   AAATGAATCAGAGAGGGCAGTCGCGGCGGGCACGGTGGGCAGCGGCTGGTCAGTGGTGGCAGA
   1                                   acggtgggcagcggctggtcggtggcatcaga
```

```
          *      o                        *       * o
 218   ATA TTG GCG ACC AGA GTA TTT AGC CTG ATT GGC AGG CGT GCA GTT TCC
  33   atg ttg gca acc aga gta ttt agc ctg att ggt agg cgt gca atc tcc
       Met Leu Ala Thr Arg Val Phe Ser Leu Ile Gly Arg Arg Ala Ile Ser
```

```
 266   ACC TCG GTG TGT GTT CGG GCC CAT GGA AGT GTT GTA AAG AGT GAA GAT
  81   acc tcg gtg tgt gtt cgg gcc cat gga agt gtt gta aag agt gaa gat
       Thr Ser Val Cys Val Arg Ala His Gly Ser Val Val Lys Ser Glu Asp
```

```
 314   TAT GCT CTC CCG AGT TAT GTG GAC CGG CGT GAC TAC CCC TTG CCC GAC
 129   tat gct ctc ccg agt tat gtg gac cgg cgt gac tac ccc ttg ccc gac
       Tyr Ala Leu Pro Ser Tyr Val Asp Arg Arg Asp Tyr Pro Leu Pro Asp
```

```
                                                          *
 362   GTG GCC CAC GTC AAG AAC CTG TCT GCC AGC CAG AAG GCC TTG AAG CAG
 177   gtg gcc cat gtc aag aac ctg tct gcc agc cag aag gcc ttg aag gag
       Val Ala His Val Lys Asn Leu Ser Ala Ser Gln Lys Ala Leu Lys Glu
```

```
                                             o  *
 410   AAG GAG AAG GCT TCC TGG AGC AGC CTC TCC ATC GGT GAG AAA GTT GAA
 225   aag gag aag gct tcc tgg agc agc ctc tcc att gat gag aaa gtt gaa
       Lys Glu Lys Ala Ser Trp Ser Ser Leu Ser Ile Asp Glu Lys Val Glu
```

```
        *  **                                   o   *
 458   CTG CAC GAC CTT AAG TTC AAG GAG AGC TTC GCT GAG ATG CGC AGG AGC
 273   ctg tac cgc ctt aag ttc aag gag agc ttc gcc gag atg aac agg agc
       Leu Tyr Arg Leu Lys Phe Lys Glu Ser Phe Ala Glu Met Asn Arg Ser
```

```
           o **   *  *     o                         *
 506   ACA A-C AGG TGC CAG ACA GTG GTG GGC GCG GCC ACG TTC TTC ATC GGC
 321   aca aat gag tgg aag acg gtg gtg ggc gcg gcc atg ttc ttc atc ggc
       Thr Asn Glu Trp Lys Thr Val Val Gly Ala Ala Met Phe Phe Ile Gly
```

```
                          *           *   *
 553   TTC ACC GCG CTC CTC --- ACC TGG GAG AAG CGC TGT GTG TAC GGC CCC
 369   ttc acc gcg ctc ctc ctc atc tgg gag aag cac tat gtg tac ggc ccc
       Phe Thr Ala Leu Leu Leu Ile Trp Glu Lys His Tyr Val Tyr Gly Pro
```

(B)

```
 598   ATC CTG CAC ACC TTT GAA GAG GAG TGG GTG GCC AAG CAG ACC AGG AGG
           Leu                                              Arg
       Ile Pro His Thr Phe Glu Glu Glu Trp Val Ala Lys Gln Thr Lys Arg
```

```
 646   ATG CTC GAC GTG AAG GTG GCC CCC ATC CAG GGC TTC TCA GCC AAG CGG
           Val                                              Arg
       Met Leu Asp Met Lys Val Ala Pro Ile Gln Gly Phe Ser Ala Lys Trp
```

```
 694   GAC TAC GAC AAG AAC AAG TGG AAG ACG TAA GGACCCGCGGTTCCCGAGTCTGC
           Lys         Thr
       Asp Tyr Asp Lys Asn Glu Trp Lys Lys ---
```

```
 747   ACCTCACCTGTCCGGTCACGTCCATGCAGCTCTGCATGTTCACTGGAAGCGCTGTGTTGCAGC

 810   ACCAGTACTAATAAGTGTGCAGTTTACATGAAAAAAAAAAAAAAAAAGAAAGAAAGAAATGAATA

 873   AGACTAAAGTAATATGCACATTTAATAGAAGTCCACAGAAAAACACAATATTGCTTTTGCCAC

 936   TTTTTGGACTAGAAAATCTACTACTTAGGTAAAAATAGTAAAAGAAAGAAGTTAGATTTAGAA

 999   AGCATGTGATGGAGGGTGAAACAGCCAAAAATCTC
```

Fig. 3. Nucleotide sequences of part of pBCO4.2. Nucleotides in the pseudogene sequence (upper case) are numbered beginning with the SalI site. The poly(A) addition site and poly(A)-rich region are indicated by a bar above the sequence. Flanking direct repeats are underlined. Shown for comparison are the nucleotide sequence of the cDNA (lower case), numbered from the most 5' nucleotide in pCOX4.419, and the amino acid sequence of bovine COX IV protein, including the predicted 22-aa N-terminal extension. The sequence of nt 247–412 of pCOX4.419 was determined by dideoxy sequencing of AvaI-PstI and RsaI fragments cloned into M13mp11. Two errors in the previously published sequence of pCOX4.419 (Lomax et al., 1984), at nt 122 and 134, are corrected here and underlined. (A) Sequence of the pseudogene to the end of the cDNA sequence. Base substitutions are indicated above the pseudogene (°, silent changes; *, replacement changes). A single nt deletion in the pseudogene (following nt position 509) shifts the reading frame. The C residues at nt positions 413–416 of the cDNA, although found in the same position in the pseudogene, could have been generated by addition of C-tails during the cDNA cloning. (B) Sequence of the pseudogene beyond the cDNA insert of pCOX4.419. Amino acid changes are shown where they differ from the COX IV protein.

upstream frameshift mutation. Interestingly, the CTC sequence is repeated three consecutive times in the cDNA. Loss of such short repeated sequences has been observed previously in noncoding DNA (Efstratiadis et al., 1980; Moore, 1983; Foran et al., 1985) and interpreted as examples of slipped mispairing (Streisinger et al., 1966) during DNA replication.

Comparison of the nucleotide changes between the pseudogene and the subunit IV cDNA can be used to estimate the time of divergence of the pseudogene. It shows high homology to the bovine liver cDNA, with a 107-bp region of complete identity. In the 416 nt that can be compared, there are 29 single nt substitutions and two small deletions. If the deletions are scored as single events, the homology is 92.5%. Applying to nucleotide changes in the pseudogene the value of 7×10^{-9} substitutions/nt/yr for silent mutations (Perler et al., 1980), we estimate the date of origin of the bovine pseudogene as about 10.7 Myr ago. The bovine pseudogene is thus not expected to be present in other mammals, which diverged from cows earlier than 10.7 Myr ago. Furthermore, it appears to have been accumulating unselected changes for most or all of its existence, based on examining the positions within codons where nt changes have occurred. The changes (7, 8 and 10 changes for codon positions 1, 2 and 3, respectively) are statistically random with respect to position within codons. In addition, the fraction of replacement changes in the pseudogene (0.68) vs. the bovine cDNA is not significantly different ($P < 0.05$) (Freeman and Tukey, 1950) from the value of 0.78 calculated for the pseudogene assuming nucleotide changes were wholly random.

Further support for a relatively recent origin comes from comparison of our bovine liver cDNA sequence with that of a human liver *coxIV* cDNA (Zeviani et al., 1987). The coding region homology over the 384 nt that could be compared is 87%. Since this homology between coding regions is less than that between the bovine pseudogene and cDNA, it supports the above calculation that the bovine pseudogene has appeared since the divergence of cows and man. A direct test would be to see if the pseudogene is present in other members of *Bos*, such as yak, but absent from goat, which split from cow roughly 17 Myr ago (Gentry, 1978).

Only 32 bp of 5'-untranslated sequence are available for determining the number of nucleotide changes in a noncoding region; thus, the time of origin of 17.9 Myr ago estimated from the four differences observed in this region may not be significantly different from 10.7 Myr. These are probably maximum times since the evolution rate of the subunit IV gene suggested by comparison with human subunit IV (Zeviani et al., 1987) would have caused about 10% of the changes observed to have taken place in the expressed gene rather than the pseudogene.

Correlative evidence from the analysis of other pseudogenes and their cognate expressed genes shows that the position of the direct repeat preceding the pseudogene often falls within a few bases of the 5' end of the expressed gene transcript (Rogers, 1985; Weiner et al., 1986). If this correlation applies in the present case, a 5'-untranslated region of 54 or 55 nt is predicted, only 23 nt longer than the 5'-untranslated region of the bovine cDNA clone (Fig. 3). At the 3' end, the direct repeat overlaps the poly(A) tract. The length of the 3'-untranslated region of the pseudogene is virtually the same as that in a human liver subunit IV DNA (Zeviani et al., 1987).

(c) Characterization of a subunit IV gene that may be expressed

Comparison of the restriction maps of λBCO4.5 and λBCO4.2 (Fig. 1) indicated that they contained different genomic regions. Furthermore, λBCO4.5 contained sequences homologous to only the 5' subregion probe of the bovine cDNA, not to the 3' subregion probe. Three lines of evidence suggested that λBCO4.5 contains the 5' region of a *coxIV* expressed gene: (i) Southern analysis; (ii) genomic blot analysis, and (iii) preliminary nucleotide sequence analysis of exons I and II.

Exons were localized by Southern hybridization. Several restriction enzymes, including *Pst*I, *Sma*I, and *Ava*I, cleave the hybridizing region of λBCO4.5 to generate two fragments (Fig. 4), whereas these enzymes do not cleave the 5' end of the cDNA. One exon was localized to the 375-bp 5' *Sst*I-*Ava*I fragment of pBCO4.5A and the second exon to the 700-bp *Xba*I fragment (Fig. 4). The two exons are
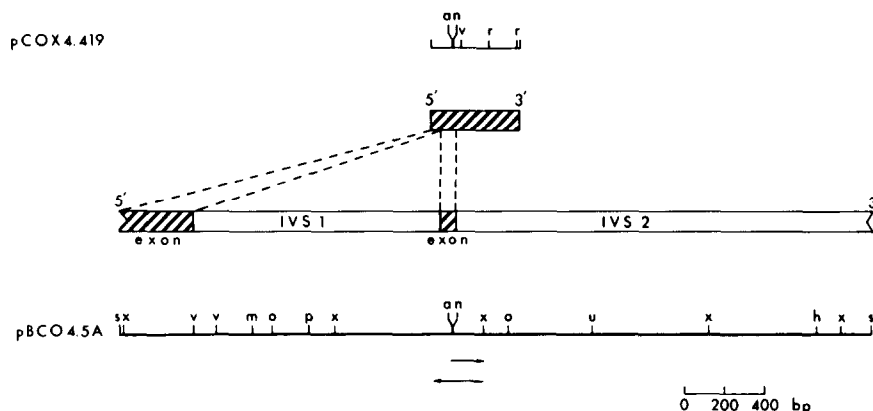
Fig. 4. Restriction map of pBCO4.5A, the 3.7-kb *Sst*I subclone of λBCO4.5. Restriction enzymes: a, *Apa*I; h, *Hin*dIII; m, *Sma*I; n, *Nco*I; o, *Xho*II; p, *Pst*I; r, *Rsa*I; s, *Sst*I; u, *Stu*I; v, *Ava*I; x, *Xba*I. For comparison, the map of the *coxIV* cDNA, pCOX4.419, is also shown. Regions in pBCO4.5A that hybridize to the cDNA probe are indicated by hatching. The size of the more 5' exon has not been precisely determined. The orientation of COX IV-coding sequences determined by nucleotide sequencing is indicated. The strategy for dideoxy sequencing of exon II is shown by arrows below the figure. IVS1 and IVS2 represent introns.

thus separated by at least 1.2 kb not present in the cDNA.

Preliminary sequence data on the 375-bp *Sst*I-*Ava*I fragment (Fig. 4) indicated that exon I encoded the 5'-untranslated region of the *coxIV* gene. Homology with the bovine pseudogene sequence began at nt 166 of pBCO4.2 (4 bp downstream from the end of the first direct repeat) and extended through the 5'-untranslated region to within 1 bp of the initiation methionine codon. Exon II, the first coding exon, thus extends from the A preceding the initiator ATG codon through the presequence to codon 3 of the mature protein (nt 32–106 of the cDNA). The 5' or donor splice junction of both exons I (CAg:GTGAG—) and II (ATG:GTAAGT—) agree well with the consensus sequence —AG:GT(A/G)AGT— (Padgett et al., 1986). (The splice junction in exon I is based on sequencing one strand, and the lower-case g indicates a position of uncertainty, although it agrees with the cDNA and pseudogene sequences.) This positioning of introns in the bovine *coxIV* gene is thus far consistent with the hypothesis of Gilbert (1977) that exons encode different functional domains of proteins.

Southern hybridization indicated that neither λBCO4.2 nor λBCO4.5 suffered rearrangement during construction of the genomic library. For both clones, internal fragments that hybridize to the cDNA are also observed in genomic digests (Fig. 5, A and B). In addition, sites for *Mbo*I, which was utilized to construct the library, are not found in the pseudogene sequence.

(d) Are separate *coxIV* genes expressed in different tissues?

Several observations have suggested that at least some of the nuclear encoded subunits are present as tissue-specific isozymes: (i) enzyme from different tissues of the same organism shows differences in kinetic properties (Merle and Kadenbach, 1982), apparent size (Merle and Kadenbach, 1980a, b; Kadenbach et al., 1982), reactivity with polyclonal antibodies (Kuhn-Nentwig and Kadenbach, 1985), and, for subunit IV, with monoclonal antibodies (Nakagawa et al., 1985); (ii) the N terminus of some subunits (but not subunit IV) is related but different in sequence between liver and heart in both beef and pig (Kadenbach et al., 1983), and (iii) a recently recognized class of fatal infantile mitochondrial myopathies shows the specific absence of cytochrome oxidase activity in muscle, although normal activity is usually present in other tissues examined (Minchom et al., 1983; for a review, see DiMauro et al., 1985).
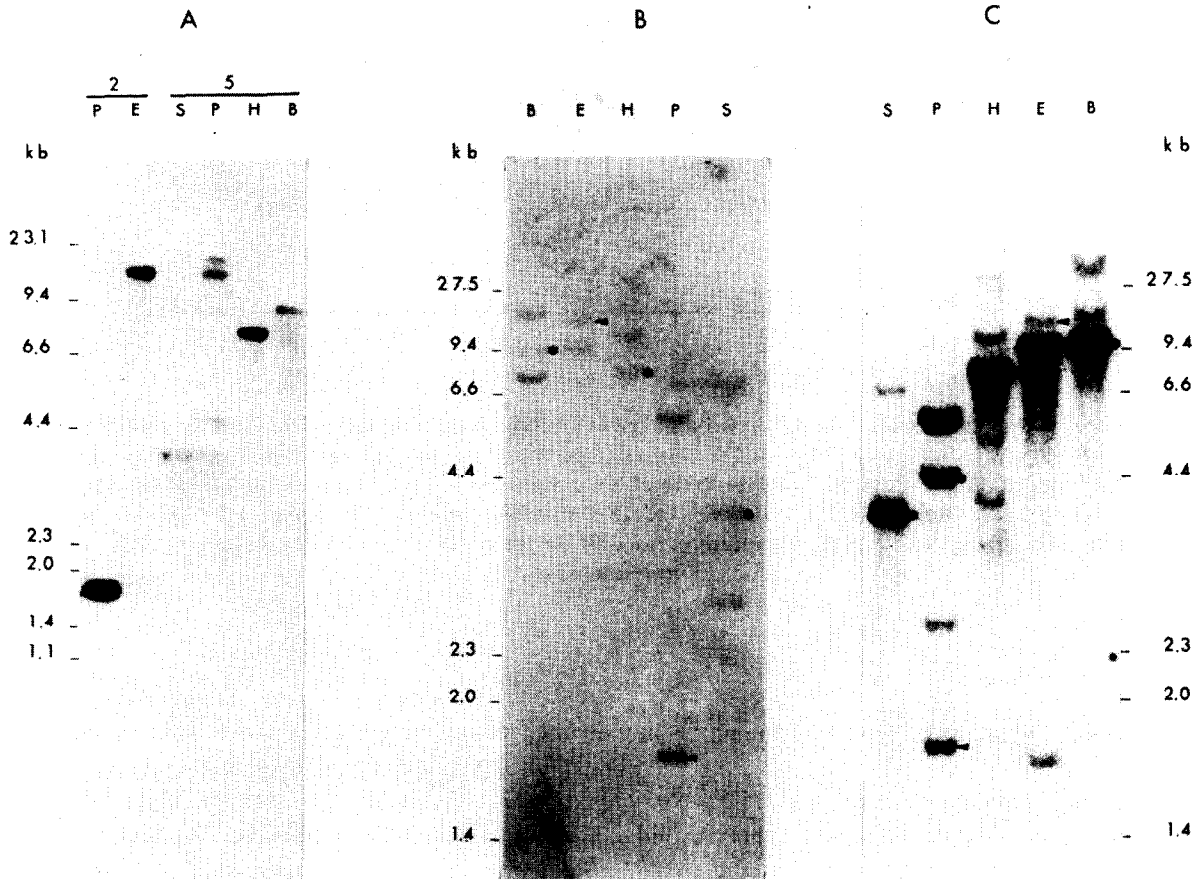
Fig. 5. (Panel A) Southern hybridization of the bovine *coxIV* cDNA insert to *Eco*RI (E) and *Pst*I (P) of λBCO4.2 and *Bgl*II (B), *Hin*dIII (H), *Pst*I (P) and *Sst*I (S) digests of λBCO4.5. The larger *Pst*I-hybridizing fragment from λBCO4.5 contains sequences from the right vector arm and appears as a doublet due to incomplete melting of a small fragment attached at λcos. M_r standards are indicated on the left margin. (Panel B) Southern hybridization of the bovine *coxIV* cDNA insert to total bovine DNA digested with *Bgl*II (B), *Eco*RI (E), *Hin*dIII (H), *Pst*I (P) and *Sst*I (S). Hybridizing genomic DNA fragments identical to those cloned in λBCO4.2 are indicated with arrowheads, those identical to fragments in λBCO4.5 are indicated with asterisks. M_r standards are indicated on the left margin. (Panel C) Southern hybridization of the 5' *Sst*I-*Nco*I fragment derived from pBCO4.5 (see Fig. 4) to total bovine genomic DNA digested with *Bgl*II (B), *Eco*RI (E), *Hin*dIII (H), *Pst*I (P) and *Sst*I (S). Hybridizing fragments identical to those in λBCO4.2 and λBCO4.5 are indicated with arrowheads and asterisks, respectively. M_r standards are indicated on the right margin

Our Southern data, however, were most consistent with the existence of one expressed gene. The cDNA insert (Fig. 4a) and the 5' *Sst*I-*Nco*I fragment (containing exons I and II and intron I) from pBCO4.5A (Fig. 4b) were each used to probe blot transfers of genomic DNA. The subunit IV cDNA hybridized to two to four genomic fragments (Fig. 5B). The 5' *Sst*I-*Nco*I fragment from pBCO4.5A hybridized to two to seven fragments (Fig. 5C). The more intensely hybridizing bands in each lane of Fig. 5C represent

the genomic region containing the 5' end of the gene.

The above interpretation of Southern data is based on the size correlation of hybridizing genomic fragments with fragments present in λBCO4.2 and λBCO4.5 and the expectation that genomic regions colinear with the intron portion of the 5' probe would hybridize more efficiently. The two most informative digests (*Pst*I and *Sst*I) clearly distinguish between genomic fragments arising from the pseudogene and the 5' end of the

expressed gene. For example, the 1.7-kb *Pst*I fragment containing the pseudogene (Fig. 5A) was clearly detected when total genomic DNA was probed with either the cDNA or the 5' end of the expressed gene (Fig. 5, B and C). The two strongly hybridizing *Pst*I bands in panel C must represent the two genomic fragments containing exons I and II. The larger band (5.8 kb) probably contains exon II because the region of homology to the cDNA is sufficiently long (74 bp) to be detected with the cDNA probe (panel B), whereas the region of homology between the cDNA and exon I is only 31 bp. Thus, the 4.3-kb fragment containing exon I was not detected in total genomic DNA (Fig. 5B) probed with the cDNA. Similarly, the 3.7-kb *Sst*I fragment containing the 5' end (exons I and II and intron I) of the expressed gene can be seen in *Sst*I digests of λBCO4.5 and in total genomic DNA probed with either the cDNA or the 5' probe (panels A, B, and C). The 7.4-kb *Sst*I fragment could be assigned to the pseudogene since it hybridized to both probes; the smaller (2.7 and 1.9 kb) *Sst*I fragments must contain additional exons homologous to sequences contained in the cDNA.

Most of the hybridizing fragments in the other digests can be accounted for similarly by the hypothesis that one pseudogene and one expressed gene are present in the bovine genome. Although the presence of a family of highly similar expressed genes is not ruled out by the number of bands, the similar hybridization intensity of expressed and pseudogene bands when probed with the cDNA (Fig. 5B) makes this possibility unlikely. In addition, Northern analysis has shown a single transcript of identical size in human liver, muscle, and HeLa cells (Zeviani et al., 1987).

Mechanisms other than multiple genes could account for many of the tissue-specific differences described above. For instance, the immunological differences may be due to the alteration of antigenic determinants by post-translational modification of the protein, as in isozymes of triosephosphate isomerase (Brown et al., 1985). A second possibility is that alternate pathways exist for the processing of the transcript of a single gene, as for example with mouse α-amylase (Hagenbuchle et al., 1981). The sequence analysis of subunit IV clones from tissue-specific libraries, now in progress, is likely to help select among these possibilities.

REFERENCES

Anderson, S., De Bruijn, M.H.L., Coulson, A.R., Eperon, I.C., Sanger, F. and Young, I.G.: Complete sequence of bovine mitochondrial DNA: conserved features of the mammalian mitochondrial genome. J. Mol. Biol. 156 (1982) 683–717.

Anderson, S., Bankier, A.T., Barrell, B.G., De Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G.: Sequence and organization of the human mitochondrial genome. Nature 290 (1981) 457–465.

Azzi, A.: Cytochrome *c* oxidase: towards a clarification of its structure, interactions and mechanism. Biochim. Biophys. Acta 594 (1980) 231–252.

Baxter-Gabbard, K.L.: A simple method for the large-scale preparation of sucrose gradients. FEBS Lett. 20 (1972) 117–119.

Benton, W.D. and Davis, R.W.: Screening λgt recombinant clones by hybridization to single plaques in situ. Science 196 (1977) 180–182.

Birnboim, H.C. and Doly, J.: A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res. 7 (1979) 1513–1523.

Brown, J.R., Daar, I.O., Krug, J.R. and Maquat, L.E.: Characterization of the functional gene and several processed pseudogenes in the human triosephosphate isomerase gene family. Mol. Cell. Biol. 5, (1985) 1694–1706.

Blin, N. and Stafford, D.W.: A general method for isolation of high molecular weight DNA from eukaryotes. Nucl. Acids Res. 3 (1976) 2303–2308.

Capaldi, R.A., Malatesta, F. and Darley-Usmar, V.M.: Structure of cytochrome *c* oxidase. Biochim. Biophys. Acta 726 (1983) 135–148.

DiMauro, S., Bonilla, E., Zeviani, M., Nakagawa, M. and DeVivo, D.C.: Mitochondrial myopathies. Ann. Neur. 17 (1985) 521–538.

Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Spritz, R.A., DeRiel, J.K., Forget, B.G., Weissman, S.M., Slightom, J.L., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C. and Proudfoot, N.J.: The structure and evolution of the human beta-globin gene family. Cell 21 (1980) 653–668.

El-Gewely, M.R. and Helling, R.B.: Preparative separation of DNA-ethidium bromide complexes by zonal density gradient centrifugation. Anal. Biochem. 102 (1980) 423–428.

Foran, D.R., Johnson, P.J. and Moore, G.P.: Evolution of two actin genes in the sea urchin *Strongylocentrotus franciscanus*. J. Mol. Evol. 22 (1985) 108–116.

Freeman, M.F. and Tukey, J.W.: Transformations related to the angular and the square root. Ann. Math. Stat. 21 (1950) 607–611.

Gentry, A.W.: Bovidae. In Maglio, V.J. and Cooke, H.B.S. (Eds.), Evolution of African Mammals. Harvard University Press, Cambridge, MA, 1978, pp. 540–572.

Gilbert, W.: Introns and exons: playgrounds of evolution. In Axel, R., Maniatis, T. and Fox, C.F. (Eds.), Eukaryotic Gene Regulation, ICN-UCLA Symposia on Molecular and Cellular Biology, Vol. 14. Academic Press, New York, NY, 1977, pp. 1–12.

Gutweniger, H., Bisson, R. and Montecucco, C.: Hydrophobic photolabelling of the yeast cytochrome *c* oxidase subunits in contact with lipids. Biochim. Biophys. Acta 635 (1981) 187–193.

Hagenbuchle, O., Tosi, M., Schibler, U., Bovey, R., Wellauer, P.K. and Young, R.A.: Mouse liver and salivary gland alpha-amylase mRNAs differ only in 5′ nontranslated sequences. Nature 289 (1981) 643–646.

Hudspeth, M.E.S., Ainley, W.M., Shumard, D.S., Butow, R.A. and Grossman, L.I.: Location and structure of the *var1* gene on yeast mitochondrial DNA: nucleotide sequence of the *40.0* allele. Cell 30 (1982) 617–626.

Kadenbach, B., Hartmann, R., Glanville, R. and Buse, G.: Tissue-specific genes code for polypeptide VIa of bovine liver and heart cytochrome *c* oxidase. FEBS Lett. 138 (1982) 236–238.

Kadenbach, B., Ungibauer, M., Jarausch, J., Buge, U. and Kuhn-Nentwig, L.: The complexity of respiratory complexes. Trends Biochem. Sci. 8 (1983) 398–400.

Kuhn-Nentwig, L. and Kadenbach, B.: Isolation and properties of cytochrome *c* oxidase from rat liver and quantification of immunological differences between isozymes from various rat tissues with subunit-specific antisera. Eur. J. Biochem. 149 (1985) 147–158.

Lewin, A.S., Gregor, I., Mason, T.L., Nelson, N. and Schatz, G.: Cytoplasmically made subunits of yeast mitochondrial F1-ATPase and cytochrome *c* oxidase are synthesized as individual precursors, not as polyproteins. Proc. Natl. Acad. Sci. USA 77 (1980) 3998–4002.

Lomax, M.I., Bachman, N.J., Nasoff, M.S., Caruthers, M.H. and Grossman, L.I.: Isolation and characterization of a cDNA clone for bovine cytochrome *c* oxidase subunit IV. Proc. Natl. Acad. Sci. USA 81 (1984) 6295–6299.

Maniatis, T., Fritsch, E.F. and Sambrook, J.: Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982.

Mason, T.L. and Schatz, G.: Cytochrome *c* oxidase from bakers' yeast, II. Site of translation of the protein components. J. Biol. Chem. 248 (1973) 1355–1360.

Maxam, A.M. and Gilbert, W.: Sequencing end-labeled DNA with base-specific chemical cleavages. Meth. Enzymol. 65 (1980) 499–560.

Merle, P. and Kadenbach, B.: The subunit composition of mammalian cytochrome *c* oxidase. Eur. J. Biochem. 105 (1980a) 499–507.

Merle, P. and Kadenbach, B.: On the heterogeneity of vertebrate cytochrome *c* oxidase polypeptide chain composition. Z. Physiol. Chem. 361 (1980b) 1257–1259.

Merle, P. and Kadenbach, B.: Kinetic and structural differences between cytochrome *c* oxidases from beef liver and heart. Eur. J. Biochem. 125 (1982) 239–244.

Mihara, K. and Blobel, G.: The four cytoplasmically made subunits of yeast mitochondrial cytochrome *c* oxidase are synthesized individually and not as a polyprotein. Proc. Natl. Acad. Sci. USA 77 (1980) 4160–4164.

Minchom, P.E., Dormer, R.L., Hughs, I.A., Stansbie, D., Cross, A.R., Hendry, G.A.F., Jones, O.T.G., Johnson, M.A., Sherratt, H.S.A. and Turnbull, D.M.: Fatal infantile mitochondrial myopathy due to cytochrome *c* oxidase deficiency. J. Neurol. Sci. 60 (1983) 453–463.

Moore, G.P.: Slipped-mispairing and the evolution of introns. Trends Biochem. Sci. 8 (1983) 411–414.

Nakagawa, M., Miranda, A.F., Moggio, M., Bonilla, E. and DiMauro, S.: Demonstration of tissue-specific cytochrome *c* oxidase (COX) with monoclonal antibodies. Neurology 35 (1985) 95.

Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S. and Sharp, P.A.: Splicing of messenger RNA precursors. Annu. Rev. Biochem. 55 (1986) 1119–1150.

Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. and Dodgson, J.: The evolution of genes: the chicken preproinsulin gene. Cell 20 (1980) 555–566.

Power, S.D., Lochrie, M.A., Patterson, T.E. and Poyton, R.O.: The nuclear-coded subunits of yeast cytochrome *c* oxidase, II. The amino acid sequence of subunit VIII and a model for its disposition in the inner mitochondrial membrane. J. Biol. Chem. 259 (1984) 6571–6574.

Proudfoot, N.J. and Brownlee, G.G.: 3′ non-coding region sequences in eukaryotic messenger RNA. Nature 263 (1976) 211–214.

Rogers, J.H.: The origin and evolution of retroposons. Int. Rev. Cytol. 93 (1985) 187–279.

Rubin, C.M. and Schmid, C.W.: Pyrimidine-specific chemical reactions useful for DNA sequencing. Nucl. Acids Res. 8 (1980) 4613–4619.

Sanger, F., Nicklen, S. and Coulson, A.R.: DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74 (1977) 5463–5467.

Schmelzer, E. and Heinrich, P.C.: Synthesis of a larger precursor for the subunit IV of rat liver cytochrome *c* oxidase in a cell-free wheat germ system. J. Biol. Chem. 255 (1980) 7503–7506.

Schmelzer, E., Northemann, W., Kadenbach, B. and Heinrich, P.C.: Cell-free synthesis of a larger molecular weight precursor of cytochrome *c* oxidase subunit V from rat liver and

the distribution of its mRNA between free and membrane-bound polysomes. Eur. J. Biochem. 127 (1982) 177–183.

Soares, M.B., Schon, E., Henderson, A., Karathenasis, S.K., Cate, R., Zeitlin, S., Chirgwin, J. and Efstratiadis, A.: RNA-mediated gene duplication: the rat preproinsulin I gene is a functional retroposon. Mol. Cell. Biol. 5 (1985) 2090–2103.

Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E. and Inouye, M.: Frameshift mutations and the genetic code. Cold Spring Harbor Symp. Quant. Biol. 31 (1966) 77–84.

Weiner, A.M., Deininger, P.L. and Efstratiadis, A.: Nonviral retroposons: genes, pseudogenes, and transposable elements

generated by the reverse flow of genetic information. Annu. Rev. Biochem. 55 (1986) 631–661.

Woychik, R.P., Camper, S.A., Lyons, R.H., Horowitz, S., Goodwin, E.C. and Rottman, F.M.: Cloning and nucleotide sequencing of the bovine growth hormone gene. Nucl. Acids Res. 10 (1982) 7197–7210.

Zeviani, M., Nakagawa, M., Herbert, J., Lomax, M.I., Grossman, L.I., Sherbany, A., Miranda, A., DiMauro, S. and Schon, E.: Isolation of a cDNA clone encoding subunit IV of human cytochrome c oxidase. Gene 55 (1987) 205–217.

Communicated by J.L. Slightom.