ATP Synthase from Bovine Mitochondria: Complementary DNA Sequence of the Mitochondrial Import Precursor of the γ-Subunit and the Genomic Sequence of the Mature Protein†

Mark R. Dyer,‡ Nicholas J. Gay,§ Steven J. Powell,‖ and John E. Walker*

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K.

Received September 29, 1988; Revised Manuscript Received December 28, 1988

ABSTRACT: The γ-subunit of mitochondrial ATP synthase is part of the extrinsic membrane sector of the enzyme F_1-ATPase. It is a nuclear gene product. Complementary DNA clones encoding a precursor of the protein have been isolated from a bovine library. The initial partial clone was identified with a mixture of 32 synthetic oligonucleotides designed from the known protein sequence (Walker et al., 1985), and this isolate was then used to screen the library again in order to find a complete cDNA. The DNA sequence of a clone that encodes the entire mature protein has been established, and the deduced protein sequence agrees exactly with that determined by direct sequence analysis of protein isolated from bovine hearts (Walker et al., 1985). At the 3′ ends of two independently isolated clones, alternative polyadenylation sites have been observed; otherwise, the DNA sequences of the clones are concordant. In common with many other mitochondrial proteins encoded in nuclear genes, the deduced protein sequence has an N-terminal extension that is absent from the mature protein. These presequences direct the protein to its appropriate mitochondrial compartment and are removed during the import process. The cDNA clone has been employed to isolate bovine genomic clones containing the gene for the γ-subunit. From them, the DNA sequence has been established of a region encoding the mature protein and six amino acids in the presequence, but not the remainder of the proposed import sequence. This sequence extends over almost 10 kb and is divided into eight exons. Intron B between exons I and II contains a sequence that is related to long interspersed repetitive elements (LINEs) that have been described in other mammals. Human LINEs are usually flanked by directly repeated sequences with a poly(A) tract at their 3′ ends, and these features are present in the bovine LINE which is truncated. This sequence contains an open reading frame encoding part of a protein that is closely related to a protein encoded in mouse LINEs, to reverse transcriptase, and to DNA binding proteins. We have also made a preliminary investigation by DNA hybridization of the number of sequences related to the bovine gene in both the bovine and human genomes. Under the experimental conditions employed, one fragment hybridized in digests of bovine DNA, and two to four bands were detected in digests of human DNA; these latter fragments have originated from either expressed genes or pseudogenes. Thus, it appears that at least in cows some of the subunits of mitochondrial ATP synthase may have single expressed genes whereas other components (the α-subunit and the dicyclohexylcarbodiimide-reactive subunit, for example) present in the same enzyme complex have at least two expressed genes that are regulated differently in various tissues (Gay & Walker, 1985; Walker et al., 1989).

The γ-subunit of mitochondrial ATP synthase is part of F_1-ATPase, the extrinsic membrane sector of the enzyme. It appears to be a globular protein, and its role in the bovine enzyme is unknown. In common with all but 2 of the 13 or so of the proteins that make up the bovine ATP synthase (Fearnley & Walker, 1986; Walker et al., 1987a), it is encoded in nuclear DNA. The corresponding mRNAs are translated on cytoplasmic ribosomes, usually as longer precursors with N-terminal presequences which serve to direct them into the organelle (Schatz & Butow, 1983). Investigations of bovine cDNAs for subunits of ATP synthase have shown that at least two of the proteins are each encoded by more than one expressed gene. For example, the bovine dicyclohexylcarbodiimide-reactive proteolipid subunit, a membrane component of ATP synthase, has two different expressed genes, known as P1 and P2 (Gay & Walker, 1985). The P1 and P2 genes code for precursors of the proteolipid that differ in their N-terminal import sequences, but removal of the presequence produces an identical mature polypeptide. The two genes appear to be expressed in different ratios in various bovine tissues (Gay & Walker, 1985). In cows and also in humans, the P1 and P2 genes are part of large multigene families which also contain spliced pseudogenes (Dyer et al., 1989; Dyer & Walker, 1989). The α-subunit of bovine mitochondrial ATP synthase forms part of the F_1 assembly and also has at least two different expressed genes (Walker et al., 1989). It appears that one gene is expressed in heart (Walker et al., 1985, 1989) and the other in liver (Breen, 1988). Other components of ATP synthase may have multiple genes. Hybridization experiments on restriction digests of human and bovine genomic DNA with probes for the d- and b-subunits, and for the oligomycin sensitivity conferral protein, detected several hybridizing bands (Walker et al., 1987c,b). However, these experiments detect

† The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J02850.

* To whom correspondence should be addressed.

‡ Supported by an MRC Research Studentship.

§ Present address: University Biochemical Laboratory, Tennis Court Rd., Cambridge, U.K. Supported by an MRC Postdoctoral Research Training Fellowship.

‖ Present address: ICI Pharmaceuticals Division, Northwich, Cheshire, U.K.

0006-2960/89/0428-3670$01.50/0

both expressed genes and pseudogenes. In contrast, similar experiments suggest that the bovine and human ATPase inhibitor proteins and F6 subunits have single-copy genes (Walker et al., 1987b).

As discussed in this paper, the γ-subunit of ATP synthase also apparently has a single gene in the bovine genome, while in humans it may have multiple genes and (or) pseudogenes. Complementary DNA clones have been isolated for this subunit from a bovine library, and their DNA sequences have been determined. The bovine cDNA has been employed in hybridization experiments with restriction digests of bovine and human DNA. One band was detected in the former and two to four in the latter. They have also been used to isolate clones containing a segment of the bovine genome covering the part of the gene that codes for the mature γ-subunit. This region has been sequenced and is divided into eight exons that are distributed over almost 10 kb of DNA.

Materials and Methods

Screening the cDNA Library. In order to identify a full-length cDNA clone encoding the γ-subunit of bovine ATP synthase, the cDNA library derived from heart and liver RNA (Gay & Walker, 1985) was rescreened with a "prime-cut" probe (Farrell et al., 1983) prepared from the original cDNA isolate (Walker et al., 1985). This partial clone encompasses amino acids 149–272 of the mature γ-protein and probably also extends through to the 3′ end of the mRNA up to the poly(A) tract. The probe employed was an *Eco*RI fragment and contains bases 576–960 of the completed sequence (see Figure 2), the *Eco*RI site at the 5′ end of the probe being in the polylinker of the vector. The inserts in positively hybridizing recombinants were released by digestion with the restriction enzymes *Eco*RI and *Bam*HI, and the digestion products were analyzed by electrophoresis in 0.6% agarose gels. Restriction fragments were recovered from gels and cloned into either M13mp8 or M13mp9, and the DNA sequences were determined in regions adjacent to the cloning sites. In the case of clone pBov-γ1, a *Bam*HI fragment of approximately 1.0 kb was subdigested with *Hin*fI. The ends of the resulting fragments of DNA were repaired, and the products were cloned into the *Sma*I site of M13mp8.

Preparation and Screening of a Bovine Genomic Library. The preparation of bovine liver DNA has been described previously (Walker et al., 1987b). A phage library of partial *Sau*3AI fragments of bovine genomic DNA was made in the vector λ2001 (Karn et al., 1984; Dyer et al., 1989). Approximately 10^6 recombinants were screened for clones containing the gene for the γ-subunit of mitochondrial ATP synthase using procedures for hybridization and screening the library described by Dyer and Walker (1989). A "prime-cut" hybridization probe containing nucleotides 576–815 of the cDNA was employed in these experiments. The radiolabeled probe was released from the M13 template DNA by digestion with the restriction enzyme *Eco*RI.

DNA Sequence Analysis. DNA sequences were determined by the dideoxy chain termination method (Sanger et al., 1977) as modified by Biggin et al. (1983). All sequences were determined minimally at least once in both senses of the DNA. Three regions of sequence were completed by the use of unique synthetic oligonucleotide primers, 17 bases in length, made with the aid of an Applied Biosystems 380B automated oligonucleotide synthesizer. The DNA sequence was "compressed" in three sections, but these problems were resolved by the substitution of deoxyinosine triphosphate for deoxyguanosine triphosphate in the sequencing reaction mixtures (Mills & Kramer, 1979). Further details of procedures employed for sequencing DNA and for compilation and analysis of DNA sequences are described in Dyer and Walker (1989).

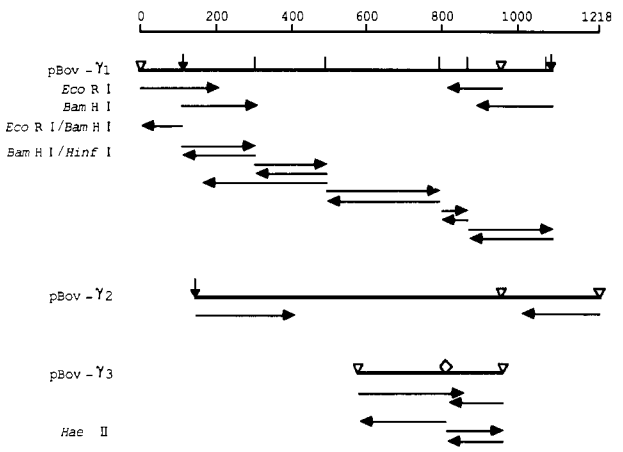


FIGURE 1: Sequence analysis of cDNA clones encoding a possible precursor of the γ-subunit of ATP synthase from bovine mitochondria. The thick lines represent the inserted DNAs in the isolates pBov-γ1-2 and the *Eco*RI fragment characterized from isolate pBov-γ3. The *Eco*RI site at the 5′ end of this fragment is in the polylinker of the vector. The inserts in pBov-γ1 and pBov-γ2 are flanked by an *Eco*RI site (▽) and a *Bam*HI site (↓) in the polylinker of the vector. The horizontal arrows represent the extent and show the sense of the sequences that were determined. The positions of *Hin*fI sites that were important in the sequencing of pBov-γ1 are indicated by vertical bars, and a *Hae*II site (◇) was used to generate fragments in the sequencing of pBov-γ3. The scale is in bases.

Hybridization with Human and Bovine Genomic DNA. This was performed as described previously (Walker et al., 1987b) by the method of Southern (1976). Samples of human and bovine DNA (20 μg/digest) were digested with the restriction enzymes *Bam*HI, *Eco*RI, *Hin*dIII, *Kpn*I, *Xba*I, *Sac*I, and *Pst*I (all purchased from New England Biolabs). The digests were fractionated by electrophoresis in 0.6% agarose gels, transferred to nitrocellulose filters, and hybridized with radioactively labeled "prime-cut" probes (Farrell et al., 1983). These probes were nucleotides 114–309 and 487–790, respectively, from the bovine cDNA. They were released from their M13 vectors by digestion with *Eco*RI. Conditions for prehybridization, hybridization, washing, and autoradiography can be found in Walker et al. (1987b).

Results and Discussion

Complementary DNA Cloning and Sequence Analysis of the Bovine γ-Subunit. In work described previously (Walker et al., 1985), a partial cDNA for the γ-subunit of bovine ATP synthase was isolated with a mixed oligonucleotide probe, and from this clone, the sequence of an *Eco*RI fragment encoding the C-terminal region of the protein (residues 149–272) was determined (see clone pBov-γ3 in Figure 1). The insert in this clone must also have contained an *Eco*RI–*Bam*HI fragment probably extending the sequence in a 3′ direction, but this fragment was not observed in restriction digests of the clone, presumably because it was too small to be visible by ethidium bromide staining after agarose gel electrophoresis. So, in order to isolate a full-length cDNA clone for the γ-subunit, the bovine cDNA library was rescreened with the *Eco*RI fragment from pBov-γ3. Two positively hybridizing recombinants, pBov-γ1 and pBov-γ2, were isolated. The inserts in the cloning vector that was employed to make the library are flanked by an *Eco*RI site and a *Bam*HI site. Upon digestion with *Eco*RI alone, a fragment of about 1 kb was released from pBov-γ1; *Bam*HI alone also released a fragment of about the same size, whereas digestion of the plasmid with

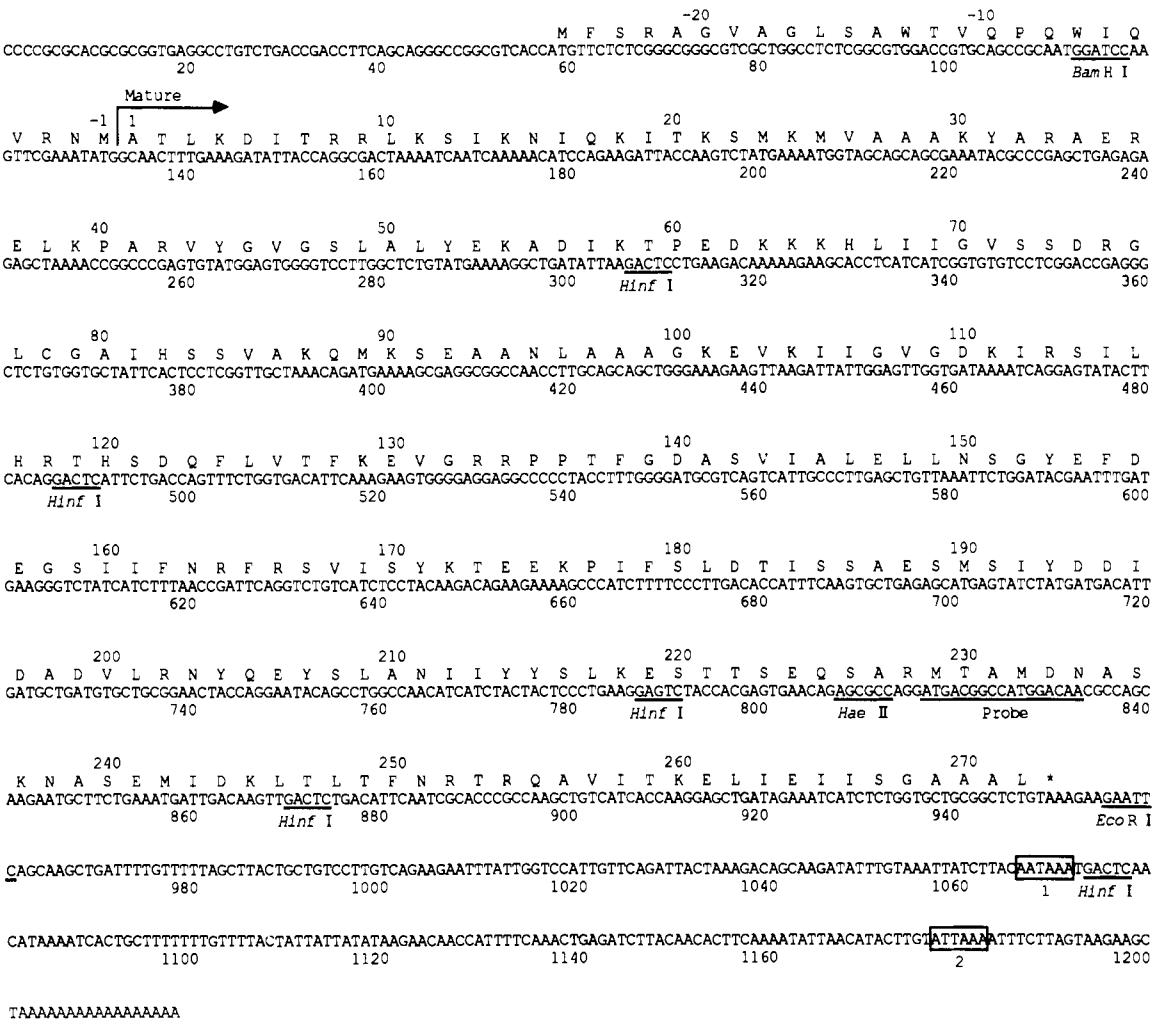


FIGURE 2: DNA sequence of cDNAs encoding a possible precursor of the γ-subunit of ATP synthase from bovine mitochondria. Amino acid 1 is the N-terminal alanine residue of the mature protein. The possible mitochondrial import presequence is numbered from −25 to −1. Various restriction endonuclease sites that were used in cloning and sequencing experiments are indicated. The nucleotide sequence that hybridized with the synthetic oligonucleotide probe in the isolation of clone pBov-γ3 is underlined, and boxes have been place around the two polyadenylation signals. Clone pBov-γ1 extends from nucleotide 1 to 1083 and is followed by A_{13}, clone pBov-γ2 covers bases 142–1218, and the *Eco*RI fragment in pBov-γ3 corresponds to bases 576–962.

both enzymes together produced *Bam*HI–*Eco*RI fragments of about 0.8 and 0.1 kb. These various restriction fragments were cloned into appropriate M13 vectors, and flanking sequences were determined, and from these data, the arrangement of restriction fragments shown in Figure 1 could be deduced. The sequence of the inserted DNA in this recombinant was completed by sequence analysis of *Hinf*I fragments derived from the *Bam*HI fragment. These experiments together were sufficient to produce a complete sequence for the recombinant DNA in both of its senses, with the exception of one overlap through the *Hinf*I site at bases 486–490; this is provided by the known protein sequence of the γ-subunit (Walker et al., 1985).

Digestion of isolate pBov-γ2 with *Eco*RI and *Bam*HI released an *Eco*RI fragment estimated to be about 0.4 kb and an *Eco*RI–*Bam*HI fragment with an estimated size of about 0.8 kb. Sequence analysis at the ends of these fragments showed that the insert in pBov-γ2 did not extend as far in a 5′ direction as that in pBov-γ1 but that it extended further in a 3′ direction.

The sequence of the bovine cDNA presented in Figure 2 is 1218 nucleotides long. It is a composite of the sequences of pBov-γ1 and pBov-γ2. They finish at bases 1083 and 1218 with runs of A_{13} and A_{17} respectively, at their 3′ ends, both of these 3′ sequences presumably originating from the 3′ poly(A) tails in mRNAs. The poly(A) tracts are preceded by the sequences AATAAA and ATTAAA, respectively, both of which can serve as signals for polyadenylation of transcripts (Proudfoot & Brownlee, 1976). Alternative polyadenylation sites have been noted with transcripts of the heart isoform of the α-subunit of ATP synthase (Walker et al., 1989), and also in the case of the mitochondrial phosphate carrier (Runswick et al., 1987). Over the regions of the overlapping clones, including clone pBov-γ3, the DNA sequences that have been determined are entirely in agreement with each other. The sequence of bases 1083–1218 in clone pBov-γ2 has been determined only in one sense of the DNA, but the corresponding sequence has been determined in both senses in the genomic DNA (see below), and the cDNA and genomic sequences are in total agreement.

Deduced Protein Sequence. The protein sequence for the mature γ-subunit of bovine ATP synthase derived from the cDNA sequence agrees exactly with that described previously (Walker et al., 1985). This earlier sequence was determined almost entirely by direct sequence analysis of protein isolated from bovine heart mitochondria, with the exception that the overlap between amino acids 164–166 was established by DNA sequence analysis of the clone pBov-γ3. The 5′ end of the cDNA sequence suggests that the N-terminal alanine residue of the mature protein could be preceded by a processed mi-

tochondrial import sequence of 25 amino acids. This proposed precursor sequence has a number of characteristics that have been found to be associated with mitochondrial import sequences; is has a net basic charge and contains no acidic residues, and an arginine residue is found in the precursor close to the site where cleavage would occur to produce the mature form of the protein (von Heijne, 1986). The length of the precursor is defined by the proposed initiator methionine codon. This is preceded by an in-phase potential termination codon, TGA, at bases 28–30, and no other ATG triplets are present in the intervening sequence. Yet, despite these characteristics, some doubt surrounds the authenticity of this proposed import sequence. The N-terminal alanine residue of the mature protein is immediately preceded by a methionine residue, which in the absence of other evidence could also be considered as a potential translational initiator, and would thereby produce a mitochondrial protein with no processed import sequence. This is uncommon, but at least two other example are known. They are the ADP/ATP translocase from several species [see Cozens et al. (1989)] and the d-subunit of bovine ATP synthase (Walker et al., 1987). A more serious worry, however, is our failure so far to isolate a bovine genomic clone that extends through this region. Characterized clones have a 5′ end at the *Bam*HI site at bases 113–118 of the bovine cDNA sequence, and none has been identified that extends in a 5′ direction beyond this site, which also encompasses a *Sau*3AI site within it. It is possible that this region contains an unusually high concentration of *Sau*3AI sites that would lead to its absence from genomic libraries that have been generated from partial *Sau*3AI fragments such as the one we have constructed (see below).

Gene Cloning and DNA Sequence of the Gene for the Mature γ-Subunit. The bovine genomic library was screened with a hybridization probe containing a segment of the coding sequence from the bovine cDNA for the γ-subunit. Stable duplexes were formed between the probe and with DNA from two different recombinants, named λγA and λγB. These recombinants were rescreened and grown in liquid culture, and their DNA was isolated. Southern hybridization experiments performed on the phage DNA indicated that they both contained identical inserts of bovine genomic DNA.

DNA from recombinant λγA was digested with the restriction enzymes *Sac*I and *Nco*I, fractionated by electrophoresis in a 0.6% agarose gel, and hybridized to a "prime-cut" probe containing nucleotides 576–815 of the bovine cDNA. The *Sac*I site at the 5′ end of this fragment was contained in the polylinker of λ2001, and the genomic sequence begins at an adjacent *Sau*3AI site. This probe hybridized with a 5.9-kb *Sac*I–*Nco*I fragment. It was thought likely that the end of the fragment generated by *Nco*I corresponded to the site for this enzyme that had been shown to be present in the cDNA, and so codons for the 41 C-terminal amino acids of the protein were not expected to be present in the fragment. Its DNA sequence was determined by the random strategy using cloned fragments of sonicated DNA (Bankier & Barrell, 1983). It contained codons for six amino acids of the presumptive mitochondrial import presequence and sequence coding for amino acids 1–231 of the mature protein. As anticipated, the 3′ end of the fragment did correspond to the *Nco*I site in the cDNA, and so in order to obtain the 3′ region of the gene, it was necessary to identify and sequence an overlapping DNA fragment from the insert of λγA. An *Nci*I site was present in the sequence about 450 bp from the 3′ end of the existing sequence at this point. However, digestion of the insert in λγA with this enzyme alone produced an overlapping restriction fragment of about 8 kb, and it was possible that this extended far beyond the 3′ end of the gene. Therefore, a smaller *Nci*I–*Eco*RI fragment (3 kb) was chosen for further sequence analysis, although it was evident that this was likely to extend no further than the *Eco*RI site detected in the 3′ region of the bovine cDNA. Its DNA sequence was determined by the random strategy, and it was found to contain the missing region which codes for the C-terminal end of the protein. However, as anticipated, the DNA sequence of the gene still lacked sequence for the 3′ untranslated region found in the mRNA. In order to obtain a genomic DNA fragment to extend this end of the gene, a *Xba*I digest of DNA from λγA was hybridized with a segment of the cDNA containing the 3′ untranslated region of the bovine mRNA (nucleotides 956–1218 in Figure 2), and 4.3-kb *Xba*I fragment formed a stable duplex with the probe. It was sequenced completely by the random approach and contains the sequence for the 3′ untranslated region of the mRNA. Only part of the sequence of this DNA fragment from bases 8223–10000 is presented in Figure 3.

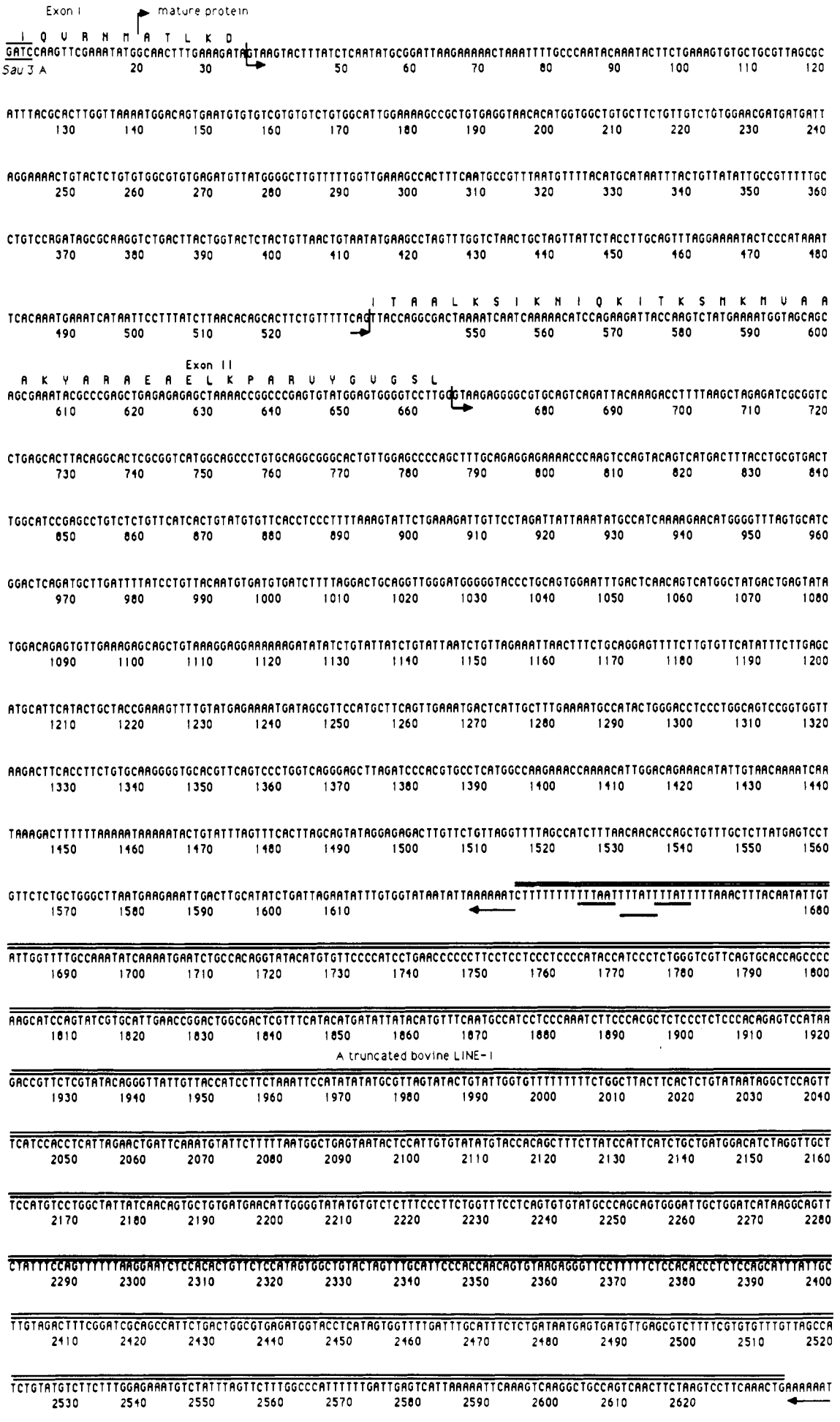
The inserted bovine genomic DNA in λγA did not contain the 5′ region of the bovine gene, and so the library was rescreened with a hybridization probe containing nucleotides 1–171 from the bovine gene. A single recombinant, known as λγC, was identified, but Southern hybridization experiments indicated that its inserted DNA was identical with those found in λγA and λγB. In a further attempt to clone this region of the bovine gene, the genomic library was screened with a probe derived from the 5′ region of the cDNA (nucleotides 1–114 in Figure 2), but no positively hybridizing recombinant was identified. No further efforts have been made to clone the 5′ region of the bovine gene.

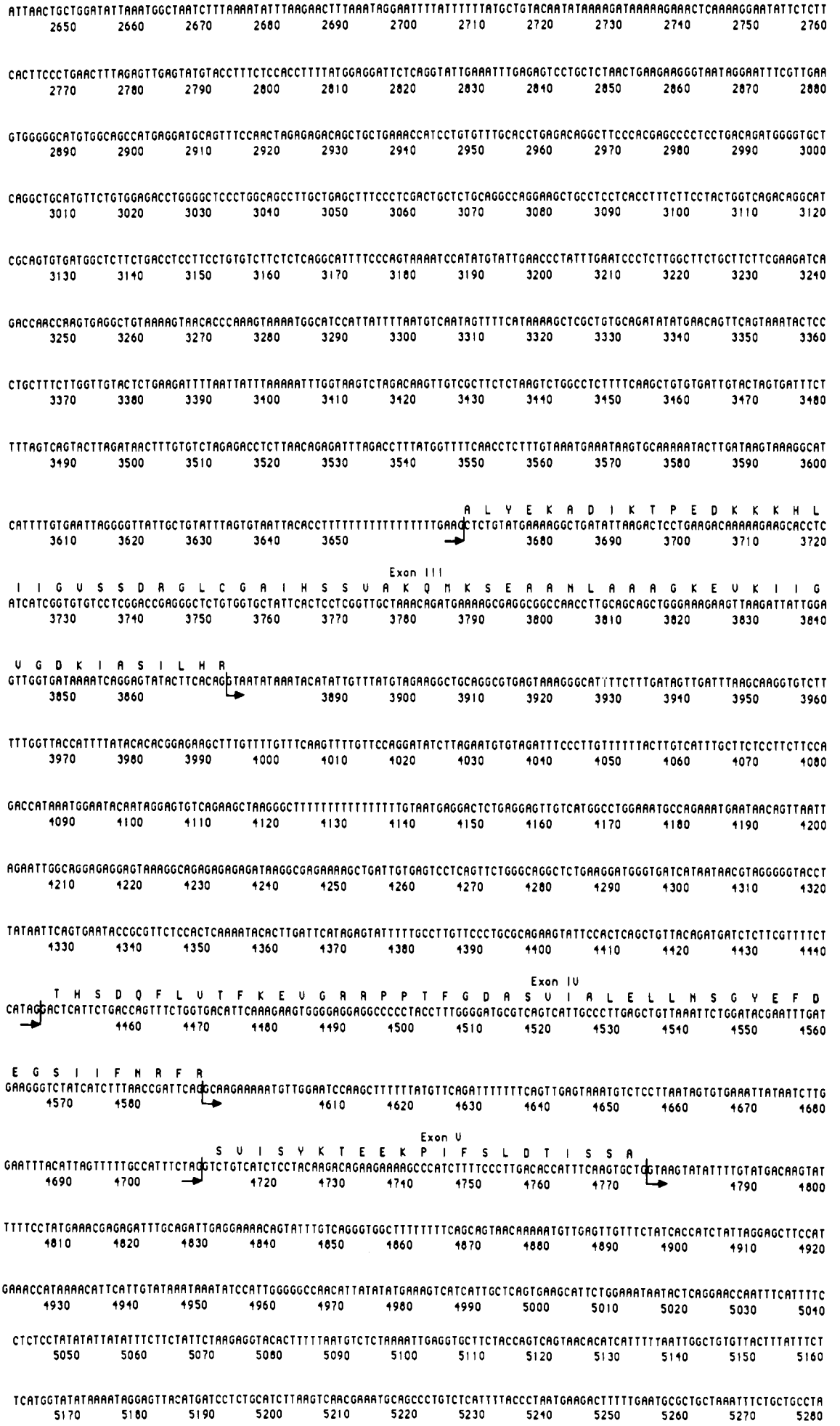
Each nucleotide in the sequence of the gene for the bovine γ-subunit of ATP synthase was determined at least 6 times on average and at least once on each strand of the DNA. The G + C content of the 10-kb fragment presented in Figure 3 is 40%, in reasonable agreement with the estimated G + C content of 42% for the bovine genome (Chargaff & Lipshitz, 1953).

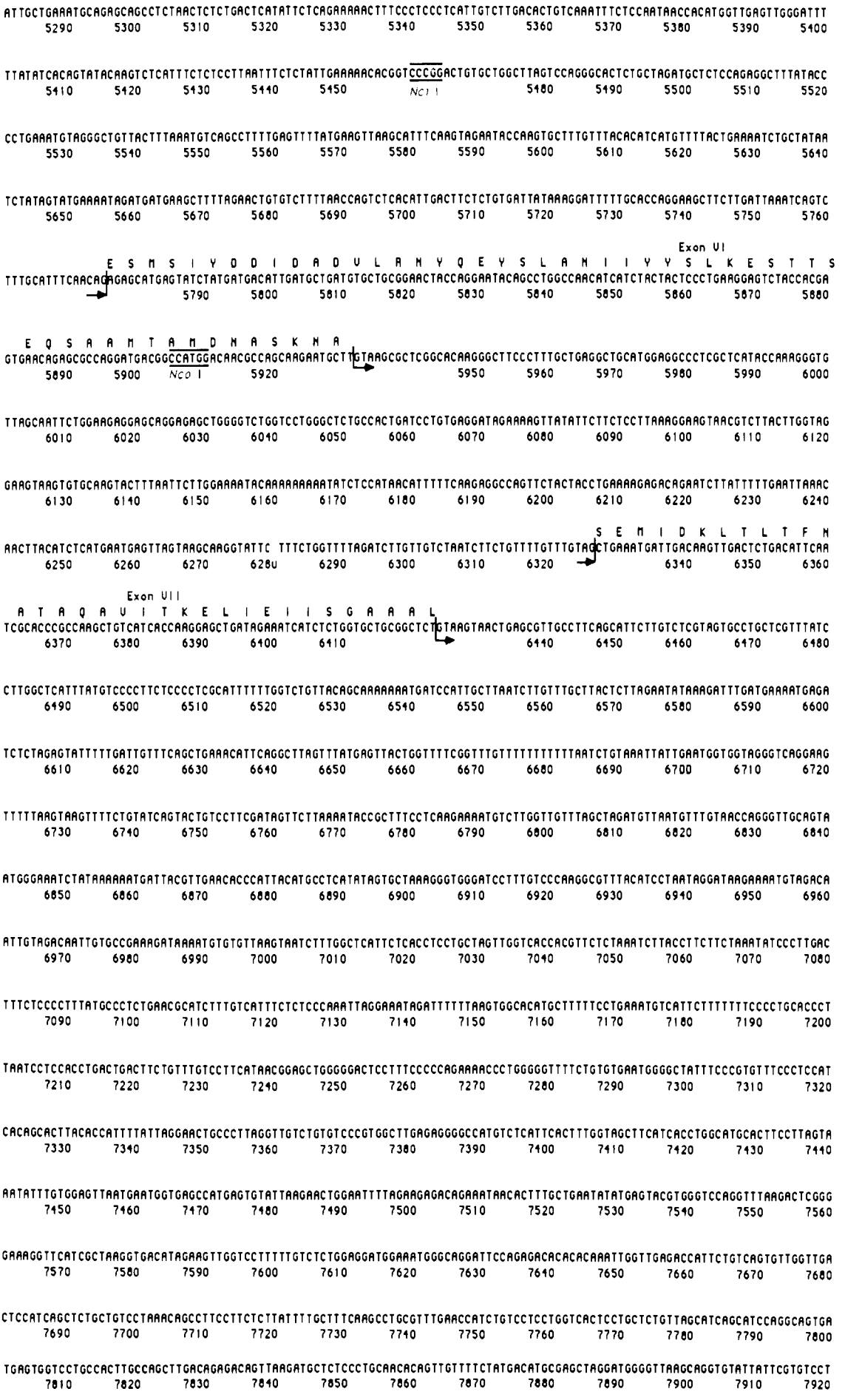
Gene Structure. (i) Identification of Exons. The exons of the bovine gene for the γ-subunit of ATP synthase were identified by comparison of the sequence with that of the bovine cDNA. Also taken into consideration when assigning the borders of individual exons were the rules for intron–exon splice sites (Breathnach & Chambon, 1981). These predict conservation of the dinucleotides GT and AG next to the 5′ and 3′ boundaries of introns, respectively.

(ii) Exon–Intron Structure. The gene for the γ-subunit of bovine ATP synthase is split into at least eight exons (see Figure 4). Since the sequence of the gene is incomplete at its 5′ end, it is possible that separate exons contain the 5′ untranslated region of the mRNA and the coding sequence for a presumptive presequence. The first exon in the DNA sequence for the bovine gene, called exon I, encodes six amino acids of a possible import peptide and the five amino acids at the N-terminus of the mature protein. The remainder of the coding sequence for the mature polypeptide is found in exons II–VII, and an eighth exon contains the sequence for the 3′ untranslated region found in the message.

With one exception, all of the introns in the bovine gene contain the canonical dinucleotides GT and AG at their 5′ and 3′ splice sites, respectively (Breathnach & Chambon, 1981; see Table I for a summary of the properties of the introns). An unusual GC sequence is found at the 5′ end of intron D. This dinucleotide has been found to replace the consensus


Exon I
mature protein
Sau 3 A
Exon II
A truncated bovine LINE-1


Exon III
Exon IV
Exon V


ATTGCTGAAAATGCAGAGCAGCCTCTAACTCTCTGACTCATATTCTCAGAAAAACTTTCCCTCCCTCATTGTCTTGACACTGTCAAATTTCTCCAATAACCACATGGTTGAGTTGGGATTT
5290 5300 5310 5320 5330 5340 5350 5360 5370 5380 5390 5400
TTATATCACAGTATACAAGTCTCATTTCTCTCCTTAATTTCTCTATTGAAAAACACGGTCCCGGACTGTGCTGGCTTAGTCCAGGGCACTCTGCTAGATGCTCTCCAGAGGCTTTATACC
5410 5420 5430 5440 5450 NCI I 5480 5490 5500 5510 5520
CCTGAAATGTAGGGCTGTTACTTTAAATGTCAGCCTTTTGAGTTTTATGAAGTTAAGCATTTCAAGTAGAATACCAAGTGCTTTGTTTACACATCATGTTTTACTGAAAATCTGCTATAA
5530 5540 5550 5560 5570 5580 5590 5600 5610 5620 5630 5640
TCTATAGTATGAAAATAGATGATGAAGCTTTTAGAACTGTGTCTTTTAACCAGTCTCACATTGACTTCTCTGTGATTATAAAGGATTTTTGCACCAGGAAGCTTCTTGATTAAATCAGTC
5650 5660 5670 5680 5690 5700 5710 5720 5730 5740 5750 5760
Exon VI
E S M S I Y Q D I D A D V L A M Y Q E Y S L A M I I Y Y S L K E S T T S
TTTGCATTTCAACAGAGGAGCATGAGTATCTATGATGACATTGATGCTGATGTGCTGCGGAACTACCAGGAATACAGCCTGGCCAACATCATCTACTACTCCCTGAAGGAGTCTACCACGA
5790 5800 5810 5820 5830 5840 5850 5860 5870 5880
E Q S A A M T A M D N A S K M A
GTGAACAGAGCGCCAGGATGACGGCCATGGACAACGCCAGCAAGAATGCTTGTAAGCGCTCGGCACAAGGGCTTCCCTTTGCTGAGGCTGCATGGAGGCCCTCGCTCATACCAAAGGGTG
5890 5900 NCO I 5920 5950 5960 5970 5980 5990 6000
TTAGCAATTCTGGAAGAGGAGCAGGAGAGCTGGGGTCTGGTCCTGGGCTCTGCCACTGATCCTGTGAGGATAGAAAAGTTATATTCTTCTCCTTAAAGGAAGTAACGTCTTACTTGGTAG
6010 6020 6030 6040 6050 6060 6070 6080 6090 6100 6110 6120
GAAGTAAGTGTGCAAGTACTTTAATTCTTGGAAAATACAAAAAAAAATATCTCCATAACATTTTTCAAGAGGCCAGTTCTACTACCTGAAAAGAGACAGAATCTTATTTTTGAATTAAAC
6130 6140 6150 6160 6170 6180 6190 6200 6210 6220 6230 6240
S E M I D K L T L T F M
AACTTACATCTCATGAATGAGTTAGTAAGCAAGGTATTC TTTCTGGTTTTAGATCTTGTTGTCTAATCTTCTGTTTTGTTTGTACCTGAAATGATTGACAAGTTGACTCTGACATTCAA
6250 6260 6270 6280 6290 6300 6310 6320 6340 6350 6360
Exon VII
A T A Q A V I T K E L I E I I S G A A A L
TCGCACCCGCCAAGCTGTCATCACCAAGGAGCTGATAGAAATCATCTCTGGTGCTGCGGCTCTGTAAGTAACTGAGCGTTGCCTTCAGCATTCTTGTCTCGTAGTGCCTGCTCGTTTATC
6370 6380 6390 6400 6410 6440 6450 6460 6470 6480
CTTGGCTCATTTATGTCCCCTTCTCCCCTCGCATTTTTTGGTCTGTTACAGCAAAAAAATGATCCATTGCTTAATCTTGTTTGCTTACTCTTAGAATATAAAGATTTGATGAAAATGAGA
6490 6500 6510 6520 6530 6540 6550 6560 6570 6580 6590 6600
TCTCTAGAGTATTTTTGATTGTTTCAGCTGAAACATTCAGGCTTAGTTTATGAGTTACTGGTTTTCGGTTTGTTTTTTTTTTTAATCTGTAAATTATTGAATGGTGGTAGGGTCAGGAAG
6610 6620 6630 6640 6650 6660 6670 6680 6690 6700 6710 6720
TTTTTAAGTAAGTTTTCTGTATCAGTACTGTCCTTCGATAGTTCTTAAAATACCGCTTTCCTCAAGAAAATGTCTTGGTTGTTTAGCTAGATGTTAATGTTTGTAACCAGGGTTGCAGTA
6730 6740 6750 6760 6770 6780 6790 6800 6810 6820 6830 6840
ATGGGAAATCTATAAAAAATGATTACGTTGAACACCCATTACATGCCTCATATAGTGCTAAAGGGTGGGATCCTTTGTCCCAAGGCGTTTACATCCTAATAGGATAAGAAAATGTAGACA
6850 6860 6870 6880 6890 6900 6910 6920 6930 6940 6950 6960
ATTGTAGACAATTGTGCCGAAAGATAAAATGTGTGTTAAGTAATCTTTGGCTCATTCTCACCTCCTGCTAGTTGGTCACCACGTTCTCTAAATCTTACCTTCTTCTAAATATCCCTTGAC
6970 6980 6990 7000 7010 7020 7030 7040 7050 7060 7070 7080
TTTCTCCCCTTTATGCCCTCTGAACGCATCTTTGTCATTTCTCTCCCAAATTAGGAAATAGATTTTTTAAGTGGCACATGCTTTTTCCTGAAATGTCATTCTTTTTTTCCCCCTGCACCCT
7090 7100 7110 7120 7130 7140 7150 7160 7170 7180 7190 7200
TAATCCTCCACCTGACTGACTTCTGTTTGTCCTTCATAACGGAGCTGGGGGACTCCTTTCCCCCAGAAAACCCTGGGGGTTTTCTGTGTGAATGGGGCTATTTCCCGTGTTTCCCTCCAT
7210 7220 7230 7240 7250 7260 7270 7280 7290 7300 7310 7320
CACAGCACTTACACCATTTTATTAGGAACTGCCCTTAGGTTGTCTGTGTCCCGTGGCTTGAGAGGGGCCATGTCTCATTCACTTTGGTAGCTTCATCACCTGGCATGCACTTCCTTAGTA
7330 7340 7350 7360 7370 7380 7390 7400 7410 7420 7430 7440
AATATTTGTGGAGTTAATGAATGGTGAGCCATGAGTGTATTAAGAACTGGAATTTTAGAAGAGACAGAAATAACACTTTGCTGAATATATGAGTACGTGGGTCCAGGTTTAAGACTCGGG
7450 7460 7470 7480 7490 7500 7510 7520 7530 7540 7550 7560
GAAAGGTTCATCGCTAAGGTGACATAGAAGTTGGTCCTTTTTGTCTCTGGAGGATGGAAATGGGCAGGATTCCAGAGACACACACAAATTGGTTGAGACCATTCTGTCAGTGTTGGTTGA
7570 7580 7590 7600 7610 7620 7630 7640 7650 7660 7670 7680
CTCCATCAGCTCTGCTGTCCTAAACAGCCTTCCTTCTCTTATTTTGCTTTCAAGCCTGCGTTTGAACCATCTGTCCTCCTGGTCACTCCTGCTCTGTTAGCATCAGCATCCAGGCAGTGA
7690 7700 7710 7720 7730 7740 7750 7760 7770 7780 7790 7800
TGAGTGGTCCTGCCACTTGCCAGCTTGACAGAGACAGTTAAGATGCTCTCCCTGCAACACAGTTGTTTTCTATGACATGCGAGCTAGGATGGGGTTAAGCAGGTGTATTATTCGTGTCCT
7810 7820 7830 7840 7850 7860 7870 7880 7890 7900 7910 7920

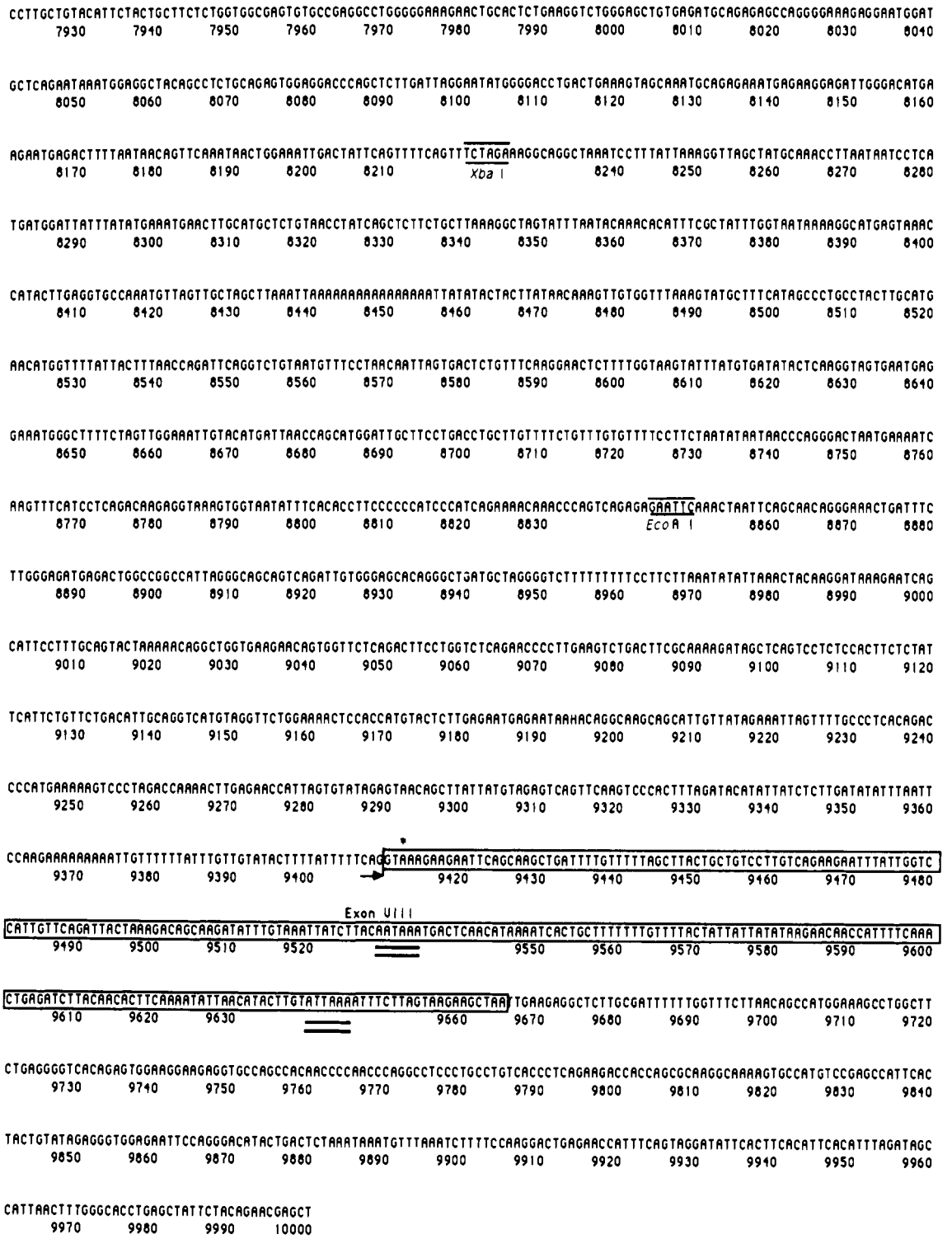


FIGURE 3: DNA sequence of a segment of bovine genomic DNA containing the coding sequence for the mature γ-subunit of ATP synthase. The nucleotide sequence is numbered from 1 to 10000. The restriction sites used in cloning the three overlapping DNA fragments are shown. Intron–exon boundaries are denoted by small arrows, and exon VIII which contains the 3′ noncoding region present in the mRNA is boxed. The doubly overlined sequence is a truncated bovine long interspersed repetitive element of the LINE-1 class (Hutchison et al., 1989), and arrows denote direct repeated sequences that flank it. The singly underlined sequences indicate potential polyadenylation signals near to the 3′ end of the LINE. Sequences that have been underlined twice are potential poly(A) addition signals for the mRNA of the γ-subunit.

sequence at 5′ splice sites in introns of other genes from animals and plants; see, for example, the chicken and duck α-globin genes (Dodgson & Engel, 1983; Erbil & Niessing, 1983) and the gene for nodulin-24 from soyabean (Katinakis & Verma, 1985). Moreover, the GC dinucleotide results in correct splicing out of introns in vitro albeit at a lower efficiency (Aebi et al., 1986, 1987).

A Long Interspersed Repeated Sequence in the Bovine Gene. The DNA sequence containing the gene for the γ-subunit of bovine ATP synthase was analyzed for the presence of repetitive elements. As part of this analysis, the nucleotide sequence was compared with those of the two classes of bovine *Alu* type repeats (Watanabe et al., 1982; Duncan, 1987) using the computer program DBCOMP (Staden, 1982), but neither class of repeat was detected. The sequence was also examined for the presence of multiple copy repetitive elements by comparison of the sequence with itself using the computer program DIAGON (Staden, 1982), but again no repeated sequences were

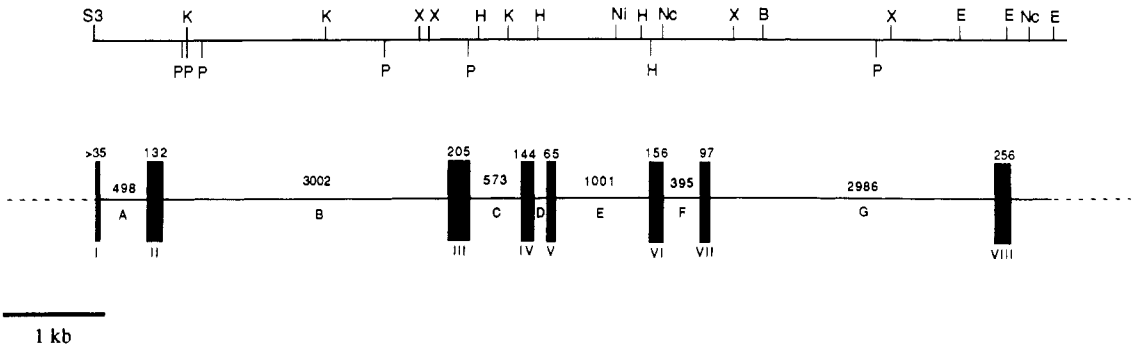


FIGURE 4: Structure of the region of the bovine gene encoding the mature γ-subunit of mitochondrial ATP synthase. Exons I–VIII and introns A–G are denoted by black boxes and solid lines, respectively. The sizes of exons and introns are given in base pairs. The extent of exon I is unknown at present. Above the gene structure are shown the sites for the various restriction enzymes that were employed either in cloning or in Southern blotting experiments (see Figure 6). B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; Nc, *Nco*I; Ni, *Nci*I; P, *Pst*I; S3, *Sau*3AI; X, *Xba*I.

Table I: Introns in the Bovine Gene for the γ-Subunit of ATP Synthase

Intron	Size	Class	Sequence	
			5' boundary	3' boundary
A	498	1	gat.aGTAAGTACTT D	TGTTTTTCAGtt.acc I T
B	3002	1	ttg.gGTAAGAGGGG L	TTTTTTGAAGct.ctg A. L
C	573	2	cac.agGTAATATAAA H R	TTTCTCATAG.g.act T
D	120	2	ttc.agGCAAGAAAAA F R	CCATTTCTAGg.tct S
E	1001	1	gct.gGTAAGTATAT A	ATTTCAACAGag.agc E S
F	395	1	gct.tGTAAGCGCTC A	TTGTTTGTAGct.gaa S A
G	2986	2	gct.ctGTAAGTAACT A L	TATTTTTCAGg.taa *
Consensus sequence			cagGTRAGT	YYYYYYYYYNCAGg

[a] The asterisk indicates a termination codon.

found. Then both strands of the DNA sequence were compared with all of the entries in the EMBL DNA sequence library. This revealed that a nucleotide sequence of about 600 bp from intron B shares a 70% identity with the 3′ region of long interspersed elements (LINEs) which have been described in genomes of other mammals (Hutchison et al., 1989), and the significance of the relationship between the two sequences was confirmed with DIAGON. In mice, for example, there are about 10^5 copies of the LINE-1 element in the haploid genome. Some are full-length sequences of about 7 kb, and others are partial sequences which are truncated by different amounts at their 5′ ends relative to the full-length element. The DNA sequences of full-length LINE-1 elements from mice (Loeb et al., 1986) and humans (Hattori et al., 1985) have similar structural features. Often they are flanked by short direct DNA repeats of 5–15 bp, and they contain a 5′ untranslated region, two open reading frames of about 1 and 4 kb, respectively, a 3′ untranslated region, and at their 3′ ends a poly(A) tract. The flanking direct repeats and the poly(A) sequence suggest that these LINES are retroposons (Hutchison et al., 1989).

The sequence detected in intron B of the gene for the γ-subunit of bovine ATP synthase is a LINE-1 which has been truncated at its 5′ end, and it is the first reported example of this type of repetitive element in the bovine genome. The bovine LINE-1 contains the 3′ region of the large open reading frame. This sequence was translated and was found to encode

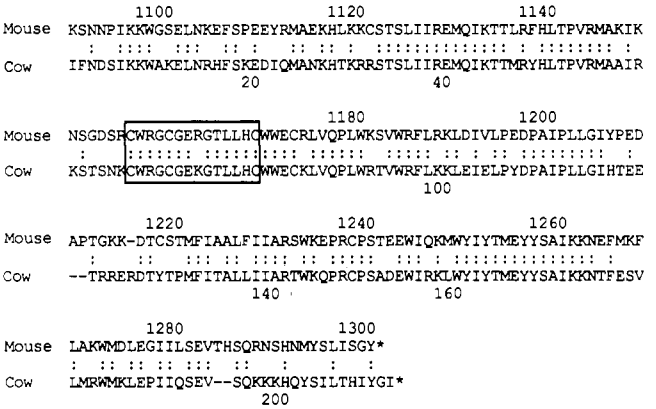


FIGURE 5: Alignment of polypeptides encoded by LINEs-1 in cows and mice. The mouse protein sequence is the C-terminal region (amino acids 1090–1300) predicted by ORF-2 in the repetitive element L1Md-A2 (Loeb et al., 1986). The computer program PRTALN (Wilbur & Lipman, 1983) was employed in making the alignment. Colons denote identities, and dashes have been introduced to improve the alignment. The asterisks represent termination codons. The boxed amino acid sequences have been suggested to resemble the amino acid motifs found in DNA binding proteins (Fanning & Singer, 1987).

a polypeptide of 210 amino acids which is related to the C-terminal region of a polypeptide encoded in the larger open reading frame of mouse LINEs-1 (see Figure 5). This latter putative polypeptide contains amino acid motifs which are

found in reverse transcriptase and DNA binding proteins (Hattori et al., 1986; Fanning & Singer, 1987), and the bovine LINE-1 protein also contains an amino acid sequence similar to a conserved sequence found in some DNA binding proteins (see Figure 5).

The 3′ untranslated region of the bovine LINE-1 sequence is not conserved, a finding which is consistent with DNA sequence studies LINE-1 members from humans and mice (Hutchison et al., 1989). In the bovine LINE-1, the 3′ untranslated region is apparently 311 bp in length and contains three potential polyadenylation signals close to its 3′ end. These sequences start 3, 9, and 14 bp unstream from the run of A residues and so are nearer than is usual to the poly(A) tail in the transcript. However, DNA sequencing studies of cDNAs for LINEs-1 in humans indicate that some potential poly(A) signals are found within 10 bp of the poly(A) tails (DiGiovanni et al., 1983). The bovine LINE-1 sequence is flanked by perfect 7 bp direct repeats (ATTTTTT). Pyrimidine-rich repeats have also been found to flank members of the mouse LINE-1 family (Hutchison et al., 1989).

Number of Bovine and Human Genes for the γ-Subunit of ATP Synthase. In an attempt to investigate the number of sequences related to the coding region for the bovine γ-subunit in both the bovine and human genomes, digests of nuclear DNA have been hybridized with segments of sequence taken from the coding region of the bovine cDNA. Under the experimental conditions used, one hybridizing band was detected in each of the digests of bovine DNA, but two or four bands hybridized with the probe in the digests of human DNA (see Figure 6). These results are consistent with the presence in the bovine genome of a single copy of the gene. Two possible explanations can be advanced to account for the more complex patterns of hybridizing bands that were detected in digests of human DNA. One explanation is that the hybridizing restriction fragments arise from the same gene, since the probe employed in the experiment encompasses bovine exons IV and V and part of exon VI, and it is conceivable that more than one site for a particular restriction enzyme is present in the corresponding regions of the human gene. An alternative explanation is that the patterns of hybridization arise from a small family of genes (maybe including pseudogenes).

It now appears from hybridization experiments that some components of mammalian mitochondrial ATP synthase possibly have single genes, whereas others have been shown by cloning and sequencing experiments as well as by hybridization studies to have more than one expressed gene. In the former category, in addition to the bovine γ-subunit, are the F6 and inhibitor subunits (Walker et al., 1987b), and in the latter category are the dicyclohexylcarbodiimide-reactive proteolipid (Gay & Walker, 1985; Dyer & Walker, 1989; Dyer et al., 1989), the α-subunit (Walker et al., 1989), and possibly also the oligomycin sensitivity conferral protein, subunit b, and subunit d (Walker et al., 1987b,c). The phenomenon of multiple genes subject to different levels of expression in various tissues applies also to other mitochondrial proteins. The transport protein, ADP/ATP translocase, has at least two expressed bovine genes, T1 and T2 (Walker et al., 1987a; Powell et al., 1989); T1 is expressed predominantly in heart muscle and T2 in smooth muscle. At least three related genes for this protein have been detected in humans; T1 and T2 (Cozens et al., 1989) are expressed in a variety of tissues and cell lines (Houldsworth & Attardi, 1987; Neckleman et al., 1987), and expression of a third related gene has been demonstrated in a cell line derived from fibroblasts that have been growth stimulated (Battini et al., 1987). In contrast, another

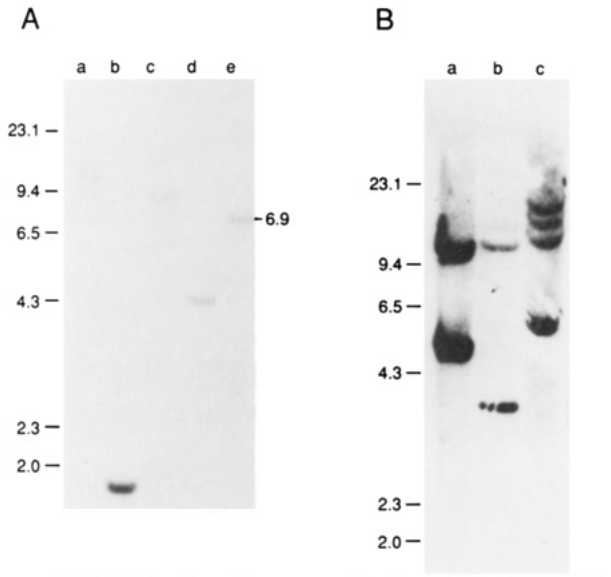


FIGURE 6: Hybridization of human and bovine DNA with cDNA probes for the gene for the γ-subunit of bovine ATP synthase. Bovine liver DNA and human placental DNA (20 μg/digest) were digested with restriction endonucleases. The fragments were fractionated by electrophoresis through a 0.6% agarose gel and were then hybridized on nitrocellulose filters to "prime-cut" probes. In part A, bovine DNA was digested with *Xba*I (lane a), *Pst*1 (lane b), *Kpn*I (lane c), *Hin*dIII (lane d), and *Bam*HI (lane e) and hybridized with a probe containing nucleotides 114–309 of the bovine cDNA (a *Bam*HI–*Hin*fI fragment). This probe corresponds to exons I and II and 26 nucleotides at the 5′ end of exon III of the bovine gene. In part B, human DNA was digested with *Sac*I (lane a), *Hin*dIII (lane b), and *Eco*RI (lane c) and hybridized to a probe containing nucleotides 487–790 of the bovine cDNA (a *Hin*fI fragment). This probe corresponds to exons IV and V and part of exon VI. The filters were washed in 0.2 × SSC at 65 °C and autoradiographed at −70 °C for 72 h. In part A (lane e), a *Bam*HI fragment of 6.9 kb is indicated; a fragment of the same size is found in the DNA sequence for the bovine gene. The positions of markers (*Hin*dIII fragments of bacteriophage λ) are indicated along the left side of each gel. The 6.9-kb *Bam*HI fragment indicated in lane e of panel A is also present in *Bam*HI digests of the recombinant λγA.

related mitochondrial transport protein, the phosphate carrier, appears to have a single bovine gene (J. E. Walker et al., unpublished results). Estimates of the number of genes for a particular protein by hybridization experiments have to be viewed with some caution for several reasons; first, the number of hybridizing bands that are detected depends upon the experimental conditions employed. For example, on the basis of such experiments, Breen et al. (1988) have proposed that there is a single bovine gene for the β-subunit of ATP synthase, whereas we observe more than one related sequences, and we have characterized a related bovine pseudogene (J. E. Walker et al., unpublished observations). Hybridization experiments would also fail to detect sequences coding for homologous proteins in cases where the sequences have diverged extensively during evolution. Another deficiency of this approach is that is cannot readily distinguish between expressed genes and pseudogenes. There appears to be no reliable substitute for cloning and DNA sequencing.

These difficulties notwithstanding, the picture that seems to be emerging is that some, but probably not all, mammalian mitochondrial proteins have more than one expressed gene. Those with essential mitochondrial function and only one gene are presumably expressed in all respiring tissues, and so are examples of "housekeeping" genes; others, for reasons that remain to be uncovered, have multiple genes that appear to be subject to different controls in various tissues.

ACKNOWLEDGMENTS

We thank C. A. Hutchinson, III, for helpful discussion of LINEs and M. J. Runswick for the preparation of Figures 1 and 2.

Registry No. cDNA (bovine ATP synthase γ-subunit), 119618-92-7; γ-subunit, 119618-94-9; mature γ-subunit, 99638-99-0; ATP synthase, 37205-63-3.

REFERENCES

Bankier, A., & Barrell, B. G. (1983) in *Techniques in Nucleic Acid Chemistry* (Flavell, R. A., Ed.) Vol. B508, pp 1–34, Elsevier, New York.

Battini, R., Ferrari, S., Kaczmarek, L., Calabrette, B., Chen, S. T., & Baserga, R. (1987) *J. Biol. Chem. 262*, 4355–4359.

Biggin, M. D., Gibson, T. J., & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3963–3965.

Breathnach, R., & Chambon, P. (1981) *Annu. Rev. Biochem. 50*, 349–383.

Breen, G. A. M. (1988) *Biochem. Biophys. Res. Commun. 152*, 264–269.

Breen, G. A. M., Holman, P. L., & Garnett, K. E. (1988) *Biochemistry 27*, 3955–3961.

Chargaff, E., & Lipshitz, R. (1953) *J. Am. Chem. Soc. 75*, 3658–3661.

Cozens, A. L., Runswick, M. J., & Walker, J. E. (1989) *J. Mol. Biol.* (in press).

Di Giovanni, L., Haynes, S. R., Misra, R., & Jelinek, W. R. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 6533–6537.

Dodgson, J. B., & Engel, J. D. (1983) *J. Biol. Chem. 258*, 4623–4629.

Duncan, C. H. (1987) *Nucleic Acids Res. 15*, 1340.

Dyer, M. R., & Walker, J. E. (1989) *J. Mol. Biol.* (in press).

Dyer, M. R., Gay, N. J., & Walker, J. E. (1989) *Biochem. J.* (in press).

Erbil, C., & Niessing, J. (1983) *EMBO J. 2*, 1339–1343.

Fanning, T. G., & Singer, M. F. (1987) *Biochim. Biophys. Acta 910*, 203–212.

Farrell, P. J., Deininger, P. L., Bankier, A., & Barrell, B. G. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 1565–1569.

Fearnley, I. M., & Walker, J. E. (1986) *EMBO J. 5*, 2003–2008.

Gay, N. J., & Walker, J. E. (1985) *EMBO J. 4*, 3519–3524.

Hattori, M., Hidaka, S., & Sakaki, Y. (1985) *Nucleic Acids Res. 13*, 7813–7827.

Hattori, M., Kuhara, S., Takenaka, O., & Sakaki, Y. (1986) *Nature (London) 321*, 625–628.

Houldsworth, J., & Attardi, G. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 377–381.

Hutchison, C. A., III, Hardies, S. C., Loeb, D. D., Sheshee, W. R., & Edgell, M. H. (1989) in *Mobile DNA* (Berg, D. E., & Howe, M. H., Eds.) A.S.M. Press, Washington, DC (in press).

Karn, J., Matthes, H. W. D., Gait, M. J., & Brenner, S. (1984) *Gene 32*, 217–224.

Katinakis, P., & Verma, D. P. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 4157–4161.

Loeb, D. D., Padgett, R. W., Hardies, S. C., Shehee, W. R., Edgell, M. H., & Hutchison, C. A., III (1986) *Mol. Cell. Biol. 6*, 168–182.

Mills, D. R., & Kramer, F. R. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 2232–2235.

Neckelmann, N., Li, K., Wade, R. P., Shuster, R., & Wallace, D. C. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 7580–7584.

Powell, S. J., Medd, S. M., Runswick, M. J., & Walker, J. E. (1989) *Biochemistry 28*, 866–873.

Proudfoot, N. J., & Brownlee, G. G. (1976) *Nature (London) 263*, 211–214.

Runswick, M. J., Powell, S. J., Nyren, P., & Walker, J. E. (1987) *EMBO J. 6*, 1367–1373.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.

Schatz, G., & Butow, R. A. (1983) *Cell 32*, 316–318.

Southern, E. M. (1975) *J. Mol. Biol. 98*, 503–517.

Staden, R. (1982) *Nucleic Acids Res. 10*, 2951–2961.

von Heijne, G. (1986) *EMBO J. 5*, 1335–1342.

Walker, J. E., Fearnley, I. M., Gay, N. J., Gibson, B. W., Northrop, F. D., Powell, S. J., Runswick, M. J., Saraste, M., & Tybulewicz, V. L. J. (1985) *J. Mol. Biol. 184*, 677–701.

Walker, J. E., Cozens, A. L., Dyer, M. R., Fearnley, I. M., Powell, S. J., & Runswick, M. J. (1987a) *Chem. Scr. 27B*, 97–105.

Walker, J. E., Gay, N. J., Powell, S. J., Kostina, M., & Dyer, M. R. (1987b) *Biochemistry 26*, 8613–8619.

Walker, J. E., Runswick, M. J., & Poulter, L. (1987c) *J. Mol. Biol. 197*, 89–100.

Walker, J. E., Powell, S. J., Viñas, O., & Runswick, M. J. (1989) *Biochemistry* (in press).

Watanabe, Y., Tsukada, T., Notake, M., Nakanishi, S., & Numa, S. (1982) *Nucleic Acids Res. 10*, 1459–1469.

Wilbur, W. J., & Lipman, D. J. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 726–730.