Characterization of the expressed genes for subunit c of mitochondrial ATP synthase in sheep with ceroid lipofuscinosis

Simon M. MEDD,* John E. WALKER*‡ and Robert D. JOLLY†

*Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K.,
and †Department of Veterinary Pathology and Public Health, Massey University, Palmerston North, New Zealand

The human and bovine genomes each contain two expressed nuclear genes, called P1 and P2, for subunit c, a hydrophobic subunit of the membrane sector, F_o, of mitochondrial ATP synthase. Both P1 and P2 encode the same mature protein, but the associated mitochondrial import sequences are different. In sheep with the neurodegenerative disease ceroid lipofuscinosis, and also in humans with Batten's disease, unmodified subunit c accumulates in lysosome-derived organelles in a variety of tissues. However, the sequences of cDNAs for P1 and P2 from sheep with ceroid lipofuscinosis were identical to those in healthy control animals. Therefore, since there was no mutation in either of the mitochondrial import sequences of subunit c in the diseased animals, ceroid lipofuscinosis does not arise from changes in an import sequence causing mis-targeting of the c subunit to lysosomes. The levels of expression of P1 and P2 genes were approximately the same in diseased and healthy animals, and so the protein is unlikely to accumulate because of excessive transcription of either gene. Transcription of a spliced pseudogene related to P2 was detected in both a control animal and a sheep with ceroid lipofuscinosis. The transcripts encode amino acids 1–31 of the P2 mitochondrial targeting sequence. In the diseased animal, an arginine replaced a glutamine in the control sequence. However, restriction fragment analysis of genomic DNA from a further 12 sheep established that the sequence differences were not linked to ceroid lipofuscinosis.

INTRODUCTION

The ceroid lipofuscinoses are a group of recessively inherited lysosomal storage diseases of children and animals, so named because of the similarity of the stored material to the lipopigments ceroid and lipofuscin (Zeman and Dyken, 1969). The disease causes blindness, seizures and dementia, culminating in premature death. The incidence of the human neuronal ceroid lipofuscinoses, or Batten's disease, has been estimated to be 1 in 12500 live births (Rider and Rider, 1988). In children, there are three main syndromes, i.e. infantile, late-infantile and juvenile, separated according to the age of onset and the clinical course of the disease (Boustany et al., 1988; Dyken, 1988).

The sheep disease, ovine ceroid lipofuscinosis, is most similar to the juvenile form of the human disease. Storage bodies of lysosomal origin accumulate in neurons and in a wide variety of other cells (Jolly et al., 1980, 1989). The material stored in sheep liver is made almost entirely of a protein identified as subunit c (Palmer et al., 1986, 1992), a component of the intrinsic membrane domain of mitochondrial ATP synthase (Sebald and Hoppe, 1981). The stored subunit c is identical in sequence to the mitochondrial protein and has been correctly processed from its mitochondrial import precursor (Fearnley et al., 1990). As in ovine ceroid lipofuscinosis, mitochondrial subunit c also accumulates in lysosomes in the late-infantile and juvenile forms of the human disease (Palmer et al., 1989).

In humans and cattle, subunit c has two expressed nuclear genes, called P1 and P2, that encode precursor proteins containing the same mature protein, but with different N-terminal presequences to specify import into mitochondria (Gay and Walker, 1985; Dyer and Walker, 1993). In addition, numerous pseudogenes related to P1 and P2 have been discovered (Dyer et al., 1989; Dyer and Walker, 1993). The N-terminal pre-sequences of P1 and P2 are sufficiently conserved and are characteristically similar to the import pre-sequences of other nuclear-encoded mitochondrial proteins to suggest that they both direct the protein to mitochondria. The cattle genes for subunit c are expressed in all tissues, but the ratio of expression levels (P1/P2) differs from 1:1 to 1:3 in tissues of mesodermal and endodermal origin respectively (Gay and Walker, 1985).

The primary genetic defect causing the ceroid lipofuscinoses is unknown, but the discovery that the associated storage bodies are made of subunit c suggested new hypotheses. For example, it is possible *a priori* that the disease could be caused by overexpression of one or both of the genes encoding subunit c, or that a mutation in the pre-sequence of one of the precursors causes the protein to be misdirected to lysosomes. As described below, both of these hypotheses have been disproved. We have cloned cDNAs for the P1 and P2 genes from healthy control sheep and from sheep with ceroid lipofuscinosis. By sequence analysis it has been shown that the coding region sequences of the P1 and P2 cDNAs from the sheep with ceroid lipofuscinosis are identical with those of the healthy sheep. We have also compared the expression of P1 and P2 genes in healthy and diseased sheep liver and kidney, and have shown that there is no gross difference in the expression of these genes between the healthy and diseased sheep. During these experiments, a mutated cDNA related to P2 was isolated from the library derived from a sheep with ceroid lipofuscinosis. It was shown that this cDNA originated from transcripts of a spliced pseudogene related to P2, and that a healthy sheep also expressed a closely related pseudogene. By restriction analysis of the genomic DNA from a further 14 sheep, it was shown that the

‡ 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. X69904, X69905 and X69906.

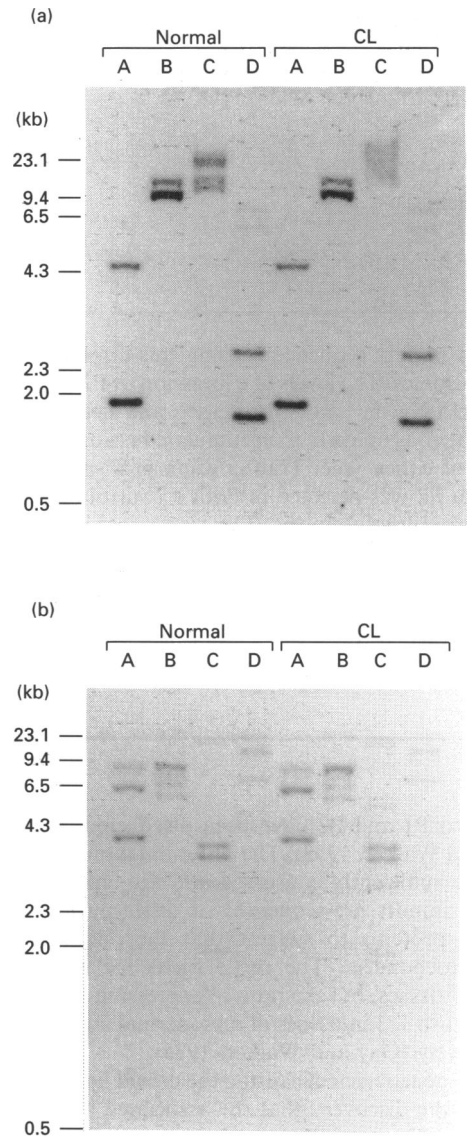

Figure 1 Hybridization of fractionated restriction enzyme digests of genomic DNA from healthy sheep and sheep with ceroid lipofuscinosis for P1 and P2 precursor proteins

DNA from healthy (normal) sheep and from sheep with ceroid lipofuscinosis (CL) were digested with the restriction endonucleases *Eco*RI, *Bam*HI, *Hin*dIII and *Sac*I (lanes A, B, C and D respectively). After electrophoresis and transfer to a membrane, the fragments were hybridized with prime-cut probes derived from the 3′ untranslated regions of the cDNAs of bovine P1 (**a**) and P2 (**b**) (nucleotides 404–558 and 406–615 respectively; see Gay and Walker, 1985). The sizes of markers are given in kilobases.

different pseudogenes were not closely linked to the gene causing ceroid lipofuscinosis.

MATERIALS AND METHODS

Isolation of genomic DNA and mRNA

Livers were removed from two 4-month-old lambs, one with ceroid lipofuscinosis as confirmed by histopathology, and the other a healthy control animal. The tissues were frozen immediately in liquid nitrogen and transported from New Zealand to Cambridge. The tissue was finely ground in liquid nitrogen and high-molecular-mass DNA (greater than 50000 bp in length) was prepared from this material (Bentley and Rabbitts, 1981). Poly(A)$^+$ RNA (Viñas et al., 1990) was purified from total sheep liver RNA (Chirgwin et al., 1979).

cDNA libraries

Double-stranded cDNA was made with a cDNA Synthesis System Plus kit (Amersham International). Its extremities were trimmed to blunt ends with T4 DNA polymerase (2.0 units/μg of original mRNA) and *Eco*RI linkers were attached. The product was ligated with λgt10 and packed *in vitro* into phage particles. Each library was amplified once, and the phage suspension was stored at 4 °C in the presence a few drops of chloroform. Amplified libraries had a titre of 1×10^{10} plaque-forming units/ml.

Two duplicate pairs of filters containing 0.5×10^6 recombinants from each library were screened on nitrocellulose filters (Benton and Davis, 1977) with 'prime-cut' probes (Farrell et al., 1983) derived from bases 1–201 of the 5′ untranslated regions of the bovine P1 cDNA and bases 1–152 of the bovine P2 cDNA (Gay and Walker, 1985). Filters were prehybridized and the hybridization probe was added (Viñas et al., 1990). They were then washed three times in 2 × SSC at 65 °C for 15 min (1 × SSC is 0.15 M NaCl and 0.015 M trisodium citrate), dried at room temperature and autoradiographed with preflashed film at −70 °C for 1–3 days. After rescreening the positive plaques, DNA was prepared from recombinants (Maniatis et al., 1982).

Circularized cDNA populations were prepared as described before (Walker et al., 1992).

Polymerase chain reactions

Sequences from both genomic DNA and total cDNA were amplified using the PCR (Saiki et al., 1988). Oligonucleotide primers for amplification were 34 bases long (25 bases complementary to the target DNA and 9 bases of a linker sequence). Primers for use in the forward direction of synthesis were made with the *Eco*RI linker TAGGAATTC, with the exception of primer *F*1 which had a *Sac*I linker, GCAGAGCTC, because the amplified product contained an *Eco*RI site. Primers to be employed in the reverse direction of synthesis were made with the *Hin*dIII linker CGAAAGCTT. Amplification reactions were performed in a solution (100 μl) containing 50 mM KCl, 10 mM Tris/HCl, pH 8.3, 1.5 mM $MgCl_2$, each dNTP (200 μM), both primers (1 μM), genomic DNA (1 μg) or cDNA (10 ng) and *Thermus aquaticus* DNA polymerase (3 units; Perkin–Elmer Cetus). Reaction mixtures were initially prepared without enzyme and preincubated at 94 °C for 2.5 min, after which the enzyme was added and the reaction mixture covered with mineral oil (100 μl; Sigma). For amplification of genomic DNA the following schedule was followed using a Techne Programmable DriBlock PHC-1: 94 °C for 2.5 min (denaturation), 55 °C for 2 min (annealing) and 70 °C for 2.5 min (synthesis) for 30 cycles, and finally a single incubation at 70 °C for 7 min. For amplification of cDNA the step times for denaturation, annealing and synthesis were 1, 2 and 2 min respectively. For amplification of cDNAs with a specific forward primer and an oligo(dT) reverse primer (25 T residues preceded by a *Hin*dIII linker), a lower annealing temperature of 50 °C was employed.

DNA sequencing

DNA fragments from λgt10 recombinants and polymerase chain

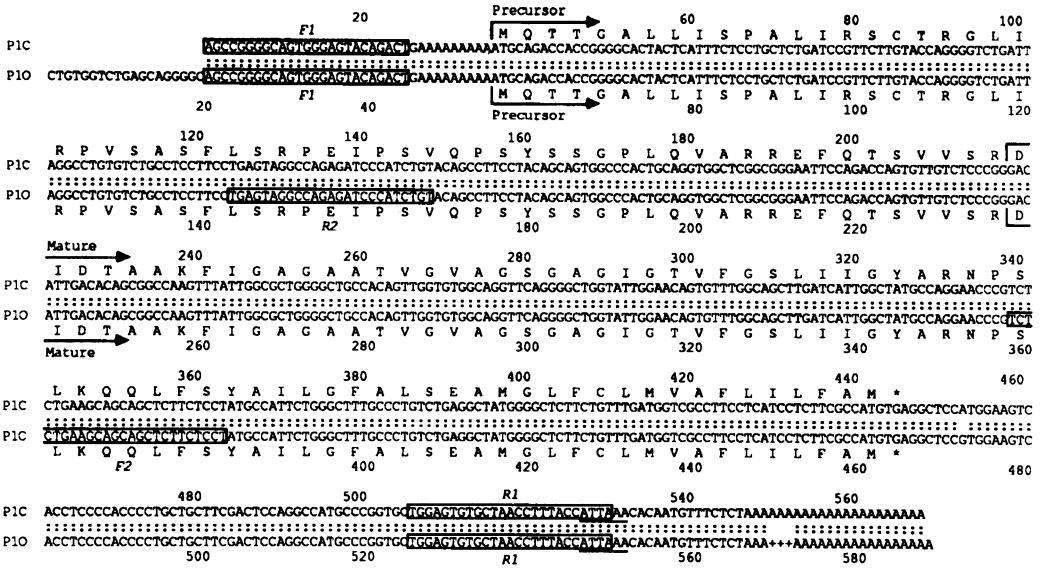

Figure 2 DNA sequences of ovine cDNAs encoding P1 precursors from healthy sheep and sheep with ceroid lipofuscinosis

The cDNA sequences are denoted P1C and P1O respectively. Colons denote identities in nucleotide sequence. Plus signs (+) indicate differences between two different isolates of P1O; one cDNA had CTC at bases 569–571, whereas AAA was present in the other isolate. Asterisks denote stop codons. Sites of translational initiation and of processing of the protein precursors to mature forms are shown. The boxed sequences show the position of primers used in the PCR. *F* and *R* refer to forward and reverse primers respectively. The underlined sequences are likely to be signals for addition of poly(A) to the 3′ end of the mRNA (Proudfoot and Brownlee, 1976).

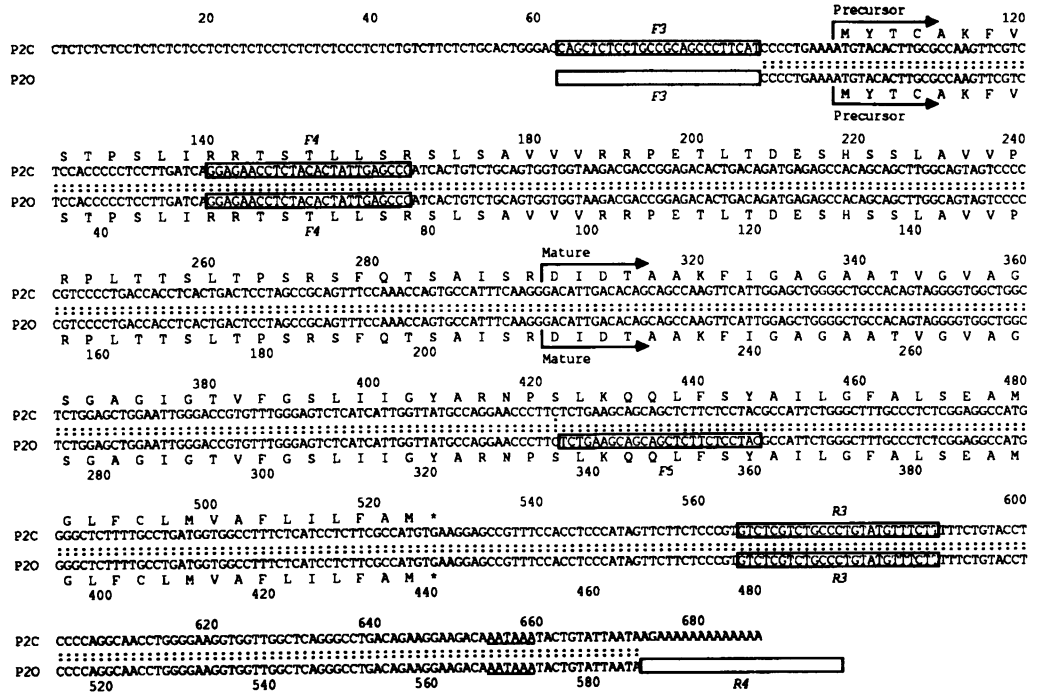

Figure 3 DNA sequences of ovine cDNAs encoding P2 precursors from healthy sheep and sheep with ceroid lipofuscinosis

The cDNAs are named P2C and P2O respectively. See the legend to Figure 2 for the meaning of the symbols. The empty boxes denote that the sequences of the primers were not determined independently. Primer R4 is an oligo(dT) reverse primer.

reaction products were cloned into M13mp18 and M13mp19 (Messing and Vieira, 1982; Norrander et al., 1983). Sequences were determined completely in both DNA senses as described before (Walker et al., 1992).

Hybridizations with RNA and DNA

RNA was transferred from agarose gels to Hybond N membranes (Amersham International) and cross-linked to the membrane by

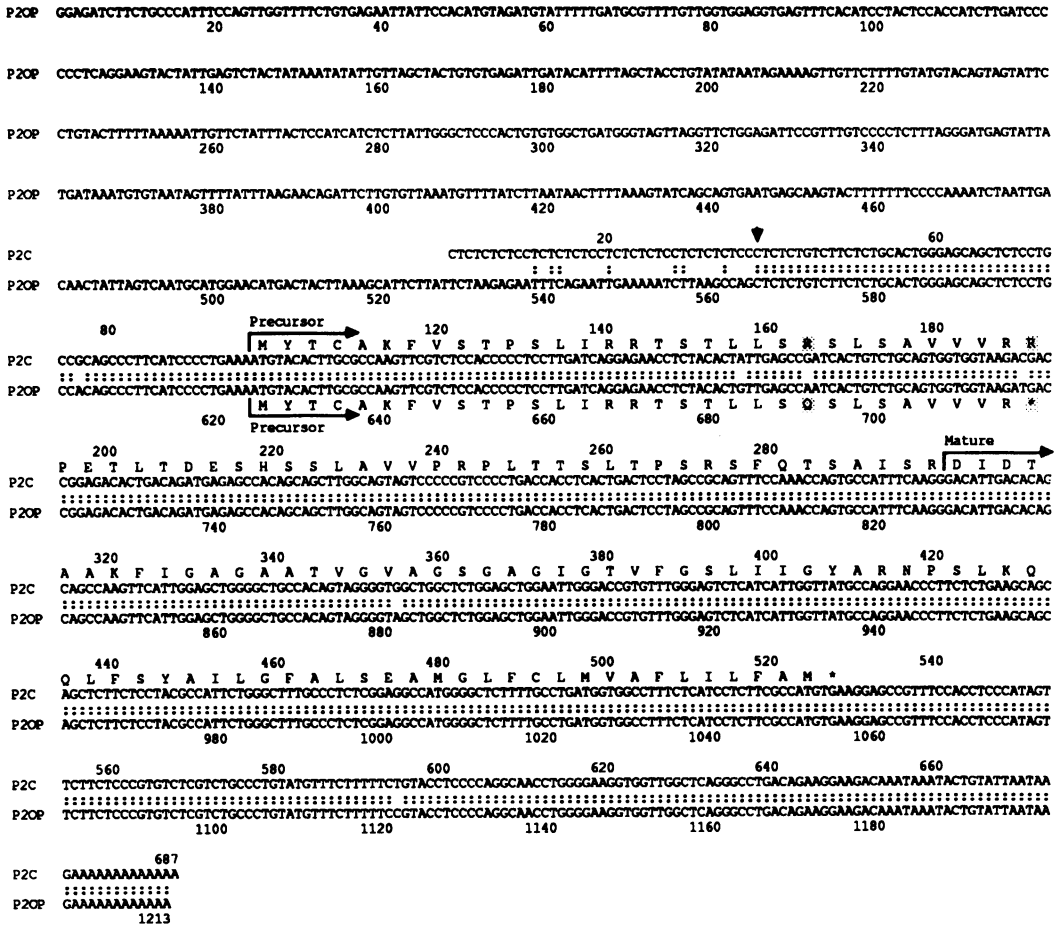

Figure 4 DNA sequences of ovine cDNAs encoding the P2 precursor from normal sheep and the P2-related pseudogene from sheep with ceroid lipofuscinosis

The cDNAs are named P2C and P2OP respectively. The vertical arrow indicates the proposed site of insertion of the pseudogene into the ovine genomic DNA. The shaded boxes contain differences in amino acid sequence. See the legend to Figure 2 for the meaning of symbols.

u.v. irradiation. The membrane was prehybridized at 65 °C for 2 h in a solution containing 6 × SSC, 5 × Denhardt's solution [Denhardt's solution contains 1 mg/ml each of polyvinylpyrrolidone, BSA (fraction V) and Ficoll], 0.5 % (w/v) SDS and sonicated salmon sperm DNA (20 μg/ml). Hybridizations were performed in the same solution containing a radiolabelled probe made by random oligonucleotide priming (Feinberg and Vogelstein, 1983). Membranes were washed twice at 65 °C in 5 × SSC/0.1 % (w/v) SDS, and autoradiographed at −70 °C for 24–96 h with preflashed film (Fuji RX 100).

Samples (10 μg) of sheep genomic DNA were digested with one of the restriction endonucleases *Eco*RI, *Bam*HI, *Hin*dIII or *Sac*I. The digested DNA was precipitated at −20 °C with ethanol and sodium acetate (final concentration 3 M). Electrophoresis of digests and transfer of separated fragments to Hybond N were performed according to Southern (1975). Conditions for prehybridization, hybridization in the presence of prime-cut probes, washing and autoradiography have been given previously (Walker et al., 1987).

RESULTS AND DISCUSSION

Number of ovine genes for subunit c of ATP synthase

The number of nuclear sequences encoding precursors of subunit c of mitochondrial ATP synthase in normal and diseased sheep was estimated by Southern blotting, using probes capable of distinguishing between the P1 and P2 genes (see legend to Figure 1). There appeared to be at least four sequences related to P1, and about seven related to P2, in the sheep genome (Figure 1). This is similar to the complexity found in the human and bovine genomes, where there are single expressed P1 and P2 genes and several spliced or partially spliced pseudogenes in each case (Dyer and Walker, 1993).

Sequences of the ovine P1 and P2 cDNAs

By library screening, cDNA clones λP1C.1, λP2C.1 and λP1O.1 were isolated from the control (C) and diseased (O) animals, but no P2 clone was obtained from the diseased animal (see below). Clone λP1O.1 lacked the 5′ non-coding region and the coding sequence for amino acids 1–7 of the precursor, but the missing sequence was obtained from circularized cDNA by using the PCR and appropriate primers (Walker et al., 1992). The sequences of the P1 cDNAs were identical through the 5′ untranslated regions and the coding regions, and only a single base change was detected in the 3′ untranslated region at base 472 (Figure 2). Full-length P2 cDNAs were amplified both from the sheep with ceroid lipofuscinosis and from the control using primers based on the 5′ untranslated region in clone λP1C.1 in

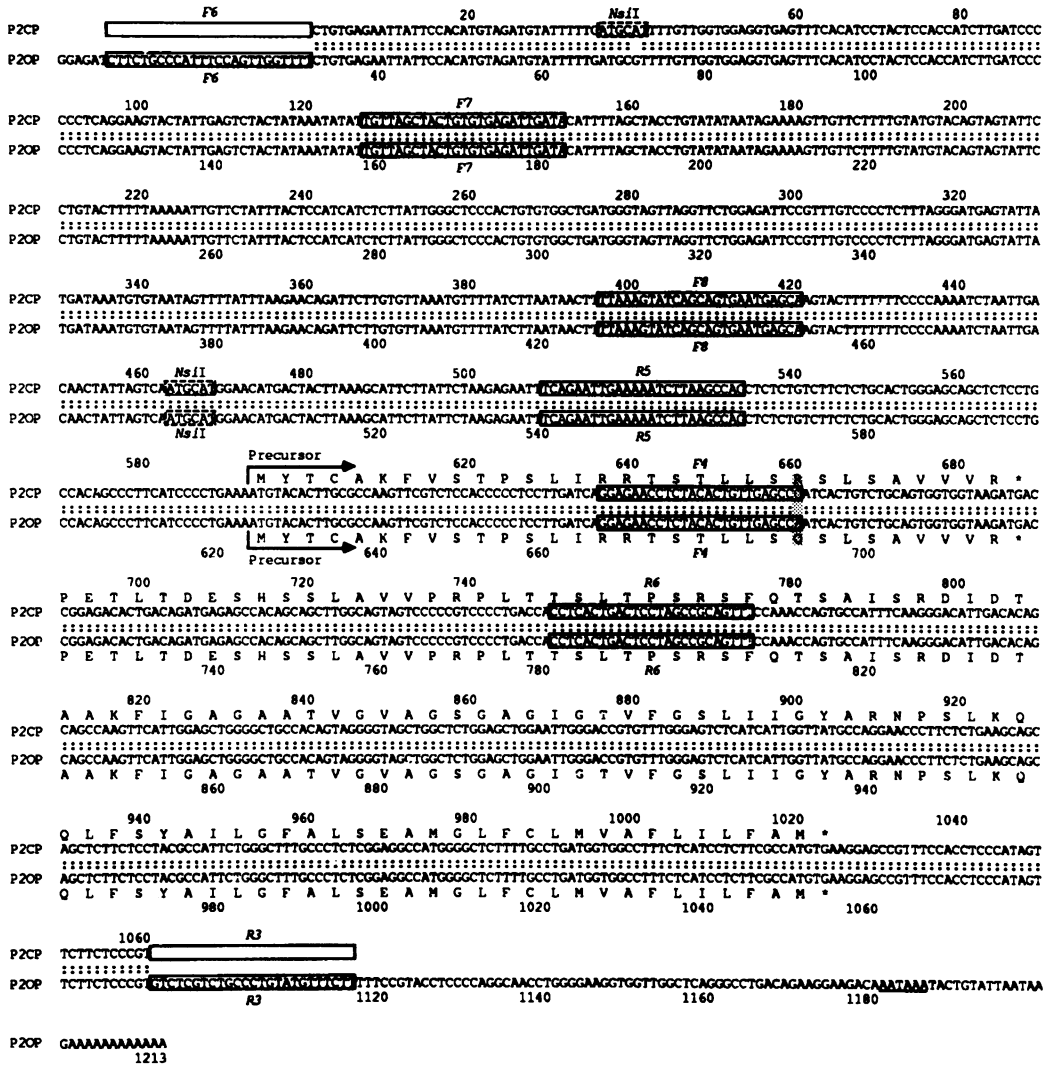

Figure 5 DNA sequences of ovine cDNAs containing expressed ovine pseudogenes related to P2 from healthy sheep and sheep with ceroid lipofuscinosis

The cDNAs are named P2CP and P2OP respectively. The positions of the primers employed in the PCR are shown, and the empty boxes represent primers for which the sequences were not determined independently. The shaded boxes indicate differences in amino acid sequence. See the legend to Figure 2 for the meaning of symbols.

PCR with circularized cDNA templates. Their sequences were identical (Figure 3). Therefore, both P1 and P2 mitochondrial import precursors of subunit c are the same in the control as in the diseased sheep, and ovine ceroid lipofuscinosis does not arise because of a mutation in the mitochondrial import sequences of this protein. This conclusion is in accord with the report that the gene causing 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 chromosomes 17 and 12 respectively (M. R. Dyer and J. E. Walker, unpublished work).

Characterization of an expressed P2 pseudogene

The screening of the cDNA library from the sheep with ceroid lipofuscinosis with a P2-specific probe yielded a single positively hybridizing clone, λP2Oϕ.1. This clone contained a sequence of 1213 bases, terminating in a 3′ poly(A) tail, with much of the sequence being closely related to the P2 cDNAs. However, it contains important differences from the P2 cDNA for the control animal isolated from the cDNA library, and for both the control and the diseased sheep isolated by PCR. First, the 5′ untranslated region was 623 bases long in the λP2Oϕ.1 sequence, but was only 96 bases in length in the P2 cDNA characterized from the control animal. Secondly, there were five nucleotide differences between the two clones; two of them changed arginines 23 and 32 in the import sequence of the P2 protein to glutamine and a stop codon respectively in the λP2Oϕ.1 sequence (Figure 4). Therefore the protein encoded by the λP2Oϕ cDNA is amino acids 1–31 of the 67-amino-acid P2 import precursor. The P2 cDNA amplified from the control animal by PCR contained not only the cDNA that had been isolated from the library (clone λP2C.1) but also a second P2-related sequence. This latter sequence, P2CP, differed from P2OP in clone λP2Oϕ.1 only at nucleotides 40 and 660 (Figure 5). Therefore two different P2 genes are transcribed in both control and diseased sheep.

The genomic DNA sequences giving rise to these mutated P2 transcripts were investigated in both the control and the diseased sheep by PCRs with forward and reverse primers based on the 5′ and 3′ ends of the recombinant sequence in λP2Oϕ.1. A product

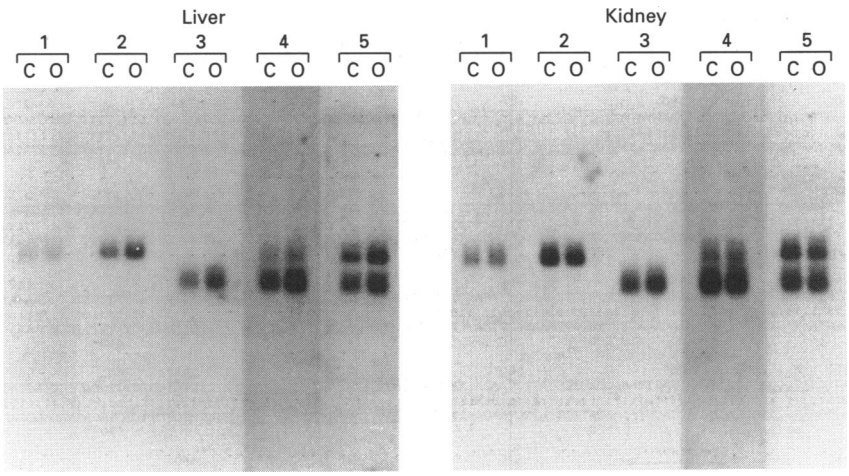

Figure 6 Hybridization of total RNA from healthy sheep and sheep with ceroid lipofuscinosis with P1- and P2-specific probes

Total RNA samples from liver and kidney are shown. The P1- and P2-specific probes are derived from bases 1–148 of the ovine P1 cDNA (see Figure 2) and from bases 96–289 of the ovine P2 cDNA (P2C in Figure 3) respectively. A third probe derived from bases 20–270 of the cDNA for the ϵ subunit of bovine ATP synthase (Viñas et al., 1990) was used as an internal control. Five identical pairs of total RNA samples from healthy sheep (C) and sheep with ceroid lipofuscinosis (O), labelled 1–5, were independently hybridized with the probes for P1 only, P2 only, ϵ only, P1 plus ϵ and P2 plus ϵ respectively.

of about 1100 bases was obtained from both control and diseased sheep, and their sequences exactly corresponded to those of the respective cDNAs shown in Figure 5. The most significant feature of these genomic sequences was that they contained no introns, and therefore the mutant cDNAs represent the transcripts of spliced P2 pseudogenes. By examination of genomic DNA from control and diseased animals it was confirmed that, as expected, both genomes also contain split genes for P2 as well as the spliced P2 pseudogenes, and that the stop codon found in the pseudogene is replaced by an arginine codon, as anticipated from the cDNA sequences. The evolution of the expressed P2 pseudogene and the origin of the P1 and P2 genes are discussed in the Appendix.

Expression of ovine P1 and P2 genes

Identical quantities of total RNA from livers and kidneys of healthy and diseased sheep were hybridized with specific probes for P1 and P2 and, as a control, with a probe for the ϵ subunit of bovine mitochondrial ATP synthase (see Figure 6). Similar levels of expression of the P1 and P2 precursors in the healthy and diseased animals were observed, and the sizes of the P1 and P2 transcripts were unchanged (Figure 6). Therefore there is no gross difference in the expression of subunit c in healthy sheep and in sheep with ceroid lipofuscinosis, disproving the hypothesis that accumulation of subunit c in lysosomes in ceroid lipofuscinosis is a consequence of overloading of the mitochondrial import pathway arising from increased expression of either the P1 or the P2 gene.

Transcripts from the spliced P2 pseudogenes were not detected in these Northern experiments, and it is concluded that they are transcribed at much lower levels than the unspliced genes. Evidence for their transcription in healthy and diseased sheep rests on the isolation of the P2 pseudogene sequence from a cDNA library, and on subsequent PCRs in which cDNA copies of transcripts derived from the pseudogenes were repeatedly isolated from total cDNA (Figure 5). Similar PCRs with genomic DNA templates amplified not only the same pseudogene sequences, but also several other mutated P2 sequences, probably

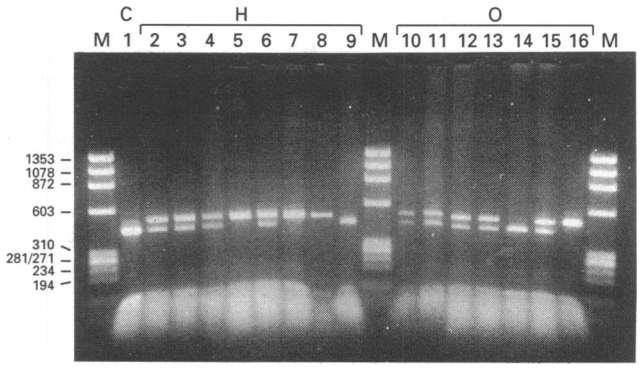

Figure 7 Restriction enzyme analysis of the sequences of genomic DNA representing the expressed P2-related pseudogenes of healthy sheep, sheep with ceroid lipofuscinosis and sheep heterozygous for the disease

The amplified DNA samples were digested with *Nsi*I and fractionated on a 1% agarose gel. The numbers along the top of the gel denote the various sheep. Sheep no. 1 was a healthy control animal (shown by C), sheep 2–9 were heterozygous (H) for ovine ceroid lipofuscinosis and sheep 10–16 had the disease (shown by O). M denotes an *Hae*III restriction digest of phage ϕX174 DNA, with the sizes of the fragments in bases.

derived from other spliced P2 pseudogenes. Therefore the detection of the single spliced pseudogene sequence by PCRs on cDNA populations was unlikely to be a consequence of contamination of the template with bovine genomic DNA.

Since the cDNA sequence of the expressed P2 pseudogene in the sheep with ceroid lipofuscinosis was different from the sequence in the control animal, the possibility arose that the pseudogenes were linked to the genetic cause of the disease. Therefore blood samples were taken from a further six sheep with ceroid lipofuscinosis and from eight sheep known to be heterozygous for the disease. The P2 expressed pseudogene was amplified from these samples and also from DNA from the two sheep investigated previously. The products were digested with the restriction enzyme *Nsi*I. This enzyme was chosen because a sequence difference between the pseudogene in the control and

the diseased animal creates an *Nsi*I site in the control sequence (Figure 7), and therefore digestion with this enzyme distinguishes between them, producing a 428 bp fragment from the control (lane 1 in Figure 7) and a 492 bp fragment from the sheep with ceroid lipofuscinosis (lane 16 in Figure 7). Some animals known to have ceroid lipofuscinosis were found that were homozygous for both of the two previously determined pseudogene sequences, and others were heterozygous for these genes. Therefore there is no linkage between the expressed P2 pseudogene and the gene causing ovine ceroid lipofuscinosis.

Possible causes of ceroid lipofuscinosis

After these investigations, the most plausible hypothesis for the basis of the ceroid lipofuscinosis is that the disease arises from a lesion in the degradative pathway of subunit c. It is thought that lysosomes digest mitochondria by a process of autophagocytosis (Pfeifer, 1987), but turnover rates of proteins of the mitochondrial inner membrane differ widely (Hare and Hodges, 1982), implying that selective turnover mechanisms may operate within the mitochondrion. However, the turnover rate of subunit c has not been measured, and it is not known whether mitochondrial proteases are involved in its degradation, nor if its degradation takes place in lysosomes.

In seeking the basis of ceroid lipofuscinosis, the distinctive properties of the c subunit of ATP synthase should also be borne in mind. First, ATP synthase is an abundant component of mitochondrial membranes, and probably has 10–12 copies of the c subunit per enzyme complex. Therefore, on a molar basis, the c subunit is one of the most abundant proteins in the inner membrane. It is also one of the most hydrophobic proteins that has been studied, and it falls into the category of proteolipid (a protein with the solubility properties of a lipid), being readily soluble in chloroform/methanol mixtures and insoluble in aqueous solutions. Therefore any mechanism for degrading the subunit would have to be able to cope with these physicochemical properties. However, subunit c is by no means the only proteolipid in mitochondrial inner membranes, and at least 15 different proteins have been found in chloroform/methanol extracts of mitochondria (Fearnley and Walker, 1986, 1987). None of these other mitochondrial proteolipids are found in the storage bodies associated with ceroid lipofuscinosis. This observation may indicate that specific factors operate in the degradation and turnover of the c subunit.

REFERENCES

Bentley, D. L. and Rabbitts, T. H. (1981) Cell **24**, 613–623
Benton, W. D. and Davis, R. W. (1977) Science **196**, 180–182
Boustany, R. N., Alroy, J. and Kolodny, E. H. (1988) Am. J. Med. Gen. Suppl. **5**, 47–58
Chirgwin, J. M., Przbyla, A. F., MacDonald, A. J. and Rutter, W. J. (1979) Biochemistry **18**, 5294–5299
Dyer, M. R. and Walker, J. E. (1993) Biochem. J. **293**, 51–64
Dyer, M. R., Gay, N. J. and Walker, J. E. (1989) Biochem. J. **260**, 249–258
Dyken, P. R., (1988) Am. J. Med. Gen. Suppl. **5**, 69–84
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. and Walker, J. E. (1986) EMBO J. **5**, 2003–2008
Fearnley, I. M. and Walker, J. E. (1987) Biochemistry **26**, 8247–8251
Fearnley, I. M., Walker, J. E., Martinus, R. D., Jolly, R. D., Kirkland, K. B., Shaw, G. J. and Palmer, D. N. (1990) Biochem. J. **268**, 751–758
Feinberg, A. P. and Vogelstein, B. (1983) Anal. Biochem. **132**, 6–13
Gardiner, R. M. (1992) Am. J. Hum. Genet. **42**, 539–541
Gay, N. J. and Walker, J. E. (1985) EMBO J. **4**, 3519–3524
Hare, J. F. and Hodges, R. (1982) J. Biol. Chem. **257**, 3575–3580
Jolly, R. D., Janmaat, A., West, D. M. and Morrison, I. (1980) Neuropathol. Appl. Neurobiol. **6**, 195–209
Jolly, R. D., Shimada, A., Dopfemer, I., Slack, P. M., Birtles, M. J. and Palmer, D. N. (1989) Neuropathol. Appl. Neurobiol. **15**, 371–383
Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
Messing, J. (1983) Methods Enzymol. **101**, 20–78
Messing, J. and Vieira, J. (1982) Gene **19**, 269–276
Norrander, J., Kempe, T. and Messing, J. (1983) Gene **26**, 101–106
Palmer, D. N., Barns, G., Husbands, D. R. and Jolly, R. D. (1986) J. Biol. Chem. **261**, 1773–1777
Palmer, D. N., Martinus, R. D., Cooper, S. M., Midwinter, G. G., Reid, J. C. and Jolly, R. D. (1989) J. Biol. Chem. **264**, 5736–5740
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
Pfeifer, U. (1987) in Lysosomes: Their Role in Protein Breakdown, pp. 3–59, Academic Press, London
Proudfoot, N. J. and Brownlee, G. G. (1976) Nature (London) **263**, 211–214
Rider, J. A. and Rider, D. L. (1988) Am. J. Med. Gen. Suppl. **5**, 21–26
Saiki, R. K., Gelfand, D. H., Stoffel, S., Scarf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Ehrlich, H. A. (1988) Science **239**, 487–491
Sebald, W. and Hoppe, J. (1981) Curr. Top. Bioenerget. **12**, 1–64
Southern, E. M. (1975) J. Mol. Biol. **98**, 503–517
Viñas, O., Powell, S. J., Runswick, M. J., Iacobazzi, V. and Walker, J. E. (1990) Biochem. J. **265**, 321–326
Walker, J. E., Gay, N. J., Powell, S. J., Kostina, M. and Dyer, M. R. (1987) Biochemistry **26**, 8613–8619
Walker, J. E., Arizmendi, J. M., Dupuis, A., Fearnley, I. M., Finel, M., Medd, S. M., Pilkington, S. J., Runswick, M. J. and Skehel, J. M. (1992) J. Mol. Biol. **226**, 1051–1072
Zeman, W. and Dyken, P. R. (1969) Pediatrics **44**, 570–583

APPENDIX

Evolution of the expressed P2 pseudogene and the origin of the P1 and P2 genes

Simon M. MEDD and John E. WALKER

The approximate elapsed time (T_o) since the divergence of two related sequences was calculated according to Miyata and Yasunaga (1981). The sequence difference (K) is defined as the number of mismatches per nucleotide site of two aligned sequences. It is assumed that nucleotide substitution follows a Poisson distribution, that it is equiprobable for all four nucleotides, and that multiple and back mutations may take place at each site. The rate of substitution (v) per nucleotide site per year is given by:

$$2vT_o = -\tfrac{3}{4}\ln[1-\tfrac{4}{3}K], \text{ where } K < \tfrac{3}{4}.$$

Estimation of the ages of the pseudogenes

From fossil evidence it is thought that humans diverged from cattle and sheep approx. 68 million years ago, and that sheep and cattle diverged from each other approx. 50 million years later (Romero-Herrera et al., 1978). Using these T_o values, and the cDNA sequence differences between humans and cattle, humans and sheep, and cattle and sheep for both P1 and P2 cDNAs (Table A1) evolutionary rates (v) for the P1 and P2 precursors were calculated (Table A2). These calculations showed that the rate of fixation of mutations in both P1 and P2 is slower between