Sequences of members of the human gene family for the c subunit of mitochondrial ATP synthase

Mark R. DYER* and John E. WALKER†
Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K.

Subunit c is an intrinsic membrane component of ATP synthase, and in mammals it is encoded by two expressed nuclear genes, P1 and P2. Both genes encode the same mature c subunit, but the mitochondrial import pre-sequences in the precursors of subunit c are different. The DNA sequences of the human P1 and P2 genes are described. They occupy about 3.0 and 10.9 kb respectively of the human genome, and both genes are split into five exons. The human genome also contains about 14 related spliced pseudogenes, and the sequence of one such pseudogene related to P2 is described. Sequences flanking the 5′ ends of the human P1 and P2 coding sequences each contain a CpG-rich island. Potential promoter elements (TATA and CCAAT boxes) are present in the 5′ sequences of the P1 gene, but not that of P2, although there is no direct experimental evidence to show the involvement of these sequences in transcription of the genes.

INTRODUCTION

Bovine mitochondrial ATP synthase is a membrane-bound complex of 14 different polypeptides (Walker et al., 1991), and the c subunit (also known as the dicyclohexylcarbodi-imide-reactive proteolipid) is an essential part of the proton channel in the membrane sector (Sebald and Hoppe, 1981). It is a hydrophobic protein of 75 amino acids, probably folded into a hairpin of two transmembrane α-helices linked by a β-turn near the membrane surface. A carboxyl group essential for functioning of the proton channel, and the site of reaction of dicyclohexylcarbodi-imide, is situated near to the middle of the C-terminal α-helix. In mammals, *Neurospora crassa* (Jackl and Sebald, 1975) and *Aspergillus nidulans* (Turner et al., 1979), but not in *Saccharomyces cerevisiae* (Macino and Tzagoloff, 1979), subunit c is a nuclear gene product, synthesized on cytoplasmic ribosomes as a precursor with an N-terminal extension. The extension directs the protein into the mitochondrion and is cleaved during import. However, the proteolipid is highly unusual, if not unique, amongst nuclear-encoded mitochondrial proteins in having two different precursors derived from separate genes (Gay and Walker, 1985). Both cDNAs for the precursors contain a segment coding for the same mature proteolipid, but the N-terminal presequences, although related, differ extensively. The 3′ non-coding regions of their cDNAs are only weakly related and so each can be employed as a specific hybridization probe (Gay and Walker, 1985). As described here, we have isolated and sequenced the human P1 and P2 genes. They are members of a complex gene family that includes numerous spliced pseudogenes for P2, and probably for P1 also. The expressed P1 gene is distributed over about 3.0 kb of DNA and the human P2 gene occupies about 10.9 kb of the genome. Both contain four introns at equivalent positions. Interest in this gene family has been increased by the recent finding that in the fatal human disease ceroid lipofuscinosis, or Batten's disease, subunit c accumulates in lysosomes (Palmer et al., 1992).

MATERIALS AND METHODS

DNA hybridization

Digests of human DNA prepared from a placenta (Walker et al., 1987) were fractionated by electrophoresis in 0.6% agarose gels, and fragments were transferred to nitrocellulose filters (Southern, 1975). The filters were incubated at 65 °C, first for 1 h in a solution containing 6 × SSC (1 × SSC is 0.15 M NaCl and 0.015 M trisodium citrate), 0.2% BSA (fraction V), 0.2% polyvinylpyrrolidone, 0.5% *N*-laurylsarcosine and sonicated salmon testis DNA (100 mg/ml), and then for 15–20 h in the presence of radioactive 'prime-cut' probes (Farrell et al., 1983) dissolved with 10% dextran sulphate in the same solution. The filters were washed four times for 30 min each at 65 °C in either 0.2 or 2 × SSC, each containing 0.5% *N*-laurylsarcosine. Autoradiographs of filters were exposed with an intensifying screen at −70 °C for either 1–7 days (genomic DNA) or 1–3 h (phage DNA).

Screening of genomic libraries

The human genomic libraries SH, AT5 (LeFranc et al., 1986) and RPMI (Forster et al., 1987), consisting of partial *Sau*3A fragments cloned into the *Bam*HI site of λ2001 (Karn et al., 1984), were gifts from Dr. T. H. Rabbitts. Plaques (approx. 5×10^5) were produced on *Escherichia coli* Q358 grown on 20 cm diameter agar plates. Phage were transferred sequentially to two nitrocellulose filters per plate, and each library was screened (Benton and Davis, 1977) with the two prime-cut probes. DNA was prepared from recombinant phages (Maniatis et al., 1982) grown in 500 ml cultures of *E. coli* Q358.

Sub-cloning and DNA sequencing

A 4.4 kb *Bam*HI fragment containing part of the human P1 gene was excised from λHP1.9, sonicated and sub-cloned into the

* Present address: Sandoz Ltd., CH-4002 Basle, Switzerland.
† To whom correspondence should be addressed.
The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under accession nos. X69907, X69908 and X69909.

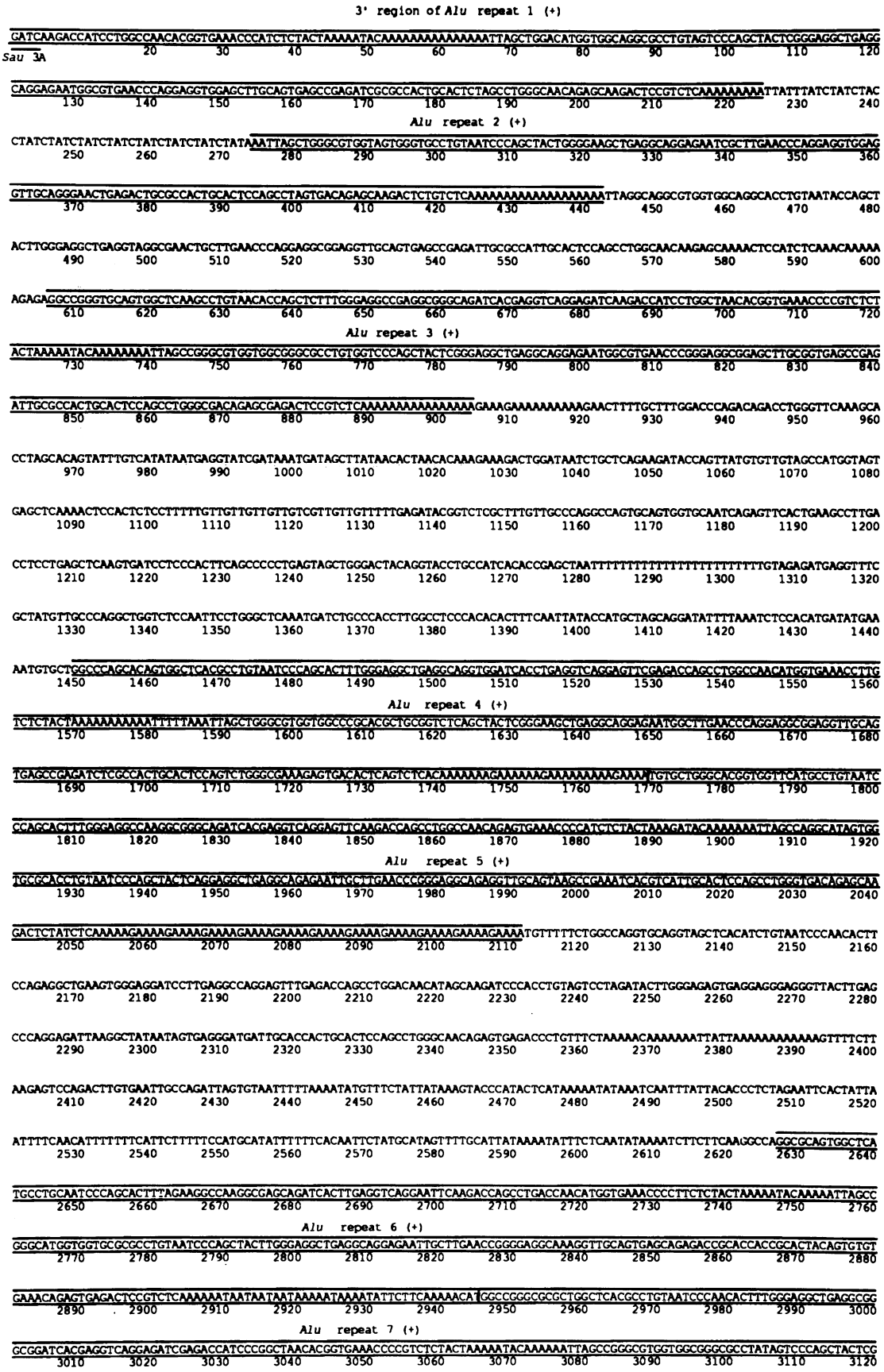

Figure 1 **For legend see page 55.**

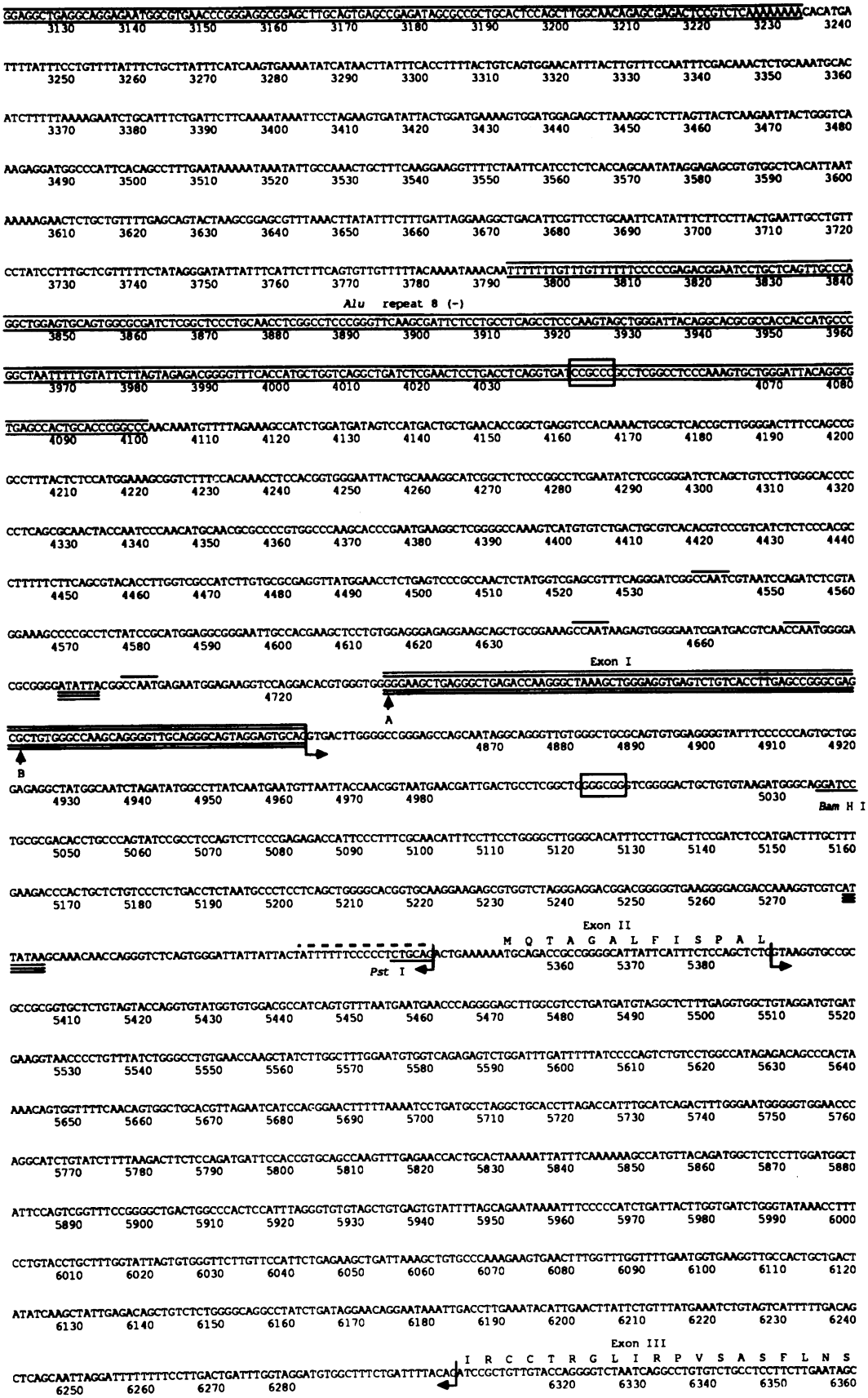

Figure 1 **For legend see page 55.**

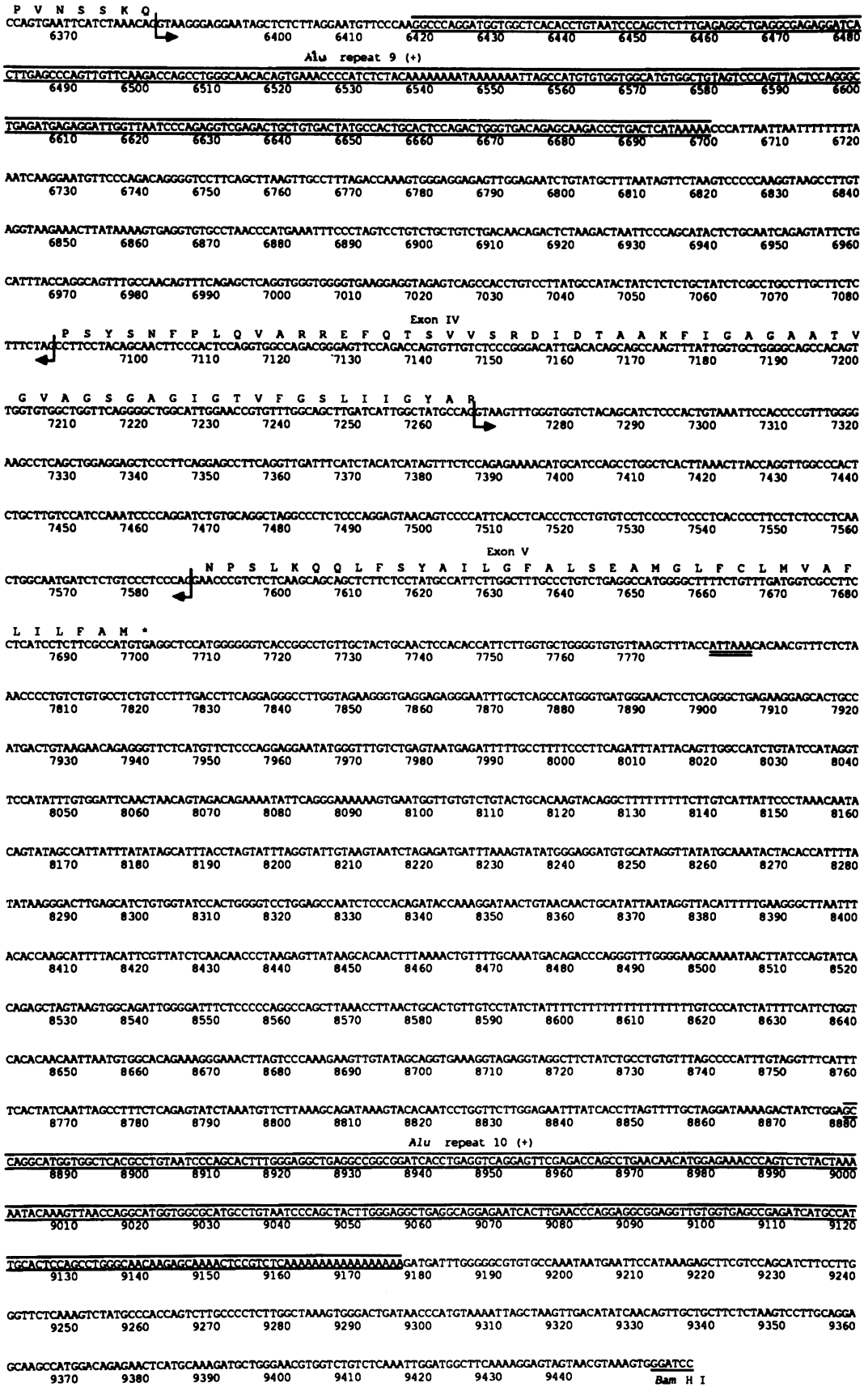

Figure 1 For legend see opposite.

*Sma*I site of M13mp8 (Deininger, 1983). A 5.3 kb *Xho*I–*Pst*I fragment extended the sequence in a 5′ direction. Similarly, a 12.2 kb *Sac*I fragment in λAT5P2.1 appeared to contain at least a substantial part of the human P2 gene. It was amplified in pUC12 (Messing, 1983), sonicated and sub-cloned into the *Sma*I site of M13mp8 (Deininger, 1983). Subsequently the sequence was extended beyond the 5′ end of this fragment by sequencing an overlapping 3.8 kb *Pst*I fragment.

DNA sequences were determined at least once in both senses of the DNA, and on average five and six times in P1 and P2 respectively, by the modified dideoxy chain termination method (Sanger et al., 1977; Biggin et al., 1983). Problematic sequences were resolved by substituting either deoxyinosine triphosphate (Mills and Kramer, 1979) or deoxy-7-deazaguanosine triphosphate (Mizusawa et al., 1986) for dGTP in the sequencing reactions. Data were compiled with programs DBAUTO and DBUTIL (Staden, 1982) and analysed with ANALYSEQ (Staden, 1985). Sequences were aligned with programs NUCALN and PRTALN (Wilbur and Lipman, 1983).

RESULTS AND DISCUSSION

Characterization of the genes

Attempts to clone the human P1 and P2 genes were hampered by the presence in the genome of numerous related spliced pseudogenes. In consequence, almost all recombinants identified by screening of genomic libraries contained spliced pseudogenes. A similar obstacle had been encountered previously in cloning the bovine genes, and no recombinant containing the expressed P1 gene was identified (Dyer et al., 1989). In the case of the human P1 gene, but not the P2 gene, this problem was surmounted by rescreening restriction digests of positive clones with a second probe derived from the 5′ region of the P1 bovine cDNAs, and searching for recombinants in which the 5′ and 3′ probes hybridized either with the same large fragments in various digests, or with more than one fragment in the same digest. Thus, isolate λHP1.9 from the SH library was found to contain only large restriction fragments (> 3 kb) that hybridized with both the 5′ and 3′ probes, indicating that the hybridizing sequences were distributed in several kilobases of DNA. Amongst hybridizing fragments in λHP1.9 was a *Bam*H1 fragment of 4.4 kb, and a fragment of this size also was detected in digests of human DNA (results not shown). It was sequenced and proved to contain the expressed human P1 gene.

A clone containing the human P2 expressed gene was isolated by probing with a sequence at the 5′ end of sequence determined in the bovine P2 gene (Dyer et al., 1989) which is not present in the bovine P2 cDNA clone. This sequence is now known to be part of intron A of the bovine and human P2 genes. Large restriction fragments (> 12 kb) in recombinant λAT5P2.1 hybridized both with this probe and with the P2 probe derived from the 3′ end of the bovine cDNA. A 12.2 kb *Sac*I fragment from λAT5P2.1 was sequenced and contained the expressed gene. A fragment of similar size hybridized with the P2 probes in a digest of human DNA (see Figure 6).

The human P1 and P2 genes

The human genomic sequences containing the expressed P1 and P2 genes are 9457 and 15016 bases in length respectively (Figures 1 and 2). Their G+C contents are 47.8% (9.4 kb sequence) and 46.9% (15 kb sequence). There is one ambiguity in the P1 gene at nucleotide 9444, where an A residue was found in one clone, and a G residue in two others. It is assumed that the correct assignment is G. Nucleotide sequences of partial cDNAs for human P1 and P2 have been described (Farrell and Nagley, 1987), but both differ from the corresponding genomic sequences at several positions (see legend to Figures 1 and 2). The P1 cDNA clone confirms that poly(A) is added after nucleotide 7799.

The protein sequences of the human and bovine P1 and P2 precursors contain an identical mature c subunit. Assuming that the sites of cleavage of the pre-sequences are the same as in the bovine proteins, the human pre-sequences of P1 and P2 are 61 and 66 amino acids long respectively. In contrast to the mature proteins, the pre-sequences are not conserved. Those of the P1 proteins are the same length, but the sequences differ in 11 amino acids. The bovine P2 pre-sequence is two amino acids longer than the human homologue, and the sequences differ in 17 amino acids.

The human P1 and P2 genes are both divided into five exons (Figure 3). In common with the rather narrow range of exon lengths observed in other eukaryotic genes (Naora and Deacon, 1982), their sizes range between 29 and 259 bp (Table 1). Introns B–D in both genes are found at almost identical positions to those in the bovine P2 gene (Dyer et al., 1989). In the human P1 and P2 genes (and also in bovine P2), exons I are in the 5′ non-coding region, and exons II correspond to the rest of the 5′ non-coding regions present in the mRNAs and to a region encoding part of the import pre-sequences. The rest of the pre-sequences are encoded in exons III and part of exons IV, which code for the N-terminal 38 amino acids of the mature protein. In none of the three genes is an intron found at the boundary between the processed import sequence and the mature protein. In contrast, the import sequence and the 5′ non-coding region of the mRNA of another mitochondrial protein, subunit IV of cytochrome *c* oxidase, are encoded in separate exons (Bachman et al., 1987), and an intron separates almost all of the DNA coding for the import sequence of the human β-subunit of ATP synthase from the region coding for the N-terminal end of the mature protein (Ohta et al., 1988).

Intron D in the human pre-proteolipid genes almost certainly does correspond to a boundary between structural domains in subunit c. It interrupts the sequence coding for Arg-Asn-Pro-, which is believed to form a β-turn outside the lipid bilayer and

Figure 1 DNA sequence of a fragment of human DNA containing the P1 gene

Exon I (marked with double lines above and beneath) is homologous to sequences in the 5′ regions of a bovine processed pseudogene (Dyer et al., 1989) and of an ovine P1 cDNA (Medd et al., 1993) from sites A and B respectively. Protein sequences are shown over exons II–V, and the small arrows denote exon–intron boundaries. The part of intron A (marked with a broken line; nucleotides 5322–5340) is very similar to nucleotides 1–31 of bovine P1 cDNA (Gay and Walker, 1985). The sequence differences in the 5′ regions of a bovine pseudogene and of bovine and ovine cDNA can be explained by two alternate transcription initiation sites in the human sequence, corresponding to two TATA boxes (triple underlines; nucleotides 4688–4693 and 5279–5285). CCAAT boxes have a single line above them, and the sequence GGGCGG and its complement are boxed. The doubly underlined sequence (nucleotides 7780–7785) is a polyadenylation signal, and poly(A) is added between nucleotides 7800 and 7802 (Farrell and Nagley, 1987). The *Alu* repeats on the displayed and complementary DNA strands are denoted by (+) and (−) respectively. The *Sau*3A site at the 5′ end of the insert in λHP1.9 is shown, as are the *Pst*I and *Bam*HI sites used in cloning this sequence. A partial human P1 cDNA sequence (Farrell and Nagley, 1987) covers the coding nucleotides from 6372 to 7785. At the positions corresponding to nucleotides 7126, 7128, 7129, 7702 and 7785, Farrell and Nagley (1987) report C, G, C, A and T in the cDNA. In addition, the cDNA sequence lacks 33 nucleotides that are present in the gene sequence, from bases 7711 to 7743.

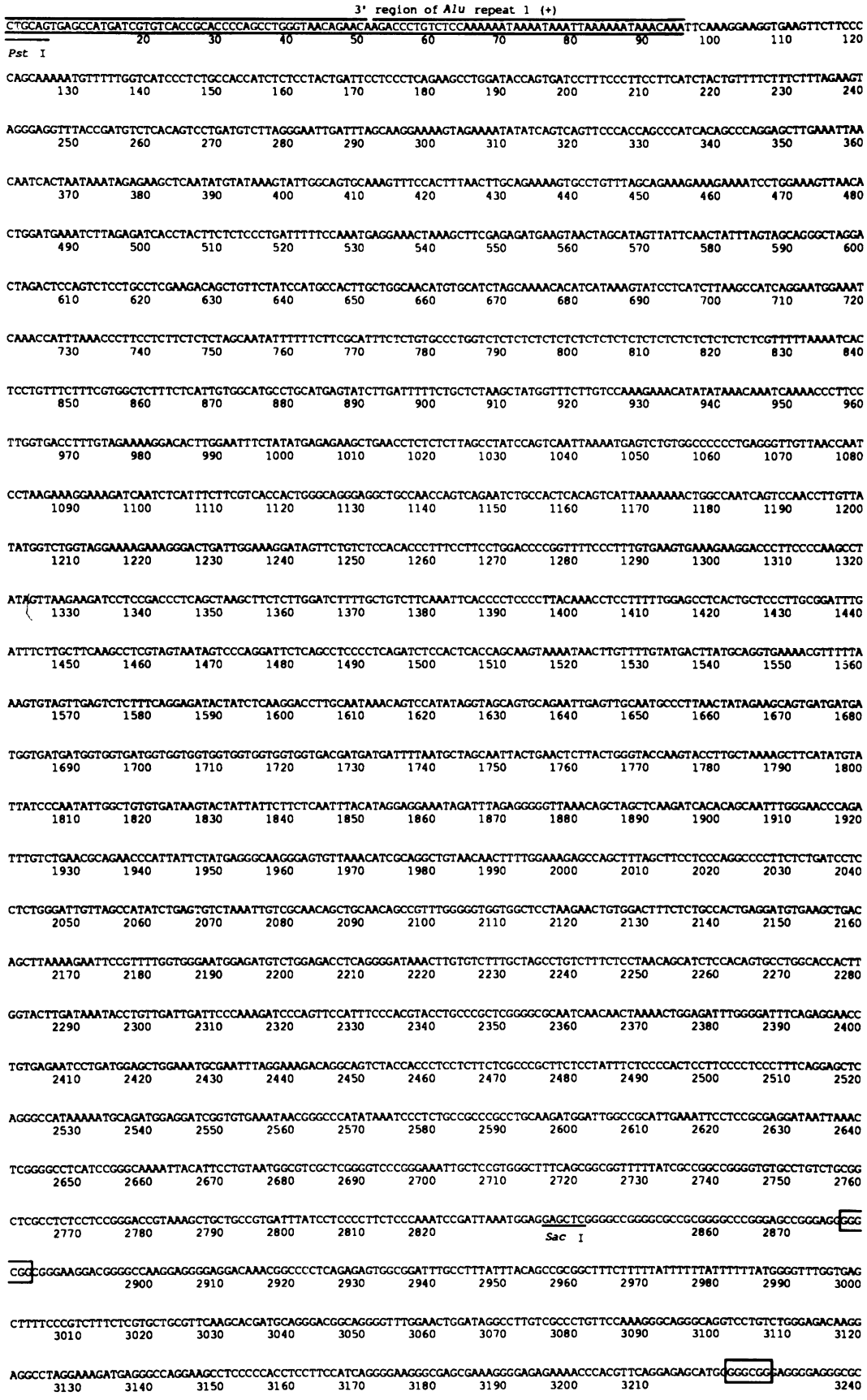

Figure 2 **For legend see page 60.**

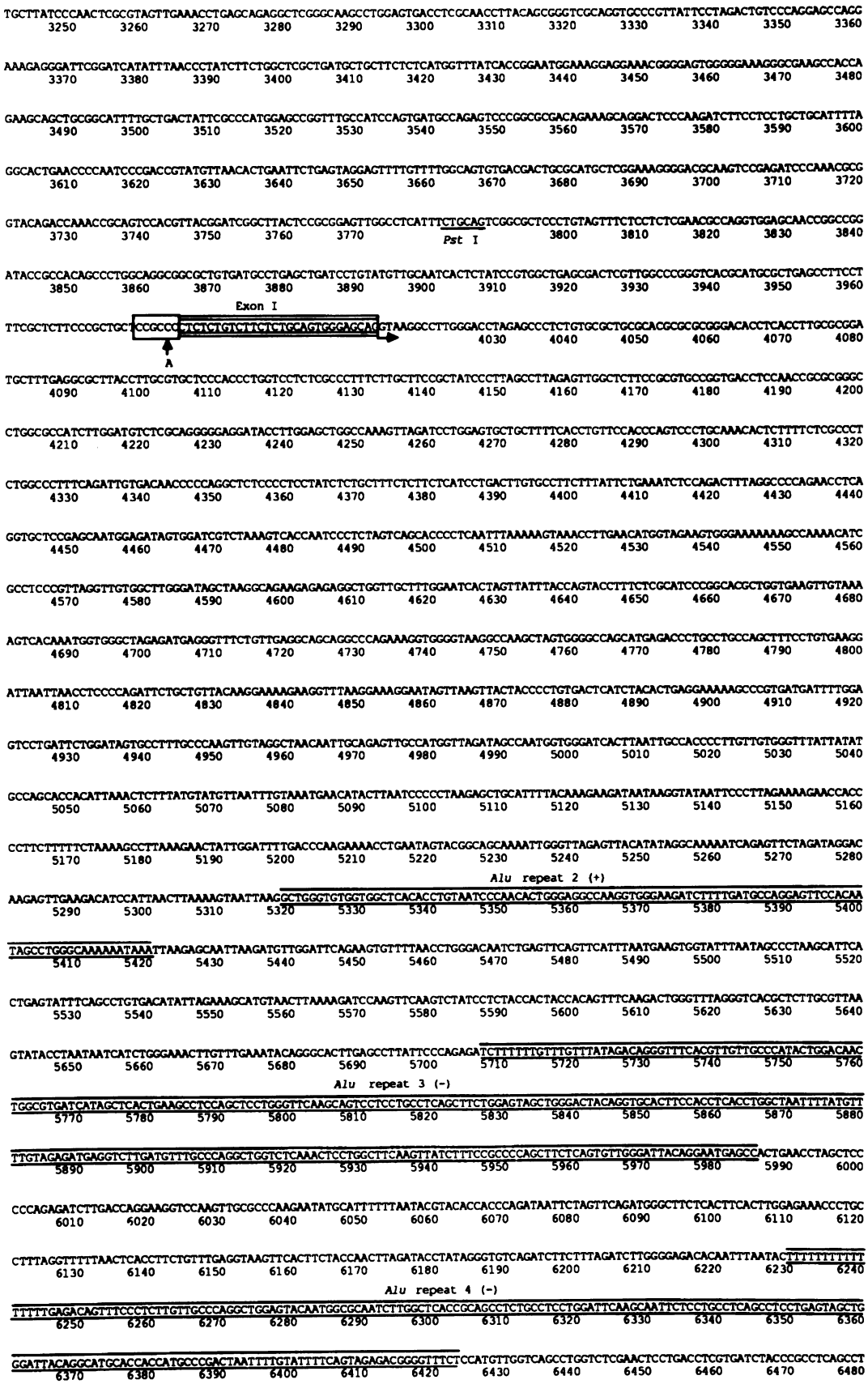

Figure 2 For legend see page 60.

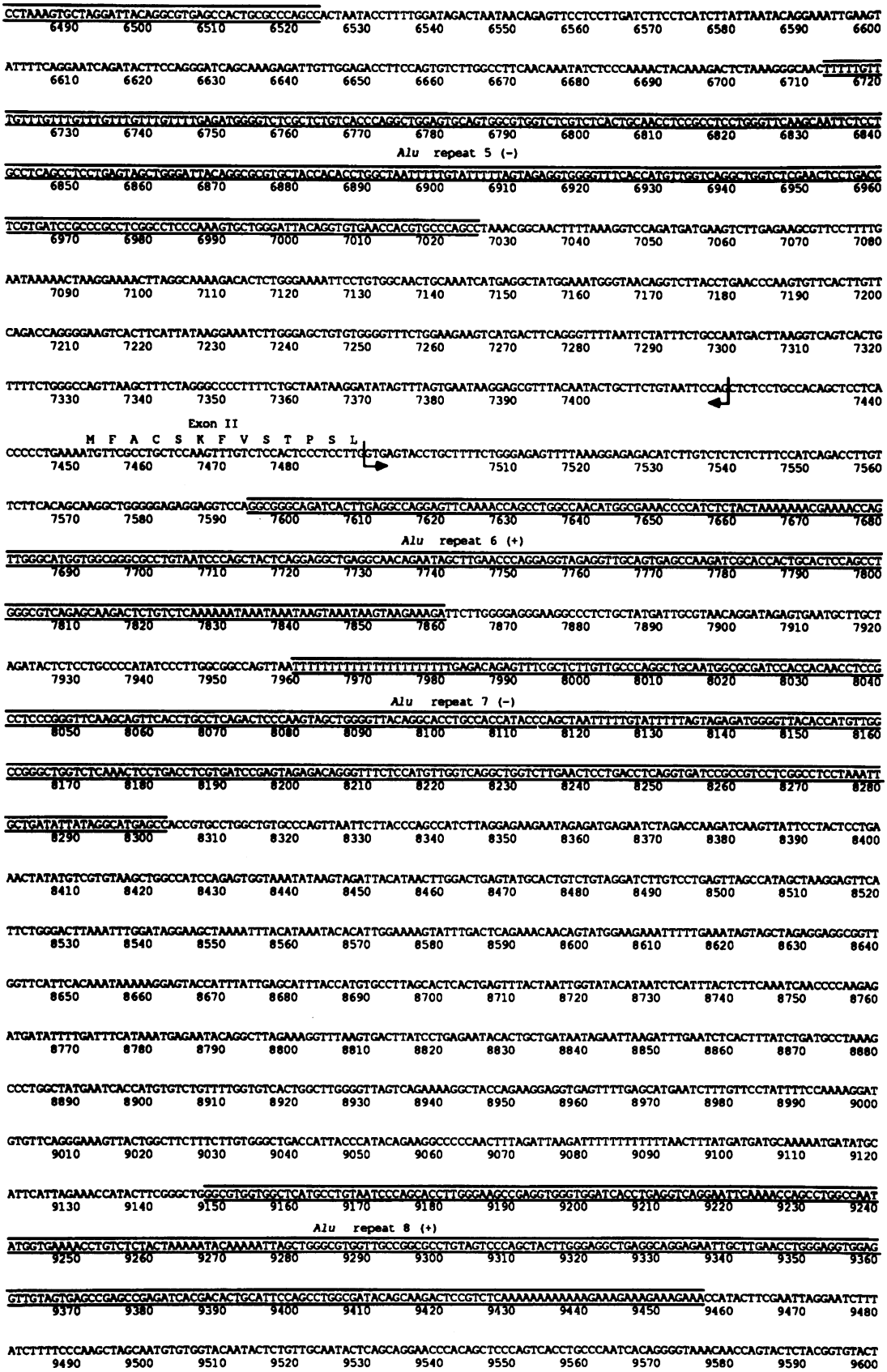

Figure 2 **For legend see page 60.**

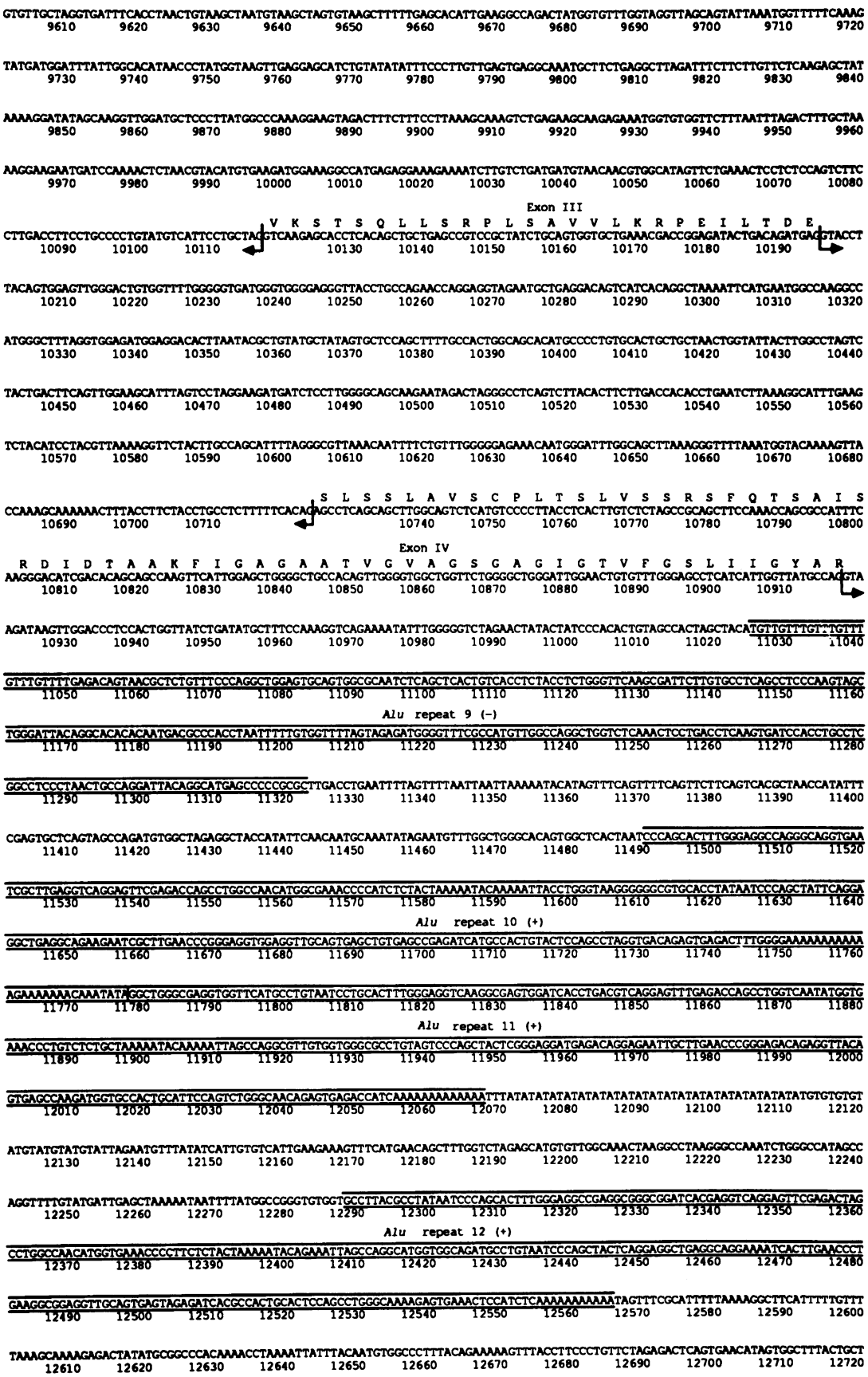

Figure 2 For legend see page 60.

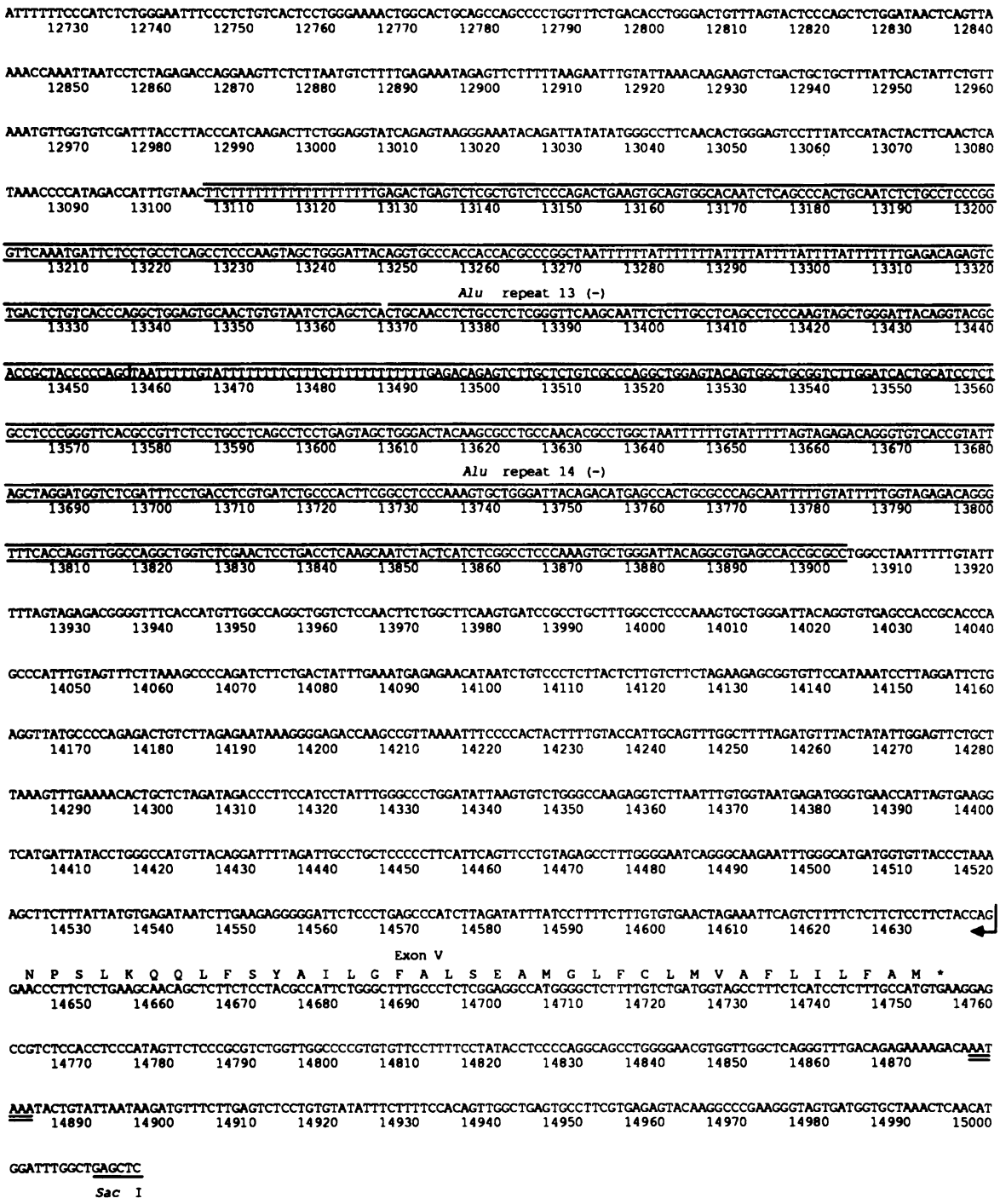

Figure 2 DNA sequence of a fragment of human DNA containing the P2 gene

Exon I (double lines above and below) is homologous to the 5′ sequence of a human P2 processed pseudogene described here, and to an ovine P2 cDNA (Medd et al., 1993) from the site marked by A. Protein sequences are shown above exons II–V, and exon–intron boundaries are denoted by small arrows. In the proposed promoter region are three copies (in boxes) of the sequence GGGCGG and its complement. The doubly underlined sequence (nucleotides 14878–14883) is a polyadenylation signal (Proudfoot and Brownlee, 1976), although the exact position of polyadenylation is not known. The *Alu* repeats on the displayed and complementary DNA strands are denoted by (+) and (−) respectively. The *Pst*I and *Sac*I restriction enzyme sites used in the cloning of this region are shown at the extremities of the sequence. The partial human P2 cDNA sequence (Farrell and Nagley, 1987) corresponds to the coding sequence from bases 10860–14822. The cDNA is reported to have the additional sequences CGGCTCTCA and TCA at its 5′ and 3′ extremities, but they are not in the genomic sequence.

to link its two transmembrane α-helices (Sebald and Hoppe, 1981). The presence of introns in segments of DNA coding for links between transmembrane α-helices has been observed in genes for other intrinsic membrane proteins, including bovine and human rhodopsins (Nathans and Hogness, 1984), mouse band III protein from the red cell membrane (Kopito et al., 1987) and ADP/ATP translocase (Cozens et al., 1989). It is consistent with the general view that exons often encode structural domains of proteins (Gilbert, 1978; Blake, 1979).

The nucleotide sequences adjacent to the 5′ and 3′ boundaries

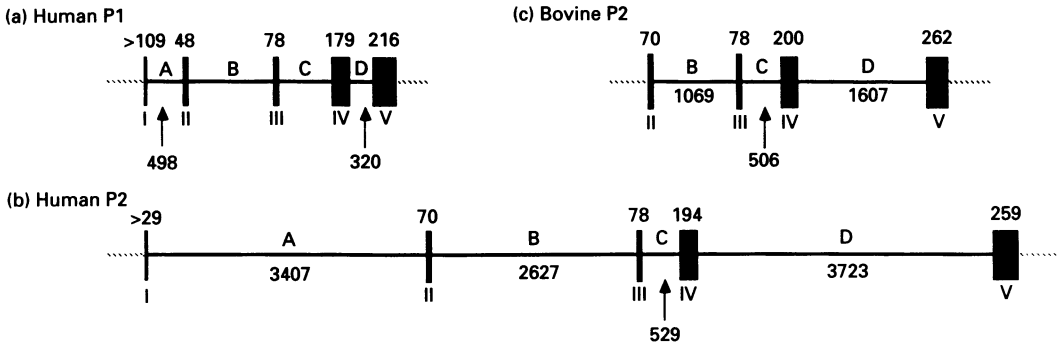

Figure 3 Structures of human P1 and P2 genes and the bovine P2 gene for the precursors of the c subunit of mitochondrial ATP synthase

In the human genes, exons I–V and introns A–D are represented by solid boxes and continuous lines respectively. The sizes of exons and introns are given in bp. Human P1 may have two promoters, one to initiate transcription from the 5′ end of exon I, and the second close to the 5′ boundary of exon II. The transcriptional initiation sites have not been determined experimentally. The known bovine P2 gene sequence does not extend into exon I (Dyer et al., 1989).

Table 1 Exon sizes in genes for the c subunit of mitochondrial ATP synthase

	Exon length (bp)				
Gene	I*	II	III	IV	V
Human P1	(109)	48†	78	179	216
Human P2	(29)	70	78	194	259
Bovine P2	–	70	78	200	262

* Parentheses indicate that the lengths of exons I have not been determined accurately, and that these are minimal estimates based upon cDNA sequences.

† It is suggested in the text that a sequence in intron A could promote transcription. If this is true in humans then exon II can also be 67 bp long.

of all of the introns in the human P1 and P2 genes, and also in the bovine P2 gene, are conserved (Table 2). They begin with the dinucleotide GT and end with AG, and so agree exactly with the consensus sequences adjacent to splice junctions (Breathnach and Chambon, 1981). Furthermore, the conservation extends for an additional 8–10 bp from the splice junctions in the sequences of the introns, and these extended sequences also agree with the consensus for sequences around splice sites (Mount, 1982). The classes of exon–intron boundary within homologous exons in both human genes (and also in the bovine P2 gene) are conserved (see Table 2). Extensive sequences are conserved within introns of human and bovine P2 genes (results not shown), indicating that they may be under evolutionary constraint.

There are probably more than 10^5 *Alu* repeats in the human genome, representing 5–6% of its DNA (Rinehart et al., 1981). They are usually about 300 bp long, and are dimeric structures

Table 2 Introns in mammalian pre-proteolipid genes

				Sequence	
Gene	Intron	Size (bp)	Class	5′ boundary	3′ boundary
Human P1	A	498	–	gtg.cag.GTGACTTGGG	CCCTCTGCAG.act.gaa
Human P2	A	3407	–	gag.cag.GTAAGGCCTT	GTAATTCCAG.ctc.tcc
Human P1	B	915	0	gct.ctg.GTAAGGTGCC	GATTTTACAG.atc.cgc
				A L	I R
Human P2	B	2627	0	tcc.ttg.GTGAGTACCT	TTCCTGCTAG.gtc.aag
				S L	V K
Bovine P2	B	1069	0	tcc.ttg.GTGAGTACCC	TTCCGGCTAG.atc.agg
				S L	I R
Human P1	C	706	0	aaa.cag.GTAAGGGAGG	CTCTTTCTAG.cct.tcc
				K Q	P S
Human P2	C	529	0	gat.gag.GTACCTTACA	TTTTTCACAG.agc.ctc
				D E	S L
Bovine P2	C	506	0	gat.gag.GTACCTTACA	TTCTTCACAG.agc.cac
				D E	S H
Human P1	D	320	2	gcc.ag.GTAAGTTTGG	TCCCTCCCAG.g.aac
				A R	N
Human P2	D	3723	2	gcc.ag.GTAAGATAAG	CTTCTACCAG.g.aac
				A R	N
Bovine P2	D	1607	2	gcc.ag.GTAAGATGGG	CCCCTCCCAG.g.aac
				A R	N
Consensus sequence				cagGTAAGT	YYYYYYYYYNCAGg

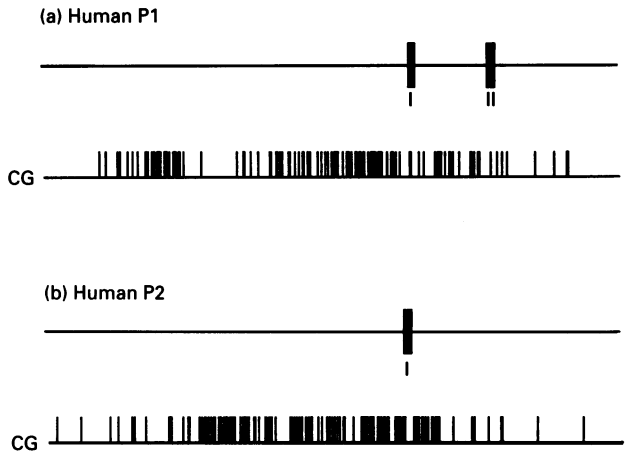

Figure 4 Distribution of the dinucleotide CpG in the 5′ regions of the human P1 and P2 genes

The vertical lines mark each CpG in (**a**) nucleotides 2250–6250 of the human P1 gene, and (**b**) nucleotides 1500–5500 of the human P2 gene. The horizontal and solid lines indicate non-coding regions and exons respectively.

which have apparently formed from internal deletions and dimerizations of 7SL RNA (Ullu and Tschudi, 1984). The two segments of human genomic sequence encompassing the P1 and P2 genes (Figures 1 and 2) contain 10 and 14 examples respectively, some in introns and others in flanking sequences. In each DNA sequence four of the repeats are clustered in pairs. *Alu* repeat 2 in intron A of the human P2 gene is exceptional. It is 102 bp long and contains only the 3′ monomeric unit. The B1 family of repeated DNA sequences in rodents have similar structures (Rogers, 1985).

Transcription of P1 and P2 genes

Within their 5′ regions and extending over exons I, the human P1 and P2 genes have CpG-rich islands (Bird, 1986) of 2 and 1.5 kb long respectively (Figure 4). Transcription probably initiates in these islands, but the transcriptional start sites for neither gene have been determined experimentally. However, the 5′ sequences determined in the P1 and P2 cDNAs in cows (Gay and Walker, 1985) and sheep (Medd et al., 1993) help to pin-point these sites. Since processed pseudogenes are believed to have arisen by a process that involved reverse transcription of mRNAs (Rogers, 1985; Weiner et al., 1986), further clues are to be found in the

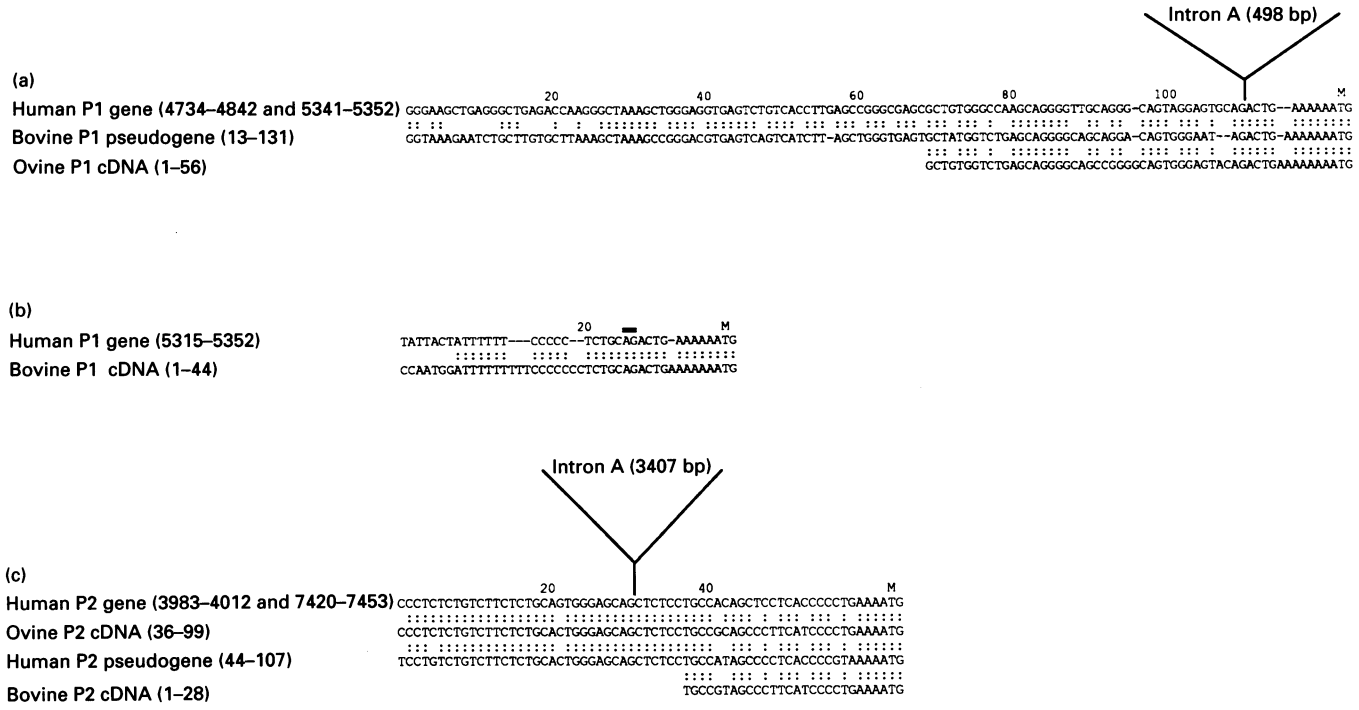

Figure 5 Comparisons of DNA sequences in the 5′ non-coding regions of the human P1 and P2 genes, in the bovine and ovine cDNAs and in related pseudogenes

The positions of the sequences in the determined sequences are given in parentheses on the left. Identities are denoted by colons (:). The positions of the translational initiator methionines are denoted by M. In (**a**), part of the human genomic sequence is aligned with the 5′ region of a bovine P1 processed pseudogene immediately following its 5′ flanking direct repeat sequence (Dyer et al., 1989), and with the 5′ untranslated region of an ovine liver cDNA for P1 (Medd et al., 1993). The position of intron A is shown. In (**b**), a sequence in the 5′ untranslated region of a bovine P1 cDNA is aligned with a different sequence in the human gene that is found adjacent to the 5′ boundary of exon II (see Figure 1). The dinucleotide AG with a bar above it could be used as a 3′ splice site in a putative human pre-mRNA initiated upstream of exon I. This would result in an mRNA similar in structure to the ovine mRNA. If, as proposed, a second promoter is found in the sequence preceding exon II, then the 3′ region of intron A (P1 gene) codes for the 5′ untranslated region of a human mRNA that is related to the bovine P1 mRNA from heart. In (**c**), the human P2 gene is compared with the 5′ untranslated region of an ovine liver cDNA. The latter contains a run of T and C residues at nucleotides 1–35 which is not related to either the human genomic sequence or the human P2 processed pseudogene. It is possible either that this TC-rich sequence is a cloning artefact, or that the ovine sequence is unrelated over this stretch. The remainder of the 5′ untranslated region of the sheep P2 mRNA is aligned with the human genomic sequence. A sequence from a human P2 pseudogene (see Figure 7) immediately downstream from its 5′ flanking repetitive sequence is also shown, as is the entire 5′ untranslated region present in a bovine heart P2 cDNA (Gay and Walker, 1985). In the human genomic sequence the position of intron A is indicated. These proposals concerning the transcription of the P1 and P2 genes have not yet been tested by transcriptional mapping studies.

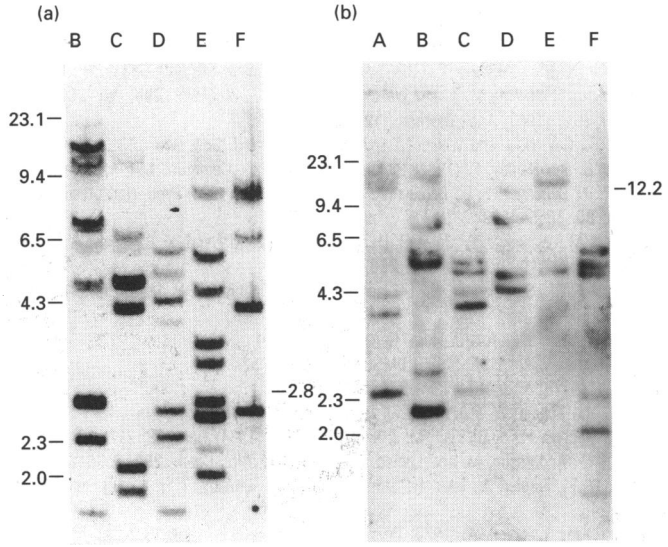

Figure 6 Hybridization of human DNA with specific DNA probes for the pre-proteolipid genes P1 and P2

The probes are nucleotides 404–558 and 406–615 of the bovine cDNAs for P1 and P2 respectively (Gay and Walker, 1985). Human placental DNA (20 μg) was digested with the restriction enzymes *Bam*HI (lane A), *Eco*RI (lanes B), *Hin*dIII (lanes C), *Nco*I (lanes D), *Sac*I (lanes E) and *Xba*I (lanes F). The fragments were fractionated by electrophoresis in a 0.6% agarose gel and then were hybridized on nitrocellulose filters to prime-cut probes for P1 (**a**) and P2 (**b**). The filters were washed in 0.2 × SSC at 65 °C and then autoradiographed at −70 °C for 72 h. In (**a**), lane B, an *Eco*RI fragment of 2.8 kb is observed; subsequently the DNA sequence of human P1 was found to contain an *Eco*RI fragment of this size. In (**b**) lane E, a *Sac*I fragment of 12.2 kb is indicated; a fragment of the same size was sequenced from the DNA of λAT5P2.1. Marker and fragment sizes are in kb.

sequences immediately downstream of the 5′ flanking repeated sequences of a human P2 and a bovine P1 processed pseudogene (Dyer et al., 1989; see Figures 5a and 5b).

In the human P1 gene, the available information (Figure 5a) indicates the presence of two independent transcriptional initiation sites. These alternative promoters could be used to regulate expression of the gene in various tissues. The transcription of the human P2 gene appears to be simpler. All of the available information (Figure 5c) is consistent with a single transcriptional initiation site in the vicinity of nucleotide 3984.

The 3′ limits of transcription of the human P1 and P2 genes are more readily discerned. Human P1 has the uncommon polyadenylation signal, ATTAAA (Berget, 1984; Martini et al., 1986), which is also used in the bovine P1 gene (Gay and Walker, 1985). The more usual polyadenylation signal, AATAAA (Proudfoot and Brownlee, 1976), is found 122 bp and 125 bp respectively after the termination codons in both the bovine and human P2 genes. Poly(A) addition to the human transcripts probably occurs within 11–13 nucleotides, to the 3′ side of these sequences.

Number of human genes for P1 and P2

Previous studies of bovine cDNAs (Gay and Walker, 1985), together with the work presented in this paper, have shown that both the human and bovine genomes contain at least two expressed genes for the dicyclohexylcarbodi-imide-reactive proteolipid subunit of mitochondrial ATP synthase. In addition, numerous spliced pseudogenes have been detected in both animals, and these observations are consistent with the complex Southern blots obtained with digests of both bovine and human

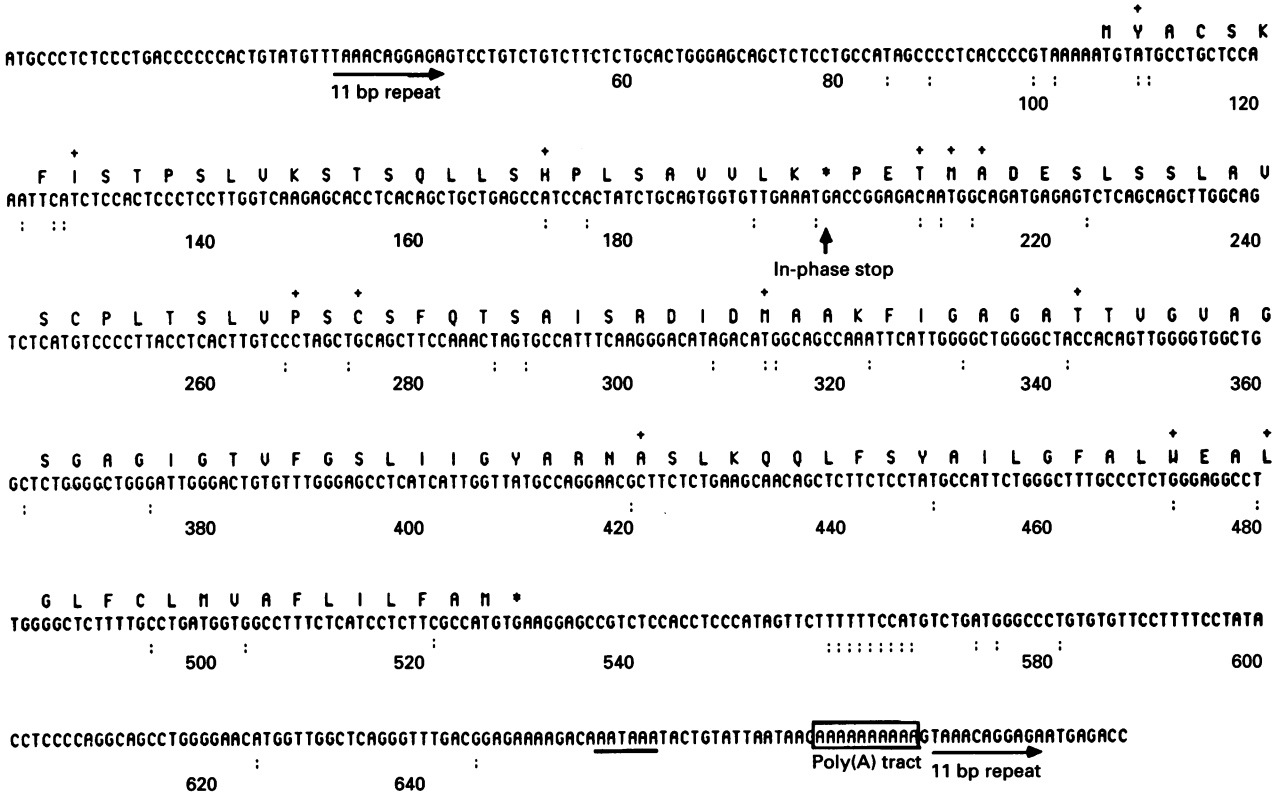

Figure 7 Sequence of a human processed pseudogene for the mitochondrial pre-proteolipid P2

Colons and crosses indicate the 50 differences in nucleotide sequence and 13 differences in protein sequence respectively between the pseudogene and the coding regions and protein sequence of human P2. The position of an in-phase stop codon is indicated by a vertical arrow. The underlined sequence is a poly(A) addition signal (Proudfoot and Brownlee, 1976; Gay and Walker, 1985). The following poly(A) tract is boxed. The direct 11 bp repeated sequences which flank the pseudogene are indicated by horizontal arrows.

DNA (see Figure 6). During the course of the cloning and sequencing experiments described above, the complete sequence of a P2 pseudogene (Figure 7) was determined from the overlapping recombinants λHP2.8 and λHP2.13. Several features of this sequence support the view that it arose by reverse transcription of the P2 mRNA, followed by recombination into the human genome. For example, the sequence is flanked by two direct 11-nucleotide repeats, and the direct repeat at the 3′ end of the pseudogene is preceded by a potential polyadenylation signal and the sequence A_{10}. Also, the pseudogene sequence differs in 50 nucleotides from the human P2 cDNA sequence deduced from the gene. This causes 13 substitutions in the amino acid sequence and introduces an in-phase stop codon. As described in the following paper (Medd et al., 1993), an intronless P2 pseudogene in the sheep genome is transcribed, and an intronless human gene encoding phosphoglycerate kinase has been shown to express the protein, but only in testis (McCarrey and Thomas, 1987). Therefore it is conceivable that some of the other processed P1 and P2 sequences in the human genome may not be pseudogenes, as we have tended to assume, but may be functional retroposons also.

The work described in this paper has a direct bearing on the fatal disease, ceroid lipofuscinosis, found in man and other mammals. In the juvenile and late-infantile forms of the human disease, and in the sheep disease (Fearnley et al., 1990), the affected individuals accumulate large amounts of the c subunit of mitochondrial ATP synthase in lysosomes. The accumulated material appears to be chemically identical to the protein normally found in mitochondria (Palmer et al., 1992). In diseased sheep the P1 and P2 cDNAs are identical in sequence to those from normal animals, and the amounts of mRNAs for both P1 and P2 are unaffected in the diseased animals (Medd et al., 1993). Therefore the disease appears not to involve mutation of the coding sequences of the P1 and P2 genes. Similar investigations have not been conducted in humans, but the gene for the juvenile form of human ceroid lipofuscinosis maps to the long arm of chromosome 16 (Gardiner, 1992), whereas the human P1 and P2 genes are on human chromosomes 17 and 12 respectively (M. R. Dyer and J. E. Walker, unpublished work).

We thank Dr. T. H. Rabbitts (of this laboratory) for supplying us with samples of human genomic libraries. M. R. D. was supported by an M.R.C. research studentship and an M.R.C. research training Fellowship.

REFERENCES

Bachman, N. J., Lomax, M. I. and Grossman, L. I. (1987) Gene **55**, 219–229
Benton, W. D. and Davis, R. W. (1977) Science **196**, 180–182
Berget, S. M. (1984) Nature (London) **309**, 179–182
Biggin, M. D., Gibson, T. J. and Hong, G. F. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 3963–3965
Bird, A. P. (1986) Nature (London) **321**, 209–213
Blake, C. C. F. (1979) Nature (London) **277**, 598
Breathnach, R. and Chambon, P. (1981) Annu. Rev. Biochem. **50**, 349–383
Cozens, A. L., Runswick, M. J. and Walker, J. E. (1989) J. Mol. Biol. **206**, 261–280
Deininger, P. L. (1983) Anal. Biochem. **129**, 216–223
Dyer, M. R., Gay, N. J. and Walker, J. E. (1989) Biochem. J. **260**, 249–258
Farrell, L. B. and Nagley, P. (1987) Biochem. Biophys. Res. Commun. **144**, 1257–1264
Farrell, P. J., Deininger, P. L., Bankier, A. and Barrell, B. G. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 1565–1569
Fearnley, I. M., Walker, J. E., Jolly, R. D., Martinus, R. D., Kirkland, K. B., Shaw, G. J. and Palmer, D. N. (1990) Biochem. J. **268**, 751–758
Forster, A., Huck, S., Ghanem, N., LeFranc, M. P. and Rabbitts, T. H. (1987) EMBO J. **6**, 1945–1950
Gardiner, R. M. (1992) Am. J. Hum. Genet. **15**, 539–541
Gay, N. J. and Walker, J. E. (1985) EMBO J. **4**, 3519–3524
Gilbert, W. (1978) Nature (London) **271**, 501
Jackl, G. and Sebald, W. (1975) Eur. J. Biochem. **54**, 97–106
Karn, J., Matthes, H. W. D., Gait, M. J. and Brenner, S. (1984) Gene **32**, 217–224
Kopito, R. R., Andersson, M. and Lodish, H. F. (1987) J. Biol. Chem. **262**, 8035–8040
LeFranc, M. P., Forster, A., Baer, R., Stinson, M. A. and Rabbitts, T. H. (1986) Cell **45**, 237–246
Macino, G. and Tzagoloff, A. (1979) Proc. Natl. Acad. Sci. U.S.A. **76**, 131–135
Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) in Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring Harbor, New York
Martini, G., Toniolo, D., Vulliamy, T., Luzzatto, L., Dono, R., Viglietto, G., Paonessa, G., D'Urso, M. D. and Persico, M. G. (1986) EMBO J. **5**, 1849–1855
McCarrey, J. R. and Thomas, K. (1987) Nature (London) **326**, 501–505
Medd, S. M., Walker, J. E. and Jolly, R. D. (1993) Biochem. J. **293**, 65–73
Messing, J. (1983) Methods Enzymol. **101**, 20–78
Mills, D. R. and Kramer, F. R. (1979) Proc. Natl. Acad. Sci. U.S.A. **76**, 2232–2235
Mizusawa, S., Nishimura, S. and Seela, F. (1986) Nucleic Acids Res. **14**, 1319–1324
Mount, S. M. (1982) Nucleic Acids Res. **10**, 459–472
Naora, H. and Deacon, N. J. (1982) Proc. Natl. Acad. Sci. U.S.A. **79**, 6196–6200
Nathans, J. and Hogness, D. S. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 4852–4855
Ohta, S., Yohda, M., Ishizuka, M., Hirata, H., Hamamoto, T., Otawara-Hamamoto, Y., Matsuda, K. and Kagawa, Y. (1988) Biochim. Biophys. Acta **933**, 141–155
Palmer, D. N., Fearnley, I. M., Walker, J. E., Hall, N. A., Lake, B. D., Wolfe, L. S., Haltia, M., Martinus, R. D. and Jolly, R. D. (1992) Am. J. Med. Genet. **41**, 561–567
Proudfoot, N. J. and Brownlee, G. G. (1976) Nature (London) **263**, 211–214
Rinehart, F. P., Ritch, T. G., Deininger, P. L. and Schmid, C. W. (1981) Biochemistry **20**, 3003–3010
Rogers, J. H. (1985) Int. Rev. Cytol. **93**, 187–279
Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
Sebald, W. and Hoppe, J. (1981) Curr. Top. Bioenerget. **12**, 2–64
Southern, E. M. (1975) J. Mol. Biol. **98**, 503–517
Staden, R. (1982) Nucleic Acids Res. **10**, 4731–4751
Staden, R. (1985) in Genetic Engineering: Principles and Methods (Setlow, J. K. and Hollaender, A., eds.), pp. 67–114, Plenum Publishing Corporation, New York and London
Turner, G., Imam, G. and Kuntzel, H. (1979) Eur. J. Biochem. **97**, 565–571
Ullu, E. and Tschudi, C. (1984) Nature (London) **312**, 171–177
Walker, J. E., Gay, N. J., Powell, S. J., Kostina, M. and Dyer, M. R. (1987) Biochemistry **26**, 8613–8619
Walker, J. E., Lutter, R., Dupuis, A. and Runswick, M. J. (1991) Biochemistry **30**, 5369–5378
Weiner, A. M., Deininger, P. L. and Estratiadis, A. (1986) Annu. Rev. Biochem. **55**, 631–661
Wilbur, W. J. and Lipman, D. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 726–730

Received 21 December 1992/28 January 1993; accepted 2 February 1993